Literacy Studies: Perspectives from Cognitive Neurosciences, Linguistics, Psychology and Education

R. Malatesha Joshi
Catherine McBride *Editors*

Handbook of Literacy in Akshara Orthography

Springer

Literacy Studies

Perspectives from Cognitive Neurosciences, Linguistics, Psychology and Education

Volume 17

While language defines humanity, literacy defines civilization. Understandably, illiteracy or difficulties in acquiring literacy skills have become a major concern of our technological society. A conservative estimate of the prevalence of literacy problems would put the figure at more than a billion people in the world. Because of the seriousness of the problem, research in literacy acquisition and its breakdown is pursued with enormous vigor and persistence by experts from diverse backgrounds such as cognitive psychology, neuroscience, linguistics and education. This, of course, has resulted in a plethora of data, and consequently it has become difficult to integrate this abundance of information into a coherent body because of the artificial barriers that exist among different professional specialties.

The purpose of this series is to bring together the available research studies into a coherent body of knowledge. Publications in this series are of interest to educators, clinicians and research scientists in the above-mentioned specialties.

Some of the titles suitable for the Series are: fMRI, brain imaging techniques and reading skills, orthography and literacy; and research based techniques for improving decoding, vocabulary, spelling, and comprehension skills.

More information about this series at http://www.springer.com/series/7206

R. Malatesha Joshi • Catherine McBride
Editors

Handbook of Literacy in Akshara Orthography

Springer

Editors
R. Malatesha Joshi
Department of Teaching, Learning, and Culture, College of Education and Human Development
Texas A&M University
College Station, TX, USA

Catherine McBride
Department of Psychology
The Chinese University of Hong Kong
Hong Kong, China

ISSN 2214-000X ISSN 2214-0018 (electronic)
Literacy Studies
ISBN 978-3-030-05976-7 ISBN 978-3-030-05977-4 (eBook)
https://doi.org/10.1007/978-3-030-05977-4

Library of Congress Control Number: 2019935530

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Contributors

Adeetee Bhide Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Learning Research and Development Center, Pittsburgh, PA, USA

Peter T. Daniels Independent Scholar, Jersey City, NJ, USA

Vasanta Duggirala Department of Linguistics, Osmania University, Hyderabad, Telangana, India

Hemanga Dutta Department of Linguistics and Contemporary English, School of Language Sciences, The English and Foreign Languages University (EFLU), Hyderabad, Telangana, India

John Everatt School of Teacher Education, College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand

Seema Gautam School of Teacher Education, College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand

Alexandra Gottardo Psychology Department, Wilfrid Laurier University, Waterloo, ON, Canada

Xuejun Ryan Ji Department of Educational and Counselling Psychology and Special Education, Faculty of Education, The University of British Columbia, Vancouver, British Columbia, Canada

R. Malatesha Joshi Department of Teaching, Learning, and Culture, College of Education and Human Development, Texas A&M University, College Station, TX, USA

Chandrika Mathur Rishi Valley Institute of Teacher Education, Rishi Valley Education Centre, Krishnamurti Foundation India (KFI), Chittoor District, Andhra Pradesh, India

Catherine McBride Department of Psychology, The Chinese University of Hong Kong, Hong Kong, China

Brigid McNeill School of Teacher Education, College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand

Amna Mirza Child and Youth Studies, Brock University, St. Catharines, ON, Canada

Ramesh Kumar Mishra Center for Neural and Cognitive Sciences, School of Medical Sciences, University of Hyderabad, Hyderabad, Telangana, India

Sonali Nag Department of Education, University of Oxford, Oxford, UK

The Promise Foundation, Bangalore, India

Pooja Nakamura American Institute for Research, Washington, DC, USA

Balambigai Narayanan Department of Speech, Hearing & Communication, National Institute for Empowerment of Persons with Multiple Disabilities (Divangjan) (NIEPMD), Chennai, India

Mimisha Nesan School of Teacher Education, College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand

Rauno Parrila Department of Educational Studies, Macquarie University, NSW, Sydney, Australia

Charles Perfetti Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Learning Research and Development Center, Pittsburgh, PA, USA

Keerthi Ramanujan Laboratory for Communication Science, Faculty of Education, University of Hong Kong, Hong Kong, China

Theeraporn Ratitamkul Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

Amir Sadeghi Department of English Language Teaching, Islamic Azad University, Damavand, Iran

Sunaina Shenoy Department of Psychology and Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

Department of Special Education, University of New Mexico, Albuquerque, NM, USA

Nandini Chatterjee Singh Language, Literacy and Music Laboratory, National Brain Research Centre, Gurgaon, Haryana, India

Shruti Sircar Department of Linguistics & Contemporary English, School of Language Sciences, English and Foreign Languages University, Hyderabad, Telangana, India

Sumathi T. A. R & D Engineer(Project), Language, Literacy and Music Laboratory, National Brain Research Centre, Gurgaon, Haryana, India

Shaher Banu Vagh ASER Centre, New Delhi, India

Pratham, New Delhi, India

Richard K. Wagner Department of Psychology and Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

Brendan S. Weekes Laboratory for Communication Science, Faculty of Education, University of Hong Kong, Hong Kong, China

Marasinghe A. D. K. Wijaythilake Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada

Heather Winskel Psychology, School of Health and Human Sciences, Southern Cross University, Coffs Harbour, NSW, Australia

Part I
Introduction

Introduction: Handbook of Literacy in Akshara Orthography

R. Malatesha Joshi and Catherine McBride

Abstract The authors provide a summary of the chapters from the handbook on literacy on different aksharas from the Indic writing system

Keywords Akshara orthography · Assessment · Indian subcontinent · Instruction · Literacy · Reading · Spelling

The United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2015) has emphasized the importance of literacy for a better life. – This includes not only affording everyone the ability to read and write, but the utility of literacy for eradicating poverty, reducing child mortality, and making peace and stability possible around the world. In the United States, the National Institutes of Health (NIH; Sweet, 2004) has considered illiteracy a public health issue, given that more than 50% of American adolescents with criminal problems, individuals with a history of substance abuse, and those living on welfare have reading problems. Considering the negative impact of illiteracy on the individual, on society, and on each nation, various disciplines have explored the many causes of illiteracy. These include factors such as home background, oral language development, exposure to literacy atmosphere at home and school, and type of instruction in the classroom.. In addition to these factors, type of orthography, can also be a factor in one's literacy development. For example, a seminal study by Seymour, Aro, and Erskine (2003) found that it typically takes approximately 2 years of formal instruction to become proficient in decoding for English-speaking children. In contrast, it usually takes only 1 year of formal instruction for achieving proficiency in some European orthographies, such as Spanish, Italian, and German. To be clear, the number of

R. M. Joshi (✉)
Department of Teaching, Learning, and Culture, College of Education and Human Development, Texas A&M University, College Station, TX, USA
e-mail: mjoshi@tamu.edu

C. McBride
Department of Psychology, The Chinese University of Hong Kong, Hong Kong, China
e-mail: cmcbride@psy.cuhk.edu.hk

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_1

years for proficient word reading as noted by Seymour et al. (2003) and others corresponds to decoding and does not refer to reading comprehension.

Before we proceed further, it is important to define the different terms we use in this volume. *Language*, generally, refers to the spoken form, whereas *orthography* refers to the visual representation of language and includes phonological, morphological, and orthotactics principles. *Writing systems* refer to how the written format is represented in terms of linguistic units, such as alphabetic writing systems (e.g., English, Spanish, and Arabic), syllabic systems (including Japanese Kana), morphosyllabic systems (such as Chinese and Japanese Kanji), and alpha-syllabic systems (e.g., Korean Hangul). *Script* is the graphemic format in which writing is represented, such as the Roman script or the Arabic script. Thus, the English orthography is an alphabetic writing system represented in the Roman script. Based on the size of linguistic units, Coltheart (1984) classified writing systems into three broad categories: alphabetic, syllabic and logographic. Coltheart further divided the syllabic writing system into those with a phonetic representation in the syllable, such as in the Devanagari script of many Indian languages, and those without phonetic representation in the syllable, such as in Japanese Kana. Examples of alphabetic writing systems are English, Spanish, and Russian. Chinese characters was traditionally conceptualized as an example of a logographic system. However, more recently, the classification has been reevaluated. Now Chinese is referred to as a morphosyllabic writing system rather than a logographic system (Joshi & Aaron, 2006).

Importantly, it has been suggested that not even all alphabetic writing systems, despite their classifications as such, are not all that similar in characteristics. For instance, even though English, Spanish, and Arabic are considered to represent alphabetic writing systems, there are many differences among these three orthographies in terms of grapheme-phoneme correspondences (GPC), direction of writing (English is written from left to right and Arabic is written from right to left), and morphological and syllable aspects. Further, these alphabetic orthographies can be further classified as transparent or opaque. Transparent orthographies have a fairly close correspondence between phonemes and graphemes while relatively opaque orthographies may have one to many correspondences between phonemes and graphemes. Therefore, it is better to view these orthographies on a continuum from opaque to transparent rather than as belonging to one or the other category. These and other complexities of the relationship between orthographies and literacy acquisition are highlighted by the Orthographic Depth Hypothesis (ODH) (Katz & Frost, 1992) and the Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005), among others. Indeed, Share and Daniels (2016), and Daniels and Share (2017) suggest classifying orthographies on ten different dimensions, including linguistic distance, visual complexity, spelling constancy despite morphophonemic alternation, omission of phonological elements, allography, dual purpose letters, ligaturing, and inventory size. Thus, it is imperative to examine literacy acquisition in different orthographies from a broad perspective.

The Indic writing system, primarily used in Indian subcontinent and a few other south-east Asian counties, has been classified as a syllabic writing system with identifiable phonetic elements. However, Karanth (2006) classified it as an alpha-syllabary system while Rimzhim, Katz, and Fowler (2014) considered it as predominantly alphabetic. Importantly, two recent reviews from Share and Daniels noted that the orthographic underpinnings of processing of akshara are unique in a variety of ways; indeed, an akshara orthography is neither alphabetic nor syllabic and should be given a separate status. Even though akshara orthographies are used by two billion people, there has not yet been a systematic attempt to bring the features, research findings, and future directions of akshara into a coherent volume. We have, therefore, tried to bring the available research on the akshara orthography together in a coherent fashion in this volume.

The volume begins with an excellent historical perspective on the Indic writing system by Daniels. The author explains how Brahmi script spread throughout the Indian subcontinent and southeast parts of Asia. The chapter also provides an overview of the historical origin of Brahmi script derived from Kharosti and explains why the Indic writing system should be considered as a separate category for its special characteristics, such as the inherent vowel in the consonant syllable. Ramanujan and Weekes demonstrate the long endurance of the akshara system, despite the fact that some of the principles and structures of it might have changed in its evolution.

The next chapter, by Nag and Narayanan, focuses on Tamil, which is one of the oldest languages in India., Tamil is a Dravidian language from South India; it also has the largest number of speakers compared to other Indian languages in India. Nag and Narayanan outline phonological, morphological, and orthographic principles of Tamil and explain the literacy instruction for it, which is primarily grounded on 'Activity Based Learning' (ABL). Tamil Nadu, where Tamil is the official language of the state, also has a history of strong educational programs.

Another Dravidian language used mostly in South India, is Malayalam, spoken primarily in the state of Kerala. Interestingly, among all of the Indian states, Kerala boasts a particulalry high literacy rate. Thus, in their chapter, Nesan, Sadeghi, and Everatt, outline the linguistic structures of Malayalam and highlight the importance of instruction in it, specifically emphasizing metalinguistic concepts. In the next chapter, Nakamura, Joshi, and Ji, focus on English spelling skills in South India. After summarizing the linguistic properties of Kannada and Telugu, two Dravidian languages, these authors present results from the data collected among very low income children. Results of Quantile Regression analyses have shown that the spelling skills in English are influenced by children's proficiency in Kannada (or Telugu). The following chapter contrasts hearing-impaired and typically developing children. This chapter focuses on another south Indian Dravidian language, namely, Telugu. Here, Duggirala, after examining both normal hearing and hearing-impaired schoolchildren, found that a syllable deletion task was more difficult than other cognitive linguistic tasks, including syllable substitution and lexical decision tasks, among these children. The results were explained in terms of sonority principles and specific syllabic features of Telugu orthography. These chapters described above all collectively represent some research on the four major Dravidian languages from South India.

In contrast, the next chapters focus primarily on languages from Northern India. Northern Indian languages, referred to as Indo-Aryan languages, share some similarities in the arrangement of aksharas and some grammatical features with the south Indian Dravidian languages. Thus, in their chapter, Gautam, Everatt, Sadeghi, and McNeill, applied the Simple View of Reading (SVR) to two Indo-Aryan languages spoken mostly in northern part of India. Punjabi is one of the very few tonal languages in India; it is written in the Gurumukhi script, while Hindi is written in the Devanagari script. In this study, in both Punjabi and Hindi, both decoding and listening comprehension predicted reading comprehension. Additionally, orthographic knowledge also made an independent contribution to reading comprehension, suggesting that the complex orthographic structure of Indic writing systems has to be taken into consideration while examining Indian languages. Literacy acquisition in Bengali, another Indo-Aryan language, is described by Sircar and Nag in their subsequent chapter. Even though Bengali, like most other languages belonging to the Indic writing system, is transparent, a longer duration of explicit formal instruction is required to master the literacy skills for this orthography, especially for writing. The next chapter also focuses on phonological and orthographic characteristics of a prominent language. This Indo-Aryan language from northern India is Assamese, which has unique features. Dutta outlines some of these characteristics in phonology and orthography and suggests some recommendations to make literacy acquisition in Assamese a little easier.

The languages we have summarized thus far are mainly used in India. In the next few chapters, languages used outside of India are considered., Of course, the closest country outside of India is Sri Lanka, where Sinhala is used. Sinhala is an Indo-Aryan language, written in a highly cursive script; it has the characteristic of 'diglossia.' Wijaythilake and Parrila, thus, outline the characteristics of the Sinhala orthography and summarize their results from some preliminary studies. They note that, similar to findings from India, in Sri Lanka also, children need 3–4 years of formal instruction to master the decoding skills and writing tends to be more difficult than reading. Another Indic writing system used outside of India is Thai. Although Thai follows the basic principles of the Indic system, it also has some unique characteristics, such as no spaces between words. Winskel and Ratitamkul provide a general framework of the Thai writing system and literacy development and point out that, in Thai, similar to many other orthographies based on Indic writing system, spelling and writing development lag behind reading.

Despite the fact that close to two billion people use the Indic writing system, good assessment procedures for it have not been developed yet. The difficulty of assessment in the Indic writing system is further complicated by the fact that the vast majority of children using this system may be learning to acquire literacy skills in a language that is, in fact, different from their home language. Additionally, India has adopted a so-called "three-language" formula, which means that by the time a student is in grade 5, s/he should be taught literacy skills in three languages. With this background, Vagh and Nag provide a framework for the assessment procedures in the Indic system, especially in India, by suggesting that different tasks be included in such assessments, taking into consideration various factors such as the complex orthographic structure of the languages, home background, and exposure to different languages. Singh and Sumathi report results from a large scale study of school children from grades 1–5 from Hindi-speaking and Marathi-speaking locations. Their results showed the importance of phonological processing especially at the syllable level, in addition to the importance of oral language development.

Similar to the lack of scientifically-based assessment procedures, evidence-based instructional procedures are also lacking in akshara based languages. Mathur and Nag, after outlining some existing programs which are not evidence-based, provide steps for systematic instruction based on home language, language of instruction, and other research methods that have been found to be successful in other languages; one such example is dialogic reading. Bhide and Perfetti describe their experiment designed to teach Marathi aksharas among adults. Interestingly, their study showed the importance of copying, a neglected aspect in teaching of literacy skills.

English language-learning is also an important issue throughout India and abroad. Thus, Mirza and Gottardo describe performances of 9–10 year-old English-Hindi bilinguals living in Canada on various literacy related tasks. These children are relatively strong in English, having grown up in Canada. Results showed that, in addition to phonological processing, vocabulary knowledge also explained word reading in both English and Hindi in these students. In the next chapter, Mishra describes results from an eye-tracking study of English-Hindi bilinguals and shows that even highly proficient second language users activate translations when listening to spoken words. In the following chapter, Shenoy and Wagner focus on the impact of socio-economic status on children's literacy development in Kannada-English bilinguals aged 7–10 years. In that study, while only 13% of the schoolchildren from high-income school performed poorly on literacy tasks in both languages, more than 50% of the children from low-income schools performed poorly in those languages. Another interesting observation was that in these bilingual classrooms, teachers in low-income schools switched languages during instruction, while teachers in high-income schools did little code-switching.

India has had a long tradition of linguistic analyses starting with Panini (sixth to fourth century B.C.E) (Coward & Raja, 1990), who systematically conducted the analyses of noun compounds and morphology, by using an idealized mathematical model. Indeed, Panini's model forms the bases of the modern day linguistic theories. Yet, very little research has been conducted examining literacy development in this Indic writing system. This volume is an attempt to integrate a few of the research studies conducted so far as well as to promote further research studies. The akshara orthography is unique, and the learning context in India makes literacy acquisition particularly challenging. Here, there are so many children learning to read in a language that is different from their home language, in addition to the three-language formula imposed by the government. Furthermore, even though the akshara orthography is fairly transparent, the visual complexity of the aksharas make it hard to master decoding skills. Indeed, it might take 4–5 years of formal instruction to be proficient in akshara decoding (Nag, 2007; Nakamura, Joshi, & Ji, 2018).

Additionally, crowded classrooms, lack of classroom materials, and poor teacher training also may contribute to problems with literacy development.

Our gratitude and our optimism are far-reaching. First, we wish to thank various reviewers of the volume as well as Ms. Jolanda Voogd and Ms. Helen van der Stelt of Springer publishers for their help with the preparation of this volume. The credit for producing the chapters which make up the volume, of course, goes to our contributors. Each chapter represents an important voice in the academic journey to understand the akshara system, which is vast in its influence and still as of yet greatly underrepresented in research studies on literacy. Our greatest hope in producing this volume is that it will encourage more and more research on both the akshara writing system itself and on countries such as India, Sri Lanka, Laos, Thailand, and others that use this system. Particularly important is for researchers worldwide to recognize the diversity in writing systems and learning environments that are associated with variability in reading and writing development and impairment. We invite more and more researchers from around the world to contribute their academic voices in helping a global audience to understand, explain, and facilitate literacy development globally. UNESCO's (2015) appreciation of the importance of literacy for a better life should be greatly enhanced as literacy researchers focus on the many systems contributing to literacy variability in and around India and beyond.

References

Coltheart, M. (1984). Writing systems and reading disorders. In L. Henderson (Ed.), *Orthographies and reading* (pp. 67–80). New York, NY: Taylor & Francis.

Coward, H. G., & Raja, K. K. (1990). *Encyclopedia of Indian philosophies* (Vol. V). New Delhi, India: Motilal Banraasidass Publishers.

Daniels, P. T., & Share, D. L. (2017). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading, 22*(2018), 101–116. https://doi.org/10.1080/10888438.2017.1379082.

Joshi, R. M., & Aaron, P. G. (Eds.). (2006). *Handbook of orthography and literacy*. Mahwah, NJ: Lawrence Erlbaum Associates.

Karanth, P. (2006). The kagunita of Kannada – Learning to read and write an Indian alphasyllabary. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 389–404). Mahwah, NJ: Lawrence Erlbaum Associates.

Katz, L., & Frost, L. (1992). Reading in different orthographies: The orthographic depth hypothesis. In R. Frost & L. Katz (Eds.), *Orthography, phonology morphology, and meaning* (pp. 67–84). Amsterdam: Elsevier.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22. https://doi.org/10.1111/j.1467-9817.2006.00329.x.

Nakamura, P. R., Joshi, R. M., & Ji, X. (2018). Investigating the asymmetrical roles of syllabic and phonemic awareness in akshara processing. *Journal of Learning Disabilities, 51*, 499–506. https://doi.org/10.1177/002221941771820.

Rimzhim, A., Katz, L., & Fowler, C. A. (2014). Brāhmī-derived orthographies are typologically Āksharik but functionally predominantly alphabetic. *Writing Systems Research, 6*, 41–53. https://doi.org/10.1080/17586801.2013.855618.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.

Share, D., & Daniels, P. (2016). Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz, and Fowler (2014). *Writing Systems Research, 8*, 17–31. https://doi.org/10.1080/17586801.2015.1016395.

Sweet, R. W. (2004). The big picture: Where we are nationally on the reading front and how we got there. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 13–44). Baltimore, MD: Paul H. Brookes.

United Nations Educational, Scientific, and Cultural Organization. (2015). *Education for all (EFA) global report*. Paris, France: Author.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29. https://doi.org/10.1037/0033-2909.131.1.3.

Indic Scripts: History, Typology, Study

Peter T. Daniels

Abstract Indic scripts are the writing systems descended from the third-century BCE Brahmi script, which was used throughout the realm of Emperor Aśoka. They are used by nearly two billion people from Pakistan through India, Tibet, Sri Lanka, and Southeast Asia, and in former times parts of Indonesia and the Philippines and along the Silk Road; Indic scripts have influenced scripts as far away as the Horn of Africa. Most of them are characterized by the distinctive typological property of basic characters denoting consonants plus the unmarked vowel (usually /a/), with the other vowels denoted by marks attached to the base characters; and in many, consonant clusters (not respecting syllable boundaries) are denoted by ligatures of two (or more) consonant characters. This chapter describes the descent and distribution of the forms, the varying typologies over space and time, and the investigations by which their properties became known to linguistic scholarship.

Keywords Abugida · Brahmi · Decipherment · Indic scripts history · Indic scripts typology · Kharoṣṭhi · Writing system typology

Prologue (DLS)[1] *It is widely acknowledged that the writing system in which literacy is acquired has a profound influence on the course of literacy acquisition (e.g., Daniels & Share, 2017; Frost, 2012; Perfetti & Harris, 2013; Seymour, Aro, & Erskine, 2003; Verhoeven & Perfetti, 2017; Ziegler & Goswami, 2005). It follows that an appreciation of the characteristics of and the differences among the world's writing systems is essential for the literacy researcher. Regrettably, the psychology of reading has persisted in adhering to an obsolete taxonomy of writing systems that recognizes just three varieties—alphabets, syllabaries, and logographies (Gelb,*

[1] The author is immensely grateful to David L. Share (University of Haifa) for the Prologue and Epilogue that situate the chapter within the context of this book.

P. T. Daniels (✉)
Independent Scholar, Jersey City, NJ, USA
e-mail: ptdaniels@uchicago.edu

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_2

1952; Taylor, 1883). This tripartite scheme, despite having been discredited by contemporary scholars (e.g., Coulmas, 2009; Daniels, 1990, 2018; DeFrancis, 1989; Gnanadesikan, 2017; Joyce & Borgwaldt, 2011; Rogers, 2005), remains alive and well in the psychological and educational literature (see, for example, Rayner, Pollatsek, Ashby, & Clifton, 2012, Chapter 2). Like the proverbial square peg in a round hole, the classic tripartite approach does not seem to accommodate Indic scripts; the dilemma of how to characterize and classify Indic scripts is still evident in ongoing debates over terminology (see, e.g., Bright, 2000; Rimzhim, Katz, & Fowler, 2013; Share & Daniels, 2015). Are the Brahmi-derived scripts alphabetic? syllabic? or a mixture of the two, as the popular but controversial term "alphasyllabary" denotes? In this chapter, Daniels discusses the origins, decipherment, dispersion, and typological diversification of Indic scripts, which he shows represent a unique category of writing system he refers to as an abugida.

Introduction

The Indic writing systems originated in the Kharoṣṭhi and Brahmi scripts of the third quarter of the first millennium BCE. The Kharoṣṭhi script was found in the far northwest of the Indian subcontinent, where it was used for Gandhari Prakrit, and subsequently along the Silk Road to Inner Asia, where it persisted to perhaps as late as the mid first millennium CE. Brahmi, the refinement of Kharoṣṭhi, appeared nearly simultaneously throughout the rest of the Subcontinent during the reign of the Mauryan emperor Aśoka, in the middle of the third century BCE; the emperor recorded his edicts in local Prakrits (Dani, 1986; Salomon 1998; Sircar, 1965, 1970–71).

Over the next two millennia, Brahmi diversified into countless local scripts within India—ten of them received standardized forms with the coming of printing in the late sixteenth to the early nineteenth centuries—and varieties were taken to mainland and island Southeast Asia by Hindus and Buddhists, at first to write Sanskrit and Pali (Filliozat, 1953; Holle, 1999).[2] There they flourished as vehicles for local languages and developed into distinctive scripts for the languages of many Southeast Asian peoples. Nearly two billion people could employ Indic scripts today, if a goal of universal literacy can be met.

[2] There is only one survey intended for a general audience (Hosking & Meredith-Owens, 1966). The two comprehensive albeit outdated English-language histories of writing include fairly sizable accounts describing the development of and relationships among the forms of Indic scripts (Diringer, 1968, pp. 1:237–239, 257–307. 310–351; Jensen, 1969, pp. 361–406).

Decipherment and Study of the Earliest Indic Scripts

As early as 1821, Ulrich Friedrich Kopp (1762–1834) had considered and rejected a historical connection between the Indic and Ethiopic scripts (Kopp, 1819–1821, p. 2: 348). He knew Indic script only from certain publications by Sir William Jones in the previous decades, though a few full, albeit crude, alphabets had been given in Diderot and d'Alembert's *Encyclopédie* in 1763.[3]

Sanskrit texts were known in Europe by the eighteenth century, but in India itself (where the humid climate did not favor the preservation of ancient manuscripts) only inscriptions on solid materials, including coins, could be found throughout the country, some of considerable age. Many could be read by pandits, but some could not. They proved to stand at the beginning of the craft of writing in India.[4]

Fifty years had passed since Sir William Jones's now-famous conjecture regarding the ancestry of the Indo-European languages ("… sprung from some common source …").[5] British hegemony over the Indian subcontinent had been extended from Bengal around to Bombay, but was by no means yet complete. In June 1837, James Prinsep, Secretary of that same Asiatic Society of Bengal which Jones had headed, announced the decipherment of the script which was to prove the ancestor of all the alphabets[6] of South and Southeast Asia (including the Devanagari and all the state and national scripts currently in use). His work had begun three years earlier, the semicentenary of the founding of that Society by Sir Warren Hastings.

Prinsep was not a scholar by profession. Born in London in 1799, he had trained in architecture and chemistry and apprenticed to the assay master of the Royal Mint (responsible for the quality of metal used). He was also skilled in drawing. He, and six of his brothers, pursued careers in India. He arrived in Calcutta a month after his 20th birthday, where he became Assistant to H. H. Wilson, afterward the compiler of the first great dictionary of Sanskrit. After a year he moved to Benares and managed the new mint. After a decade the Benares Mint was closed and Prinsep returned to Calcutta and began his association with the Asiatic Society. When in 1832 Wilson was made Professor of Sanskrit at Oxford, Prinsep succeeded him as Assay Master and in most of his public offices and the Secretaryship of the Society, despite no great ability in Sanskrit. He reorganized the Society's publications into a monthly

[3] *Planches*, vol. 2, pt. 1, section "Caractères et alphabets," pls. XV, XVII–XXII.

[4] The undeciphered "Indus Valley script" (Parpola, 1994) does not bear on the history of Indic writing. A suggestion that the Indus objects do not carry writing at all (Farmer, Sproat, & Witzel, 2004; Sproat, 2014) is not persuasive.

[5] The account of Prinsep's decipherment of Brahmi lightly adapted here was presented at the American Oriental Society's 1987 annual meeting, in recognition of the sesquicentennial. As the last time his methods had been reviewed in detail was in Rudolf Hoernle's *Centenary Review* (1884; neither Georg Bühler [1896, translated as Bühler, 1904] nor Ernst Windisch [1917] in their contributions to the *Grundriss der Indo-Arischen Philologie* has much to say about them)—over a hundred years earlier—I considered it appropriate to recall them.

[6] Before the late nineteenth century, there was no concept of "types" of writing system, and "alphabet" was used of all.

journal. In his official capacity he put through currency reform and standardization but seems to have devoted much of his time thereafter to the area where it and his Society duties overlapped, namely, Indian numismatics (Arbuthnot, 1896).[7] The study of Indian coins and inscriptions led him to an interest in inscriptions in general; furthermore it was he who engraved or lithographed many of the plates illustrating the *Journal* himself. His last published work was a study of Bactrian coins that included his decipherment of the Kharoṣṭhi script.[8]

Materials Available for the Decipherment

Brahmi

Emperor Aśoka caused to be erected throughout India (as well as "rock inscriptions") a number of "pillar inscriptions" bearing a standardized text of his "edicts." Three eventually became available to the decipherer. The first to become known to Prinsep was the "Allahabad column," bearing a number of inscriptions in various scripts, some readable with little effort based on ones closer to scripts still in use. Careful copies were made by Lieutenant T. S. Burt, Engineer, and transmitted to Prinsep, who drew them, reduced, onto lithographic stones for reproduction in the *Journal of the Asiatic Society of Bengal* [*JASB*] (Burt, 1834). A second pillar inscription was found at Mathiah near Bettiah (Hodgson, 1834); a third, from Delhi, had been published on several plates accompanying Harington & Colebrooke, 1799.[9] Early in 1837, a series of related edicts carved in rock began to be discovered and copied (Prinsep, 1838a).

[7] It may be noted that Prinsep was not the only numismatist/decipherer. Jean-Jacques Barthélemy (Daniels, 1988) in the mid eighteenth century was the Keeper of the Royal Coin Collection in Paris, and was the very first scholar to decipher an unknown script—he is credited with no fewer than three decipherments. (And the archeologist narrator of James Michener's novel *The Source* credits his career to having collected coins, rather than stamps, as a boy.) Unlike Barthélemy, however, Prinsep was never entirely forgotten by the discipline he brought into being; his name is usually mentioned in, for instance, books on the origin of Brahmi, but almost never with clear or accurate references to his publications.

[8] Despite its title, Thomas (1858) collects only Prinsep's numismatic essays, and for the most part they are not reprinted verbatim. Thomas chose to omit most of the passages detailing the decipherments; he was focused on creating a reference work on Indian coins and had no scruples about silently deleting superseded passages by Prinsep and inserting his own observations (in brackets), incorporating two further decades of research: Thomas may well have written more than half the contents of the volumes himself.

[9] As is their unfortunate wont, Google Books did not unfold the plates, so the on-line version is unusable. Harington's introduction notes that the unknown character of this inscription is also found on the Allahabad pillar, of which a specimen is included, and it is not possible to determine how many of the seven relevant plates depict the Delhi and how many the Allahabad inscription.

Kharoṣṭhi

The only examples of Kharoṣṭhi available for study by the 1830s were the legends on Bactrian coins.

> It must be remembered that the only incontestible authority for the determination of a vowel or consonant is, its constant employment as the equivalent of the same Greek letter in the proper names of the Bactrian kings. (Prinsep, 1838c, p. 639 = Prinsep apud Thomas, 1858, p. 2: 127)

Prinsep himself engraved plate after plate reproducing dozens of Bactrian coins, which are for the most part well reproduced in Thomas's re-edition.

Decipherment

Brahmi

Prinsep contributed numerous articles to the *JASB* each year, on an amazing variety of topics, but only three bearing on the decipherment of Brahmi, in March and October 1834 and June 1837. The first is "Notes on Inscription No. 1 of the Allahabad Column" (1834a). He suggests on both internal and external evidence that this inscription used a very archaic alphabet: it is simple in form; it is almost unique (only two like it are known); it occupies the first place on the monument; and even in the earliest (Persian) mention of the pillar, its origin is unknown. There are reasons both for and against considering the language to be Sanskrit.

The important contribution of this article is the following:

> I have taken the trouble of analyzing carefully the whole of the inscription …, classifying those forms which seemed to be derived from the same radix …. I soon perceived that each radical letter was subject to five principal inflections. (1834a, p. 117)

These inflections corresponded to those in the already deciphered second inscription. He also added figures for the frequency of each letter (Fig. 1). (An adjoined table of the frequencies of letters from a page of Sanskrit proved irrelevant, because of varying subject matter.) It is possible that if Prinsep had known more Sanskrit he might have been able to read the inscription straightaway; he suggests that some of the letters might be identifiable with some of those in the second inscription, but declines even to set them down.

Prinsep's second article, "Note on the Mathiah Láth inscription" (1834b), was occasioned by the publication of the second pillar. It was obvious that it was in the same script as Allahabad No. 1; but upon closely comparing them and the column at Delhi—it is not clear why this was not noticed earlier—, they turned out to be three copies of the same text. Comparison enabled Prinsep to throw out some of the anomalous letters from his earlier chart, ascribing them to copyists' errors. He was also able to make some suggestions based on phonotactics: two particular elements of what seemed to be compound letters were likely to be *y* and *s*; these suggestions

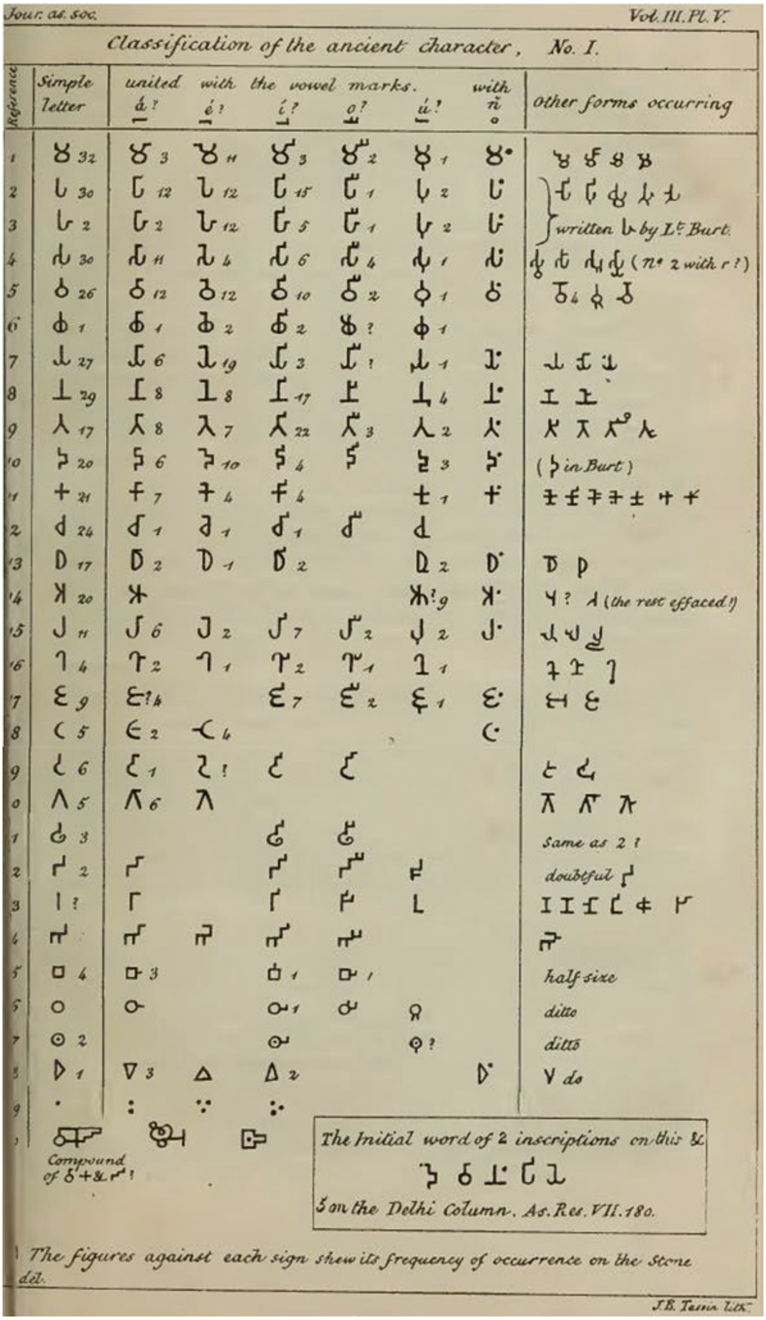


Fig. 1 Prinsep 1834a, pl. V

Fig. 2 Prinsep 1837a, p. 476

proved to be correct. Furthermore, in the Delhi column the words are separated from each other by blank spaces, enabling certain letters to be identified as inflections; this is interesting from the point of view of the history of writing—why should one copy of an inscription divide between words while its contemporaries did not? Finally, some mistakes in the arrangement of the plates could be corrected.

A fourth copy of the inscription then turned up: directly following that second article, Prinsep had published an inscription from Sanchir near Bhilsá,[10] in what he could see was a less than satisfactory copy, along with a wish for a better one (1834c). When this hope was finally fulfilled, almost three years later, it proved to be in the No. 2 script and fully readable (1837a). But it brought with it an inestimable bonus: 23 short inscriptions in the Brahmi script, which appeared to be graffiti. Here it was a helpful coincidence that Prinsep himself was the lithographer. He says that while he was arranging and copying the lines, he noticed that they all ended with the same two letters. These were thus all little formulas—probably either obituary or votive. And these two letters often followed, at the end of the preceding word, the letter already identified as *s*—which, he had just realized, from newly read coin legends, was a genitive inflection. Then, in the leap of intuition that characterizes the great decipherer, he realized that the two-letter word was *dánam* 'gift',

> teaching me the very two letters, *d* and *n*, most different from known forms, and which had foiled me most in my former attempts. Since 1834 also my acquaintance with ancient alphabets had become so familiar that most of the remaining letters in the present examples could be named at once on reinspection. (1837a, p. 461)

By the time the whole article was set in type, a font for the new alphabet had been cut (Fig. 2). Every letter had been correctly identified.

Prinsep then contributed an edition, translation, and commentary on the pillar inscriptions (1837b). It was accompanied by the text set in the newly prepared type. Note that it was not Prinsep, but the Honourable G. Turnour, who identified the author, "Piyadasi," as Aśoka (Turnour, 1837, pp. 1054–1057). He did this, in an appendix to his own treatment of the pillar inscriptions, by citing Buddhist annals from Ceylon. His treatment was published at the end of the same year; he observed that he Paliized, Prinsep Sanskritized. Prinsep inserted his own signed footnotes

[10] The acute accent was often used, like the macron today ($\bar{a}$), to mark a long vowel.

into Turnour's article, defending his own opinions and deriding Turnour's[11]; but he immediately capitulated on the identification of the king ("his happy discovery of PIYADASI's identity.–ED." (p. 1061 n. †).

Prinsep's final statement on Brahmi came in 1838 when he edited the rock inscriptions from Girnar (1838a, 1838b) and included plates and an essay on the history of the alphabet from Brahmi to Devanagari (1838b, pp. 271–276).

How, then, did Prinsep accomplish what more learned scholars of Sanskrit had not? The leap of intuition mentioned above, of course, was a big part of it. The absolute *sine qua non* of any decipherment is accurate copies of the inscriptions in question.[12] With the publication of the Allahabad column, this was available, so Prinsep decided to chart the letters—a natural reaction for a physical scientist![13] Of the six other factors leading to successful decipherment that I have isolated (Daniels, 1988, pp. 435–436), three involve physical bilinguals and so are not relevant here,[14] but the others apply: the unknown was in a familiar language, its identity was known, and the script was related to known ones.

Kharoṣṭhi

The decipherment of what Prinsep called the "Indo-Bactrian alphabet" was more straightforward, precisely because the considerable supply of coins of the appropriate eras provided him with a considerable number of bilingual coin legends with the names of Greek potentates on the obverse. His first attempt (Prinsep, 1835, pp. 329–336) compared letters in "Pehlevi," i.e., Sassanian inscriptions and coins, though not Pahlavi manuscripts; the "Zend," or Avestan, alphabet is also given (after Burnouf, 1833)[15]—which language he seems to take as equivalent to "Páli" (see below).

Those earlier interpretations were less than satisfactory, and after his triumph with the Aśokan script, Prinsep turned back to the Bactrian (1838c). With a fine dramatic sense, Prinsep begins with a lengthy list of the coins he used in his studies and the gentlemen who had made them available to him (as indeed he had in the earlier article).

[11] "We consider it a duty to insert this paper, just received, in the same volume with our version of the inscription, adding a note or two in defence of the latter where we consider it still capable of holding its ground against such superior odds!—ED." (Prinsep apud Turnour, 1837, p. 1049 n. *)

[12] We take this for granted today, with the photographer constantly in attendance, but in 1837 photography was just being invented.

[13] I do not know why the Delhi column text, available for nearly four decades, had not been an adequate stimulus. Perhaps it was because even then, as now, the Indian alphabets were taught as some number of basic signs to which vowel markers are added, rather than in a two-dimensional array such as is used for Ethiopic (Section "Ethiopic" below); and it simply never occurred to a learned Sanskritist to arrange all the letters into such a scheme.

[14] For completeness, these are: the existence of bilingual inscriptions, which contain proper names, with one-to-one correspondences of letters within them.

[15] On these and other Iranian scripts see Skjærvø, 1996.

He then turns to the language to be used in interpreting the Bactrian legends on the reverses of the coins. Prinsep uses two pieces of evidence to argue that the vernacular language of the Bactrian kingdom was "Páli" (what today is called a Prakrit)[16]: First, a coin identifies 'The great king EUCRATIDES, son of HELIOCLES and KANLODICE.' The latter cannot be read as a Greek name KANLODIKE but does seem to correspond to

> the Sanskrit name कमलाधिका *Kamaládhiká*,—meaning 'superior to *Kamalá*, or VENUS, (alias 'fairer than the lily.') This name in the vernacular of the present day would be pronounced exactly as the Greek legend has it, *kauñla* a lily, *kauñládhikí*.
>
> EUCRATIDES then was the son of a Greek officer married to a lady of the country, whom we may set down as of Hindu parentage and language; and we may thence argue that a dialect mainly derived from the Sanskrit was then used in Bactria, or at least in the Panjab, as in the present day. (1838c, p. 638)

Second,

> we can now also adduce a *Páli* inscription in the old character procured by Captain BURNES from the northern side of the great chain of mountains, near *Badakshán*; … to say nothing of the *Páli* reverses of the Agathocles and Pantaleon coins from the same region.
>
> The natural inference is that we should seek the explanation of the legends on the reverses of the Bactrian coins rather through the medium of *Páli* or *Zend*, as I attempted in 1835, than as has been preferred by M. JACQUET of Paris, through the medium of Syriac and Chaldaic, with what success 1 have not the means of judging*. (Prinsep, 1838c, p. 638)
>
> *It will be proper here to notice that in 1836, M. JACQUET, obligingly forwarded to me a lithographed page of his readings of the Bactrian alphabet and names. In the modifications I now propose, however, I do not borrow one letter from his list, because in fact he has followed quite another track. … It was this which led to the expression of doubt of my own former alphabet, and to the just satire thereon in the Meerut Magazine.[17]

Prinsep had already recognized that the copy and the wax impression of the Eucratides coin were both faulty. He received better materials in time to add an erratum leaf to the same number of the journal acknowledging that the name

> became very clearly ΛΑΟΔΙΚΗΣ the genitive of a genuine Greek female name. … My speculations therefore of the Indian origin of EUCRATIDES' mother fall to the ground; and the reader is requested to pass over them. The unregal station of his parents still remains a matter of probability, on the grounds urged in the text ["it is evident from the absence of title and diadem that [HELIOCLES] was a private person"].[18]

---

[16] "Pali" is used nowadays for a Middle Indo-Aryan language associated particularly with Theravada Buddhism; it is first attested somewhat later than the considerable variety of Prakrits whose earliest records are from the time of Aśoka.

[17] *Meerut Universal Magazine* was a satirical publication that brought out at least four volumes beginning in 1835. It has not been possible to discover how it dealt with Prinsep's work.

[18] Thomas (1858, p. 2: 127), on the grounds of the immediately corrected misreading, omits both of Prinsep's arguments save for Eucratides' non-royal parentage, thus offering no account at all of Prinsep's insight that the Bactrian inscriptions ought to be interpreted in terms of Prakrit. Prinsep's was a far from unique example of a correct conclusion arrived at by way of a mistaken assumption.

It may further be noted that where Prinsep was able to use accurate Bactrian type in Calcutta, Thomas in England admittedly had to make do with approximations, thus further distorting Prinsep's work.

Having satisfied himself as to the language to be used in his interpretations, Prinsep confidently lays out the Bactrian, i.e., Kharoṣṭhi,[19] characters and their Greek equivalents, noting that the plethora of consonants corresponding to T and Δ is only to be expected in a Sanskritic language with a full complement of cerebrals (retroflexes) and aspirates. He includes a list of all the proper names found on the obverses with the Bactrian, i.e., Prakrit, equivalents from the reverses. His last step is to interpret all the additional words, such as *mahárájasa rájadhirájasa mahatasa* = ΒΑΣΙΛΕΩΣ ΒΑΣΙΛΕΩΝ ΜΕΓΑΛΟΥ 'great king of kings' (which in 1835 he had read as *malakao kakkao malako*).

Theories of Origin

The leading studies of the origins of writing in India are by von Hinüber (1990) and Falk (1993); see the review article by Salomon (1995).

Kharoṣṭhi

It had become clear by the end of the nineteenth century, with the abundance of comparative materials by then available, that the model for Kharoṣṭhi was to be sought in the Aramaic of the fifth century BCE (Bühler, 1898, pp. 92–114). Aramaic was the chancery language of the Achaemenid (Persian) Empire, used as we now know as far east as Bactra, the capital of ancient Bactria, modern Balkh, Afghanistan. Recently, a good-sized corpus has been discovered of letters written there in Imperial Aramaic (Naveh & Shaked, 2012). Their language and script are indistinguishable from those of near-contemporary documents from southern Egypt and from the Achaemenid administrative centers at Babylon and Susa (Daniels, in press). The Bactra materials come from the third quarter of the fourth century BCE—one of them even with a date in a regnal year of Alexander the Great equivalent to June 8, 324 BCE—and they show that Bühler's suppositions, based on minimal evidence, were correct.[20] Achaemenid dominion over the region began ca. 500 BCE, so that the opportunity to develop Kharoṣṭhi extended across some two and a half centuries until the script's first attestation ca. 250. The attested lettershapes exhibit sufficient difference from the model to suggest that the adaptation might have occurred rather

[19] The ancient names "Brahmi" and "Kharoṣṭhi" were properly assigned to the scripts only some decades later (Bühler, 1898, pp. 23–24; 1904, p. 2). I use the modern terms here so as to avoid having to list the multitude of designations employed by the various nineteenth-century authors.

[20] The existence of these documents shows that an argument by Harry Falk (1993) for the very late invention of Kharoṣṭhi, as summarized by Salomon (1995, p. 275b), does not hold: "Kharoṣṭhi, according to Falk, must have been created at one stroke at some time later, not before 325 B.C. (p. 104). The argument for this date is based on the theory that the new script could only have originated when the professional monopoly of the Aramaic scribe-bureaucrats of the Achaemenid empire (cf. pp. 78–81) had broken down in the wake of the Greek conquest."

earlier during that period than later, affording time for several generations of scribes to unintentionally alter the shapes to suit their own esthetic impulses.

Brahmi

Prinsep, though no philologist ("as I was the first to analyze those unknown symbols and shew their accordance with the system of the Sanscrit alphabets in the application of the vowel marks …, so I should be now rewarded with the completion of a discovery I then [1834] despaired of accomplishing for want of a competent knowledge of the Sanscrit language"; 1837a, p. 452), pooh-poohed a suggestion that what would eventually be called Brahmi might have descended from the Greek.[21]

During the middle of the nineteenth century, most Western scholars found it difficult to credit any suggestion of indigenous origin of so sophisticated a script (see section "Typology of Indic scripts" below) as Brahmi. Resemblances were found, on the whole not imaginary, to several Semitic scripts. The competing proposals were summarized and assessed by Isaac Taylor (1883, pp. 2: 304–323), who is followed here.[22] He lists three possibilities: Greek influence, Semitic origin, and indigenous origin, and takes them up in order.

In Taylor's day, not many data were available on either the Indian side or on the side of the potential sources or inspirations of the Indic scripts, so perforce much of his discussion concerns what little was known of contacts between the two sides. As for Greek influence (Taylor wrongly lists Prinsep among those of this opinion), it "may be dismissed without serious examination, as it is beset by difficulties, both chronological and phonological, of a most formidable nature" (p. 306),[23] despite the assertion that several Brahmi letters "may possibly have been derived from the Greek" (p. 303 n. 2).

Semitic Scripts

The nineteenth century saw suggestions of the derivation of both Kharoṣṭhi and Brahmi from three potential Semitic sources:

---

[21] "Mr. Stirling has suggested as a remarkable circumstance that many letters of the No. 1 type resemble Greek characters, and he instances the 'ou, sigma, lambda, chi, delta, epsilon, and a something closely resembling the figure of the digamma.' This resemblance is, however, entirely accidental, and the genus of the alphabet can I think be satisfactorily shewn to have no connection whatever with the Greek" (1834a, p. 117).

[22] In some ways Taylor's account of Indian writing has still not been superseded, but the attitude of many of his successors can be seen from a remark in a popular book from the turn of the last century: "Those who care to pursue a subject yielding to few in dryness will find it summarised in the tenth chapter" of Taylor's *Alphabet* (Clodd, 1900, p. 192).

[23] How different the conclusion might have been had those scholars known of the Iranian language Bactrian, which was written with the Greek alphabet (Sims-Williams, 2000–2007)!

> India, prior to the third century B.C., had been in commercial or political connection with Phœnicia, Babylonia, and Arabia: from any one of these regions the art of alphabetic writing might have been transmitted.
>
> Benfey's conjecture[24] that it came direct from the Phœnicians is open to fatal objections. The trade of the Phœnicians with India, which commenced in the time of Solomon, ceased as early as the year 800 B.C. *If the alphabet had been communicated at this early period a variety of Indian scripts would in all probability have sprung up during the long interval which elapsed before the time of Asoka, whereas, in the 3rd century B.C. a uniform alphabet prevailed over a vast Indian area.* (Taylor, 1883, p. 312 [emphasis added])

The italicized statement is immensely important, and Taylor may have lost sight of it elsewhere in the discussion.

Evidence was then decades in the future of the widespread use of Aramaic, both the language and the script, in the first millennium BCE in Mesopotamia and especially in the Achaemenid Empire, so the only variety of Aramaic writing that was available for consideration was still that of the Iranian inscriptions that Prinsep had tried to use (see at n. 15 above). Taylor attributes the suggestion of a "Persian" source of Brahmi to Burnell, who, however, suggests only that it be considered, and that there is insufficient evidence for deciding between Phoenician and Aramaic. Burnell does, though, point out a number of difficulties with the already popular derivation from South Arabian, known at the time as "Himyaritic"[25] (1878, pp. 8–9).

It was the Himyaritic connection, though, that won the day, as early as the first careful investigation of the question (Weber, 1856). In Weber's time, very little was known of South Arabian inscriptions, but within a generation, tables could be drawn up displaying similarities between their letters and those of Brahmi (Deecke, 1877). Bühler (1898, p. 54), however, rejects both Deecke's and Taylor's comparisons and sets out reasonable principles for seeking a source script[26]—and settles on Phoenician. There is, moreover, an additional difficulty besides those raised by Burnell and Bühler. The earlier authors set great store on the fact that a number of Brahmi letters (for sounds not found in Semitic: aspirates, retroflexes, sibilants) were elaborations of other letters for "simpler" sounds, and conveniently they arrived at 22 basic shapes—precisely the number of letters available in Phoenician

[24] In fact Benfey's "conjecture" was a mere throw-away line that can be rendered "The Phoenicians were certainly long before [Solomon's reign] intermediaries in the trade between India and the West, and as they, most probably, brought writing to India, they and perhaps Egypt itself moved various other cultural traits to and fro." (Benfey, 1864, p. 170).

[25] "The great difficulty is to show that the people of S. W. Arabia were in a position to furnish India with the elements of an alphabet so early as the fourth century B.C. … It must also be recollected that the Himyaritic alphabet did not mark the vowels, as its derivative, the Æthiopic alphabet does" (Burnell, 1878, p. 8). The first reliable synthesis of South Arabian materials was still well in the future (Hommel, 1893). A recent review of archeological work on Arabian–Indian contacts shows virtually no evidence of inscriptional interchange (Seland, 2014).

[26] To wit: "(1) The comparison must be based on the oldest forms of the Indian alphabet and actually occurring Semitic signs of one and the same period. (2) The comparison may include only such irregular equations, as can be supported by analogues from other cases, where nations have borrowed foreign alphabets. (3) The comparisons must show that there are fixed principles of derivation" (Bühler, 1898, p. 55).

or Aramaic (though not in "Pehlevi"). South Arabian, however, has 29 letters, including two additional dentals and no fewer than three for sibilants. Why would Brahmi have had just 22 basic forms?

Indigenous Origin

For Taylor and his contemporaries, "indigenous origin" meant invention of writing *ex nihilo*, by way of a pictographic beginning, as could be assumed for cuneiform (, for Egyptian,)[27] and for Chinese. No trace of any such background could be found in India.[28] There were, though, several scholars who insisted that no foreign influence be sought: Cunningham, first editor of the *Corpus Inscriptionum Indicarum*, proposed pictographic origins of the Brahmi characters in parts of the body, adducing Egyptian parallels (1879, pp. 52–61 and pl. XXVIII, reproduced as Fig. 3)[29]; Dowson (1881) describes the many ways in which Brahmi suits the Indic languages well but does not explain why it could not have had the Semitic origin favored by his contemporaries; Thomas (1858, p. 2: 43; 1871, pp. 420, 421f. n. 2) insists that because the cerebral letters[30] had been borrowed from the Dravidians (scil. the Dravidian languages),[31] Brahmi could not possibly have been based on any foreign, e.g., Semitic, script, though he offers no supporting argument (he shows that the aspirates were differentiated from the unaspirated letters within Brahmi).

Taylor's discussions of relationships between Kharoṣṭhi and Brahmi are inconsistent and occasionally counterfactual. On the one hand, the two scripts must have developed from two different Semitic sources (p. 2: 301 #1); on the other, Brahmi influence on Kharoṣṭhi can be detected (p. 2: 303 n. 2). Taylor was hampered by the uncertain or unreliable datings of the earliest inscriptional materials.

But once we have recognized that the existence of Kharoṣṭhi predates that of Brahmi by perhaps more than two centuries, it seems most reasonable not to go hunting among inscriptional Semitic sources for possible antecedents of Brahmi, but to accept that the most reasonable hypothesis is that Brahmi is a refinement of Kharoṣṭhi that was prompted by the sophistication of contemporary phonological

[27] To my way of thinking, writing was not independently invented in Egypt (Daniels, 2018, pp. 141–142).

[28] The Indus Valley signs were not yet known, except for a single seal that had been found at Harappa (see Fig. 3 herein, lower right corner), but had they been, the complete absence of intermediate forms would have been a powerful argument against them as a possible origin of Kharoṣṭhi or Brahmi.

[29] Several authors ridiculed the suggestion (e.g., Dowson, 1881, p. 115; Taylor, 1883, p. 307 n. 1). Cunningham read the Harappan seal (n. 26) as Brahmi (p. 61 & pl. XXVIII) but without realizing it is a seal to be impressed on a yielding surface and thus to be read in mirror-image.

[30] The distinction between "sound" and "letter" was not consistently made until comparatively recent times (Abercrombie, 1949).

[31] The borrowing is undoubted, not least because there are no retroflexes in the intimately related Iranian languages (except those in direct contact with Indo-Aryan ones). An overly clever argument has been made that the retroflexes are an inner-Indo-Aryan development (Hamp, 1996).

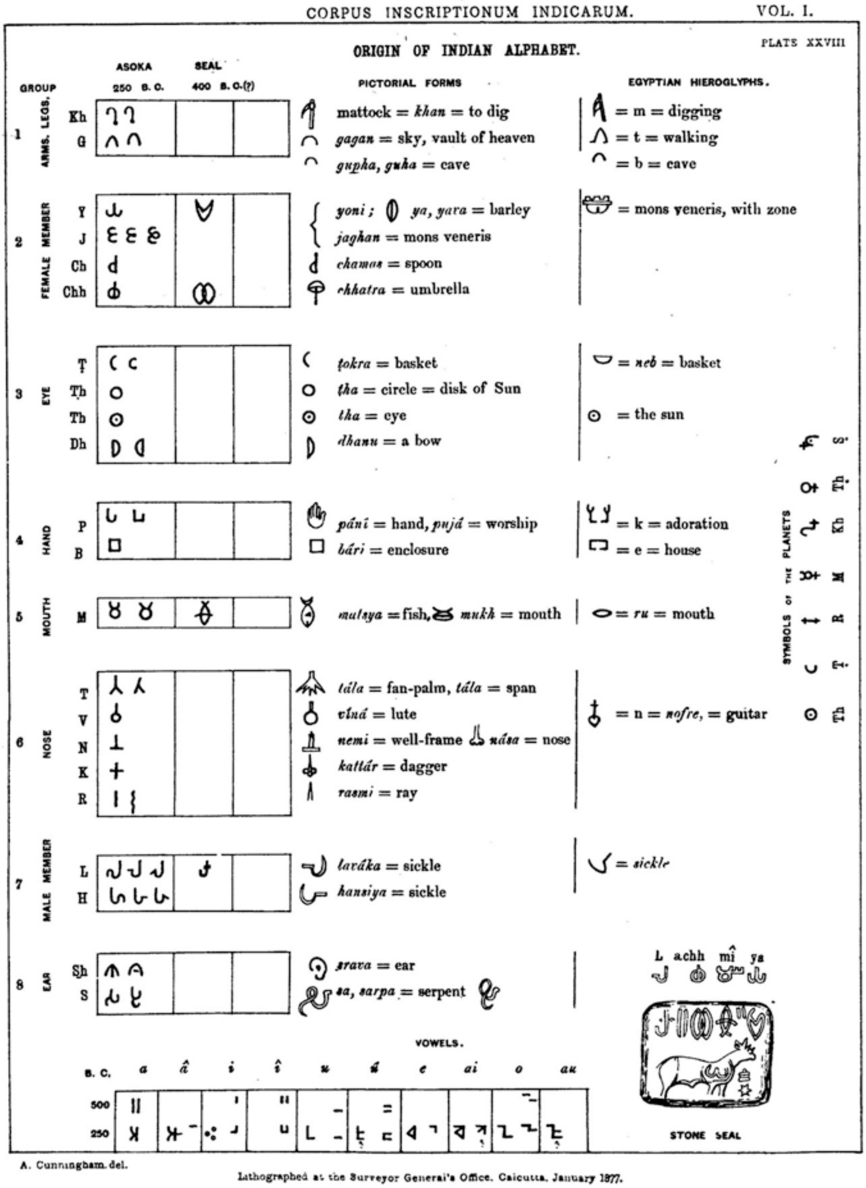


Fig. 3 Cunningham 1879, pl. XXVIII

analysis (Allen, 1953). This approach accounts for the features enumerated in section "Refinement in Brahmi". Salomon (1995, p. 278b), following Falk (1993), recognizes that

> there is every reason to think that Brahmi did not exist before the third century B.C., and that it was created then on the basis of a loose adaptation of one or more pre-existent Semitic scripts, with Kharoṣṭhi playing at least a partial role,

and in successive publications seems to come closer to endorsing the indigenous-origin view (cf. Salomon, 2007, p. 88).

Historico-Geographical Dispersal of the Indic Scripts

Just as with languages, scripts, too, can diverge if the communities using them are not in frequent (written) communication with each other. Nowhere is this clearer than in the Indian subcontinent, where tremendous diversity prevailed across the wide expanses. The divergence considered here concerns the shapes of the characters; divergences in the way the scripts represent their languages are sketched in section "Alterations in descendant scripts" below. Sivaramamurti (1948, pp. 56–154, figs. 20–65) drew up uniform diagrams of the development of each character down to their assumption of modern forms (though excluding the regional developments from Nāgarī, section "North Indic" below).[32] The outline used here is that of Sircar (1970–71, pp. 113–125, 131–133), which is largely followed in the publications of Richard Salomon.

Early Brahmi (from Aśoka's reign to the 1st c. BCE) was strikingly uniform across the entire area of its use, from Kandahar in the far northwest to Rampurya in the northeast and Siddhapura in the south.[33] Such uniformity is the hallmark of two situations: recent introduction of a script, and a strong central administration with effective scribal training. Middle Brahmi (1st c. BCE–3rd c. CE) was the script of the Kuṣāṇa age in the north and of the Later Śātavāhanas and the Ikṣvākus in the south. Late Brahmi (4th–6th c. CE) served the Gupta age in the north and the Early Pallavas and others in the south. Some Late Brahmi characters are so different from their Early Brahmi forebears that their ancestry can be verified only by consulting the intermediate Middle Brahmi forms. It was during these periods that the "serifs" at the tops of the characters began to display calligraphic exuberance, resulting in "triangle-headed" and "box-headed" scripts and later the "shell character."

In the southern region, the transitions between the stages of Brahmi occurred somewhat later. This is in accord with the principle of historical dialectology that peripheral linguistic areas tend to be more conservative, central areas more innova-

[32] Dani (1986, p. 9) remarks: "Efforts have been made to trace the evolution of the regional scripts and bring them down to modern times. An elaborate application of this principle is seen in [Sivaramamurti, 1948], in which he has tried to put on different pages the changing forms of every letter and then reduce them into abstract shapes apart from the context in which they appear. This method is no doubt suitable for showing the development of the forms in a museum gallery, and the book is at best a faithful record of that method." Cf. n. 34.

[33] Though an incipient distinction between North and South can be identified, in that southern inscriptions show that for the distinctive Old Tamil sounds *ṉa ṟa ḻa*, three extra letters were derived from Aśokan Brahmi letters.

tive. At the end of the periods named above, Late Brahmi developed into the ancestral Telugu/Kannada script.

In both north and south, comparative tables selected either chronologically or geographically make it easy to see how the sometimes highly baroque modern forms of the characters emerged gradually from perhaps originally unconscious tendencies in penwork—such as extending vowel markers to make them more distinctive—that then provided opportunities for esthetically inclined scribes to beautify the forms. Horizontal lines could bend down or up, vertical lines could be separated or joined to the characters they accompanied, loops could become closed or open as the pen flew (Dani, 1986, pls. III–XXII).[34] The succession can even be traced in a series of examples from contemporary scripts. The Brahmi mark for *e* is a horizontal line extending to the left from the top of the stalk of the consonant character 𑀓𑁂,[35]

के ਕੇ કે কে କେ கெ കെ කෙ కె ಕೆ,[36]

and that for *ā* is its opposite on the right, 𑀓𑀸:

का ਕਾ કા কা କା கா കാ කා కా ಕಾ.

Silk Road

The Silk Road was not a single highway leading merchants and travelers from Southwest and South Asia to East Asia. Rather, it was a network of roads skirting the desert of the Tarim Basin (north of Tibet) to both the north and the south. Over the past century or so, thousands of documents in many languages have been recovered from storage or hiding places that were abandoned during the vicissitudes of history (Hitch, 2010).

Kharoṣṭhi

In recent years, ample caches of Kharoṣṭhi manuscripts, recording Gandhari Prakrit (and occasionally Sanskrit and Tocharian), have surfaced. They show that the script was used more extensively than had been realized, on both the northern and southern routes of the Silk Road (Glass, 2000).

[34] Each of Dani's plates occupies four octavo pages, displaying a dozen or so complete alphabets from a specific chronological and geographic range. His remit extends to the eighth century and as far as the Sanskrit and earliest native inscriptions of insular and mainland Southeast Asia.

[35] The examples are *ke* and *kā* taken from, in order left to right, Devanagari, Gurmukhi, Gujarati, Bengali, Oriya, Tamil, Malayalam, Sinhala, Telugu, and Kannada.

[36] This sequence shows how the preposing of certain vowel marks, which exercises some psychologists of reading (Kandhadai & Sproat, 2010), came about.

Brahmi

The development of Brahmi in Inner Asia is summarized by Sander (2005, p. 135). As Sanskrit came to replace Prakrit, Brahmi replaced Kharoṣṭhi, beginning with a Kuṣāna form of Middle Brahmi, which yielded to a Gupta type of Late Brahmi. Early Turkestan Brahmi served Sanskrit and Tocharian B (Kuchean) beginning in the fourth to fifth centuries on the northern Silk Road, and Sanskrit and (Old) Khotanese on the southern Silk Road.[37] The distinctive Turkestan varieties of Brahmi then began to appear: In the north, what used to be called "Slanting Gupta," which Sander renamed North Turkestan Brahmi, was also used for Tocharian A (Agnean) and for Tumshuqese in the seventh century, and for (Iranian) Sogdian and the Turkic language Uyghur in the 9th. This variety remained in use until the fourteenth century. In the south, South Turkestan Brahmi (formerly "Upright Gupta") began to appear in the seventh century, still writing Sanskrit and Khotanese, and in the ninth to tenth centuries Chinese was also found; after that date, the script was no longer used.

North Indic

We continue to follow Sircar (1970–71, pp. 117–123). In northern India, from the sixth to the tenth centuries, the development of Late Brahmi called Siddhamātr̥kā was widely used.[38] It provided the basis for many regional developments: in the northeast the Tibetan (7th c., but cf. section "Tibetan"), in the northwest the Kashmiri or Sāradā (8th c.), and throughout the north the Nāgarī (10th c.). The Gurmukhi script (see sections "Notation of tone: Gurmukhi, Burmese, Thai, Lao" and "Rationalization of vowel notation: Gurmukhi" below) is an adaptation of Sāradā, and the similar Lahṇḍā is a commercial cursive script. In the northeast, a derivative of Siddhamātr̥kā called Gauḍī (10th–14th c.) yielded the group of scripts that eventually (both 14th c.) jelled into Bengali and Oriya, which last took the "Nepalese hook" to the extreme, turning it into the characteristic "umbrella" (other regional varieties did not become typographic standards).

In more modern times, there are certain differences between western and eastern varieties of Nāgarī. Devanagari is the modern outcome, used for Hindi and Marathi *inter alia*, and has become the default script for Sanskrit. Gujarati is a "slightly cursive" version (missing the connecting headstroke) in the west, and Kāyethi is a similar cursive in the east. Moḍī of the Marathi-speaking area is an "extremely cursive" variety of western Nāgarī.

[37] Tumshuqese (attested until the fourteenth century) and Khotanese are the (Iranian) Saka languages.

[38] This style is known as Siddham to the Chinese and to modern calligraphers (Stevens, 1988).

Tibetan

The need to write Tibetan seems to have accompanied the translation of Buddhist texts from Sanskrit. Sircar (1970–71, p. 116) took Tibetan to be a seventh-century derivative of Siddhamātr̥kā, but the most recent investigation finds the Tibetan script to have been based on a version of Nāgarī used in Nepal in the eighth century (van Schaik, 2011). Kubla Khan commissioned from a Tibetan lama in his court a script that could write all the languages of the Mongol Empire. Named hPags pa for its creator, it was introduced in 1268 and served, albeit mostly for Chinese, for just a century; the letter shapes are based on Tibetan.[39]

South Indic

In the south (Sircar, 1970–71, pp. 123–125), Late Brahmi developed into Telugu/Kannada on the one hand and (via Pallava) Grantha/Tamil on the other; there was no intermediate stage parallel to the Siddhamātr̥kā of the north. Early Telugu/Kannada script was used in the sixth to tenth centuries, and the Middle varieties in the 10th–14th, after which the two began to diverge.

Still farther south, the Grantha script developed from Brahmi for literary use and Tamil Brahmi for everyday use. Eventually the Tamil script replaced the Grantha, but the additional letters in modern Tamil writing needed for Sanskrit are called the "Grantha letters" (Steever, 1996). A third early derivative of Brahmi used for Tamil and Malayalam was Vaṭṭeḻuttu, which, not being well attested, led to some confusion among the earlier epigraphists. In the fourteenth to fifteenth centuries it developed into the extremely cursive Koleḻuttu, while about the same time the Malayalam script proper developed from Grantha.

Early Brahmi may have been carried to Sri Lanka as early as Aśoka's time, and the Sinhalese script seems to have developed in place over some two millennia, thus apart from the subsequent split into North and South Indic scripts.

Southeast Asia

The script history of mainland Southeast Asia has not, to my knowledge, received a comprehensive account; the references in the most easily accessible sketch (Court, 1996) comprise almost exclusively general histories of the region. In grossest outline, the earliest known inscriptions, found in present-day Vietnam, Thailand, and Malaysia, are in Sanskrit (and also Pali and Tamil), dating to the fourth century CE or so, and the script appears to be the Pallava of southern India. Subsequently,

[39] hPags pa letters probably inspired the shapes of the letters of the Korean alphabet devised in 1446 (Ledyard, 1997).

Cham, Khmer, and Malay began to be written, and local scripts began to diverge. Siddhamātr̥kā is again implicated, in connection with Buddhist influence; by the late ninth century, Hindu inscriptions are found at Angkor Wat in a distinctively Khmer script.

The mainland descendants of Pallava appear to fall into three branches corresponding to three ancient languages, Mon, Khmer, and Champa. By around 1500, the Mon script had developed into the Burmese. The Thai and Lao group branched off from Khmer quite early on, and the three languages now have distinctive standard scripts. The modern Cham script serves a language that has no modern nation.

Island Southeast Asia is in comparatively better scholarly shape,[40] with a compendious survey (de Casparis, 1975) and ongoing research by Miller (2010, 2014a, 2014b). Miller innovatively considers typological and orthographic traits, such as reduplicated letter names and the order in which vowels are listed, and the use of a "killer" mark for a vowelless consonant, above and beyond the shapes of letters. According to de Casparis, the earliest inscriptions in Indonesia are in Pallava script, in two recognizable stages (beginning in the mid fourth and mid seventh centuries), the latter of which gave rise to the Kawi (used for Old Javanese beginning in the mid eighth century) that is ancestral to Javanese script. Javanese and other regional scripts are seen from the mid thirteenth century.

Miller contributes the following summary of his current understanding[41]: The Pallava-based Kawi yielded the Javanese-Balinese group, used also for Malay in Sumatra. But most of the variety of Indonesian scripts springs from contact with Gujarati merchants, who introduced informal Nāgarī script to Malay-speaking eastern Sumatra. Adapted to Malay, it was taken up in simpler form by the Bugis of southwest Sulawesi around 1400, spreading from there to the northern Philippines (e.g. Bisayan and Tagalog) before the advent of the Spanish in 1521. Meanwhile, the Sumatran script, after undergoing several innovations, split into a Batak northern group and a southern group including Lampung in the far south, *Surat Ulu* 'upstream script' of Rejang around the Musi River, and *Incung* to its northwest at Mt. Kerinci. The much simplified Bugis-Makassarese script gave rise in the 1600s to derived scripts Ende and Bima in Flores and Sumbawa, respectively (two islands directly south); and, around 1600, an eclectic hybrid of South Sumatran and Javanese letters with Bugis-Makassarese script gave rise to "Old" Makassarese script.

[40] Taylor (1883, pp. 2: 340–341, 359–361) already had available the vast array of alphabets collected by Holle in 1877 (1999) and made a valiant, though largely wrong, attempt to organize them historically.

[41] Reflected also in Daniels, 2018, p. 193, Map 5.

Ethiopic

In the last centuries BCE, Sabaean administration crossed the mouth of the Red Sea to the Horn of Africa and left some inscriptions in their South Arabian language, using the South Semitic consonantal script (an *abjad*, section "Terminology"). By the fourth century CE, the script was used for the local Semitic language Gɨʿɨz; and in the middle of that century, a remarkable thing happened: King Ezana converted to Christianity—and his inscriptions before that date were with the unvoweled inherited script; but after that date, they were fully vocalized. The vowels are not written with separate letters, as would be likely if they had received the idea of writing vowels from Coptic or Greek missionaries. They are not written with dots or other signs (as if they could have been taught by time-traveling Syriac missionaries—who in the 4th c. had only *matres lectionis* [section "Nature of the Aramaic abjad with *matres lectionis*"] for vowel notation). Instead, they are written *in the Indic fashion* with lines, hooks, and circles appended to the letters, or with changes in letter shape.[42] It is unlikely that the same distinctive mechanism for writing vowels was independently invented: it is probable that the principle crossed the Arabian Sea with sailors, or perhaps missionaries, from a Christian community on the Indian west coast who brought some, albeit incomplete, knowledge of writing a language of India (Daniels, 1992).

Typology of Indic Scripts

For a century and a half (1838–1988), it was understood that there are three types of writing systems[43]: alphabets, syllabaries, and logographies.[44] They denote respectively *segments* (consonants or vowels, later taken to be equivalent to *phonemes*), syllables, and words.[45] So long as the Indic scripts remained a minor or nonexistent concern[46] of graphonomy (the study of writing systems), those types were adequate to classify known writing systems. They covered Hebrew and Greek, Japanese and Akkadian, Chinese and Sumerian.

---

[42] Christopher Miller (pers. comm.) detects some similarity between some of the vowel markings in Ethiopic and some in (Middle) Brahmi.

[43] The tripartite typology appears to have first been proposed in a work (Du Ponceau, 1838) that remained, however, unremarked until it was discovered a century later (Chao, 1940). The typology became graphonomic orthodoxy in the first scientific work on writing systems (Taylor, 1883).

[44] The term "ideogram" was often seen in place of *logogram*, but the incorporation of the morpheme "ideo-" wrongly suggests that such characters denote "ideas" rather than elements of language.

[45] "Words" is a pretheoretic term; what logograms denote are *morphemes*, and careful writers use *morphogram* rather than *logogram*.

[46] Cf. n. 22.

Terminology

However, the work that became the standard reference on writing (Gelb, 1952) stretched the tripartite classification beyond the breaking point (Daniels, 2000) with the Principle of Unidirectional Development (P.U.D.), claiming that "in reaching its ultimate development writing … must pass through the stages of logography, syllabography, and alphabetography in this and no other order" (p. 201). In order to make that come out right, Gelb had to insist that the Phoenician "alphabet," from which uncontroversially the Greek alphabet developed, is a syllabary; and that the Indic "syllabaries," which also developed from a Phoenician or related alphabet, are alphabets.

This is at best counterintuitive and at worst raises difficult problems with regard both to Phoenician (and its relatives such as Hebrew and Arabic) and to Indic. A solution was offered (Daniels, 1990), positing not three but five basic types and jettisoning the P.U.D. altogether. An *abjad* (from an Arabic word) is a script in which each letter designates only a consonant; the pure examples are Phoenician and the Sabaean or Himyaritic of southern Arabia. An *abugida* (an Ethiopic word)[47] is a script in which each character denotes a consonant followed by a basic vowel (usually /a/, whatever its phonetic realization may be) and the other vowels are denoted by additions to or modifications of the basic consonant symbols.[48]

In traditional Indian grammar, "[o]ne result of [the] phonological interdependence of syllable (*akṣara*) and vowel (*svara*) is that the term *akṣara* is frequently extended to mean 'vowel' [a]nd thence also, like *varna*, to mean simply 'letter'" (Allen, 1953, p. 80 with n. 9). The term *mātrā* 'quantity' originally designated what in contemporary linguistics is termed a *mora* ("something of which a long syllable consists of two and a short syllable consists of one"; McCawley, 1968, p. 58 n. 39) but was misguidedly extended to refer also to light and heavy syllables (Allen, 1953, p. 87), but not to vowels themselves. More recently, "akshara" has come to be used for written consonant-or-initial-vowel-symbol and "matra" for written other-vowel-marker. Writing had no place in Pāṇinian grammar.

---

[47] If the orders originally used for Kharoṣṭhi consonants and vowels (Salomon, 2006) had been known in 1988, this type might have been called the *arepiconu* accordingly.

[48] From time to time, scholars, particularly of Indian languages, noticed that the Indic scripts differed from other syllabaries and devised a number of terms for the type: neosyllabary, semisyllabary, pseudo-alphabet, and, most popularly, *alphasyllabary*. I reject all these terms because they suggest that the abugida is somehow an intermediate type standing between alphabet and syllabary. It is not.

Nature of the Aramaic Abjad with **Matres Lectionis**

As seen above (Section "Kharoṣṭhi"), there can be little doubt that Kharoṣṭhi is an adaptation of the Aramaic abjad brought to the East (i.e., to the northwestern frontier of India) by the Achaemenid administration. Contact could have been as early as the early fifth century BCE, not long after 500 BCE. Even in the earliest known examples of Aramaic, dating to the tenth century BCE, the way the orthography uses the abjad has differed from the Phoenician way, and by the fifth century the system had developed robustly. Certain letters for consonants—*w y ʔ h*—were *also* used to indicate the presence of long vowels and diphthongs, *ū ī aw ay*, but *ā* only at the end of a word. In this use, such letters are called *matres lectionis* ('mothers of reading' in Latin, translating a term used by medieval Hebrew grammarians). Much the most common vowel within a word, though, was *ă*, and this remained unmarked.

Influence of Pāṇinian Grammar on the Invention of Kharoṣṭhi

The Indian grammatical tradition was well developed by the time Aramaic writing came to the pandits' attention (however uncertain the dating of Pāṇini himself may be), and it cannot have been much of an intellectual stretch to recognize the utility of a consonantal writing system in which each basic character represented a consonant followed by the ordinary vowel *a* (as the Aramaic model appeared to do) with an added mark for each of the other four vowels of the language to be written, namely, Gandhari Prakrit (Salomon, 2002).

Refinement in Brahmi

When we recognize that Brahmi is a refinement of Kharoṣṭhi and not an independent invention (Section "Indigenous origin"), we can explain several of the divergences from its putative models.

- Most striking is the severely geometrical shapes of the letters. Rarely is the adoption of a set of characters from a foreign model accompanied by so great a change in their shapes from the model, but it is reasonable to expect someone who takes on the task of refining an existing system already in use for their language to "neaten it up" a bit.
- Change from sinistrograde to dextrograde direction is more unusual: typically it happens when a model script is used *boustrophedon* 'as the ox plows', alternating direction with each new line. But in this case it may simply be that a right-handed master scribe felt that it would be more practical to have the line of writing not obscured by the arm as the line emerged from the pen or stylus.

- Third, Brahmi includes a separate character for each phrase-initial vowel, and it provides for both short and long vowels. Was vowel length not significant in Gandhari?[49] or did it carry a greater functional load in other Prakrits? Could they even have cast an eye toward writing Vedic or Sanskrit, even though the traditional texts were to be transmitted only orally for centuries more?
- The distinctive treatment of /r/ also appears to relate to some Prakrit traits.[50]

Other features of Brahmi seem more suited to Sanskrit, such as the development of conjunct consonants—clusters were rare to nonexistent in Prakrits, and the very diversity in their notation among the descendant Indic scripts (Daniels, 2008, pp. 298–299) shows that they were barely present in early Brahmi (compare the systematic similarities across scripts of all but the most innovative vowel notations, section "Historico-geographical dispersal of the Indic scripts" above). The characteristics of Devanagari descend from the patterns set in the earliest days of Brahmi.

Alterations in Descendant Scripts

Suited though Devanagari is to notating Sanskrit—and the other scripts of India generally incorporate features, not needed for their own languages, that serve when Sanskrit is written with them—over the centuries, changes set in to better fit local scripts to their languages. They are taken up here roughly chronologically.

Additional Letters: Tamil, Tocharian, Khotanese, Etc.

As Brahmi expanded its range both to the southern parts of the Indian Subcontinent and into Inner Asia along the Silk Road, additional letters were needed to accommodate non-Indic languages for which it was adapted: Old Tamil (*ṟa ḻa ṉa*); Turkic Uyghur (*qa ɣa ẓa źa ða ʣa*); and, curiously, several Tocharian consonants for which letters already existed but new symbols were devised for consonants with *ə* (*kə tə nə pə mə rə lə śə ṣə sə*); additionally, a vowel marker was added for *ə* with consonants (Sander, 1986). The shapes of these additional letters are shown by Daniels (2018, p. 72, Table 5.8). The extensive additions for the Iranian languages Khotanese and Tumshuqese (and one known ms. in Sogdian) are set forth by Hitch (1989).

[49] Vowel length may have been lost in Gandhari before Kharoṣṭhi was devised. A possible indication that there was no such distinction is that when, eventually, a mark was added for **ā*, there was none for the other long vowels, suggesting that what was recognized was a distinction in quality ([ʌ] or [ɔ], perhaps, vs. [a]) rather than a distinction in length (R. G. Salomon, pers. comm., 30 December 2016).

[50] Cf. n. 53.

Loss of Conjuncts: Tamil

The development of Tamil orthography can be followed in the epigraphic and numismatic record (Mahadevan, 1990). Old Tamil was first written with ordinary Aśokan Brahmi, but soon the abugida principle was given up, with even short /a/ indicated with a mark and the plain consonant with the bare consonant letter. This was then given up in favor of a dot to mark vowelless consonants; for a time this dot also indicated a long vowel when placed on a vowel-marked letter, but eventually vowel length was again indicated with differing marks, and only unvocalized consonants with the dot. Tamil orthography thereafter had no need of conjunct forms of consonants. Compare Telugu మహాత్మా *Mahātmā* (త్మా is త *t* + the subscript form of మ *m* + ◌ా *ā*) with Tamil மகாத்மா.

Loss of Null Vowel: Ethiopic

In the case of Ethiopic, it may be more that those who brought some knowledge of vocalizing texts across the Arabian Sea were not fully versed in writing their own languages, or may not have realized that the Gɨʿɨz they were helping with also had consonant clusters and long consonants; but the Ethiopic script has no provision for either. The spelling (ይ)<u>ስ</u>ብር፡ can be read either *s̲ɨbɨr* 'break!' or *yɨs̲bɨr* 'may he break', with the vowelless consonant *s* written with the character for *sɨ*; መሰለ፡ can be either *masala* 'he resembled' or *massala* 'he compared'.[51] *Massala* cannot be written *መስሰለ፡.

Separation of Syllables: Tibetan

The languages for which the abugida was devised, Indo-Aryan languages, where there is no set limit on the number of syllables in a morpheme (whether base or inflection), were and are *inflecting* languages, where grammatical information is expressed in suffixes for the most part, suffixes that can communicate more than one quantum of information—such as "first person plural." Tibetan, however, is an *isolating* language like its distant relative Chinese: morphemes are usually one syllable in length, and there are normally no *sandhi* effects at syllable boundaries. When a variety of Nepali Nāgarī was adapted to Tibetan, syllable/morpheme boundaries came to be marked with a dot at the right shoulder of the rightmost consonant letter in each syllable.

[51] ሰ *sa* has a vertical stroke at the top, ስ *s(ɨ)* has a slanted stroke at the top. ሰ *sa* has rounded shoulders, ለ *la* has sloped shoulders.

Enlargement of Vowel Markers to Letters: hPags pa

The hPags pa script retains squared-up forms of the Tibetan letters, and each consonant letter represents the consonant followed by *a*, but the indicators of the vowels *i*, *e*, *o*, and *u* are no longer attachments to the consonants (above or below, in Tibetan), but separate characters—letters a bit smaller than the consonant letters—placed after the consonant letter they modify. hPags pa is written vertically (the columns read left to right), so each vowel appears below its consonant.[52]

Notation of Tone: Gurmukhi, Burmese, Thai, Lao

The existence of lexical tone distinctions in the Punjabi language is a fairly recent discovery (set forth in Gill & Gleason, 1969). Voiced aspirate consonants do not occur in Punjabi, and in the Gurmukhi script the inherited letters for them and for ‹h›, depending on their position in a syllable, are used to mark the three tones—high, mid, low (Gill, 1996).

Several Southeast Asian languages adopted or adapted for vernacular use the scripts used locally for Sanskrit. Over the centuries, various phonological phenomena developed in various ways into lexical tone, and the copious resources of the parent script are employed in different ways to record those tones.

Burmese has three tones (creaky [short], low, and high [both long]). For the three vowels that in the ancestral script had short and long versions, *a i u*, the old short-vowel markers are used for the creaky tone and the old long-vowel markers for the other two, followed by *visarga* for high tone. The mid vowels *e* and *o* use the plain markers for low tone and *visarga* again for high tone, but a single circle beneath the consonant symbol for creaky tone. The tone marking of an additional pair of vowels, low mid ɛ and ɔ, is less systematic (Wheatley, 1996).

In the Thai and Lao scripts, closely related to each other, tones are indicated by complicated interactions among three consonant classes (class 1 represents the ancestral voiceless unaspirated stops, class 2 the ancestral voiceless aspirates, and class 3 most of the other consonants [though some of the non-occlusives are distributed between classes 1 and 2 as well]), open vs. closed syllables, short vs. long vowels, and a pair of tone diacritics (Diller, 1996; Haas, 1956).

[52] It is this property of hPags pa that brought to light a significant difference between *alphasyllabary* and *abugida*, namely, that the former is a *formal* category, the latter a *functional* category (Daniels apud Daniels & Bright, 1996, p. 4 n. *). The putative inventor of the term *alphasyllabary* explained that it does not apply to hPags pa because the vowels are not indicated by appendages to the consonant letters (Bright, 2000). The inventor of the term *abugida* explains that it applies to hPags pa because the consonant letters carry an inherent vowel.

Reuse of Consonant Distinctions for Vowel Quality: Khmer

Khmer, the national language of Cambodia, has a fairly limited inventory of consonants (there is no voiced/voiceless distinction, for instance), but an impressive array of distinctive vowels. Vowel qualities are indicated by having each vowel marker denote two different sounds according as they follow ancestral voiceless or voiced consonant letters. The set of vowel markers is, moreover, augmented by additional shapes and supplemented by several diacritics (Schiller, 1996).

Loss of Consonant Distinctions: Thai

Thai, one of the most extremely historical orthographies in the world—having persisted longer without reform than probably any other orthography except Tibetan (Brown, 1985)—has no fewer than 44 consonant letters for just 27 consonant phonemes (five different letters are pronounced [kʰ]); to a limited extent the duplication is exploited in the tone notation, section "Notation of tone: Gurmukhi, Burmese, Thai, Lao" above (Lao, after extensive reform, has 27 letters for 20 consonant phonemes).

Addition of Honorific Forms: Javanese

Javanese is perhaps the most extreme example of diglossia, the situation in which the vernacular and high-status varieties of a language differ more than is typical of the world's languages (Ferguson, 1959). Unlike most languages, though, Javanese reflects the underlying social situation in its orthography. There is a set of three elaborate symbols that indicate the relative status of petitioner and addressee—one for an inferior addressing a superior, one for communication between equals, and one for a superior addressing an inferior. There is also a set of letters—though it does not include one corresponding to each of the letters in the script—functioning somewhat like the capital letters in Greek, Roman, and Cyrillic alphabets: if one or more from this set appears anywhere in a word, the word is marked as having a special status, as with initial capitals in the Western alphabets (Kuipers & McDermott, 1996, pp. 477–480).

Rationalization of Vowel Notation: Gurmukhi

In most of the Brahmi-descended scripts, there is little relationship between the symbols for word-initial vowels, save that those for long vowels are usually based on those for short vowels, e.g. Devanagari अ आ *a ā* इ ई *i ī* उ ऊ *u ū* ए ऐ *e ai* ओ औ *o aw*. In the Gurmukhi script for Punjabi, however, the symbols for all the low vowels and diphthongs (ਅ ਆ ਐ ਔ *a ā ai au*) are built on one base ਅ, those for the high front vowels (ਇ ਈ ਏ *i ī e*) on another ੲ, and those for the back rounded vowels (ਉ ਊ ਓ *u ū o*) on a third ੳ, with the differentiators being the vowel marks: with ਕ *ka*, ਕਾ ਕੈ ਕੌ ਕਿ ਕੀ ਕੇ ਕੁ ਕੂ (ਕੋ).

Script Reform: Malayalam

Until the middle of the twentieth century, the Malayalam script employed a wide and beautiful variety of vowel marks, and conjunct forms to notate consonant clusters. These included forms in which the same vowel—*u* or *ū*—was denoted with a considerable variety of calligraphic elaborations; in conjunct formations, pairs of consonants were arranged either vertically or horizontally. This elaboration proved difficult to accommodate on typewriters, which had available a limited number of possible different shapes and could handle vertical adjustments only by interrupting the flow of speed-typing. Reforms were instituted to make all *u* and *ū*, and even *r* in clusters,[53] distinct in-line characters: പു *pu*, പൂ *pū*. പൃ *pr̥*, പ്ര *pra*, and to use *virama* (the "killer" stroke) after most unvoweled consonants so that many, though not all, of the elaborate conjunct characters were rendered obsolete.[54] There survive a diacritic for doubling, ച്ച *cca*, ബ്ബ *bba*, but also duplication for doubling, പ്പ *ppa*; as well as some of the combinatory clusters, among them ക + ക = ക്ക *kka*, ണ + ട = ണ്ട *ṇṭa*, ഞ + ഞ = ഞ്ഞ *ñña* (Mohanan, 1996).

Loss of Abugida Nature: Lao

Finally, the Lao script is no longer an abugida. Every *a* is marked with a sign of its own used like the signs for all the other post-consonantal vowels.

Epilogue (DLS) *The geo-historical and typological diversification of scripts described above by Daniels mirrors the remarkable linguistic, ethnic, and religious diversity of the Indian subcontinent and its neighbors in Inner Asia, Sri Lanka, and Southeast Asia. In India alone, several hundred different native languages are spoken, with 22 recognized as "official languages," many of which have their own scripts (versus the handful of European alphabets—Roman, Greek, Cyrillic, Georgian, Armenian). This linguistic and graphonomic diversity makes literacy-learning in the Brahmi-derived scripts of Asia unique in another way: they are rarely learned in isolation. In most parts of Europe and North America, literacy-learners are typically monolinguals learning to read and write a single alphabet. By contrast, in most parts of South Asia, many speak their local (often unwritten) dialect at home, but also learn a regional standard written in a script that may differ from that of the standard national language, as well as a further additional language, English, written in the Roman alphabet. This multiplicity of language and script, so unlike the minority bilingual English Language Learning (ELL) situation in the USA, has far-reaching implications for theorizing about literacy-learning.*

---

[53] In most modern Indian scripts, cluster-initial *r* is written with a hook above the top right corner of the following consonant(s), cluster-final *r* with a stroke beneath: Devanagari र्न्त *-rnta*, न्त्र *-ntra*. The character र *ra* is the narrowest in the set.

[54] Few of the older forms have been included in contemporary Unicode-compliant Malayalam fonts, and so cannot be shown here, but a number of them can be seen in (pre-Unicode) Mohanan, 1996.

References

Abercrombie, D. (1949). What is a 'letter'? *Lingua, 2*, 54–63.

Allen, W. S. (1953). *Phonetics in ancient India* (London Oriental series 1). London, UK: Oxford University Press.

Arbuthnot, A. J. (1896). Prinsep, James (1799–1840). In S. Lee (Ed.), *The dictionary of national biography* (Vol. 16, pp. 395–396). London, UK: Oxford University Press.

Benfey, T. (1864). Indien und Aegypten. *Orient und Occident, 3*(1), 168–170.

Bright, W. (2000). A matter of typology: Alphasyllabaries and abugidas. *Studies in the Linguistic Sciences* (Urbana)*, 30*, 63–71.

Brown, J. M. (1985). *From ancient Thai to modern dialects and other writings on historical Thai linguistics*. Bangkok, Thailand: White Lotus.

Bühler, G. (1898). *On the origin of the Indian Brāhma alphabet, together with two appendices on the origin of the Kharoṣṭhī alphabet and the origin of the so-called letter-numerals of the Brāhmī* (2nd ed.). Strasbourg, Germany: Trübner.

Bühler, G. (1904). *Indian paleography* (Grundriss der Indo-Arischen Philologie und Altertumskunde, vol. 1 section 2). (J. F. Fleet, Ed., & G. Bühler, Trans.) *Indian Antiquary*, Appendix.

Burnell, A. C. (1878). *Elements of South-Indian palæography from the fourth to the seventeenth century A.D* (2nd ed.). London, UK: Trübner.

Burnouf, E. (1833). *Commentaire sur le Yaçna.* Paris, France.

Burt, T. S. (1834, March). A description, with drawings, of the ancient stone pillar at Allahabad called Bhim Sén's Gadá or Club, with accompanying copies of four inscriptions engraven in different characters upon its surface. *Journal of the Asiatic Society of Bengal, 3*(3), 105–113.

Chao, Y. R. (1940). A note on an early logographic theory of Chinese writing. *Harvard Journal of Asiatic Studies, 5*(2), 189–191.

Clodd, E. (1900). *The story of the alphabet*. New York, NY: Appleton.

Coulmas, F. (2009). Evaluating merit — The evolution of writing reconsidered. *Writing Systems Research, 1*(1), 5–17. https://doi.org/10.1093/wsr/wsp001.

Court, C. (1996). The spread of Brahmi script into Southeast Asia. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 445–449). New York, NY: Oxford University Press.

Cunningham, A. (1879). *Inscriptions of Asoka* (Corpus Inscriptionum Indicarum 1). Calcutta, India: Office of the Superintendent of Government Printing.

Dani, A. H. (1986). *Indian palaeography* (2nd ed.). New Delhi, India: Munshiram Manoharlal.

Daniels, P. T. (1988). 'Shewing of hard sentences and dissolving of doubts': The first decipherment. *Journal of the American Oriental Society, 108*, 419–436.

Daniels, P. T. (1990). Fundamentals of grammatology. *Journal of the American Oriental Society, 110*, 727–731.

Daniels, P. T. (1992). Contacts between Semitic and Indic scripts. In A. Harrak (Ed.), *Contacts between cultures: Selected papers from the 33rd International Congress of Asian and North African Studies, Toronto, August 15–25, 1990*, vol. 1: *West Asia and North Africa* (pp. 146–152). Lewiston, NY: Edwin Mellen.

Daniels, P. T. (2000). Syllables, consonants, and vowels in West Semitic writing. *Lingua Posnaniensis, 42*, 43–55.

Daniels, P. T. (2008). Writing systems of major and minor languages. In B. Kachru, Y. Kachru, & S. N. Sridhar (Eds.), *Language in South Asia* (pp. 285–308). Cambridge, UK: Cambridge University Press.

Daniels, P. T. (2018). *An exploration of writing*. Sheffield, UK: Equinox.

Daniels, P. T. (in press). Aramaic Documents from Achaemenid Bactria: Connections to the West—and the East—and the Future. In A. Kaplony & D. Potthast (Eds.), *From Barcelona to Qom: Documents from the Medieval Muslim World* (Islamic History and Civilization). Leiden: Brill.

Daniels, P. T., & Bright, W. (Eds.). (1996). *The world's writing systems*. New York, NY: Oxford University Press.

Daniels, P. T., & Share, D. L. (2017). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading, 22* (2018), 101–116. https://doi.org/10.1080/10888438.2017.1379082.

de Casparis, J. G. (1975). *Indonesian palaeography: A history of writing in Indonesia from the beginnings to c. A.D. 1500* (Handbuch der Orientalistik 3.4.1). Leiden, Netherlands: Brill.

Deecke, W. (1877). Ueber das indische Alphabet in seinem Zusammenhange mit den übrigen südsemitischen Alphabeten. *Zeitschrift der Deutschen Morgenländischen Gesellschaft, 31*, 598–612.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu, HI: University of Hawaii Press.

Diller, A. (1996). Thai and Lao writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 457–466). New York, NY: Oxford University Press.

Diringer, D. (1968). *The alphabet: A key to the history of mankind* (3rd ed., R. Regensburger (Ed.), 2 vols. [text and plates]). New York, NY: Funk & Wagnall's.

Dowson, J. (1881). The invention of the Indian alphabet. *Journal of the Royal Asiatic Society, NS 13*(1), 102–120.

Du Ponceau, P. S. (1838). *A dissertation on the nature and character of the Chinese system of writing in a letter to John Vaughan, esq.* (Transactions of the Historical and Literary Committee of the American Philosophical. Society 2, pp. vii–xxxii, 1–123). Philadelphia, PA: M'Carty and Davis.

Falk, H. (1993). *Schrift im alten Indien: Ein Forschungsbericht mit Anmerkungen* (ScriptOralia 56). Tübingen, Germany: Narr.

Farmer, S., Sproat, R., & Witzel, M. (2004). *The collapse of the Indus-script thesis: The myth of a literate Harappan civilization.* Retrieved from Electronic Journal of Vedic Studies: http://www.ejvs.laurasianacademy.com/ejvs1102/ejvs1102article.pdf

Ferguson, C. (1959). Diglossia. *Word, 15*, 325–340.

Filliozat, J. (1953). Paléographie. In L. Renou & J. Filliozat (Eds.), *L'Inde classique: Manuel des études indiennes* (pp. 665–712). Paris, France: Imprimerie Nationale.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35*(5), 263–279.

Gelb, I. J. (1952). *A study of writing: The foundations of grammatology*. Chicago, IL: University of Chicago Press.

Gill, H. S. (1996). The Gurmukhi script. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 395–398). New York, NY: Oxford University Press.

Gill, H. S., & Gleason, H. A. (1969). *A reference grammar of Punjabi* (2nd ed.). Patiala, India: Punjabi University, Department of Linguistics.

Glass, A. (2000). *A preliminary study of Kharoṣṭhi manuscript paleography.* M.A. Thesis, University of Washington, Dept. of Asian Languages and Literature. Retrieved from http://andrewglass.org/ma.php

Gnanadesikan, A. E. (2017). Towards a typology of phonemic scripts. *Writing Systems Research, 9*(1), 14–35. https://doi.org/10.1080/17586801.2017.1308239.

Haas, M. (1956). *The Thai system of writing*. Washington, DC: American Council of Learned Societies.

Hamp, E. P. (1996). On the Indo-European origins of the retroflexes in Sanskrit. *Journal of the American Oriental Society, 116*(4), 719–723.

Harington, J. H., & Colebrooke, H. (1799). Translation of one of the inscriptions on the pillar at Delhee, called the Lāt of Feerōz Shah. *Asiatick Researches, 7*, 175–182.

Hitch, D. A. (1989). Brāhmī. In *Encyclopædia Iranica* (Vol. 4, pp. 432–433). Retrieved from http://www.iranicaonline.org/articles/brahmi-indian-script.

Hitch, D. A. (2010). Aramaic script derivatives in Central Eurasia. *Sino-Platonic Papers, 198*. Retrieved from http://sino-platonic.org/complete/spp198_aramaic_script.pdf

Hodgson, B. H. (1834, October). Notice of some ancient inscriptions in the characters of the Allahabad Column. *Journal of the Asiatic Society of Bengal*, 3(10), 481–483.

Hoernle, A. F. (1884). *Centenary review of the Asiatic Society of Bengal from 1774–1883, part 2: Archæology, history, literature, &c*. Calcutta, India: Asiatic Society of Bengal.

Holle, K. F. (1999). Table of old and new Indic alphabets: Contribution to the paleography of the Dutch Indies. (C. Molony & H. Pechler, Trans.). *Written Language and Literacy, 2*(2), 167–246.

Hommel, F. (1893). *Süd-arabische Chrestomathie: Minäo-sabäische Grammatik – Bibliographie – Minäische Inschriften nebst Glossar*. Munich, Germany: G. Franz.

Hosking, R. F., & Meredith-Owens, G. M. (1966). *A handbook of Asian scripts*. London, UK: British Museum.

Jensen, H. (1969). *Sign, symbol and script* (3rd ed.). (G. Unwin, Trans.) London, UK: George Allen & Unwin.

Joyce, T., & Borgwaldt, S. R. (2011). Typology of writing systems: Special issue introduction. *Written Language and Literacy, 14*(1), 1–11. https://doi.org/10.1075/wll.14.1.01joy.

Kandhadai, P., & Sproat, R. (2010). Impact of spatial ordering of graphemes in alphasyllabic scripts on phonemic awareness in Indic languages. *Writing Systems Research, 2*(2), 105–116. https://doi.org/10.1093/wsr/wsq009.

Kopp, U. F. (1819–1821). *Bilder und Schriften der Vorzeit*. Mannheim, Germany: Author.

Kuipers, J. C., & McDermott, R. (1996). Insular Southeast Asian scripts. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 474–484). New York, NY: Oxford University Press.

Ledyard, G. K. (1997). The international linguistic background of the Correct Sounds for the Instruction of the People. In Y.-K. Kim-Renaud (Ed.), *The Korean alphabet* (pp. 31–87). Honolulu, Hawaii: University of Hawaii Press.

Mahadevan, I. (1990). Orthographic systems in early Tamil writing. *Journal of the Institute of Asian Studies, 8*(1), 35–47.

McCawley, J. D. (1968). *The phonological component of a grammar of Japanese* (Janua Linguarum Series Practica 5). The Hague, Netherlands: Mouton.

Miller, C. R. (2010). A Gujarati origin for scripts of Sumatra, Sulawesi and the Philippines. In N. Rolle, J. Steffman, & J. Sylak-Glassman (Eds.), *Proceedings of the thirty-sixth annual meeting of the Berkeley Linguistics Society February 6–7, 2010* (pp. 276–291). Berkeley, CA: Berkeley Linguistics Society. Retrieved from http://escholarship.org/uc/item/4q6906nz.

Miller, C. R. (2014a). Devanagari's descendants in North and South India, Indonesia and the Philippines. *Writing Systems Research, 6*(1), 10–24. https://doi.org/10.1080/17586801.2013.857288.

Miller, C. R. (2014b). A survey of indigenous scripts of Indonesia and the Philippines. In *Proceedings of the international workshop on endangered scripts of Island Southeast Asia* (pp. 1–49). Tokyo, Japan: University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa.

Mohanan, K. P. (1996). Malayalam writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 420–425). New York, NY: Oxford University Press.

Naveh, J., & Shaked, S. (2012). *Aramaic documents from ancient Bactria (fourth century BCE) from the Khalili collections*. London, UK: The Khalili Family Trust.

Parpola, A. (1994). *Deciphering the Indus script*. Cambridge, UK: Cambridge University Press.

Perfetti, C. A., & Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. *Language Learning and Development, 9*(4), 296–316.

Prinsep, J. (1834a, March). Note on inscription no. 1 of the Allahabad column. *Journal of the Asiatic Society of Bengal, 3*(3), 114–123.

Prinsep, J. (1834b, October). Note on the Mathiah Láth inscription. *Journal of the Asiatic Society of Bengal, 3*(10), 483–487.

Prinsep, J. (1834c, October). Second note on the Bhilsá inscription. *Journal of the Asiatic Society of Bengal, 3*(10), 488–490.

Prinsep, J. (1835, June). Further notes and drawings of Bactrian and Indo-Scythic coins. *Journal of the Asiatic Society of Bengal, 4*(6), 327–348.

Prinsep, J. (1837a, June). Note on the facsimiles of inscriptions from Sanchí near Bhilsa, taken for the Society by Captain Ed. Smith, Engineers; and on the drawings of the Buddhist monument presented by Captain W. Murray, at the meeting of the 7th June. *Journal of the Asiatic Society of Bengal, 6*(6), 451–477.

Prinsep, J. (1837b, July). Interpretation of the most ancient of the inscriptions on the pillar called the lát of Feroz Sháh, near Delhi, and of the Allahabad, Radhia and Mattiah pillar, or lát, inscriptions which agree therewith. *Journal of the Asiatic Society of Bengal, 6*(7), 566–609.

Prinsep, J. (1838a, February). Discovery of the name of Antiochus the Great, in two of the edicts of Asoka, king of India. *Journal of the Asiatic Society of Bengal, 7*(2), 156–166.

Prinsep, J. (1838b, March). On the edicts of Piyadasi, or Asoka, the Buddhist monarch of India, preserved on the Girnar rock in the Gujerat peninsula, and on the Dhauli rock in Cuttack; with the discovery of Ptolemy's name therein. *Journal of the Asiatic Society of Bengal, 7*(3), 219–283.

Prinsep, J. (1838c, July). Additions to Bactrian numismatics, and discovery of the Bactrian alphabet. *Journal of the Asiatic Society of Bengal, 7*(7), 636–655.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of reading* (2nd ed.). New York, NY: Psychology Press.

Rimzhim, A., Katz, L., & Fowler, C. A. (2013). Brāhmī-derived orthographies are typologically Āksharik but functionally predominantly alphabetic. *Writing Systems Research, 6*(1), 41–53. https://doi.org/10.1080/17586801.2013.855618.

Rogers, H. (2005). *Writing systems: A linguistic approach* (Blackwell textbooks in linguistics). Malden, MA: Wiley-Blackwell.

Salomon, R. G. (1995). On the origin of the early Indic scripts. *Journal of the American Oriental Society, 151*(2), 271–279.

Salomon, R. G. (1998). *Indian epigraphy*. New York, NY: Oxford University Press.

Salomon, R. G. (2002). Gāndhārī and the other Indo-Aryan languages in the light of newly-discovered Kharoṣṭhī manuscripts. In N. Sims-Williams (Ed.), *Indo-Iranian languages and peoples* (Proceedings of the British Academy 116, pp. 119–134). Oxford, UK: Oxford University Press.

Salomon, R. G. (2006). Kharoṣṭhī syllables used as location markers in Gandhāran Stūpa architecture. In P. Callieri (Ed.), *Architetti, capomastri, artigiani: L'organizzazione dei cantieri e della produzione artistica nell'Asia ellenistica, Studi offerti a Domenico Faccenna nel suo ottantesimo compleano* (Serie Orientale Roma 100, pp. 181–224). Rome, Italy: Istituto Italiano per l'Africa e l'Oriente.

Salomon, R. G. (2007). Writing systems of the Indo-Aryan languages. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (Routledge language family series, pp. 67–103). London, UK: Routlege.

Sander, L. (1986). Brahmi scripts on the eastern Silk Roads. *Studien zur Indologie und Iranistik, 11–12*, 159–192.

Sander, L. (2005). Remarks on the formal Brāhmī script from the southern Silk Route. *Bulletin of the Asia Institute, NS 19*, 133–144.

Schiller, E. (1996). Khmer writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 467–473). New York, NY: Oxford University Press.

Seland, E. H. (2014). Archaeology of trade in the western Indian Ocean, 300 BC—AD 700. *Journal of Archaeological Research, 22*, 367–402.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.

Share, D. L., & Daniels, P. T. (2015). Aksharas, alphasyllabaries, abugidas, alphabets, and orthographic depth: Reflections on Rimzhim, Katz, and Fowler (2014). *Writing Systems Research, 8* (2016) (1), 17–31. https://doi.org/10.1080/17586801.2015.1016395.

Sims-Williams, N. (2000–2007*). Bactrian documents from northern Afghanistan* (3 vols. Studies in the Khalili Collection 3; Corpus Inscriptionum Iranicarum 2.3). London, UK.
Sircar, D. C. (1965). *Indian epigraphy*. Delhi, India: Motilal Banarsidass.
Sircar, D. C. (1970–71). Introduction to Indian epigraphy and palaeography. *Journal of Ancient Indian History, 4*(1–2), 72–136.
Sivaramamurti, C. (1948). *Indian epigraphy and South Indian scripts* (Bulletin NS General Section 3/4 [1952]). Madras [now Chennai]: Madras Government Museum.
Skjærvø, P. O. (1996). Aramaic scripts for Iranian languages. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 515–535). New York, NY: Oxford University Press.
Sproat, R. (2014). A statistical comparison of written language and nonlinguistic symbol systems. *Language, 90*(2), 457–481.
Steever, S. B. (1996). Tamil writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 426–430). New York, NY: Oxford University Press.
Stevens, J. (1988). *Sacred calligraphy of the East*. Boulder, CO: Shambhala.
Taylor, I. (1883). *The alphabet: An account of the origin and development of letters*. London, UK: Kegan Paul, Trench.
Thomas, E. (Ed.). (1858). *Essays on Indian antiquities, historic, numismatic, and palæographic, of the late James Prinsep, F.R.S*. London, UK: John Murray.
Thomas, E. (1871). Comments on recent Pehlvi decipherments. With an incidental sketch of the derivation of Aryan alphabets. And contributions to the early history and geography of Tabaristán. Illustrated by coins. *Journal of the Royal Asiatic Society, NS, 5*(2), 408–459.
Turnour, G. (1837, December). Further notes on the inscriptions on the columns at Delhi, Allahabad, Betiah, &c. *Journal of the Asiatic Society of Bengal, 6*(12), 1049–1064.
van Schaik, S. (2011). A New look at the Tibetan invention of writing. In Y. Imaeda, M. T. Kapstein, & T. Takeushi (Eds.), *New studies of the old Tibetan documents: Philology, history and religion* (Old Tibetan documents online monograph series 3, pp. 45–96). Tokyo: Tokyo University of Foreign Studies, Research Institute for Languages and Cultures of Asia and Africa. Retrieved January 27, 2014, from Old Tibetan Documents: https://otdo.aa.tufs.ac.jp.
Verhoeven, L., & Perfetti, C. A. (Eds.). (2017). *Reading acquisition: Cross-linguistic and cross-script perspectives*. Cambridge, UK: Cambridge University Press.
von Hinüber, O. (1990). *Der Beginn der Schrift und frühe Schriftlichkeit in Indien* (Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse 1989/11). Mainz: Akademie der Wissenschaften und der Literatur.
Weber, A. (1856). Ueber den semitischen Ursprung des indischen Alphabetes. *Zeitschrift der Deutschen Morgenländischen Gesellschaft, 10*, 389–406.
Wheatley, J. K. (1996). Burmese writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 450–456). New York, NY: Oxford University Press.
Windisch, E. (1917). *Geschichte der Sanskrit-Philologie und indischen Altertumskunde*, part 1 (Grundriss der Indo-Arischen Philologie und Altertumskunde, vol. 1 section 1B). Berlin, Germany: de Gruyter.
Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29.

What Is an Akshara?

Keerthi Ramanujan and Brendan S. Weekes

Abstract The answer to the question of what is an akshara is complex. To understand the concept of akshara in Indic languages, it is necessary to consider multiple levels of meaning. In historical terms, the akshara is an example of the resilience of the human invention that we call writing. In purely linguistic terms, the concept can be defined as an example of an abugida writing system, which is by no means restricted to Asian languages. In sociocultural terms, akshara is beyond any simple description. Akshara can be viewed as connected to the metaphysical world in the liturgy of ancient philosophies.

Keywords Akshara · Alphasyllabary · Devnagari · Indian languages · Indic orthography · Sanskrit · Writing systems

Akshara is a Sanskrit word that refers to syllable. The syllable enjoys a high status in the Hindu Vedas, transmitted through an oral tradition. Akshara is the minimal articulatory unit for Indic languages. In Sanskrit, the word literally means that which is not destructible (destructible = *kshara*) in reference to the idea that sound (and its constituent syllables) is inherently indestructible.

Aksharas are typical of writing systems that are derived from the common parent, Brahmi (Bright, 1996, 2000). Brahmi is a writing system that dates to as far the fifth to third century BC (Patel, Pandey, & Rajgor, 2007; Vaid & Gupta, 2002) and is considered the common ancestor of several Indic scripts. Interestingly, the languages that these Indic scripts represent do not necessary belong to one common language family. For example, Punjabi is an Indo-Aryan language spoken in Northwest India. Its akshara Brahmi-derived script is called Gurmukhi. Tamil on the other hand, is a Dravidian language spoken in South India and Sri Lanka. The Tamil script can also be traced back to the Grantha branch of the ancient Brahmi.

K. Ramanujan · B. S. Weekes (✉)
Laboratory for Communication Science, Faculty of Education, University of Hong Kong, Hong Kong, China
e-mail: keerthi.ram@connect.hku.hk; weekes@hku.hk

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_3

Despite belonging to several different language families, Indic scripts seem to share a common written ancestor.

Akshara is a unit of speech. In writing, it is represented as a vowel and consonant with an inherent schwa or as a consonant with a vowel marker, known as matra, which modifies the sound the consonant represents. This functional definition does not fully capture the linguistic uniqueness of Indic writing systems. Does one consonant-matra cluster equate to one akshara symbol – making the writing system a syllabary? Or do the matras/vowel markers also function as independent subunits, making the writing system more like an alphabet? Writing of vowels as appendages rather than as independent letters (as in alphabetic systems) is a more parsimonious way to represent sounds compared to a syllabary, which has unique graphemes to represent consonant-vowel combinations. However, this leaves open the definition of an akshara symbol in modern writing.

Origins

Linguists agree on the resilience of Brahmi writing systems in the languages of South and South-East Asia. A remarkable feature of these systems is longevity. All modern writing systems used in South Asia today are derived in part from Brahmi and derivations can be found across South East Asia – from Myanmar, Thailand, Indonesia, and the Philippines. Brahmi writing systems that evolved to record the various Indic languages, specifically, were developed in a way that would capture the language's natural syllabification. This may be less true for non-Indic spoken languages. The Brahmi script family is depicted in Fig. 1.

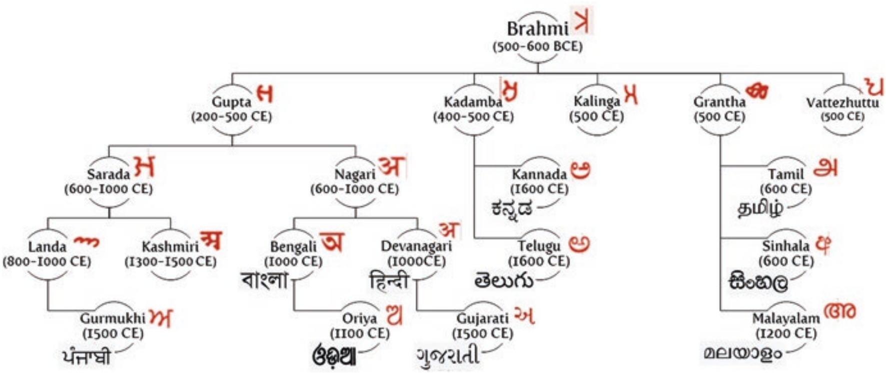


Fig. 1 Family tree and timeline of Brahmi based scripts. The first vowel "A" is depicted in red next to the names of the script. The names of the present-day languages of the Indian subcontinent that use these Brahmi-derived scripts are indicated at the bottom left corner. English transliterations are as follows: Punjabi (Gurmukhi), Bangla (Bengali), Odia (Oriya), Hindi (Devanagari), Gujarati, Kannada, Telugu, Tamizh (Tamil), Sinhala, Malayalam. (Figure adapted from: http://www.ancientscripts.com/sa_ws_cmp.html, Lawrence Lo, 2012)

Sanskrit is an ancient language believed to having originated around the second millennium BC (Benware, 1974; Burrow, 2001). Despite its near extinct status today, Sanskrit is a liturgical language in literature, music, and scripture, and is of spiritual importance across India. Other writing systems that evolved to write Sanskrit have become extinct. Kharosthi a writing system that developed in northwestern India possibly in the fourth to fifth century BC (Salomon, 1996, 1998) disappeared by the beginning of the third century BC (Trigger, 2004), replaced by the Brahmi systems elsewhere in the Indian subcontinent (Chakrabarti, 2006).

Sanskrit was the language of the Brahminic liturgy. Prakrit vernaculars are derived from the Indo-Aryan family and typically dominate the Northern part of modern day India. Dravidian languages are spoken in the South and the script is derived from a Brahmi predecessor called Grantha. Sanskrit was the language of prestige, liturgy, literary arts, and scholarship and may well have been the lingua franca (Keown & Prebish, 2013), especially amongst the Brahmanical or the literate classes but also co-existing with other vernaculars (Deshpande, 2011) just as Latin was in medieval Europe. It is interesting that while alphabets developed to write Indo-European languages in Europe, the akshara-based scripts continued to endure as the principal writing system for the languages of the Indian subcontinent, despite well-established and flourishing cultural and commercial connections between the east and west. This illustrates how the distinctive features of spoken languages can constrain the evolution and refinement of written languages within one language family.

The earliest Sanskrit texts date back to the Rigveda at the second millennium BC (Meier-Brügger, 2003). The first grammatical documentation of a spoken language was written down for Classical Sanskrit (Ashtaadhyayi) by the scholar Paanini in 500 BCE (Houben, 1996; Keith, 1993). Inscriptions of Brahmi-based scripts can be found in Sri Lanka, Myanmar, and Indonesia due to the adoption of Hinduism that spread as far as East Asia when some Buddhist monks attempted to record Vedic scriptures in Sino-Tibetan languages.

Features of Brahmi Writing Systems

Brahmi writing systems have been classified as alpha-syllabaries (Nag, 2014; Pandey, 2007, 2014). Alpha-syllabaries separate vowels and consonants within large units (syllables) rather than small units (phonemes) but are not alphabets, as symbols cannot be reduced to constituent phonemes (Gnanadesikan, 2011). According to some linguists, a defining feature of Brahmi writing systems is use of an *abugida* – a category of segmental writing systems in which consonant–vowel sequences are written as a unit and each of the units is based on a consonant letter with vowel notation as a secondary and optional feature (Daniels, 1996). This concept resembles Gnanadesikan's (2011) pithy definition of akshara symbols as "writing that specifies a vowel sign written as appendages to a consonant symbol." By this definition, an abugida is a suitable term for many scripts including

Semitic-Ethiopic scripts and Canadian Aboriginal syllabic scripts. Abugida consist of a syllable and a consonant cluster but differ from syllabaries because abugida symbols can be divided into consonants and vowels whereas a syllabary cannot (Gnanadesikan, 2011). An *abugida* is therefore a more precise description of the Brahmi-based alpha-syllabaries because it defines consonants and vowels as a cluster. The features of all Brahmic writing systems, do conform to abugida. The Akshara symbols are defined as the basic units of Brahmi writing (Padakannaya & Mohanty, 2004). Defining features of Brahmi are (a) symbols for syllabic (initial) vowels; (b) consonant symbols but also one inherent vowel schwa; (c) vowels in CV combinations denoted by vowel markers/matras; (d) consonant clusters represented by ligatures; and (e) an inherent vowel can be muted by a diacritic mark (known as virama). Akshara symbols meet all these criteria. However, there is a certain aspect of the akshara that is often glossed over: aksharas refer to the syllabic units of speech. The symbols/graphemes are merely written forms that were developed to depict acoustic syllables in print. The Vedas, for example, place a high degree of importance on sound (shabda) because of the oral method of its transmission and learning. The akshara as a unit of sound is of great importance in the context of Vedic literary meters (known as chanda, in Sanskrit). An example of a Vedic literary meter is the Anushtup chanda. A verse in this meter has exactly four stanzas with 32 aksharas (sound syllables), with eight akshara (syllables) er stanza. Breaking up of literary verses into akshara level also serves as prosody and stress based-retrieval cue to aide phonological recall for reciting sequences such verses.

Brahmi writing systems contrast with Protosemitic (Northern) ancestors (Canaanite, Phoenician and early Hebrew) and modern cousins (Cyrillic, Greek, Roman) because the Brahmi family evolved into dozens of visually distinct scripts, many of which are very much in use today.

What an Akshara Is and Is Not

Akshara are the basic units of writing found in writing systems of the Indian subcontinent. Akshara symbols represent several syllable compositions including CV, CVV, CCV, CCVV, CCCV, CCCVV, V and VV (long vowel). Akshara symbols are sub-syllabic because they represent the onset, or onset plus nucleus or nucleus alone. Uniquely, the coda in syllables is depicted in writing systems as a subsequent akshara symbol.

Akshara encode both syllable and matra values. In speech, a matra can be defined as a unit of prosodic marking, – matras are used to denote both shorth and long vowels. These are marked in akshara symbols at the sub-syllabic level. A syllable has a vowel that can be preceded and/or followed by a consonant but might also be a vowel only e.g., /a/. A consonant that precedes a vowel in a syllable is the onset, and the consonant that follows a vowel is the coda. A vowel is also referred to as the 'nucleus' or 'peak' of a syllable. The combination of onset plus nucleus can be called a 'body' while nucleus plus coda can be called a 'rime.' Syllables in Sanskrit derived languages also have quantity, which is referred to as 'heavy' or 'light.' A syllable ending with a consonant (closed) is considered heavy (guru) whereas a syllable ending in a short vowel (open) or a stand-alone vowel is light (lagu). Akshara symbols represent onset, onset plus nucleus, or nucleus alone but not coda alone, which is always positioned in a subsequent symbol (Patel, 2004). This is different from alphabetic systems, although some systems do vary on this feature e.g., Cyrillic. Sequences of akshara symbols can be arranged linearly to form a 'word'.

Akshara symbols represent phonological units that resemble syllables (or precisely mora). However, akshara symbols differ from other depictions of mora e.g., Japanese kana, because they reduce to the sub-syllabic consonant and vowel structure. Akshara symbols also represent vowels and consonants as in alphabetic systems. However, akshara cannot be called alphabetic, as they never represent phonemes. Akshara are organized to represent diacritics and ligatures positioned around a root consonant, making the 'alphabet' more similar to Arabic or Hebrew than European alphabetic writing systems. Although all Brahmi scripts can be defined as abugida, members of the Brahmi family use a variety of visually distinct symbols. This is to support the phonology of the languages they encode. The size of the phonemic repertoire is different for the various Indic languages that use a Brahmi-based script. For example, Hindi and Malayalam both have nearly the same set of phonemes as Sanskrit. Therefore, a Sanskrit sentence or phrase may be transcribed in Devanagari and Malayalam script with high fidelity. Tamil and Sanskrit do not have a large shared phoneme set – so many Sanskrit words cannot be accurately transcribed into Tamil, using Tamil's native script. The purpose of all writing systems is to support a spoken language (literary or vernacular) by representing the phonemes of the language. To be able to read and write a script in all Indic languages, it is necessary to learn the grapheme-phoneme associations of the language using akshara symbols. As Brahmic systems are a written language family with cousins in Northern and Southern India as well as other parts of South East Asia, they can be defined according to the use of abugida but vary in terms of scripts.

Cultural Context

The development of akshara symbols in Brahmi script testifies to a sophisticated phonetic knowledge acquired in antiquity. Prakash and Joshi (1995) argue that the concept of akshara was in place before widespread development of writing throughout South Asia. Indeed, they propose that the structure of Brahmi was intentionally crafted from early Semitic writing systems and then nativized for the Sanskrit language. Spoken syllables most certainly were available before the invention of any writing system. But the fact that a wide variety of Asian languages (indic and non-indic) spanning a large geographical region, all make use of the abugida writing design, is quite remarkable.

Diversity of scripts allows for the preservation and continuation of core values in a linguistically, culturally and politically diverse environment that is also geographically vast and varied. Whether such diversity in writing systems needs a preservation order in the twenty-first century is an issue worth pondering about, since in the age of globalization and high technology with today's societies attempting to converge and minimize differences. However, linguistic and cultural diversity is so enshrined in Modern India that official scripts (albeit a common writing system) number over twenty, and documents are routinely translated into several languages and scripts. It is the tolerance of diversity that is key to the preservation of akshara symbols today. In a way, the akshara based writing systems have indeed lived up to their name as indestructible since, despite socioeconomic and geopolitical changes through time, they are resilient. It is interesting to consider how electronic means of communication (texting) might preserve the grammar and script of the spoken languages in the coming century. Will the Akshara symbols be resilient even in the face of pressures to communicate with a global audience in our ever increasingly interconnected world?

Scripts in Contact – Urdu and Hindi

An interesting question to ponder is why related languages in the subcontinent would adopt different writing systems. Linguists agree that Hindi and Urdu are equal descendants of the Indo-Aryan branch of the Indo-European family with a common historical origin (Ahmad, 2008), and they exist in the same territories in Northern India, often in the same street or even building. There are differences between languages in phonology and vocabulary, however. Pure Hindi is more Sanskritized and lacks many of the phonemes in Urdu (borrowing from Arabic, Turkish, Persian, and Sanskrit). Mutual intelligibility is due to the fact that modern day vernacular Hindi uses a lot of Urdu words, and Hindi and Urdu share a large vocabulary. This has resulted in what some refer to as Hindustani (Siddiqi, 1994) – a vernacular that is a mix of both languages with pure Hindi and pure Urdu at the extremes of the spectrum. There are several Sanskrit-derived or Prakrit-derived words in Hindi not occurring in Urdu, and Urdu has words derived from Arabic, Persian and Turkish – that are not found in Hindi. Why have these "scriptages" evolved and coexisted for centuries?

A divergence of Hindi and Urdu likely first occurred during the period of Mughal rule over a substantial part of present day India (Rao, 2010). Hindi and Urdu took shape as a composite mix of regional dialects in and around Delhi, while influenced by Persian, the language of the rulers, as well as older languages and dialects used around Delhi such as Prakrit and Apabhramsh (Rao, 2010). According to Rao, the terms Hindvi (or Hindavi) and Urdu were used to refer to a newly evolving language: The term Hindvi was used to designate the vernacular of the people of Hind (a region near the Indus river), while the term Urdu originated from the Turkish word for army encampment. These terms were used to describe a single language until the seventeenth century, by which time the name Hindustani (meaning Indian) gained currency. The official sanctioning of Urdu/Hindustani as a court language by the British fuelled many sociopolitical dynamics, leading to the separation of Modern Standard Hindi from the parent Urdu/Hindustani language. The gulf between Hindi and Urdu widened as Hindi speakers adopted words from Sanskrit as a source to expand Hindi vocabulary, thereby distancing it from Urdu (dominated by Perso-Arabic vocabulary). These tensions are reflected in the written texts wherein borrowed Arabic script (of the usurped Mughals) and akshara symbols were adapted by the occupied to sustain their own ideologies (Hinduism and Islam) in parallel worlds of liturgical texts, music, and philosophy co-existing in adjacent neighborhoods. Sulatana has translated the Qur'an into Sanskrit (2010) To be able to read the Qur'an, phoneme-grapheme mappings and Arabic itself (in terms of vocabulary) must be known, due to the abjad's context-based approach to reading. Indian Muslims learn the Arabic (in a limited manner and context) to read Qur'anic verses. This is because Hindi/Devanagari does not have graphemes to represent Arabic phonemes. This is why Sultana translated Arabic Qur'an into Sanskrit and did not transliterate it. According to a personal interview (November 2016, Deogarh, India) it took more than12 years to complete this work because finding Sanskrit for Qur'anic words was difficult (Fig. 2).

The survival of Urdu depends in part on preservation of liturgical texts written in Arabic. A Hindi-speaker from India, as well as Urdu-speakers in Pakistan, are mutu-

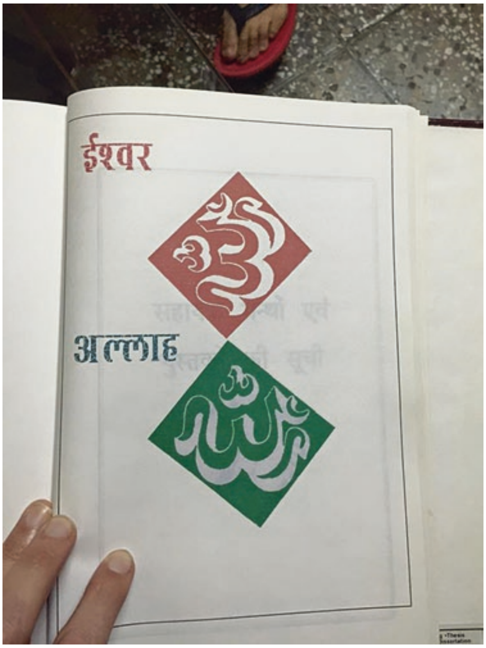


Fig. 2 Qur'an written in Devanagari script. Red Font says: Ïshvar = God in Sanskrit/Hindi; Red calligraphy is Devanagari symbol for "AUM" stylized in the manner of Arabic calligraphy; Green Font says Allah = God in Quran; Green calligraphy is Arabic script for "Allah"

ally intelligible to each other. But to be a self-defined native Urdu speaker signifies a specific linguistic and cultural identity.

Final Remarks

The evolution of all writing systems is marked by the tendency to graft dominant written codes on to native spoken languages. No spoken language affords a script naturally. The result is almost inevitably arbitrary links between printed text and phonology. Such adaptations can be seen in writing systems today, e.g., umlauts in Germanic languages, digraphs in English (and French), optional diacritic markers in Hebrew and Persian, and phonetic radicals in Chinese characters.

In terms of a morphographic script (Chinese), it is notable that characters can be shaped into many spoken Sino-Tibetan languages including Cantonese in Hong Kong, as well as Japanese and Korean, but this always results in irregular mappings between orthography and phonology. A logograph always represents a syllable, but it is not usually marked for pronunciation at the sub-syllabic or sound-tonal level. Therefore, to correctly read a logograph character in any Chinese language, 'lexical' knowledge of the specific representations of syllable sounds is required (Weekes, 2012). It is possible that logographs will have similar pronunciations in different Chinese languages due to shared morpho-phonology. The same is true for Arabic script used in Persian or Urdu wherein the same orthographic symbols are pronounced according to the dominant language (see Bakhtiar & Weekes, 2015; for discussion of reading Arabic abjad in Persian). Morphographic scripts, e.g., Chinese characters and Japanese Kanji, contain thousands of written symbols representing morphemes (meanings) directly and syllables obliquely. In all writing systems, the productivity of the script depends on how effectively the sounds of a language can be depicted at the sub-syllabic level. Alphabetic scripts e.g., English, Greek, and Russian, have a limited number of sub-syllabic symbols and grapheme to phoneme correspondences to be learned. Exception spellings are the result of a Latin script imposed on unrelated languages (English and French).

The arbitrary nature of writing systems is illustrated by comparisons in Arabic and akshara to depict the same words in Urdu and Hindi. This is similar to adaptation of Cyrillic and Roman alphabets to read Serbo-Croatian and Kanji and Kana to read Japanese words. Comparisons between bi-script monolinguals can be informative in terms of the transparency-opacity continuum. Brahmi writing systems are transparent. That is, the symbol affords a pronunciation consistently and there is a requirement to infer missing vowels from context. Urdu also affords transparent pronunciations when context is added but pronunciation of an abjad is not always consistent and depends on context (as in Hebrew and Persian). Symbol-sound mappings in all Brahmic systems using abugida are consistent, predictable, accurate, and faithful transcriptions of spoken words. (Nag, 2014).

Conclusion

Western scholars tend to view literacy through a lens of alphabet supremacism. It is important to remind scholars of the arbitrary relationships between writing systems and spoken languages. This has been the rule for as long as humans have had written languages to communicate beyond the present. One exception to the rule comes from attempts to transcribe spoken languages with an invented script, e.g., missionary scripts developed in Africa and South America to depict the sounds in the entire language (Gnanadesikan, 2011). One interesting conjecture is that to survive, a writing system does not need to resemble spoken language at all. It is more likely (historically) for extinction to follow attempts to codify a local vernacular. However, the structure of the writing system whether abjad, akshara, or alphabet remains the same precisely because it is linked to a deeper, stronger cultural identity. Uniquely, the subcontinent is a living laboratory to study distinct scrips and languages in close contact with each other. Surrounded by orthographic hegemons of Arabic, Cyrillic and Latin to the West and North and Sino-Tibetan writing systems to the East, no region on earth has been able to incubate and sustain as many diverse scripts. Geography is one likely factor since divisions between cultures, languages and their writing systems seem to coincide with impenetrable natural obstacles and barriers such as the Himalaya to the north, arid plains and deserts to the west, marshy wetlands to the east and oceans to the south. However, to understand the survival of akshara based writing systems it is necessary to put spoken and written languages of South Asia into a cultural context. Linguistic diversity in different scripts allows the DNA of shared wisdom to survive at least as it is perceived by a majority of the nearly 2 billion inhabitants.

Writing systems can only survive if they are generative, productive and resilient. Brahmi writing systems meet all these criteria. The key question is why did akshara symbols survive for so long (fourth century BC to the twenty-first century AD)? 2500 years is a long time for a cultural system to survive; only Chinese characters have a longer history. The answer cannot be reduced to one explanation. It is particularly informative to consider how a writing system can remain intact in the context of competitor systems that surround it, often within just a few miles. This contrasts with the hegemons of Arabic, Cyrillic, Latin and Chinese writing that will dominate large parts of the globe in the twenty-first century.

References

Ahmad, R. (2008). Scripting a new identity: The battle for Devanagari in nineteenth century India. *Journal of Pragmatics, 40*(7), 1163–1183. https://doi.org/10.1016/j.pragma.2007.06.005.

Bakhtiar, M., & Weekes, B. (2015). Lexico-semantic effects on reading in Persian: Does age of acquisition have an effect? *Memory & Cognition, 43*(2), 298–313.

Benware, W. A. (1974). *The study of indo-European vocalism in the 19th century: From the beginnings of Whitney and Scherer: A critical-historical account*. Amsterdam, Netherlands: Benjamins.

Bright, W. (1996). The Devanagari script. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems*. New York, NY/Oxford, UK: Oxford University Press.

Bright, W. (2000). A matter of typology: Alphasyllabaries and abugidas. *Studies in the Linguistic Sciences, 30*(1), 1–11.

Burrow, T. (2001). *The Sanskrit language*. New Delhi, India: Motilal Banarsidass.

Chakrabarti, D. K. (2006). *The Oxford companion to Indian archaeology: The archaeological foundations of ancient India, stone age to AD 13th century*. New Delhi, India: Oxford University Press.

Daniels, P. T. (1996). The invention of writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems*. New York, NY/Oxford, UK: Oxford University Press.

Deshpande, M. (2011). Efforts to vernacularize Sanskrit: of success and failure. In J. Fishman & O. Garcia (Eds.), *Handbook of language and ethnic identity: The success-failure continuum in language and ethnic identity efforts*. New York, NY: Oxford University Press.

Gnanadesikan, A. E. (2011). *The writing revolution: Cuneiform to the internet*. New York, NY: Wiley.

Houben, J. E. M. (1996). *Ideology and status of Sanskrit: Contributions to the history of the Sanskrit language*. Leiden, Netherlands: E.J. Brill.

Keith, A. B. (1993). *A history of Sanskrit literature*. New Delhi, India: Motilal Banarsidass.

Keown, D., & Prebish, C. S. (2013). *Encyclopaedia of Buddhism*. Hoboken, NJ: Taylor & Francis.

Lo, L. (2012). *South Asian writing systems comparison*. http://www.ancientscripts.com/sa_ws_cmp.html.

Meier-Brügger, M. (2003). *Indo-European linguistics*. Berlin, Germany: Walter de Gruyter.

Nag, S. (2014). Alphabetism and the science of reading: From the perspective of the akshara languages. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00866.

Padakannaya, P., & Mohanty, A. K. (2004). Indian orthography and teaching how to read: A psycholinguistic framework. *Psychological Studies, 49*(4), 262–271.

Pandey, P. (2007). Phonology – orthography interface in Devanāgarī for Hindi. *Written Language and Literacy, 10*(2), 139–156.

Pandey, P. (2014). Akshara-to-sound rules for Hindi. *Writing Systems Research, 6*(1), 54–72. https://doi.org/10.1080/17586801.2013.855622.

Patel, P. G. (2004). *Reading acquisition in India: Models of learning and dyslexia*. New Delhi: Sage.

Patel, P. G., Pandey, P., & Rajgor, D. (2007). *The Indic scripts: Palaeographic and linguistic perspectives*. New Delhi, India: D.K. Printworld.

Personal Interview. (2016, November). *Dr Raisa Sultana held in Deoband.* Agra.

Prakash, P., & Joshi, R. M. (1995). Orthography and reading in Kannada: A Dravidian language. In I. Taylor & D. R. Olsen (Eds.), *Scripts and literacy: Reading and learning to read alphabets, syllabaries and characters*. Dordrecht, Netherlands: Kluwer.

Rao, C. (2010). Morphology in word recognition: Hindi and Urdu. PhD thesis. Texas A&M University.

Salomon, R. (1998). *Indian epigraphy: A guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages. South Asia research*. New York, NY: Oxford University Press.

Salomon, R. G. (1996). *Brahmi and Kharoshthi. The world's writing systems* (pp. 373–383). New York, NY: Oxford University Press.

Sanskrit translation of Holy Qur'an. (2010, August 15). *The Milli Gazette*, p. 18. Retrieved from http://www.milligazette.com/

Siddiqi, M. T. (1994). *Hindustani-English code-mixing in modern literary texts*. Madison, WI: University of Wisconsin—Madison.

Trigger, B. G. (2004). Writing systems: A case study in cultural evolution. In S. D. Houston (Ed.), *The first writing: Script invention as history and process* (pp. 60–61). Cambridge, UK: Cambridge University Press.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*(1–3), 679–690. https://doi.org/10.1006/brln.2001.2556.

Weekes, B. S. (2012). Acquired dyslexia and dygraphia across scripts. *Behavioural Neurology, 25*, 159–163.

Part II
Akshara Orthographies: A Brief Description

Orthographic Knowledge, Reading and Spelling Development in Tamil: The First Three Years

Sonali Nag and Balambigai Narayanan

Abstract Tamil (tamiḻ) is a South Dravidian language with a contained orthography in contrast to the other more extensive akshara orthographies of South Asia. In this chapter we briefly describe the Tamil orthography, introducing the rich vocabulary available in the language to describe the symbol set. Similar descriptions are given for Tamil phonology and morphology. We also focus on Tamil teaching before presenting research findings on early literacy development. Our survey data suggest that important milestones in the journey to mastery of word reading and spelling include learning about the native and grantha consonants and unconnected diacritic markers, especially those with ambiguous visual elements such as the *thunaikaal*. A discussion on writing routines, diglossia and children's spontaneous writing is also included.

Keywords Tamil · Reading · Spelling · Orthographic knowledge · Diglossia · Writing routines · Contained orthography · Literacy instruction

Introduction

Tamil (tamiḻ) belongs to the South Dravidian language family. Tamil has the largest number of speakers compared to the other literary South Dravidian languages of Telugu, Malayalam, and Kannada. Tamil is spoken in parts of India, Malaysia, Singapore, Sri Lanka, the Fiji Islands, and an extensive diaspora elsewhere. The

S. Nag (✉)
Department of Education, University of Oxford, Oxford, UK

The Promise Foundation, Bangalore, India
e-mail: sonalinag@t-p-f.org; sonali.nag@education.ox.ac.uk

B. Narayanan
Department of Speech, Hearing & Communication, National Institute for Empowerment of Persons with Multiple Disabilities(Divangjan) (NIEPMD), Chennai, India
e-mail: balambigai.n@gmail.com

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_4

southern state of Tamil Nadu in India has the largest concentration of Tamil speakers which with a population of 96 million (Census India, 2015) is larger than most countries in the world.

Tamil has an uninterrupted literary history of more than 2000 years with some of the earliest cave and potshard inscriptions dated to the second century BCE. The history of Tamil is typically seen as spanning works belonging to Old Tamil, Middle Tamil and Modern Tamil. The earliest treatise on Tamil grammar, orthography and theoretical linguistics is entitled *Tolkaappiyam*. While the date and authorship of this ancient text is somewhat disputed, one school of thought is that *Tolkaappiyam* is perhaps a series of anthologies compiled over several centuries.

The following sections will cover descriptions of Tamil orthography, phonology, and morphology interspersed with our preliminary research findings in the areas of early reading and spelling development. We also focus on Tamil teaching practices and describe in particular detail the first year of literacy instruction. The formal and written form of Tamil or the ‘high’ literary Tamil is called centamiḻ (‘pure Tamil’) whereas the vernacular, ‘low’ variety is called koṭuntamiḻ (the harsh or ‘crooked Tamil’). As with other languages imbued with diglossia—Bengali and Hindi for example—very little information is available about the impact of diglossia on Tamil language and literacy acquisition. This chapter will touch briefly on issues of Tamil diglossia and literacy acquisition and flag this area as a gap that needs systematic research attention.

Tamil Orthography

The Tamil orthography is an alphasyllabary belonging to the Brahmi group of scripts. Tamil-Brahmi, or Damili, predates the Prakrith-Brahmi scripts and includes unique symbols to capture the sounds of the Tamil language (Rajan & Yateeskumar, 2013). Similar to other Brahmi-derived orthographies, the symbol units in Tamil are written in symbol blocks. The *akshara* is minimally a vowel (/V/) or consonant (/C/), and a consonant-vowel combination. The mid-low vowel /ʌ/[1] is found inherent in the consonant, and for all other vowels a diacritic marks the vowel in a CV akshara. Unlike other Brahmi-based systems, conjoint symbols to represent consonant clusters are rare in Tamil. The Tamil orthography is considered mostly transparent although both historic and current influences have introduced several instances of opacity in the system.

[1] We represent Tamil phonemes and akshara sound values using the alphabet of the International Phonetic Association (IPA), 1999.

Fig. 1 Recent shifts in the form of select Tamil akshara
Note: Examples of akshara from Old Tamil are in the first line. In the second line are the same akshara written using changes initiated by the orthography reform (contemporary Tamil). The last line gives the sound value of each akshara

Diachronic and Synchronic Features

The ancient script of Tamil Brahmi shares features with Sinhala Brahmi and predates Asokan Brahmi of North India. The features of Tamil Brahmi seem to have been best suited for the phonological characteristics of Old Tamil. Over time the symbol pairs [ɳ] and [n], [ɾ] and [r], and [ɻ] and [ɭ] found in formal Tamil and in spelling have merged in spoken, informal Tamil. Another more recent influence to the orthography is the changing of the visually complex symbol shapes, perhaps led by two forces: One is typological, linked to the advent of the printing press but crystallised as font decisions for computer technology, and the other of orthography reform led, most notably, by the Government of Tamil Nadu. Figure 1 gives examples of these shifts. A curious possibility we will consider towards the end of this chapter is whether these trends to simplify and linearise the symbol set have inadvertently increased the learning demand for the young learner.

The Symbol Set and a Vocabulary to Describe it

Tamil orthography (tamil̲ariccuvaṭi[2]) consists of ten vowels and two diphthongs (together called uyirel̲uttu), and eighteen consonants (meyyel̲uttu). The vowels can be further divided into the five long and short vowels (the neṭil and kur̲il). Consonants are divided into hard plosives (vallin̲am) which includes /k, s, ʈ, t̪, p, r/, the soft nasals (/n, n̪, ŋ, ɲ, ɳ, m/, mellin̲am), and medium consonants (iṭaiyin̲am) which include /j, ɾ, ʋ, l, ɭ, ɻ/. The symbol ஃ (called āyutael̲uttu) is mainly used as a pre-fix to native symbols to accommodate sounds in loan words (ஃ + ப் = f). In addition to this are the *grantha* symbols (e.g., /ʤ, s, ʂ, h/) to accommodate the phonology of Sanskrit which was a dominant source for loan words into Old Tamil. As is evident from the above description, Tamil has a rich vocabulary for sets of symbols in the orthography. Similarly, labels are available for intra-symbol elements. Examples of

[2] We use the Tamil Lexicon Notation (1939) collated by the University of Madras to represent Tamil linguistic terms. See Appendix for the linguistic terms written in Tamil along with their IPA transcription and a simple definition.

labels for the Tamil symbol unit include puḷḷi (், 'dot'), tuṉaikkāl (ா 'extra leg'); ottaikkompu (ெ 'one horn'); reṭṭaikkompu (ே, 'two horns'); reṭṭaicuḻi 'n' (ண, '2 loops n') and mūṉṟucuḻi 'n', (ண, '3 loops n'). An important point to note for emergent orthographic knowledge is that these labels for visual elements are generously used in early instruction.

Nag (2007, 2017) describes the world's orthographies by their orthographic breadth. Within this formulation, orthographies may be placed along a continuum that is contained to extensive in symbol repertoire: Cyrillic, Hebrew, Arabic and the Roman-derived alphabets with small symbol sets fall on the contained end of the continuum, while Chinese, with thousands of symbols, falls in the extensive end. The akshara orthographies fall somewhere in between and amongst them some have a smaller symbol set. Tamil may be considered a *relatively more contained orthography* when compared to the other more extensive akshara orthographies (usually symbol sets of 400 to 700). Ten vowel and two diphthong symbols combine with eighteen consonants to form 216 CV combinations. Added to this is the limited set of *grantha* akshara and the symbol ஃ described above. As mentioned earlier, conjoint consonant clusters are rare in Tamil, and the exceptions are ஸ்ரீ and க்ஷ (/ʃr̪iː/ and /kʂʌ/, respectively). For other instances of consonant clusters, found mainly in loan words, the first consonant gives up the inherent vowel by taking on the schwa suppressing marker (puḷḷi, ்), making the number of completely new looking akshara beyond the basic set less common in Tamil. The linearised arrangement of the CCV akshara block has a second implication: Tamil is unlike neighbouring Kannada and Telugu which have a rich non-linear arrangement with the first and second consonants joined-up or ligatured to each other.

Surface Features of Vowel Diacritics

The description in the research literature on symbol blocks within the akshara writing system has been dominated to date by those akshara orthographies that have a preference for joined-up or ligatured markers (e.g., Bengali, Hindi, Gujarati, Kannada and Marathi). These forms of representation are arguably later inventions of Brahmi. The surface architecture of earlier forms of Brahmi—the Tamil Brahmi and Sinhala Brahmi from fifth and sixth century BCE for example—had many more disconnected elements and a less expansive use of the visuospatial space. Modern Tamil has retained this preference for unconnected elements and this defining surface feature is seen most prominently in the vowel diacritic system of the orthography (see Fig. 1).

The primary form of vowels is of full size and written in-line (e.g., உ, /u/; ஒ, /o/). The secondary form of the same vowels is a diacritic and is either a fraction of the primary form (இ - ி, /i/) or a visually distinct unit (ஆ - ா, /ɑː/). Vowel diacritics of /ɑː, e, eː, o, oː/ and diphthongs /ʌj/ and /ʌʊ/ sit in-line and without a connected point with the base consonant (e.g., பெ, பா and பொ; /pe/, /pɑː/, and /po/, respectively) while the vowel markers /pu/ and /puː/ are off-the-line but ligatured to the base (பு and பூ, respectively). Some vowel diacritics have one and others have two unconnected elements, such as in the examples above with /e/, /ɑː/ and /o/. The learning task here is therefore quite different from the learning task in other akshara systems where diacritics are ligatured or connected to a point in a base consonant. In Tamil, novice readers should learn to gather up the right number of unconnected elements and not treat each as separate akshara.

The location and shape of a vowel marker within the akshara symbol block is entirely predictable except for diacritic markers for /u/ and /uː/. The diacritic marker for vowel /u/ always begins from the 3rd quadrant but the shape depends on the base consonant (e.g., கு, சு, and னு; /ku/, /su/, and /nu/, respectively). More variable in form is the vowel /uː/ (e.g., சூ and நூ; /suː/ and /n̪uː/). Taken together, the vowel diacritic system in the Tamil orthography is mostly predictable in location within a symbol unit.

Opaque Features

Tamil orthography should be considered partially opaque for several reasons. At the level of the vowel, the few instances where ligaturing rules are needed are not consistent (see examples for the vowel diacritic u and u: above). Secondly, some vowels are visually confusable with some consonants (diacritic vowel form ை and ண, /ʌj/ - /nʌ/; primary vowel form ஐ and ஜ, /ʌj/ - /ʤʌ/). Third, two vowel diacritics share visual elements with <ɑ:>, <e>, and <e:> (ா, ெ, and ே); these are <o> (ொ) and <o:> (ோ). Added to this, some unconnected elements become more complex because when read as a concatenating whole they signify one sound value but if read in isolation each carries a different sound value (e.g., கௌ is /kʌʊ/, but கெ seperate from ள are /ke/ and /ɭ/). Readers probably use the context to decide which unconnected elements must be gathered up to form the akshara unit. Finally, even though the Tamil orthography is less non-linear in visuospatial arrangement than many other akshara systems, the non-linearity that does exist is because some vowel diacritics occur before, above and below the consonant, or on both the sides of the consonant (e.g., பெ, பி, பு, and பொ; /p/ + /e/, /i/, /u/, and /o/). Diacritics written off-the-line include the extended line below the consonant /u/ in /pu/ (பு), /ju/ (யு), and /ʋu/ (வு), the all-encompassing curl below the consonant /u:/ in /pu:/ (பூ), /ju:/ (யூ), and /ʋu:/ (வூ), and the minor curl above or in-line for /i:/ in /ki:/ (கீ), /si:/ (சீ), /li:/ (லீ) and /ʈi:/ (டீ). These features of the orthography could arguably make akshara processing more demanding.

Drop in transparency at the level of the consonant is because the Tamil orthography does not differentiate between voiced and voiceless stops. The symbol க represents /kʌ/, /gʌ/ or /hʌ/ and the symbol ச represents /sʌ/, /t͡ʃʌ/ or /ʃʌ/; in other words, the affricate and fricative are represented by the same symbol. A novice reader must learn to use context to decode these akshara. Going beyond the single symbol level, a further point of ambiguity is with phonological changes in sound values within words. For example, when the consonant /r/ (ற்) geminates it is pronounced as /ttr/ as in /ʋettri/ (வெற்றி) but when it follows a nasal consonant it becomes a /dr/ as in /sendrɑːn/ (சென்றான்). And finally, spellings within words may be opaque due to diglossia. For example, the written form of /kɑːleːlʌ/ (‘in the morning’) is <kɑːlʌjjil> (காலையில்), and <ɾeɳɖu> represents the spoken word /iɾʌɳɖu/ (‘two’, இரண்டு) (more details related to diglossia are given in the next section).

In summary, Tamil’s surface organization is more linearised than most other akshara orthographies. There is reduced stacking of markers compared to orthographies like Kannada, and no subscripts. The defining feature of the akshara block in Tamil is markers mainly placed to the right and left of the base, and use of the upper and lower quadrant of the visuo-spatial space.

Tamil Phonology

A description of the phonological system of Tamil must begin with a note on whether the description is about phonology as transcribed in written Tamil or phonology as found in the multiple spoken varieties of the language. This distinction is important because Tamil is a strongly diglossic language and apparent differences across scholars in their account of the phonology of the language are primarily because of the form described; and till recently, the focus was mainly on written Tamil. We closely follow the diglossia by first describing Tamil phonology as captured in the orthography followed by accounts of the spoken varieties. Our description is brief and hence selective; for a more complete description see Annamalai and Steever, 1998, Christdas, 1988, and Ramasamy, 2010.

Written Tamil has eighteen consonants which include five plosives (/k, s, ʈ, t̪, p/), six nasals (/n, n̪, ŋ, ɲ, ɳ, m/), two lateral approximants (/l, ɭ/), two trills/taps (/r, ɾ/) and three central approximants (/ɻ, j, ʋ /) (c.f., Yang, 2011). In addition, written Tamil includes five consonants that can be traced back to early borrowings from Sanskrit (/ʤ, ʂ, s, h, kʂ/). Written Tamil has four front vowels (/i, iː, e, eː/), two central vowels (/ʌ,ɑː/), four back vowels (/u, uː, o, oː/) and the two diphthongs, /ʌj/ and /ʌʋ/. All vowels in Tamil can occur in all positions except the diphthong /ʌʋ/ which does not occur in the word-final position. Among the consonants, /k, t̪, n̪, p, m, s, ʋ, ɲ, j, ɾ, l / can occur in the word-initial position; /n, ɳ, m, j, ɾ, l, ɭ , ɻ / can be found in the word-final position and all eighteen consonants can appear in the word-medial position. Alveolar and retroflex sounds do not occur in the word-initial position (Christdas, 1988) although Annamalai and Steever (1998) show occurrence of initial retroflexes in what they describe as native onomatopoeia. Consonant sequences are found in the word-medial position and these are typically nasals with stops /ŋg , ɲd͡ʒ , ɳɖ, nr, n̪t̪, mp/ in words such as <maː**ŋg**ɑːj>, <pʌ**mp**ʌɾʌm>, and <kʌ**ɳɖ**ʌm> (‘raw mango’, ‘spinning top’, and ‘continent’, respectively). Turning to consonant clusters, all consonants in the native Tamil repertoire except for /ɾ / and / ɻ / can geminate as in words like /pʌkkʌm/, /pʌʈʈʌm/, and

/t͡ʃeʋʋɑːj/ ('near', 'kite', and 'Tuesday', respectively). Some consonants are realised both in short and long duration (e.g., /k/ and /kk/). The long consonant is called a geminate. Three consonantal clusters are combinations of geminates /kk, t͡ʃt͡ʃ, t̪t̪, pp, nn, mm, n̪n̪/ preceded by /ɾ/, /j/ or /ɻ/. Native Tamil phonology does not allow clusters in the word-initial and word-final position, but loan words bring the exception to this rule (e.g., '**gl**obal'- க்லோபல், '**gl**ass' – க்ளாஸ், 'ba**nk**' – பாங்க், and 'pa**rk**' – பார்க்).

Tamil is an agglutinating language, and an important aspect of word-level phonology is around the morpheme boundary in inflected words (e.g., root + past tense marker in verbs) and compound words (e.g., *Chennai* + *pattinam* = *Chennaippattinam*, Chennai city). Adjacent morphemes trigger morpho-phonological processes with a strong influence of a rich set of phonological principles called *sandhi.* One abiding influence is to maintain sonority. When the morpheme ends in a vowel and the next word begins with a consonant, then there is a geminate (see example above of *Chennaippatinam*). Another combination at the boundary is a morpheme-final vowel followed by a morpheme-initial vowel. In such instances of consecutive vowels the consonant /j/ or /ʋ/ is inserted with the choice of consonant constrained by the value of the final vowel (e.g., *paravai* ('bird') + *ai* (accusative) is *paravai**y**ai* but *puu* ('flower') + *ai* (accusative) is *puu**v**ai*). Turning next to the occurrence of consonant sequences across morphemic boundaries, the examples are several: with nasals such as /ɳk, ɳp, nk, np/ in words like *thira**nk**al, ka**nk**aanippu* and *paya**np**aduthu* ('skills', 'tracking', and 'apply' respectively), retroflexes like /ʈk, ʈs, ʈp/ in words like ***utk**ondu*, *naa**tk**al* and ***utp**attu* ('feed', 'days', and 'subject', respectively), and trills in words like *aasiriya**rk**al*,***urp**aththi* and *pa**rk**kal* ('teachers', 'manufacture', and 'teeth', respectively). Similarly, there are several instances involving approximants at the morphemic boundary such as /lk, lp, lʋ, ɭk, ɭʋ, ɻʋ/ in words like *kappa**lk**al, sey-a**lp**aadu, and na**lv**aravu* ('ships', 'function', and 'welcome', respectively).

The Unaspirated-Voiceless and the Aspirated-Voiced Pattern

The Tamil phonological system comprises unaspirated and voiceless sounds and the orthography has dedicated akshara for these sounds. However, since aspirated and voiced sounds do occur in loan words, these are accommodated in the orthography by doubling the use of the available akshara set. Thus, akshara for unaspirated consonants are used to also represent the corresponding aspirated consonant and the akshara for voiceless consonants for the voiced counterpart (e.g., akshara க for /kʌ/, /gʌ/). This opaque feature of the orthography does not necessarily hinder decoding of words because, for the most part, voicing changes do not bring a change in meaning, although there are exceptions because of loan words (e.g., /bɑːʋʌm/ - /pɑːʋʌm/, 'expression' – 'sin', where *baavam* is a borrowing).

There is considerable debate on the characterization of voicing in Tamil phonology. While the classical text, Tolkaappiyam, does not differentiate between voiced and voiceless consonants and has named both types under the valliṉam consonant set, others propose that voiced consonants are specific to loan words. What is undisputed however is the number of positional constraints in Tamil phonology. For example, stop consonants in the initial position and geminates in the medial position should be voiceless (e.g., *saththam*, sound) but the consonant is voiced if in the medial position and followed by a homorganic nasal (e.g., *kaagam*, crow). The voicing rule for the stop consonants in the medial position is not straight forward. While the rule could be mostly true for the retroflex (/ʈ/), it is inconsistent for other stop consonants (an example of the lack of clarity is in the name *Prakalad* where the pronunciation of the akshara <ka> is also rendered as *Pragalad* and the Sanskrit borrowing, *Prahalad*).

Spoken Varieties

The description of Tamil phonology thus far has been with the formal, literary variety. In the spoken varieties, the phoneme inventory and positional constraints change. In Spoken Malaysian Tamil for example, the number of central vowels increases dramatically with use of not just the low central /a/ and /aː/ but also high central /ʉ/ and mid central /ʌ/ and / ə/. Voiced consonants also appear, as do other variations in the consonant repertoire (Ramasamy, 2010). Similarly, within the southern state of Tamil Nadu in India, the phoneme count varies across the dialects of the region. Keane (2004) reports 21 phonemes in the spoken Tamil varieties near Madras (the current Chennai area). Within this Madras Tamil variety are included fricatives, affricates, and voiced counterparts for plosives and an ongoing reduction in distinction on the /r/ and /ɾ/, and the nasals /n̪/ and /n/. Annamalai and Steever (1998) report 22 consonants including allophones, while other counts are more extensive. For example, Christdas (1988) demonstrated 35 phonemes in the Kanyakumari Tamil variety. A further characteristic of the spoken varieties across the region is that the distinction between phonologically-close consonants (such as l, ɭ, ɻ) is often blurred.

The ambiguity with voicing in Tamil phonology appears to make a difference to the pattern of acquisition of literacy. For example, in a comparison of simple frequencies of voicing in primary school children, Malini (1993) found variability in oral language sampled during casual speech, careful speech, word list reading and passage reading. In the word-initial position, use of voicing was more frequent than voiceless consonants in casual speech, but the opposite was true for passage reading. In other words, the rules of written Tamil were followed more carefully for word-initial phonology when children dealt with text. In the intervocalic word-medial position, voiceless velar fricatives were more common than voiced velar plosives in both causal speech and passage reading. It appears that children were over applying the rule by devoicing to word internal plosives. One important trend from the above data is that the mapping of voiced and voiceless plosives to orthography is a pressure point in Tamil reading for some children. We found some evidence of this in a reading survey of 120 children in Grades 2 and 3 in three schools in Karaikudi (Tamil Nadu, India). Three words—<uːɾd̪i>, <kɑːɳbi>, <sʌŋgʌd̪i>, ‘vehicle’, ‘show’, and ‘topic’, respectively—were chosen for analysis because all had the potential for devoicing errors on sounds appearing in the non-initial position in the word. We found 23.33% of children reading <uːɾd̪i> as /uːɾt̪i/, 9.16% reading <kɑːɳbi> as /kɑːɳpi/, and 5% reading <sʌŋgʌd̪i> as /sʌŋgʌt̪i/. These devoicing errors are made by a relatively small percentage of children, but the numbers are significant enough to suggest that fine-grained representation of voicing-devoicing is perhaps an important aspect of phonological knowledge in Tamil, and in turn of Tamil orthographic knowledge. It could be that children unfamiliar with the target words sounded out the written string using the default voiceless option for the ambiguous akshara (i.e., <d > read as /t/, <b > as /p/, <g > as /k/).

The trends from these exploratory reading studies suggest that it is reasonable to expect that phonological characteristics of the spoken varieties will also be a potential source of error in spelling words in standard literary Tamil. But it is also possible that classroom instruction may reduce its effects. In a Grade 3–4 narrative writing programme in Kannada for example, an (unintended) outcome was a reduction of diglossia effects on children’s spelling (Nag, 2013). The purpose of communicating for an audience seemed to perhaps increase use of the literacy variety in children’s writing. This pattern appears in the one other study we found on spelling performance in Tamil by Aaron and Joshi (2005): Most errors were on the retroflex and secondary vowel diacritic markers and characteristics of casual, conversational speech such as palatalization and cluster reduction were not seen; children appeared to always use the literary variety to inform their spelling. These interesting trends in contexts of Tamil diglossia clearly need research attention.

Tamil Morphology

Tamil is a morphologically rich language. Nouns and verbs in Tamil can be simple or compound and differ based on what they stand for. Adjectives and adverbs may be derived and there are several instances of compounding in the language. This section will describe features of noun and verb morphology to illustrate the morphological, morphophonological and morphosyntactic features of Tamil (for a comprehensive review see Krishnamurti 2003; for a comparison with other Dravidian languages see Amritavalli, 2017).

Nouns

Nouns (called *Peyarccol* in Tamil) are classified as rational and irrational, as in other Dravidian languages. Rational nouns include humans and the Gods, whereas irrational nouns are everything else, living and non-living. Rational nouns can be further classified based on gender and number (masculine, feminine, and neuter; singular and plural). Nouns in Tamil are inflected for number, gender, and case. The basic form of the plural marker is /–kʌɭ/, which is geminated to /–kkʌɭ/ if following a long vowel. The singular forms are unmarked. Examples of gender markers are /–kɑːɾʌn/ and /–kɑːɾi/ in words such as *samayal.kaaran* and *samayal.kaari* (male cook and female cook).

Case markers are bound morphemes suffixed to the root noun. Excluding the benefactive and vocative case described by linguists, the marking for the other seven cases are illustrated with the Tamil word for fruit, *pazham* (/pʌɻʌm/):

- The nominative case has no marker. Thus while the noun suffixed with other case markers will have varied final syllables (e.g., /pʌɻʌt̪t̪il/ and /pʌɻʌt̪t̪oːɖu/; 'in the fruit' and 'with the fruit'), the nominative case is the noun stem /pʌɻʌm/, 'fruit'.
- The accusative is marked by /-ʌj/. In the case of a rational and definite object noun phrase, the accusative marker is obligatory, but in the case of the irrational and indefinite noun phrase, it is optional (e.g., /pʌɻʌt̪t̪ʌj/).
- The instrumental and sociative case has multiple markers. The suffix /–ɑːl/ indicates causation and thus is instrumental (e.g., /pʌɻʌt̪t̪ɑːl/, 'with the fruit') and suffixes /–oɖu/, /–oːɖu/, and /–uɖʌn/ indicate connection and thus are sociative markers (e.g., /pʌɻʌt̪t̪oːɖu/, 'with the fruit').
- The dative case is indicated by /–ku/, /–kku/, and /–ukku/. While the suffix –kku occurs after the vowels /i, iː, ʌj, u/, and /–ku/ occurs on nouns ending with euphonic increments /in/ and /an/, /–ukku/ follows all the other nouns (e.g., /pʌɻʌt̪t̪ukku/, 'for the fruit').
- The ablative case of separation or motion is indicated by /–iɾun̪d̪u/. Inanimate forms take the marker /–il/ whereas animate nouns take the marker /–iɖʌm/ followed by /–iɾun̪d̪u/ (e.g., /pʌɻʌt̪t̪iliɾun̪d̪u/, 'from the fruit').
- A variety of suffixes communicate genitive case. Suffixes /–ʌd̪u/ and /–uɖʌjjʌ/ occur with nouns to make a complex noun phrase; suffix /–in/ occurs with nouns which function as noun modifier, its functions are no different from euphoric increments (e.g., /pʌɻʌt̪t̪inuɖʌjjʌ/, 'of the fruit').
- Locative case markings are most commonly /–il/ occurring with both inanimate and animate nouns, and /-iɖʌm/ with only animate nouns (e.g., /pʌɻʌt̪t̪il/, 'in the fruit').

A sub-class of nouns is pronouns referring to the first, second and third person. Interrogative pronouns are formed by the addition of the vowel 'e' to nouns while demonstrative pronouns prefix /ʌ–/ (/ʌʋʌn/ and /ʌʋʌɭ/, 'that' he and she), and /i–/ (/iʋʌn/ and /iʋʌɭ/, 'this' he and she). The first, second and third person forms are inflected for number but markers for gender are only carried in the third person. A feature that is shared with other languages in the Dravidian language family is that the third person plural marker differentiates between human and non-human but not gender or number. The 'we' in Tamil could be inclusive (/n̪ɑːm/) or exclusive (/n̪ɑːŋgʌɭ/).

Adjectives in Tamil (called uriccol) can occur naturally (e.g., /n̪ʌllʌ/ and /pʌɻʌjjʌ/, ‘good’ and ‘old’) or are derived from nouns by the addition of /–ɑːnʌ/ and /–uɭɭʌ/ (e.g., /ʌɻʌgɑːnʌ/, /irʌkkʌmuɭɭʌ/, ‘beautiful’ and ‘sympathetic’). Similarly adverbs can either occur naturally or are derived typically from nouns.

Verbs

Verbs (called vin̪aiccol) are of two types: the finite verb with tense and person-number-gender (PNG) markers and a non-finite verb which does not mark the PNG. PNG markers attached to verbs always agree with the nouns, and examples of these morpho-syntax dependencies are:

(a) *avalpaarththaal* (‘She saw’, /ʌʋʌɭ + pɑːɾ +t̪ + t̪ + ɑːɭ/; ʌʋʌɭ – she, noun; pɑːɾ – see, verb; t̪ – sandhi; t̪ – past tense; ɑːɭ – 3rd person singular female).
(b) *avanpaarththaan* (‘He saw’, /ʌʋʌn + pɑːɾ + t̪ + t̪ + ɑːn/; ʌʋʌn – he, noun; pɑːɾ – see, verb; t̪,- sandhi; t̪ – past tense; ɑːn – 3rd person singular male).
(c) *avarkalpaarththaarkal* (‘They saw’, /ʌʋʌɾkʌɭ + pɑːɾ + t̪ + t̪ + ɑːɾkʌɭ/; ʌʋʌɾkʌɭ – they, noun; pɑːɾ – see, verb; t̪ – sandhi; t̪ – past tense; ɑːɾkʌɭ – 3rd person plural)

Verbs in Tamil, similar to other agglutinating Dravidian languages, are inflectionally packed. Verbs inflect on tense, mood, and aspect. The present tense markers are /–kir, –kkir/ (e.g., /peːsukirɑːn/, ‘he is talking’), past tense markers are /–t̪ , –t̪t̪ , –in, –n̪d̪, -i/ (/peːsinɑːn/, ‘he talked’), and the future tense markers are /-p, –pp and –v/ (/peːsuʋɑːn/, ‘he will talk’). Imperative verbs occur only with the second person. Singular is unmarked and hence is the same as the verb stem; negation, plural and plural negative are marked by /–ɑːt̪/, /–uŋkʌɭ/, /–iːɾkʌɭ/, respectively. The indicatives also take a positive and negative form.

Returning to our focus on orthographic knowledge, the multiple sandhi rules and morphophonological processes will both be important to examine, particularly since diglossia permeates these aspects of Tamil as well. For example, in our survey in Karaikudi, we recorded a distinction between the Brahmin dialect and the working class dialect in a variety of multi-morphemic words (‘to lie down’: *taaccu* versus *paduththu;* ‘didn’t you pluck?’: *killiniyono* versus *killinella*; ‘didn’t I tell?’: *sonneenoonoo* versus *sonnenla*). These and other instances of diglossia around the phonological realisation of inflections and the morphemic boundaries most probably influence development of orthographic knowledge. While preliminary evidence of difficulty on morphemic boundaries is reported by Aaron and Joshi (2005) we were unable to find any research report examining whether inflections are pressure points in Tamil literacy acquisition.

Tamil Literacy Instruction

Our discussion of Tamil literacy instruction is specific to the practices recorded in the Indian state of Tamil Nadu. This state has a strong history of universalist social policies as well as trust in public institutions (Drèze & Sen, 2013). The public funded school system is administered with comparative efficiency and program continuity when compared to other states in India. Importantly, in the context of literacy instruction, successive governments have avoided fast-track measures that have gained in popularity in many low and middle income countries (e.g., low paid, part-time teachers, cash vouchers to private schools). Instead, the early years curriculum has been strengthened through multi-party partnerships (Neisz, Krishnamurthy & Mahalingam, 2012). The early grades program in this state is called Activity Based Learning, or ABL, referring to a commitment to a child-centered, activity- filled approach to learning. First introduced in 2003, the ABL method is one of India's celebrated stories of education reform. Reports on the efficacy of ABL in Tamil Nadu are limited to thick descriptions (e.g., Anandalakshmy, 2007), commentaries (e.g., Krishnamurthy, 2011) and secondary data analysis of children's outcomes (Aslam, 2016). No controlled trials are available.

Schools in the ABL programme use government prescribed material. Tamil textbooks are supplemented with multiple teaching-learning resources. These include a set of activity cards, a two-lined book for Tamil writing practice, several story books, and locally sourced materials. There are also homework cards and evaluation cards. The ABL programme began with a focus on activity cards but transitioned into a hybrid programme with a prescribed textbook and activity cards that mirror the textbook and additional material for reinforcement. The sequence of activities in the instruction scheme is illustrated as a fold-out paper ladder prominently displayed, within the child's reach. Learning steps for Grades 1 and 2 are together on one ladder and there is another ladder covering learning targets for Grades 3 and 4. This simple illustration captures the philosophy that children can move at their own pace; children in Grade 2, for example, can continue consolidating lower level skills, while those in Grade 1 can pull ahead if they have attained a particular milestone. This pedagogical commitment to child-paced instruction is compromised in the first instance by a more recent policy change that insists that all children must move up to the next grade at the end of each academic year; no child should fail and repeat an academic year. A second level of compromise is in the individual interpretations of the ABL programme within different schools (more details below).

Activities for each skill level (or steps on the ladder) may be seen to roughly correspond to a chapter in a traditional textbook. Each grade has multiple milestones, and in each milestone, there are multiple activities. Activity icons cover discussion, storytelling, singing, drawing and coloring, homework, and term tests. There are also subject-specific logos such as reading a poem, writing practice for akshara and word games. Activity cards are colour coded, with the cards for the Tamil language in blue (mathematics is maroon, science green, social science orange, and English pink). Finally, when a milestone is reached, a star logo on the ladder indicates that it is time for evaluation. In addition the logo of a book with a bouquet interspersed throughout the programme signals that the child can take a break from the textbook and activity cards to read books in the class library.

The first introduction to the akshara set is through visually similar but phonologically distinct akshara (e.g., ட, ப, ம; /ʈʌ, pʌ, mʌ/). These are taught as singletons, embedded in mono- or bisyllabic words, tagged to pictures, and through copying practice (see Table 1). The prescribed scheme for Grade 1 focuses on consonants with the inherent vowel /ʌ/, consonants with the vowel diacritic /ɑː/ and consonants as a phonemic unit. The introduction of the phonemic consonant in the first year is unique because akshara instruction regimes typically postpone this to the middle school years with consequences on development of both orthographic knowledge and phonological processing (e.g., Kannada: Nag, 2007). Examples of phonemic consonants (i.e. the akshara with the dot above) in Grade 1 are seen in these lists of rhyming words பட்டம், மட்டம், சட்டம், வட்டம், கட்டம் (/pʌʈʈʌm, mʌʈʈʌm, sʌʈʈʌm, ʋʌʈʈʌm, kʌʈʈʌm/; 'kite', 'level', 'law', 'circle', and 'grid', respectively), and words with the vowel diacritic /ɑː/ சாயம், காயம், தாயம்,மாயம் (/sɑːjʌm , kɑːjʌm , t̪ɑːjʌm , mɑːjʌm/; 'dye', 'injury', 'dice', and 'magic', respectively). This early introduction of the phonemic consonant is useful to initiate reading of authentic texts because several early acquired words in Tamil end in closed syllables which can only be written using a phonemic akshara. In addition, the presence of such lists suggests a focus on phonological awareness at the level of syllable and body-coda processing. Finally, a range of activities point to an oral language and print-rich curriculum for literacy learning (for a summary of the Grade1 curriculum, see Table 1).

Examining the Variety in Tamil Instruction

We sampled instruction in Grades 1 to 3 in three schools that followed the ABL methodology in Karaikudi, a Class 1 town in the southern district of Sivaganga in Tamil Nadu. Karaikudi with a population of more than a hundred thousand has literacy rates that are closely similar to the statistics of the whole state (e.g., as per Census of India, 2011, 81.5% of over 7 year olds in Karaikudi were considered as in-school or literate, the number is 80% for Tamil Nadu state). The main language of instruction in schools in our survey was Tamil (Tamil medium schools) and catered to families with equivalent socio-economic profiles. All the schools were run under private management but teacher salaries, ABL materials, children's uniforms and school meals were funded by the government. Despite the centralised ABL curriculum and controlled supply of ABL materials to schools, we found differences in literacy instruction. Each school had a different focal point from which the curriculum was delivered. While school 1 focused on the activity materials, school 2 stayed close to the textbook and school 3 transferred the focus to the blackboard. A point to note is that anecdotal evidence suggests that this variety is also present in government schools, although a systematic survey of classroom literacy practices within this sector is currently unavailable.

Table 1 A summary of the textbook for Grade 1

Term	*Oral language and print related activities*	*Singleton akshara*
1	Use of pictures: • Name and describe pictures with a focus on target sounds • Match pictures to given word labels • Label pictures Sound awareness: • Listen to monosyllabic and bisyllabic words with target sounds whose akshara will be or have been introduced • Identify words with target long or short vowels • Identify rhyming words Akshara and word reading: • Identify familiar akshara in a word, identify unfamiliar akshara in a word • Read words with target akshara, combine target akshara to form words • Read words which also have geminates and clusters of taught akshara • Write one akshara of the given words in each tile (e.g. /goːpuɾʌm/ 'ornate temple gate' - 4 tiles for <goː>, <pu>, <ɾʌ>, <m>, கோ. பு. ர. ம்) • Sequence words by the standard vowel order (e.g., *aṇil*, *aamai*, ... ('squirrel', 'tortoise',…) • Sequence pictures with target akshara in word initial position by the standard consonant-vowel order (e.g., *kaṇ, kaagam, kiḷi, kiiri*; 'eye', 'crow', 'parrot', 'mangoose') Word building and writing: • Copy akshara or words with the target akshara • Fill in akshara missing from all word positions, or in words with more than one akshara missing • Complete crosswords • Combine words to make new words • Arrange jumbled-up akshara to form words • Form new words hidden in a longer word • Substitute the medial akshara to form new words • Write short sentences of 2-3 words length with a focus on target akshara Spelling to dictation (Term 3): • Write the dictated word Reading connected texts (Term 3): • Read short phrases • Write the missing word in a sentence	The order of akshara introduction is with the inherent vowel and the vowel /ɑː/ • ட, ப, ம, ர, ழ (/ʈʌ ,pʌ ,mʌ ,ɾʌ ,ɻʌ/) • பா, மா, ய(/pɑː, mɑː, jʌ/) • க, ச, த, ந (/kʌ, sʌ, t̪ʌ ,n̪ʌ/) • ங, ற (/ŋʌ, rʌ/) • ல, வ, ன, ண (/lʌ, ʋʌ, nʌ, ɳʌ/) • ள, ஞ (/ɭʌ, ɲʌ/) These are introduced in parallel in words (see adjacent column). The above consonants are also introduced without inherent vowel (/C/, phonemic consonants) but these are always encountered in a word context
2		• அ, ஆ (/ʌ, ɑ :/– the primary form) • இ, ஈ (/i i:/) and diacritics ி and ீ with the taught consonant set, e.g., கி, கீ (/ki, ki:/) • உ, ஊ (/u, u:/) and diacritics with the taught consonant set
3		• எ, ஏ (/e, e:/) and diacritics ெ◌ and ே◌ with the taught consonant set • ஐ (/ʌj/) and diacritic ை◌ with the taught consonant set • ஒ, ஓ (/o, o:/) and diacritics ெ◌ா and ே◌ா with the taught consonant set • ஔ (/ʌʋ/) and diacritic ெ◌ள with the taught consonant set • Lists with standard vowel order, including lists with ஃ The entire akshara matrix of V, C and CV combinations

School 1: Activity Based Learning

Teachers in this school were amongst the first in the district to be trained in the ABL methodology, and had experience with activity cards without a textbook and, approximately 10 years on, were using both activity cards *and* the textbooks. Each day in school started with the teacher announcing the subject for the day. Children then individually picked up the card corresponding to the level they were in for that subject and settled down to work. Some children remembered their level but when unsure the teacher told them where to begin. In parallel, children identified their working group for the day from a set of six logos also shown on the activity card. Since two or more children could be at the same level, oftentimes children had to share a card. Mobility was not restricted; children could pick up their card and choose to write or work through the activity card at the low-level blackboard running the perimeter of the room. There was usually no group teaching unless two or more children were progressing at the same pace and were at the same level. The school also encouraged children to bring low or no-cost materials that were easily available in their neighbourhood (e.g., *rangoli* colours, vegetables, pumpkin seeds and cotton for craft work and class projects).

Teaching in this school was typically informed by guidelines implicit in the activity cards. For example, in Grade 1, writing cards provide arrows to show the formation of the akshara (e.g., for அ, the circle is shown first, followed by the curve below, followed by the horizontal line and finally the vertical line). We found these guidelines were used although not consistently. In addition, teachers also wrote on the board to show the sequence of strokes for each akshara.

School 2: Textbook Led Learning

The textbook was the main teaching resource in this school. In Grade 1, morning sessions typically had children naming and labelling pictures, and then copying and reading these picture names. These words were linked to target akshara (see Table 1). When a new word was introduced there was a lot of writing. First, the word was written with a pencil into the notebook and then practiced five times by writing with chalk onto a handheld slate. The same words were again assigned as homework to be written on a slate. Initially the teacher wrote new words in each child's notebook but 6 months on children were expected to copy independently from the blackboard. Afternoon sessions were dedicated to spoken language activities (picture description, singing, storytelling).

Ideas implicit in the ABL ladder were often renegotiated. Group activities and activities that needed locally sourced material were skipped. Opportunities for exploring the language through expressive language activities were limited. In second and third grades, any given day typically focused on the teaching of one subject with a new lesson from the Tamil textbook coming up about once a week.

The typical routine differed across teachers. One teacher taught each lesson by following a set sequence of activities: She put down on a handheld slate the textbook questions at the end of the lesson along with a standard answer and asked children to copy this into their notebooks. She occasionally included topics beyond the textbook focused on general knowledge. Another teacher had a very different sequence. She introduced the lesson one day, asked questions given at the end of the chapter the next day and, as children answered, she wrote on the blackboard. Children then copied this teacher-written answer into their notebook, carried it home, and copied both the question and the answers again three times as homework. This particular class also had daily spelling tests, and poor spellers retook the test. The second teacher appeared to have a closer engagement with what children wrote and learned from each lesson.

School 3: Blackboard Led Learning

In this school, each class had big-sized blackboards on three walls. When a new chapter was introduced, teachers wrote key words using a fraction of one blackboard. The next day the word list was moved to a board on an adjacent wall where it was left for a long time. We found that some information, depending on space and teachers' judgement of what still needed revision, could be left on one of the adjacent walls for up to a term. Since new words were constantly added to the adjacent boards, almost all the words for the term were available for a child to see. Every morning children read all the words on the board in a chorus. While choral lessons were the norm, the strategy for word-level reading changed by grade. In Grade 1, children read akshara-by-akshara (<a-ni-l > for /ʌɳil/ ('squirrel'), <si-*ng*-ga-m > for /siŋgʌm/ ('lion'). This sequence of sounding out and then blending occurred several times in a day. There were subtle exchanges during the sounding out to focus children's attention on vowel length and phonological realisation of confusable consonants. In Grade 2 the read aloud sessions did not have akshara-by-akshara sounding out. The writing on the wall and the chorus captured the teaching philosophy of this school*: A lot of reading aloud practice improves reading, and keeping word lists in the classroom allows for a long period of practice.* Approximately 90% of textbooks appeared to be covered with this method of instruction; the group activities and group projects were done to a lesser extent.

Even though there was a lot of teacher-written print in the classroom, children did not see teachers write very much. Most of children's writing was to copy from the large boards, and sometimes to write from memory. For Grade 1 the copying was of akshara and word lists; for Grade 2 it was questions and answers which were again recited every morning (e.g., *Who stands at the traffic signals? Traffic police.*). While children worked on a slate or the low-level blackboard, teachers moved around. Corrective feedback or further explanation occurred when a teacher stopped behind a child upon spotting an error. There was a lot of drilling, reciting, and copy writing practice in this school.

Taken together, we found distinct interpretations of the state mandated ABL programme in each school. For example, at the level of akshara instruction, teachers in school 1 brought focus on vowel length with intonation and an accompanying gesture. Children repeated words with this dramatisation at the appropriate point in a word. The routines were repeated when children read words to us, accompanying long vowels in words with lengthy articulation and an additional tilt of the head (e.g., /kɑːɳbi/, /mʌsɑːl ɑː/). Success rates on vowels appeared to be higher in Grade 1 compared to other schools, perhaps because that is when the practice is the most prevalent in this school (School 1, 2 and 3 at 48%, 21% and 30% respectively). However by Grade 2 all instruction strategies appeared to show similar vowel learning (Schools 1, 2 and 3: 58%, 40% and 56% respectively). The use of dramatised cues for long vowels appears to help jump start orthography-phonology mapping. But, from the limited data we have, there is no reason to believe that the sing song recitation followed in the other schools slows down symbol learning.

The similarities across schools are also worth highlighting. All schools kept a record of children's performance on state- supplied templates for formative assessment and summative assessment. All children, irrespective of their performance, compulsorily moved to the next grade at the end of each academic year. A struggling first grader had to start on activities for second grade when he moved to the next grade even if he needed more of the earlier placed activities. Equally, even if a first grader performed well, he moved to activities assigned to Grade 2 only in the next academic year. At the level of infrastructure, all classrooms had a low level blackboard running along the perimeter of the room with individual spaces for children to write. Slates were used often because the government supplied notebooks were too slim to last a full year of writing practice. In all schools teachers wrote on the board to show sequences of strokes to write the akshara and we saw feedback to individual children, but this occurred in a serendipitous fashion. Interviews with teachers and a review of classroom resources such as the charts, material written on blackboards, and children's note books showed that children were not taught the vowel diacritic in isolation but always in a symbol block with the consonant. Instead of this, teachers in all schools liberally used the rich Tamil vocabulary to describe the orthography, particularly the interior of the akshara. Finally, none of the schools expressly devalued dialect in language usage, but the tacit understanding between teacher and child was that standard Tamil pronunciation is expected when reading or spelling, and this was easily reinforced in joint read aloud sessions. This invariance in pronunciation for written Tamil by primary school teachers has been recorded by other authors as well (Geetha, 2012; Sugathapala De Silva, 1986).

Orthographic Knowledge in Tamil

Performance on single word reading and spelling requires multiple layers of orthographic knowledge (see Nag, 2017 for developmental trajectories in akshara-based decoding). Greater mastery of the rules of orthography may allow for higher accuracy and automaticity in reading and spelling. The very small body of literature that is available in the field of Tamil orthographic knowledge suggests that transparency assists in learning and that the younger reader and poor readers are particularly vulnerable to phonological errors (Aaron & Joshi, 2005; Bhuvaneshwari & Padakannaya, 2014; Nag-Arulmani, 2003). We extend this limited earlier literature with reports from two studies looking specifically at three aspects of orthographic knowledge: word reading, spelling and, for the first time in the literature, on early writing routines in Tamil.

Single Word Reading

Balambigai and Purushothama (2011) examined the role of orthographic knowledge in word reading with a group of 60 native Tamil-speaking children studying in Grade 3 in four government schools following the ABL curriculum described above. Most schools in the sample were close to the daily routines described under school 1 (Activity Based Learning). In each school, the class teacher was asked to rank order the students according to their academic performance, and for a contrastive analysis the highest achieving 10% and the lowest achieving 10% were taken for the study. Children were tested for speech and hearing difficulties, and if the results were positive they were dropped from the study. School records showed that all of the children in the high achieving group (30 children) were placed towards the upper end of the ABL ladder and the low achieving group (30 children) were at the bottom. Children from Grade 3 were chosen because that is probably when 'learning to read' changes to 'reading to learn' and this is the youngest possible age group who may be considered as proficient enough to assess several aspects of orthographic knowledge.

A word list of 71 words was taken from the Grade I Tamil textbook, and this allowed for ensuring all words were familiar to participating children in school. The list included at least two words with each vowel in the Tamil repertoire, and all consonants and consonants with vowel combinations. Thus a child who read the whole list correctly would have applied all possible akshara-level Tamil orthographic rules. The task was to quickly read words briefly exposed on a computer screen. The words were printed using the 'Baraha Tamil' font with font size adjusted automatically according to the length of the word. The longest word had five akshara (மிதிவண்டி, /mid̪iʋʌɳɖi/) though the maximum number of disconnected units was seven (ஔவையார், /ʌʋʋʌjjɑːɾ/). The word stimuli were randomized and presented in the same order. The exposure time for the brief exposure condition was controlled to be at one second duration. The interval between words (the inter-trial interval) was held constant.

The child sat comfortably in front of a computer screen on which the words were presented. Before the test, each child was instructed individually in Tamil to read words appearing on the screen as quickly and as accurately as possible. As the child read each word the vocal reaction time was recorded. This is the time between onset of the word on the screen and onset of the child's vocalisation as they read the word aloud. If an error was made, the error was noted down by the experimenter. This procedure was followed for all the 71 words before the second part of the study began. In this phase, words that each child skipped or misread during brief exposure were presented again to be read at a long exposure. This time the child was encouraged to read aloud each akshara, or spell out the word. This instruction was introduced because it could perhaps help differentiate whether the child's problem with reading a word was at the level of automaticity (that is, the child could read when given enough time but not when exposure to word was too quick) or at the level of knowledge of the orthography (that is, even with enough time to decode the child still struggled). Once again, the child's responses were noted down verbatim.

As expected, good readers read significantly more words than poor readers. When words were shown in a brief exposure, the good readers read 94.8% of words correctly while the poor readers read 65.31% correctly. The accuracy rate of the good readers was almost uniform across the group (SD = 2.78) reflecting the properties of the simple word list; an alternate list with low frequency words may have dispersed the scores even among the good readers. There was a scattering of scores among the poor readers (SD = 13.74), suggesting that, although all poor readers had lower accuracy when a word was briefly shown, some were much worse than others. Among the good readers, the majority of the words on the list (85.91%) were read correctly by all the children. Among the poor readers, all read 32.39% of words correctly and these words only had akshara with the inherent vowel, and the vowel diacritics for /i/ and /ɑː/.

To further determine the nature of orthographic knowledge, children's scores for reading at brief exposure and reading at long exposure were analysed. In the long exposure condition, good readers could read 71.81% of the earlier skipped or misread words suggesting that their greater difficulty was in decoding under the pressure of time, when given enough time they could apply their orthographic knowledge accurately. In the brief exposure condition, good readers struggled to identify words with grantha akshara, akshara with an inconsistent ligaturing rule (the vowels u and uː, e.g. கூ, ரூ), and akshara with non-linear arrangements. Most of these difficulties reduced with long exposure though accuracy remained low on words with diacritics for dipthongs /ʌʊ/ and /ʌj/, the vowel diacritic for /o/ (ெ◌ா), low frequency akshara /ɲʌ/ (ஞ) and the grantha akshara /ʤʌ/ (ஜ). In other words, some aspects of the symbol set and orthographic rules were still being mastered by the highest achieving readers in Grade 3.

The group of poor readers also improved their performance when given time in the long exposure condition. They could read correctly 67.11% of the words skipped or misread at brief exposure. Although poor readers improved their performance when they had time to decode a word, error analysis suggested shaky orthographic knowledge. More than 33.33% of poor readers had difficulty reading the following akshara at brief exposure: ahu (ஃ), ʌj (ஐ), ʋa: (வா), ʤʌ (ஜ), hʌ (ஹ), ʤi (ஜி), n̪i: (நீ), lu: (லூ), ɖu: (டூ), n̪o (நொ), mo: (மோ), ko: (கோ), ʂʌ (ஷ), pʌʋ (பௌ), se (செ), and s (ஸ்). Sixteen percent had particular difficulty reading the following akshara at brief exposure: o (ஒ), e (எ), ʌj (ஐ), ʌʋ (ஔ), su (சு), ku: (கூ), su: (சூ), ʤu: (ஜூ), mʌj (மை), pe (பெ), ne (னெ), and kʌʋ (கௌ). But even when read in leisure, the poor readers had difficulty with words containing the vowel diacritic ou (ெ◌ள), o (ெ◌ா), u: (நூ, லூ); and words with low frequency akshara (e.g., ஞ, நு), the grantha akshara (e.g., ஸ், ஜ), and the symbol ஃ. Other errors were with a disconnected unit of an akshara which lead to a change or omission in vowel diacritics, confusion among visually similar akshara, and poor knowledge of akshara in context. The latter error type is particularly interesting because it demonstrates the possible role that semantic knowledge could have to avert confusion when orthographic knowledge is shaky: as described earlier, the symbol பௌ can be read as /pʌʋ/ or /pe/ + /ɭʌ/, and must be decoded depending on the word context; thus the word may be / pʌʋɾɳʌmi/ (‘full moon’) as against the nonword /peɭʌɾɳʌmi/. Similar disambiguation is needed to read /gʌʋɾi/, /mʌʋnʌm / and /ʌʋʋʌjja:ɾ/ (‘a girl’s name’, ‘silence’ and ‘ouvaiyyar, name of a Tamil poet’) rather than the alternative orthographic possibility of /keɭʌɾi/, /meɭʌnʌm/ and /oɭʌʋʌjja:ɾ/, all of which are nonwords. Poor readers did not self-correct when their decoding strategy left them with a nonword.

The trends in this study show that poor readers in Grade 3 were yet to become adept at attending to the details of print. The vocal reaction time data mirrored this finding. Good readers were much faster in reading out a word when compared to the poor readers (see Fig. 2). The average reaction times in milliseconds for the good readers was 927.49 (SD = 77.48) and the poor readers was 1176.08 (SD = 79.07), and this was statistically a significant difference. Among the good readers, words that took the shortest time included words with primary vowels, consonants with inherent vowels or the vowel /i/ and phonemic consonants. Words that had the longest reaction time were words that included vowel diacritics with more disconnected units /o, o:, ʌʋ/, the low frequency /ʤʌ/, and the symbol ஃ. Unsurprisingly, words with more akshara also had longer vocal reaction times. Among the poor readers, words that took the shortest time were the same as for the good readers, but they still took longer when compared to the good readers. Words that took the

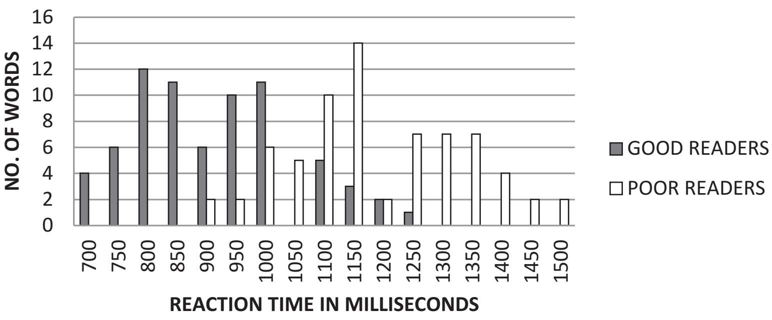


Fig. 2 The vocal reaction times of words read by Grade 3 good and poor readers following an exposure of 1 s per word (the brief exposure condition)

longest included all of the instances noted for good readers and also the /uː/ diacritic marker. In summary, greater mastery in orthographic knowledge is demonstrated with greater accuracy and speed of word recognition.

Spelling

The Tamil akshara system is particularly interesting for the study of spelling development because mastery is needed at several levels. First, as with other akshara orthographies, the number of symbols in Tamil can be expected to take time to learn. The sheer number of symbols also implies a potential for confusion when selecting the appropriate akshara for spelling, and this confusion may be at the level of closely similar sounds (phonological confusion) or similar looking symbols (visual confusion). In addition, two akshara features may reduce accuracy in spelling, namely, the presence of disconnected elements in the akshara block and non-linear arrangements. A further feature specific to Tamil orthography is the drop in transparency because one symbol represents both the voiced and voiceless consonants and the stop symbol represents the affricate and fricatives (ச for /sʌ/, /t͡ʃʌ/, /ʃʌ/, although the later can be written with a grantha consonant ஷ as well).

Finally, diglossia may disrupt spelling accuracy in Tamil. We examined the impact of all these features in the spelling of 193 native Tamil-speaking children from Grades 1 to 3. All children were from the three Karaikudi schools described above.

The words chosen were highly familiar and acquired before age five. They were balanced for parts of speech (nouns and verbs), mostly bisyllabic but some trisyllabic, all mono-morphemic and no loan words. To understand the nature of children's orthographic knowledge, an akshara-by-akshara analysis of spelled words was done after setting aside akshara that were illegible. The pattern of attainments in this sample broadly replicated uneven attainment patterns reported for the early grades in other akshara-based spelling studies (e.g., Kannada: Nag, Treiman & Snowling, 2010). Accuracy was higher on consonants than vowels and knowledge of visual features of the orthography was higher than knowledge that accurately maps phonology to spelling. All children, including the first graders, gave a spelling that broadly fitted the length of the dictated word; this is evident from the low occurrence of errors due to the addition of an extra akshara or omission of a phonemic marker within an akshara (1.75% and 6.30% respectively). Instead, the common error was of substitution (91.97%) with more than half of these substitutions on vowels rendered as a diacritic marker (56.85%).

A 'pure' visual confusion such as ை◌ written as ள was rare (0.19% of all errors), and confusion with the orthotactics of the writing system was seen in slightly less than 5% of all errors. Instead, most errors could be traced to confusion with a phonological source or a source where the role of orthography and phonology could not be separated. Among the few orthotactic errors seen, the types of confusion in order of frequency of occurrence were a) ignoring position constraints either on the vowel form (e.g., the primary vowel form after a consonant) or in writing the non-linearly arranged vowels in the same sequence as the spoken stream, and b) thinking of a disconnected element in an akshara as an akshara-like unit and substituting the full akshara with only these elements. None of the children including Grade 1 spellers made errors on ligaturing position, which, in Tamil, can become quite changeable for only two vowels (recall that vowel marker /u/ and /uː/ ligatures at different places depending on the shape of the base consonant). Unusually though, such advances in accuracy was not the case for unconnected markers suggesting that an increasing number of unconnected markers makes symbol learning more complex. The most vulnerable was the marker named tuṉaikkāl (ா), (close to 40% of all errors), because this is a visual element shared by multiple vowel markers. Errors in order of frequency were a) omission of the /ɑː/ marker, b) omission of the element in the /o/ and /oː/ marker, c) addition of the element to the /e/ marker, and d) substituting other vowel markers with the element.

The phonological errors covered the multiple sound families where distinction was blurred in speech even while retained in spelling. For example, there was confusion between the dental lateral (/l/), the retroflex lateral (/ɭ/) and retroflex approximant (/ɻ/) so that a word like /kiɻi/ was written as <kiɭi> (கிழி – கிளி). Similarly, there was confusion between the dental (/n̪/), alveolar (/n/) and retroflex (/ɳ/) nasals where, for example, the word /mʌɳi/ was written as <mʌni> (மணி – மனி). Similar trends were found on words with trills, with children mixing up the akshara for the dental tap (/ɾ /) and the alveolar trill (/r/). All of these errors were seen across all school types and across grades, with no evident influence attributable to the children's dialects. Other errors suggested that mastery over some aspects of the Tamil phonology was still emergent, such as ending a word with an illegal diphthong or vowel. But by far the most confusing for children in all schools was the short and long vowels and the diphthongs, though for the latter we saw some children approximating the word in spelling (e.g., /ʋʌj/ written as <ʋʌ.j>, வய் for வை).

A surprising finding from our survey, but in-line with trends reported in Nag (2013) and Aaron and Joshi (2005), was that the effect of diglossia on spelling accuracy was rare. Several words on our list should have attracted diglossia influences because of the spoken variety in the Karaikudi region. For example, the spoken form of <uʈkɑ ːɾ> is /ukkɑ ːɾu/ (*ukkaaru*, 'sit'), <pejʌɾ> is /peːɾ/ (*paer*, 'name') and <kɑːɳbi> is /kɑːmi/ (*kaami*, 'show'). In other words, although we began with the reasonable expectation that diglossia will be a potential source of spelling errors in Tamil, we did not find evidence for this. The data instead suggests that spelling conventions may be quickly learned, with one possible mechanism being a growing awareness among children that acceptable spellings (written Tamil) must follow a 'standard' phonological register.

Lastly, an intriguing error type, although rare in this sample (0.51%), was the use of English letter substitutions for Tamil consonants (e.g., 'b' for பி, as seen in காண்b for காண்பி, and தம்b for தம்பி). Children in this survey had minimal exposure to English in school, but the errors appear to suggest that there may have been English tutoring at home.

In summary, a survey of early grade spellings in Tamil shows overlapping advances in knowledge about multiple properties of the orthography. Children show emergent mastery over explicitly taught akshara sets for the vowels, consonants with inherent vowel, and consonants with vowel diacritics. Among the consonants with vowel diacritics, mastery is uneven, with mastery of a consistent ligatured marker among the first to emerge. Slower to learn are those vowel markers that are disconnected from the consonant, and among these, the slowest to stabilize in learning are those vowel diacritics that are ambiguous because they are shared by multiple vowel markers. Spelling studies are particularly rare in the Indic alphasyllabaries, but our limited survey has allowed us to outline the variety of parallel learning demands that may explain the nature of orthographic knowledge and spelling development in Tamil.

Writing Routines

The sequence with which children write out the word in a dictation task can shed light on the pressure points for learning the visuo-spatial arrangement of a Tamil akshara. We noted the writing routines children followed as they wrote the dictation task described above. All akshara in the word lists had been taught although schools differed in focus on writing routines and nature of practice to learn spellings (described earlier). We found that 93% of all children followed the routine of writing left to right, completing the writing of all elements of an akshara before moving to the next akshara in a word. The remaining 7% returned to an earlier written akshara to insert missed elements. Table 2 gives a case-by-case sample of unusual writing routines. Returning to write down a missed part of a word (returns) never occurred for consonants. Instead, all returns were on vowels and these vowels were always a diacritic marker and not in the primary form. Among the diacritic markers, the markers most vulnerable for returns were the ones written in-line and placed in a non-linear arrangement. This is a counterintuitive finding since the available literature in the akshara orthographies suggests that in-line nonlinearities are more vulnerable to errors.

Several points are worth noting from this pattern of writing routines: First, recall that 50% of Tamil vowel diacritics have an element of nonlinearity and are written before the base consonant. These are orthographic representations which are in mis-sequence of the speech stream (/ke/ written as /ek/). Despite the higher frequency of these arrangements, 93% of children wrote left-to-right ignoring the spoken sequence (ie. writing /ek/ even as they said /ke/). Importantly, the 7% of children who returned to insert a missed orthographic element did so more often for the missed linear elements rather than the non-linear elements (compare the last four columns, Table 2). A further point to note about returns on linear arrangements is that the tuṉaikkāl (ா) attracted the most returns in the writing routine. Since the tuṉaikkāl represents the vowel /ɑː/ and is also part of the vowels /o/ and /oː/, we propose that factors related to transparency may be more important than linearity for managing returns in writing routines. In summary, information from children's writ-

Table 2 Akshara writing during a spelling task: Returns to insert vowel diacritics

Child (Grade)	Initial written sequence	Final written output	In-line, linear arrangement		In-line, nonlinear arrangement	
			In-line, linear	Inserted unit	In-line nonlinear	Inserted unit
CH (1)	உத	உதை			✓	ை◌
KI (1)	சறு	சிறு	✓[2]	◌ி		
SR (1)	ச	செடி			✓	ெ◌
AB (2)	வ	வை			✓	ை◌
	உத	உதை			✓	ை◌
SA (2)	படு	பாடு	✓	◌ா		
RO (2)	பசா	பைசா			✓	ை◌
TH (2)	வேண்டம்	வேண்டாம்	✓	◌ா		
JA (2)	கண்	காண்பி	✓	◌ா		
ME (2)	பேடு	போடு	✓	◌ா		
MU (3)	செல்	சொல்	✓	◌ா		
	கெட்டு	கொட்டு	✓	◌ா		
	சதம்	சாதம்	✓	◌ா		
KA (3)	பெம்	பொம்மை	✓	◌ா		
TA (3)	கண்பி	காண்பி	✓	◌ா		
AR (3)	வேண்டம்	வேண்டாம்	✓	◌ா		
AN (3)	சத	சாதம்	✓	◌ா		

Note: In a sample of 193 children there were no returns to enter a missed consonant. [2]With the exception of this one instance, no child returned to insert a ligatured diacritic. All other returns were for disconnected elements of the vowel diacritic ing routines shows that orthographic knowledge related to linearities and nonlinearities are quickly learned by most children and across most Tamil akshara, while a drop in transparency slows down learning.

We also examined if there were differences in return-rate across schools. Returns to finish incomplete units were seen more often in the activity focused school (14%), than the textbook and blackboard focused schools (only 6.5% and 7%, respectively, returned). It may be the case that the focus on copywriting practice in the textbook- and blackboard-focused schools helped scaffold learning of stroke sequences. But it could also be that the children who returned to insert missing visual elements were the better readers and spellers who had spontaneously recognized errors in their writing and self-corrected, and the activity-focused schools nurtured this self-monitoring and self-correction. A comparison of errors on diacritics across the three schools suggests the later may be the case: Returns when writing are perhaps a correction strategy used by better readers and spellers. More research is needed to examine these relationships between instruction, writing routines and spelling accuracy.

End Note

In this chapter we set out to describe the development of orthographic knowledge after providing a description of Tamil orthography, phonology and morphology, and the variety in early literacy instruction. Of particular interest was the relative impact of orthography-specific transparent and opaque features and language-specific phonological and diglossia-related features. Our focus was on single word reading, spelling and writing routines of children in the first three years of literacy instruction when children are actively learning the surface layout of individual symbols and rules of the orthography. Vowel errors outnumbered consonant errors and phonological errors outnumbered errors because of visual confusion. We found orthography-specific features and ambiguity in orthography-phonology linkages intruding more than diglossia-related features. This trend is interesting since our classroom observations had not recorded specific instruction to raise children's awareness about intrusions because of the diglossia.

We acknowledge that our studies were set up to capture word level processing and it is possible that diglossia effects in Tamil may be more prominent at the level of syntax and sentence level processing. We also acknowledge that a dictation task may not be sensitive to capture diglossia effects (which may also explain the pattern of results in Aaron and Joshi, 2005). While trends from Nag (2013) suggest that increased opportunity for expressive writing reduces those spelling errors that are due to intrusions from the spoken variety, such spontaneous opportunities were rare in the schools in our survey. Our impression is that diglossia effects on spelling is perhaps more evident when children write spontaneously. Anecdotal reports from teachers seem to confirm our impression: They report that when children write memorized answers to questions set out in their book, diglossia effects are again low, but when children write stories or explain a taught concept in their own words, many words may be written as they are spoken (diglossia effect). The absence of strong diglossia effects in children's spelling is perhaps linked to two factors. First, our classroom observations had shown a strong focus on learning through writing, and this practice may have crystallized orthographic representation for literary Tamil. Second, there seemed to be no confusion in teachers' minds about the spelling of words, or when spellings reflect diglossia effects then they should be dismissed as incorrect. This assurance within the Tamil teacher community was recorded by Sugathapala De Silva (1986) in Tamil schools several decades earlier in a sociolinguistic analysis of teachers' attitudes to what the author termed the 'linguistic cleavages' due to diglossia. Tamil primary school teachers interviewed between 1974 and 1976 were unlike Sinhala, Kannada and Telugu teachers interviewed around the same time. While the latter teachers showed indecision, ambivalence or a diminishing allegiance, respectively, to teaching the 'high' literary form and accommodating diglossia influences, teachers of Tamil showed 'an unquestioned allegiance to the classic norm for formal use' (p. 305). The focus of literacy instruction was defined by all Tamil teachers as 'the ability to read and write *correctly*' (our highlight), where correctness was defined as 'the observance of the correct rules of spelling and grammar' (p. 309).The hold of Literary Tamil in literacy instruction in the early grades was more recently recorded in the areas of teacher education and children's literature in India (Geetha, 2012). The situation does not seem too different in Singapore, prompting the evocative description of the local Tamil population being 'tongue-tied' because of the continued focus on Literary Tamil in children's materials and literacy programmes (Schiffman, 2003). More research is clearly needed to understand when and how diglossia impacts Tamil literacy development, even when orthographic knowledge has been firmly guided to represent the standard language.

One orthography-specific process that demands further research attention is the processing of the visual element called tuṉaikkāl (◌ா). Errors related to the tuṉaikkāl are particularly interesting because in some instances the errors appear to be because of a confusion with long and short vowel sounds (a phonological error) but in other instances because of shaky knowledge about visual elements in an akshara (orthographic error). Disentangling the processing of this visual element may bring further insights into the phases of disambiguation in intra-symbol processing. Stepping back from the cognitive processes involved in the use of this visual element, the tuṉaikkāl is a symbol of macro-ecological factors influencing the learning task such that they increase the learning demand for the novice learner. As shown in Fig. 1, it was through orthographic reform that this element proliferated in the orthography. A second orthography-specific process we noted was the ambiguity that stems from multiple disconnected units in an akshara. Added to this are the demands placed by the voicing-devoicing phenomenon and the morpho-phonology in the language. All these issues would gain from systematic future research.

Acknowledgement This book chapter is funded by a seed grant on early literacy research to Balambigai Narayanan from The Promise Foundation, and a research grant to Sonali Nag from Brasenose College, University of Oxford. This chapter has gained from the critical comments received from Narayanan Srinivasan and K. Umaraj.

Appendix: Tamil Linguistic Terms with the Tamil Lexicon Notation (Madras Notation) and the International Phonetic Alphabet (IPA)

Srl No.	Madras Notation	IPA	Meaning
1	tamil̲	t̪ʌmiɻ	Tamil
2	centamil̲	sen̪t̪ʌmiɻ	Pure Tamil
3	koṭuntamil̲	koɖun̪t̪ʌmiɻ	Spoken Tamil
4	ṭāmili	dɑːmili	Tamil-Brahmi
5	tamil̲ariccuvaṭi	t̪ʌmiɻ ʌɾit͡ʃt͡ʃuʋʌɖi	Tamil orthography
6	uyirel̲uttu	ujiɾeɻut̪t̪u	The vowel and diphthong set
7	meyyel̲uttu	mejjeɻut̪t̪u	The consonant set
8	neṭil - kuril̲	n̪eɖil and kuril	The short and long vowel
9	vallin̲am	ʋʌllinʌm	The hard plosives
10	mellin̲am	mellinʌm	The soft nasals
11	iṭaiyin̲am	iɖʌjjinʌm	The medium consonants
12	āyutael̲uttu	ɑːjud̪ʌeɻut̪t̪u	Symbol ஃ
13	grantha		The borrowed symbol set from Sanskrit
14	puḷḷi	puɭɭi	The schwa suppression marker
15	tun̲aikkāl	t̪unʌjkkɑːl	Name for the unattached visual element ா
16	ottaikkompu	ot̪t̪ʌjkkombu	Name for the unattached visual element ெ
17	reṭṭaikkompu	ɾeʈʈʌjkkombu	Name for the unattached visual element ே
18	reṭṭaicul̲i ‘n’	ɾeʈʈʌjsuɻi *'n'*	Name for the type of visual elements in ள
19	mūn̲ṟucul̲i ‘n’	muːndrusuɻi *'n'*	Name for the type of visual elements in ண
20	canti	sʌn̪d̪i	Morpho-phonology rule
21	peyarccol	pejʌɾt͡ʃt͡ʃol	Noun
22	uriccol	uɾit͡ʃt͡ʃol	Adjective
23	vin̲aiccol	ʋinʌjt͡ʃt͡ʃol	Verb

References

Aaron, P. G., & Joshi, R. M. (2005). Learning to spell from print and learning to spell from speech: A study of spelling of children who speak Tamil: A Dravidian language. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy*. Mahwah, NJ: Lawrence Erlbaum.

Amritavalli, R. (2017). Morphosyntax of Dravidian languages. *Oxford research encyclopaedias of linguistics*.

Anandalakshmy, S. (2007). *Activity based learning: A report on an innovative method in Tamil Nadu*. Tamil Nadu Sarva Shiksha Abhiyan, National Informatics Centre. Chennai, India. Retrieved from http://www.ssa.tn.nic.in/docu/abl-report-by-dr.anandhalakshmi.pdf.

Annamalai, E., & Steever, S. B. (1998). Modern Tamil. In S. B. Steever (Ed.), *The Dravidian languages* (pp. 100–127). London, UK: Routledge.

Aslam, M. (2016). *Evaluation: more than an afterthought? India's activity-based learning programme*. Paper presented at Learning from learning assessments to leave no one behind, REAL Centre, University of Cambridge. Retrieved from https://www.educ.cam.ac.uk/centres/real/downloads/REALCentre_Conferencebooklet_2.pdf.

Balambigai, N., & Purushothama, G. (2011, December). *Role of knowledge of orthography and automaticity in differentiating good and poor readers in Tamil*. Poster presented at the symposium on language, literacy and cognitive development, The Promise Foundation, Bangalore, India.

Bhuvaneshwari, B., & Padakannaya, P. (2014). Reading in Tamil: A more alphabetic and less syllabic akshara-based orthography. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 192–201). Cambridge, UK: Cambridge University Press.

Christdas, P. (1988). *The phonology and morphology of Tamil*. (Unpublished doctoral dissertation). Cornell University, Ithaca, NY.

Drèze, J., & Sen, A. (2013). *An uncertain glory: India and its contradictions*. Princeton, NJ: Princeton University Press.

Geetha, V. (2012). Literacy and reading: A Tamil experiment. *Contemporary Education Dialogue, 9*, 63–84.

International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge, UK: Cambridge University Press.

Keane, E. L. (2004). Tamil. *Journal of the International Phonetic Association, 34*, 111–116.

Krishnamurthy, R. (2011, August 13). *செயல்வழிகற்றல் உருவானது எப்படி? புரட்சி அல்ல இது: பரிமாண வளர்ச்சி* (Ho0077 did ABL develop? It is not a revolution but an evolution.), *Dina Malar*. http://www.dinamalar.com/News_Detail.asp?Id=61541.

Krishnamurti, B. (2003). *The Dravidian languages*. Cambridge, UK: Cambridge University press.

Malini, R. H. (1993). *A sociolinguistic study of phonological variables in Tamil learning context at the primary school level* (Unpublished doctoral dissertation). Department of Linguistics, Bharathiar University, India. Retrieved from http://shodhganga.inflibnet.ac.in/handle/10603/101289.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22.

Nag, S. (2013). Low literacy attainments in school and approaches to diagnosis: An exploratory study. *Contemporary Education Dialogue, 10*(2), 197–221.

Nag, S. (2017). Learning to read alphasyllabaries. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 75–98). Amsterdam: Benjamins.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*(1), 41–52.

Nag-Arulmani, S. (2003). Reading difficulties in Indian languages. In N. Goulandris (Ed.), *Dyslexia in different languages: Cross-linguistic comparisons* (pp. 235–254). London, UK: Whurr.

Neisz, T., Krishnamurthy, R., & Mahalingam, V. (2012). *A history of the activity based learning movement in Tamil Nadu.* An independent report submitted to Sarva Shiksha Abhiyan, Tamil Nadu. Retrieved from http://docs.ehhs.kent.edu/ABLinTN.pdf.

Rajan, K., & Yateeskumar, V. P. (2013). New evidences on scientific dates for Brahmi script as revealed from Porunthal & Kodumanal excavations. *Pragdhara, 21–22*, 279–295.

Ramasamy, M. D. (2010). *Topics in the morpho-phonology of Standard Spoken Tamil (SST): An optimality theoretic study* (Unpublished doctoral dissertation). University of Newcastle upon Tyne, Newcastle upon Tyne, UK.

Schiffman, H. F. (2003). Tongue-tied in Singapore: A language policy for Tamil? *Journal of Language, Identity & Education, 2*(2), 105–125.

Sugathapala De Silva, M. W. (1986). Typology of diglossia and its implications for literacy. In B. Krishnamurti (Ed.), *South Asian Languages. Structure, convergence and diglossia* (pp. 304–311). Delhi, India: Motilal Banarasidass.

Tamil Lexicon Notation. (1939). University of Madras, India. https://archive.org/details/in.ernet.dli.2015.85194.

Yang, N. H. (2011). *Loan word adaptation in Tamil* (Unpublished master's thesis). Department of English Language and Literature, National University, Singapore.

Literacy Acquisition in the Malayalam Orthography: Cognitive/Linguistic Influences within a Multilingual Context

Mimisha Nesan, Amir Sadeghi, and John Everatt

Abstract Reading comprehension is a complex process that stems from the development of decoding and understanding the written form of a language. Reading development largely depends on the typological and orthographical features of a language. Hence, research investigating the impact of different writing systems on reading processes and acquisition is needed to inform reading models and teaching practices across language/learning contexts. Malayalam is a prominent Indic language, but has hardly been studied in reading research. Therefore, to stimulate such research, the present chapter explains the orthographical features of Malayalam, considering these in terms of cross-linguistic factors that are important for reading acquisition. The chapter then presents a review of the relevant studies in reading, focusing on akshara orthographies and those recognising metalinguistic awareness as an aspect of successful reading acquisition, particularly in multilingual contexts. The chapter ends by arguing that phoneme-based instructional strategies should be usefully applied to Malayalam, despite its akshara characteristics.

Keywords Akshara versus alphabet · Literacy acquisition · Malayalam language · Malayalam orthography · Meta-linguistic awareness · Multilingualism · Predictors of reading

M. Nesan (✉) · J. Everatt
School of Teacher Education, College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand
e-mail: mimisha.nesan@pg.canterbury.ac.nz; john.everatt@canterbury.ac.nz

A. Sadeghi
Department of English Language Teaching, Islamic Azad University, Damavand, Iran
e-mail: amir.sadeghi@canterbury.ac.nz

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_5

Introduction

This chapter discusses the Malayalam orthography and some potential cognitive/linguistic predictors of reading acquisition in this Indic language. The chapter initially provides a basis to understand the characteristics of Malayalam as an akshara orthography and the key aspects of the language. The existing published research is reviewed to provide the basis on which to inform further investigations in the area. Issues of the relationship between orthography and phonology are a primary focus of discussion, though consideration is also given to issues related to morpho-syntactic and semantic processing. The focus is on skills that potentially predict reading comprehension, but as part of this, skills that may potentially influence word recognition are also highlighted. Given the area within India in which Malayalam is learnt, the discussion also considers the acquisition of the orthography among children learning to be literate in a multilingual context.

Malayalam

Malayalam is one of the four major languages in the Dakshina Dravidian Language Family. It is mainly spoken in Southern India, predominantly in the state of Kerala, with nearly 35 million native speakers in the state (Ministry of Home Affairs, n.d. – retrieved August 2017). Etymologically, the word 'Mala' refers to 'hill' and 'Alam' to 'depths of ocean' or 'a place'. Hence, the word 'Malayalam' means 'a land lying between the Western Ghats and the Arabian Sea' or a 'hilly region'.

The Malayalam phonology has been greatly influenced by two Indic languages from two distinct language families. One is the Tamil language, from the Dravidian language family. The other is the Sanskrit language, which is also the language of religion and scholarship rooted from the Indo-Aryan language family. Malayalam combines Dravidian and Sanskrit phonology leading to the potential possession of one of the richest and most complex phonological systems among Indian languages. Malayalam consonant sounds embody nine places of articulation, including bilabial, labiodental, dental, alveolar, alveolar palatal, retroflex, palatal, velar and glottal (Jiang, 2010), and 10 manners of articulation including voiceless stop, voiceless aspirated stop, voiced stop, voiced aspirated stop, nasal, fricative, lateral, tap, frictionless continuant and finally glide (Mohanan, 1986). However, the literature presents inconsistent conclusions as to the number of sounds in Malayalam. For example, whereas Mohanan (1986) and Kumari (1972) delineated 43 consonant phonemes, Jiang (2010) reported 52 consonant phonemes, and Sadanandan (1999) stated only 38 consonant sounds. In contrast, most authors report that Malayalam contains 16 vowel sounds, which include 11 monophthongs and five diphthongs. In the Malayalam language, a diphthong is falling if it starts with a vowel of higher prominence (e.g. /au/) and ends in a vowel with less prominence (e.g. /ua/) (Jiang, 2010).

The syllable structure of Malayalam consists of an onset (consonants before the vowel sound), a nucleus (vowel sound), and a coda (consonants after the vowel sound). An onset can be up to three consonants. The nucleus can include a vowel or a diphthong. The coda can have only one consonant. The schematic representation of Malayalam syllables can be [C][C][C]V[V][C], where parenthesis denotes optional. For example, the syllable structure of the word സ്ത്രീ /st̪riː/ [woman] can be represented as CCCV. However, the CV structure is the most common syllable structure, followed by CCV.

The roots of the Malayalam script, like most of the Indian languages, can be traced back to the Brahmi script, which branched into Northern and Southern Brahmi Scripts. The North Indian languages, such as Hindi, Punjabi, Gujarati, and Bengali, utilized the Northern Brahmi-derived scripts. In contrast, languages such as Telugu, Kannada, Tamil, Malayalam and Sinhala descended from Southern Brahmi Scripts. The Grandha script, which emerged from the Southern Brahmi branch, directly influenced the evolution of the Malayalam aksharas. The advent of the printing press resulted in the decline of the Grandha script and gave rise to the modern Malayalam script. Additionally, in the 1970s and 1980s, Malayalam underwent major orthographic reforms, and the simplified version of the Malayalam script came into existence officially on 15 April 1971(Ramachandran, 1971). This chapter discusses the reformed Malayalam script, which is largely used in school text books in Kerala and other parts of the world.

Malayalam Orthography

Due to the lineage to the Brahmic script, like most of the Indian languages, Malayalam falls under the Indic writing system, which shares features of both alphabetic and syllabic scripts. Hence, unlike alphabetic scripts (e.g., English, Dutch, and German), which represent speech at the phoneme level, and syllabic scripts (such as Japanese), which represent speech at the syllable level, abugidas (aksharas), including Malayalam and other Indian and African languages, focus on consonant-vowel sequences as written units, but allows the consonant and vowel in most such units to be distinguished. In other words, each akshara represents a syllable rather than one sound (similar to a syllabic script), but the aksharas also incorporate phoneme markers. This special feature of the orthographic units, incorporating syllable-phoneme encodings, leads to them being referred to as akshara (as in this chapter) or akshram in Malayalam.

Orthographies can be classed from contained to extensive. In such a classification, scripts with fewer symbols are contained (English has 26 letters of the alphabet), whereas scripts with a large number of symbols (Chinese has more than 2500 fairly frequent characters, and thousands more that are less frequently used) fall on the extensive end of the continuum (Nag, 2007). Akshara languages are generally extensive orthographies. Bengali, Hindi and Kannada, for example, have more than

400 aksharas each. However, the Malayalam script reform in 1971 reduced the number of akshara symbols from an extensive symbol set of around 500 to around 90 units, making it a more contained orthography. The script includes vowel characters (Swaraksharam), diacritic symbols, consonants (Vyanjanaksharam), Chillaksharam or Chillus (which unlike other akshara symbols, do not include a vowel ending; e.g., ൻ /n/ and ർ /r/), and finally certain clustered and geminated consonant aksharas.

According to the psycholinguistic grain size theory (Ziegler & Goswami, 2005), based on the degree to which letters (or written characters) deviate from simple one-to-one grapheme-phoneme correspondence, the Malayalam orthography is a shallow, or transparent, orthography. Contrary to the English orthography, which is considered to have an irregular grapheme-phoneme relationship (Goswami, 2010; Share, 2008), most aksharas in Malayalam are pronounced as written, except for a few inconsistencies. For example, ന represents two sounds; one is dental /n̪/ and the other is alveolar /n/. Irrespective of these disparities in the grapheme-phoneme relationships, Mohanan (1986) argues that the orthographic representation of a word conveys almost everything that a reader needs to know in order to pronounce a word.

Vowels in Malayalam have independent and dependent forms. Vowels in their independent (or canonical) forms can occur only in the initial position of a word (e.g., the vowel ആ /ɑː/ in ആന /ɑːnɑ/). In other positions, a corresponding diacritic marker for the vowel is attached to the preceding consonant. Hence, in these cases, vowels take a dependent form.

Generally, a consonant in Malayalam can take up to three forms: (i) a consonant symbol in its primary form with inherent schwa sound (e.g., the consonant വ /vɑ/); (ii) a consonant in its geminated form (e.g., ക്ക /kkɑ/ which is the elongation of ക /kɑ/); and (iii) a consonant cluster (e.g., ന്ദ /n̪d̪ɑ/ which can also be written as ന് /n̪/ + ദ /d̪ɑ/). The reforms of the orthography mean that certain consonant clusters can be written in two ways: i.e., separate (like in English; e.g., sp) or as a singleton (as in the previous example of ന്ദ versus ന് ദ). In addition to these general forms, certain consonants, such as യ /jɑ/, വ /ʋɑ/, ര /ɾɑ/, can take a diacritic form.

Additionally, all consonant elongations (geminates), like some of the consonant clusters referred to in the previous paragraph, require a new akshara. For example, the letter ച /t͡ʃɑ/ geminates to ച്ച and 'ട' /ʈɑ/ to 'ട്ട'. However, in the case of consonant clusters, like other Indic scripts, they can be either written as separate units, simply by placing characters side by side in accordance with the sounds they represent, attaching virama (◌്) on the first letter to get rid of the vowel sound in the consonant, as in the consonant cluster ഞ് ച /ɲt͡ʃɑ/ (ഞ + ◌് + ച), or written as a single akshara unit, as in ഞ്ച /ɲt͡ʃɑ /. In the reformed Malayalam script, only a few ligatured consonant clusters are retained and others are written as separate aksharas using virama.

Like many other Indic scripts, Malayalam has a complex spatial organization, which makes it asymmetric and produces highly intricate shapes. While the English orthography has a linear organization in which vowels and consonants are placed in a linear left-to-right fashion, Malayalam consonants, like other Indian scripts, are written left to right but the ligaturing positions of the vowel signs to the consonant are non-linear and can be placed on either side of the consonant (ൈ◌ and ◌ാ), on both sides of the consonant (ോ◌ാ), and on the top of a consonant (◌്). A consonant can take up to 4 diacritic symbols. For example, the diacritic markers െ◌, െ◌, ◌്ര, ◌ം can be ligatured to ക /kɑ/ to produce ക്രൈം. In addition, certain vowel signs precede consonants in writing, whereas the reverse is the case when spoken. One example is the vowel symbol ൈ◌ /ɑɪ/ in കൈ /kɑɪ/. According to the reformed Malayalam script, the compounding of vowel signs does not bring any visual transformation to the basic consonant structure (e.g., the addition of the vowel symbol ◌ീ /ɪ::/ to the consonant ത /t̪/ is തീ /t̪i:/ and ◌ോ /ɔ:/ to ത /t̪ɑ/ is തോ /t̪ɔ:/). However, gemination and clustering of certain consonants produce some structural changes as described earlier. Alternatively, the reformed Malayalam script allows most of the consonant clusters to be written maintaining their primary forms by attaching halant/virma (in Malayalam, this is called ‘chandrakkala’ and is represented by the symbol ◌്) on the first consonant as in സ്പ /spɑ/. Here chandrakkala is ligatured to the first consonant സ to strip off the inherent vowel sound.

Another distinctive feature of Malayalam is the reflection of sandhi or the compounding of words. This feature leads to long and visually complex words, and can result in vowel and consonant deletion, in addition to assimilation or fusion within words at morpheme boundaries and/or at word boundaries. An example of external sandhi is in the word വിദ്യാഭ്യാസമില്ലെങ്കിലും, which can be translated to ‘even if no education,’ and consists of three words:

വിദ്യാഭ്യാസം+ഇല്ല+എങ്കിലും becomes വിദ്യാഭ്യാസമില്ലെങ്കിലും

Word boundary: 1 2

(/vɪdʝɑ:bʰʝɑ:sɑm/ + /illɑ/ + /eŋkiɭum/ = / vidʝɑ:bʰʝɑ:sɑmilleŋkiɭum/)

Here, at the first word boundary ◌ം /əm/ + ഇ /i/ becomes മി /mi/ and at the second word boundary ല്ല /llɑ/ + എ /e/ becomes ല്ലെ /lle/. An example of internal sandhi is അരി /ari/ + പൊടി /pɔdi/, which becomes the noun അരിപ്പൊടി /ɑrɪppɔdi/ [rice flour].

In short, the characteristics of the reformed Malayalam writing system lead to it being a more contained orthography in contrast to other typical akshara orthographies. However, it is still considered as more transparent compared to English, despite inconsistencies, which will be covered further in the subsequent section. In addition, it can be considered visuo-spatially complex (as detailed in some of the examples above) and contains a relatively large number of rules that need to be learnt to be able to combine written symbols appropriately. The latter point is also reflected in the use of morpho-phonemic inflections in the orthography. All of these features make the orthography interesting to study and may argue for it to be given special attention in the field of literacy research.

Orthographical and Phonological Inconsistencies in Malayalam

The presence of schwa is a unique characteristic of akshara orthographies, including Malayalam. The representation of schwa in a word depends on its position. While in the initial position, it is represented by അ /ɑ/, in the post-consonantal position it is inherent in the consonant. Problems related to dealing with this inconsistency in literacy acquisition have been documented in the literature. A study by Bhide, Gadgil, Zelinsky, and Perfetti (2014) examined schwa processing when orthographically marked and unmarked in adult biliterate Marathi-English readers. The study found that readers processed more words accurately in the marked writing system (English) than in the unmarked writing system (Marathi).

In addition, in light of the published literature on the Malayalam phonology, one can conclude that there are vowel sounds (such as /æ/ as in മാജിക് /mæɟikə/ [magic]) and consonant sounds (such as /pᵊ/ as in പാപം/pa:pᵊʌm/ [sin], /θ/ in പാതി /pa:θ i/ [half], /n/ as in പനി /pani/, /kᵊ/ as in പുക /pukᵊɑ / [smoke], /f/ as in ഫാൻ /fæn/ [fan], and / t͡ʃᵊ / as in പാചകം /pɑ:t͡ʃᵊkɑm/ [cooking]) that do not have corresponding orthographic units. Another instance of inconsistency between sounds and aksharas can be seen in gemination and clusters. The akshara റ /rɑ/ geminates to റ്റ /tɑ/, and in the cluster ന /nɑ/ + റ /rɑ/ becomes ന്റ /ntɑ/. In other cases, consonant clusters can represent two sounds. For example, the cluster ക്ല can represent the sounds /klɑ/ and /kɭɑ/, and another cluster ക്ര can represent the sounds /krɑ/ and /kɾɑ/.

Additional inconsistencies between spoken and written Malayalam can be perceived. For example, പന്ത് /pan̪t̪ə/ is spoken as /pan̪də/, and പാകം /pɑ:kəm/ as /pɑ:gəm/. Further, it is also important to note that the halant or chandrakkala (◌്) is added to the consonant to strip off the vowel sound. Hence, in a cluster, halant is silent. However, it functions as a diacritic marker at the final position of a word. Such an example is the word ഉപദേഷ്ടാവ് /upɑdeʂʈɑ:ʋə/ in which the first halant or chandrakkala (◌്) is silent whereas in the last position it represents the sound /ə/. Finally, the diacritic symbol ഃ, which is called *visarga*, represents the consonant /h/ sound after a vowel, but it is often silent in a word, as in ദുഃഖം /dukʰam/.

Phonological Awareness and Reading Acquisition in Malayalam

Data on literacy acquisition reveal that there is a strong relationship between phonological awareness and word reading (Chiappe & Siegel, 1999; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Share, Jorm, Maclean, & Matthews, 1984). Carlisle, Beeman, Davis, and Spharim (1999) also found that phonological awareness accounted for significant amounts of unique variance in a reading comprehension task. In line with such studies, phonological awareness has been found to exert a strong influence on literacy acquisition in akshara orthographies as well (Gafoor & Remia, 2013; Nag, 2013; Patel & Soper, 1987; Reddy & Koda, 2013).

However, despite the fact that phonological awareness is a major factor in literacy, its influence can vary depending on the typological nature of a language and script. While phonemic awareness has been reported to be a consistent predictor of later reading success in alphabetic writing systems, such as English (Bradley & Bryant, 1983), in logographic writing systems, such as Chinese, onset-rime awareness has been found to show larger correlations with reading than phonemic awareness (Siok & Fletcher, 2001). Conversely, owing to the combined alphabetic and syllabic nature of the akshara orthographies, there exist conflicting views concerning the relative superiority of phoneme versus syllable awareness in predicting literacy acquisition in such orthographies. Studies in akshara writing systems have shown that when compared to syllable awareness, phoneme awareness emerges gradually, although phoneme awareness is still an important predictor of reading development (Bhide, Gadgil, Zelinsky, & Perfetti, 2014; Nag, 2007; Nag, Snowling, Quinlan, & Hulme, 2014; Tiwari, 2011; Wijayathilake & Parrila, 2013). For example, Nag (2007), in a four-year longitudinal study of primary school students, reported that phoneme awareness was slow to emerge and that students showed a greater sensitivity to the syllable when compared with the phoneme. These features, Nag argued, were due to the unstable sound unit (syllable versus phoneme) in the orthographic representations in the language. Patel and Soper (1987) found similar results in their study of the acquisition of reading and spelling mechanisms in an akshara orthography among 120 Gujarati speaking students in grades 2 to 4. In contrast, the study by Somashekara, Das, and Jayashree (2014) investigated the relationship between phonological awareness and reading ability among Malayalam speaking children in grades 1 to 3 and found that a Malayalam phoneme detection task was easier than syllable-based tasks, which may suggest that Malayalam may need to be considered differently from other akshara orthographies. However, findings also suggest that phonological awareness development was not completed by the age of 8 years, and that phonological awareness skills are crucial at the younger ages tested but less so in later reading development. Hence, the specific influence on reading development of phoneme versus syllable awareness in Malayalam is still questionable, and requires further research.

Relationships Between Orthography and Phonology in Malayalam Literacy Acquisition

Issues of the relationship between orthography and phonology in reading development are widely studied across different orthographies, especially the principles pertaining to the mapping between the graphemes and various linguistic units, such as phonemes, syllables and morphemes. The visual distinctiveness of symbols has been shown to vary from one orthography to another, and so may influence literary acquisition across orthographies. Indeed, a range of studies on different orthographies argue that the visual demands of grapheme processing have a significant effect on grapheme recognition efficiency (Abdelhadi, Ibrahim, & Eviatar, 2011; Gupta, 2004; McBride-Chang et al., 2011; Pelli, Burns, Farell, & Moore-Page, 2006; Rao, Vaid, Srinivasan, & Chen, 2010; Su & Samuels, 2009).

Evidence from Malayalam also supports these convictions. Tiwari (2011) reports an investigation of reading acquisition in Malayalam-English biliterates learning to read and write in two distinct writing systems (i.e., akshara and alphabetic) at the same time. A total of 210 children participated in the study, comprising 30 participants from grades 1 to 7. The participants completed measures of phonological awareness, word and non-word reading, and orthographic awareness in each language. Findings revealed that phonological and orthographic knowledge in Malayalam emerged more slowly in comparison to English. The investigators claimed that this delay was probably due to the influence of the orthography to phonology mapping in Malayalam. This finding was corroborated by another study (Tiwari, Nair, & Krishnan, 2011), in which it was observed that children could attain mastery of consonants with inherent vowels in primary form (e.g., ത) and consonants with vowel diacritics (e.g., തി) by Grade 3. However, the mastery of consonant clusters (e.g., സ്ഥ /st̪ʰɑ/) could not be attained till Grade 3.

Gafoor and his colleagues (Gafoor, 2011, 2014; Gafoor & Kaleeludeen, 2008; Gafoor & Remia, 2013) investigated reading and spelling difficulties in Malayalam. In a study conducted among 100 Malayalam Grade 3 students, Gafoor and Remia (2013) identified some areas of spelling difficulties in Malayalam among elementary school students. These errors seemed to be due to similarities in script (e.g., the vowel diacritic markers ി /ɪ/ and ീ /ɪː/), errors from the influence of the colloquial language (e.g., the printed word for sweet in Malayalam is മിഠായി /miʈʰji/ but in the spoken language it is commonly pronounced as മുട്ടായി /muʈʈɑːji/), errors due to the confusion with consonant elongation, and finally errors from consonant clusters. In another study, while looking at developing tests for screening reading difficulties in Malayalam among upper primary school boys, Gafoor (2014) found that symbol-sound recognition, word fluency, phoneme segmentation and initial sound deletion were the most important variables for predicting reading difficulties among boys in grades 5 to 7.

Gafoor (2011) also conducted a more comprehensive study to analyse the level of achievement in reading, writing and arithmetic competencies among upper primary students in Kerala. A large number of upper primary school children (n = 1500) participated in the study. The results revealed that children showed greater difficulties in recognizing visually similar and phonologically similar written symbols. The study also showed that about 68% of the learners at grade 7 could not attain mastery level of the Malayalam akshara.

These studies argue for the importance of considering the relationships between Malayalam phonemes and graphemes, as well as other units of sound and text, when studying reading acquisition in this akshara. Children's knowledge of phonemes (i.e., the basic sounds of the language) and graphemes (i.e., the basic unit of the script) may be an important predictor of reading and spelling in this language as it is in many others (Hammill, 2004; Scarborough, 1998). It is also important to note that while in an alphabetic language, such as English, children master grapheme knowledge by grade 1 (Seymour, 2005), in akshara languages the development of mastery of written units can span beyond grade 4 (Gafoor, 2011; Nag, 2007; Tiwari et al., 2011).

Word, Sentence and Vocabulary Formation in Malayalam

Malayalam is a highly agglutinative language because what other languages express with helping verbs, pre/post positions and other additional words, Malayalam, like other Dravidian languages, expresses with suffixes that are added to word roots. Suffixes are added to nominal and verbal stems to mark grammatical categories. Hence, morphological variations appear to be more complex when compared to English or Hindi.

Malayalam enjoys a relatively free word order and permits freedom for sentence constituents. However, the default or unmarked order is the subject + verb order in intransitive and the subject + object + verb order in transitive sentences, where the subject appears first followed by the object with the verb normally remaining in the final position. An example sentence is അവൻ [he] എന്നോട് [me] ചോദിച്ചു [asked]. Noun modifiers precede nouns, while subordinate clauses precede the main clause. Noun modifiers occur in the noun phrase in the order possessive + demonstrative + numeral + adjective + noun, as in എന്റെ [my] ആ [that] ഒരു [one] കറുത്ത [black] പൂച്ച [cat].

Nouns are concatenated onto cases except for the nominative case and with the postpositions. Information such as negation, passive voice, interrogation, tense, and mood are conjugated to the verb, but verbs are not conjugated according to person, number and gender. It is possible to have a number of inflections to the verb root by adding suffixes in the order of base + tense + [aspect] + [mood] (items in parenthesis are optional). In some cases, an entire sentence can be expressed by a single word through the process of agglutination. Figure 1 presents an example of a conjugated verb in Malayalam.

In short, when suffixes are removed, the sentence becomes ungrammatical or semantically defective. Hence, Malayalam has morpho-syntactic features in which expression of grammatical relations is primarily indicated by concatenation and conjugations.

Malayalam boasts a large vocabulary. The language draws its rich diversity of words from Sanskrit, Urdu, Hindi, Tamil, Arabic, Persian, Portuguese, Dutch, French and English. This has resulted in a large number of synonyms. For instance, the word for moon has a Sanskrit origin word *chandran* and a Tamil origin word *thingal*, but other synonyms are *sasi, sasangan, amritakaranu, himamsu* and *seethikarananu*. For most discourse, the word chandran is commonly used, with the alternatives mostly found in poetry.

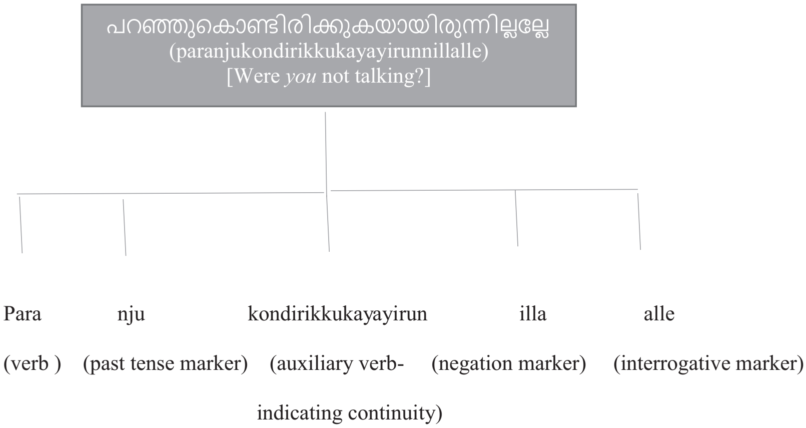


Fig. 1 An example of a conjugated verb in Malayalam

New words are formed in Malayalam mainly by processes of borrowing (words from another language; e.g., the words *mesa* [table], *savala* [onion], *thoppi* [hat] have Portuguese origins), compounding (stringing together existing words, as in *munpathram,* which combines mun + pathram [earthen pot]), derivations (forming new words from an existing word by adding prefixes or suffixes, as in *asrayam* [dependence] giving rise to *asrithar* [dependents], *asritha* [dependent woman], *asramam* [usually a secluded residency for a religious community and their guru], etc), blending (fusing of words together to make a new word as in *vidyardhi* [student] = vidya [knowledge] + ardhi [seeker]) and clipping (a process of shortening a word as in *vazhakka* (raw banana) = vazha [banana plant] + kaya [fruit]).

Impact of Agglutination and Sandhi on Automatic Word Recognition

Automatic word recognition is conceived as a characteristic of fluent reading (LaBerge & Samuels, 1974; Schadler & Thissen, 1981). Skilled readers are able to automatically access their lexicon (Acha & Perea, 2008; Arduino & Burani, 2004) and have a developed visual word recognition system which enables them to encode printed words in their mental lexicon. This results in the transition from serial letter-by-letter processing strategies to a more efficient, parallel and direct process of lexical access (Bowey & Muller, 2005; Duncan, Seymour, & Hill, 1997; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Word length effects are an important factor in investigating word recognition among beginning readers (Acha, Laka, & Perea, 2010; Acha & Perea, 2008; Bijeljac-Babic, Millogo, Farioli, & Grainger, 2004; Grainger & Jacobs, 1996) since letter-by-letter reading will be influenced by word length, but automatic reading will be more influenced by word familiarity. In this light, the typological properties of the Malayalam language require special reference. The agglutinative nature of the language (as presented in Fig. 1) makes the word longer, and compounding (sandhi) brings structural changes to words according to the morpho-phonemic, syntactic and semantic demands. For example, combining the words കത്ത് /kɑt̪ə/ [letter], എഴുതി /eɻut̪i/ [wrote] and ഇല്ല /illɑ/ [not] leads to the compound കത്തെഴുതിയില്ല /kat̪eɻut̪ijillɑ/ [did not write]. As can be seen, in the compound, the words do not retain their original form. In such instances, the strategy of retaining simple orthographic regularities is likely to be inefficient. Either the child has to learn the rules related to compounding words, and the changes in orthographic form that these require, or they may have to resort to using an indirect route in their decoding process (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). In the latter case, a more letter-by-letter decoding strategy may impede reading fluency. Further research is necessary to investigate the influence of this feature of Malayalam on reading processes.

Spoken Versus Written Malayalam and Its Implication for Reading

The Malayalam language follows different registers for speaking and writing. Correspondence between spoken and orthographic representations of phonology, grammar and vocabulary are not always direct (Prema, 2016). Prema distinguished a High and a Low variety of Malayalam. The High variety of Malayalam is used in speech on occasions such as public speaking, particularly if addressing an educated or urban audience, lectures, sermons, announcements for a general audience, news and, domains of religious, literary, philosophical and academic discourses and interviews. In writing, the High variety is used in text books, newspapers, career related writings, formal letters, invitation cards, notices, editorials, essays and narratives. The Low variety is normally used in every-day discourse.

This partial diglossia is due to the infusion of Sanskrit and Tamil in the language. Many words have both Tamil and Sanskrit derived words. For instance, the words for fire in Malayalam are തീ /t̪iː/ (Tamil loan word) and അഗ്നി /ɑgni/ (Sanskrit loan word). Since Sanskrit is considered to be a scholarly language, the Sanskrit originated words are mostly found in literary and formal writings. Caldwell (1875) rightly described Malayalam as the most Sanskritised of the Dravidian languages. In addition, the Malayalam language has many regional dialects which can be broadly classified as southern, middle, and northern Kerala dialects. The dialects also vary according to the socio, religious and economic status of the speakers. To cite examples for such differences between the standard Malayalam written register and its spoken dialects, the word for 'what' in written Malayalam are *enthanu, enthu, entha* but the spoken form of the same word can be*, enthonna, ennatha, enthutta, enthiru, enthee, enthuva* or *enthenu*. Sadanandan (1999), in her study on Malayalam phonology, indicated that Thiyya Talassery Malayalam (dialect) does not have voiced and voiceless aspirated stops. Gopinathan (1980) reported 21 grammatical differences between High and Low varieties of Malayalam. Prema (2016) argued for the need to consider both a paired lexicon (a lexicon which will comprise both High and Low pairs, as for the word ‘husband’ which has the High variety word ഭർത്താവ് /bʰɑrt̪ɑːvə/ and a Low variety word കെട്ടിയോൻ /ketᵊijɔːn/) and an exclusive lexicon (i.e., a lexicon in which a word does not have a parallel variety, as with the word അപ്രസക്തം /ɑprɑsɑkt̪ʌm/ [irrelevant] which does not have a Low variety). Moreover, this phenomenon is reflected in the Malayalam phonology. The High variety lexicon is distinct by voiced and aspirated features due to its relationship with Sanskrit. The Low versions of High variety words are achieved in speaking through phonological modifications. For example, the High variety word അഭിവൃദ്ധി /ɑbʰivridʰi/ [prosperity], can be spoken in its Low variety form /apʱivrit̪ᵊi/. Here, the voiced aspirated /pʰ/ and /dʰ/ become voiceless aspirated in the Low variety.

Further, given the context of disparity between the spoken and written forms with respect to phonology, morpho-syntax and semantics, children read in a language which can be considered as relatively unfamiliar. Hence, learners are required to negotiate mismatches between speech and print in order to become good readers. Additionally, it has been demonstrated across the reading acquisition research literature, especially in the Simple View of Reading (Gough, Hoover, & Peterson, 1996; Gough & Tunmer, 1986; Hoover & Gough, 1990), that linguistic comprehension plays a significant role in the development of reading comprehension in the child’s first language. Hence, against the backdrop of linguistic distance between the spoken and written forms, it is reasonable to hypothesize that these linguistic inconsistencies in the Malayalam language may interfere with early reading development.

Malayalam Reading Acquisition Within a Multilingual Context

The current overview of the Malayalam language and orthography, and the review of the existing research literature related to akshara literacy acquisition (including the relatively small number of investigations specific to Malayalam), argue for phonological awareness, orthographic knowledge and morpho-phonemic processing, along with vocabulary, to account for variability between individuals in their Malayalam literacy development. Each of these factors will need to be taken into account when developing models of reading and writing acquisition in Malayalam. Orthographic complexities, including large grain size, visual complexity, akshara-sound inconsistencies, morpho-phonological rules, as well as vocabulary size and differences in the use of spoken-written language, would all be predicted to influence literacy development in Malayalam. This range of factors is consistent with the multi-factorial model for skilled reading in Kannada proposed by Nag and Snowling

(2010). According to Nag and Snowling's model, reading difficulties in akshara orthographies can be considered multidimensional, such that weakness in one domain (such as poor akshara knowledge) may lead to reading difficulties, but a deficiency in multiple domains (poor akshara knowledge and weak phonological awareness) may have a greater impact on reading efficiency. The features of Malayalam argue that, in addition to the language skills that are the focus of Nag and Snowling's model, there will be a need to consider the child's abilities to deal with complexities in orthographic rules and variations in orthographic forms, as well as differing relationship between aksharas and sounds across contexts. This may lead to higher levels of influence of orthographic knowledge in reading, as well as an influence of metalinguistic skills that can support the child's awareness of when rules apply and when they do not. The difference between syllable-based versus phoneme-based correspondences when learning an akshara orthography is an obvious example, but so to is the feature of High and Low forms of Malayalam. These orthographic and linguistic skills may play a greater role in supporting acquisition than in some other orthographies (including other aksharas). Further research contrasting these potential predictors, and maybe contrasting Malayalam with other akshara and alphabetic orthographies would be useful.

Given the context of reading difficulties and akshara acquisition that the Nag and Snowling model predicts (and the current discussion proposes), literacy instruction in Malayalam should benefit from an understanding of these features and current reading models. However, a further feature of the context of learning Malayalam needs to be considered. Literacy outcomes in a multilingual context have been argued to be non-linear; see the Dynamic Model of Multilingualism by Herdina and Jessner (2002). Despite the multilingual literacy requirements (and needs) at schools in Kerala, where Malayalam is widely spoken, the present education system focuses on primarily monolingual instruction strategies, keeping language instructions rigidly separate. Although it is proper to keep a space between languages where confusion may be caused, it is also crucial to consider how cross-language transfer can be supportive rather than interfering. Consistent with this, a number of studies in akshara orthographies have reported cross-language influences on phonological awareness which suggest that skills in one can be related to skills in another language. For example, Nakamura, Koda, and Joshi (2013) investigated the contribution of decoding and reading comprehension sub-skills among Kannada and English biliterate children, tracking their development over 3 years. In this cross-language comparison, it was revealed that the weakest cross-linguistic correlation was in syllable awareness. Likewise, Gokani's 1992 study (as cited in Somashekara et al., 2014) found that children exposed primarily to Gujarati, another Indic writing system, performed better on tasks measuring syllable awareness, whereas those children who received more exposure to English performed better in phoneme awareness measures. In contrast, Somashekara et al. (2014) reported higher phoneme awareness sensitivity among those learning to read in Malayalam, which again argues for Malayalam to be considered slightly differently in terms of phonological processing from other akshara orthographies. However, Tiwari (2011) reports evidence that phonological awareness, including syllable and phonemic awareness, develops in Malayalam slower than in English, suggesting that English acquisition may support phoneme-level acquisition in Malayalam. This possibility is consistent with the findings that phonological awareness in both languages is significantly improved by grade 5. Such findings argue for the potential support in phonological processing within such a multiliterate context, and maybe the growth of metalinguistic awareness over the years (Bialystok, McBride-Chang & Luk, 2005).

Hence, since phonemic awareness is an important predictor of literacy development in akshara orthographies, including Malayalam, literacy acquisition in English may be useful to support the development of phonemic sensitivity in Malayalam despite the perceived argument by educators that akshara orthographies (such as Malayalam) should be taught via a focus on syllabic units. Such a bilingual instructional strategy may lead to phoneme awareness being developed in English and supporting subsequent use of phoneme-based strategies in Malayalam. However, links between the two orthographies will need to be made to avoid increases in confusion for some learners (particularly those with phonological difficulties).

Furthermore, given the existence of phonology-orthographic irregularities, spoken and written language inconsistencies, complex morpho-phonemic inflections, and free word order, high levels of metalinguistic awareness may be important to becoming a skilled reader in Malayalam. Metalinguistic awareness is reported as the ability to reflect on and manipulate the structural features of languages (Nagy & Anderson, 1995). Hence, a language instruction programme aiming at developing metalinguistic awareness skill may facilitate better literacy acquisition and outcomes in Malayalam. Again, further research investigating this possibility as a way of informing instructional strategies, which may benefit from the positive features of a multilingual learning context, would be beneficial.

References

Abdelhadi, S., Ibrahim, R., & Eviatar, Z. (2011). Perceptual load in the reading of Arabic: Effects of orthographic visual complexity on detection. *Writing Systems Research, 3*(2), 117–127. https://doi.org/10.1093/wsr/wsr014.

Acha, J., Laka, I., & Perea, M. (2010). Reading development in agglutinative languages: Evidence from beginning, intermediate, and adult Basque readers. *Journal of Experimental Child Psychology, 105*(4), 359–375. https://doi.org/10.1016/j.jecp.2009.10.008.

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology, 99*(2), 245–264. https://doi.org/10.1348/000712607x224478.

Arduino, L. S., & Burani, C. (2004). Neighborhood effects on nonword visual processing in a language with shallow orthography. *Journal of Psycholinguistic Research, 33*(1), 75–95. https://doi.org/10.1023/b:jopr.0000010515.58435.68.

Bhide, A., Gadgil, S., Zelinsky, C. M., & Perfetti, C. (2014). Does reading in an alphasyllabary affect phonemic awareness? Inherent schwa effects in Marathi-English bilinguals. *Writing Systems Research, 6*(1), 73–93. https://doi.org/10.1080/17586801.2013.855619.

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580–590. https://doi.org/10.1037/0022-0663.97.4.580.

Bijeljac-Babic, R., Millogo, V., Farioli, F., & Grainger, J. (2004). A developmental investigation of word length effects in reading using a new on-line word identification paradigm. *Reading and Writing, 17*(4), 411–431. https://doi.org/10.1023/b:read.0000032664.20755.af.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third-graders' silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 92*(3), 203–219. https://doi.org/10.1016/j.jecp.2005.06.005.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature, 301*(5899), 419–421. https://doi.org/10.1038/301419a0.

Caldwell, R. (1875). *A comparative grammar of the Dravidian or south-Indian family of languages* (2nd ed.). London, UK: Trübner and Co.

Carlisle, J. F., Beeman, M., Davis, L. H., & Spharim, G. (1999). Relationship of metalinguistic capabilities and reading achievement for children who are becoming bilingual. *Applied PsychoLinguistics, 20*(04), 459–478. https://doi.org/10.1017/s0142716499004014.

Chiappe, P., & Siegel, L. S. (1999). Phonological awareness and reading acquisition in English- and Punjabi-speaking Canadian children. *Journal of Educational Psychology, 91*(1), 20–28. https://doi.org/10.1037//0022-0663.91.1.20.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256. https://doi.org/10.1037//0033-295x.108.1.204.

Duncan, L. G., Seymour, P. H., & Hill, S. (1997). How important are rhyme and analogy in beginning reading? *Cognition, 63*(2), 171–208. https://doi.org/10.1016/s0010-0277(97)00001-2.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3–17. https://doi.org/10.1037//0022-0663.88.1.3.

Gafoor, A. (2011). Elementary competencies in 3R's among upper primary pupils of Kerala: A secondary analysis. *Innovations and Researches in Education, 1*(1), 51–68.

Gafoor, A. (2014). Tests screening reading difficulty in Malayalam among upper primary school boys. *Guru Journal of Behavioral and Social Sciences, 2*(1), 271–282.

Gafoor, A., & Kaleeludeen, C. (2008). Reading difficulties among upper primary school pupils in Kerala. *Journal of Studies in Teacher Education, 2*(1), 22–34.

Gafoor, A., & Remia, K. R. (2013). Influence of phonological awareness, morphological awareness and non-verbal ability on reading comprehension in Malayalam. *Guru Journal of Behavioral and Social Sciences, 1*(3), 128–138.

Gopinathan, V. P. (1980). Diglossic sistuation in Malayalam. In A. K. Karunakaran & S. Agesthialingom (Eds.), *Sociolinguistics and dialectology*. Annamalainagar, India: Annamalai University.

Goswami, U. (2010). A psycholinguistic grain size view of reading acquisition across languages. In N. Brunswick, S. McDougall, & P. de Mornay Davies (Eds.), *Reading and dyslexia in different orthographies* (pp. 23–42). New York, NY: Psychology Press.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1–13). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*(3), 518–565. https://doi.org/10.1037//0033-295x.103.3.518.

Gupta, A. (2004). Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing, 17*(1/2), 79–99. https://doi.org/10.1023/b:read.0000013823.56357.8b.

Hammill, D. D. (2004). What we know about correlates of reading. *Exceptional Children, 70*(4), 453–469. https://doi.org/10.1177/001440290407000405.

Herdina, P., & Jessner, U. (2002). *A dynamic model of multilingualism: perspectives of change in psycholinguistics*. Clevedon, UK/Buffalo, NY: Multilingual Matters.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127–160. https://doi.org/10.1007/bf00401799.

Jiang, H. (2010). *Malayalam: A grammatical sketch and a text*. Last retrieved October 2017 from http://www.owlnet.rice.edu/~hj3/pub/Malayalam.pdf

Kumari, S. (1972). *Malayalam phonetic reader*. Mysore, India: Central Institute of Indian Languages.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2.

McBride-Chang, C., Zhou, Y., Cho, J., Aram, D., Levin, I., & Tolchinsky, L. (2011). Visual spatial skill: A consequence of learning to read? *Journal of Experimental Child Psychology, 109*(2), 256–262. https://doi.org/10.1016/j.jecp.2010.12.003.

Ministry of Home Affairs, G. o. I. (n.d.). *The census of India 2011*. Office of the Registrar General & Census Commissioner, India. Last retrieved August 2017 from http://censusindia.gov.in/

Mohanan, K. P. (1986). *The theory of lexical phonology*. Dordrecht, Netherlands: D. Reidel Pub. Co.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22. https://doi.org/10.1111/j.1467-9817.2006.00329.x.

Nag, S. (2013). Akshara-phonology mappings: The common yet uncommon case of the consonant cluster. *Writing Systems Research, 6*(1), 105–119. https://doi.org/10.1080/17586801.2013.855621.

Nag, S., Snowling, M., Quinlan, P., & Hulme, C. (2014). Child and symbol factors in learning to read a visually complex writing system. *Scientific Studies of Reading, 18*(5), 309–324. https://doi.org/10.1080/10888438.2014.892489.

Nag, S., & Snowling, M. J. (2010). Cognitive profiles of poor readers of Kannada. *Reading and Writing, 24*(6), 657–676. https://doi.org/10.1007/s11145-010-9258-7.

Nagy, W. E., & Anderson, R. C. (1995). *Metalinguistic awareness and literacy acquisition in different languages*. Technical Report No. 618. Urbana-Champaign: University of Illinois, Center for the Study of Reading. Last retrieved October 2017 from https://www.ideals.illinois.edu/bitstream/handle/2142/17594/ctrstreadtechrepv01995i00618_opt.pdf

Nakamura, P. R., Koda, K., & Joshi, R. M. (2013). Biliteracy acquisition in Kannada and English: A developmental study. *Writing Systems Research, 6*(1), 132–147. https://doi.org/10.1080/17586801.2013.855620.

Patel, P., & Soper, H. V. (1987). Acquisition of reading and spelling in a syllabo-alphabetic writing system. *Language and Speech, 30*(1), 69–81. https://doi.org/10.1177/002383098703000106.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46*(28), 4646–4674. https://doi.org/10.1016/j.visres.2006.04.023.

Prema, S. (2016). *Is Malayalam a diglossic language?* Draft of paper presented at the 44th international conference of Dravidian linguists, Centre for Applied Linguistics & Translation Studies, Central University of Hydrabad, India, 16–18 June 2016. Last retrieved October 2017 from http://www.academia.edu/26401395/Is_Malayalam_a_diglossic_language

Ramachandran, P. (1971). *Malayalam script-adoption of new script for use-orders issued* (G. O. (P) 37/71/Edn). Trivandrum: Government of Kerala. Last retrieved October 2017 from http://www.unicode.org/L2/L2008/08039-kerala-order.pdf

Rao, C., Vaid, J., Srinivasan, N., & Chen, H. (2010). Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. *Reading and Writing, 24*(6), 679–695. https://doi.org/10.1007/s11145-010-9256-9.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31–74. https://doi.org/10.1111/1529-1006.00004.

Reddy, P. P., & Koda, K. (2013). Orthographic constraints on phonological awareness in biliteracy development. *Writing Systems Research, 5*(1), 110–130. https://doi.org/10.1080/17586801.2012.748639.

Sadanandan, S. (1999). *Malayalam phonology – an optimality theoritic approach* (Doctoral dissertation, Linguistics University of Southern California, Los Angeles, US). Last retrieved October 2017 from http://digitallibrary.usc.edu/cdm/ref/collection/p15799coll17/id/600186

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). Timonium, MD: York Press.

Schadler, M., & Thissen, D. M. (1981). The development of automatic word recognition and reading skill. *Memory & Cognition, 9*(2), 132–141. https://doi.org/10.3758/bf03202327.

Seymour, P. H. (2005). Early reading development in European orthographies. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 296–315). Oxford, UK: Blackwell. https://doi.org/10.1002/9780470757642.ch16.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584.

Share, D. L., Jorm, A. F., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*(6), 1309–1324. https://doi.org/10.1037//0022-0663.76.6.1309.

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology, 37*(6), 886–899. https://doi.org/10.1037//0012-1649.37.6.886.

Somashekara, H. S., Das, A., & Jayashree, S. B. (2014). Relationship between phonological awareness and reading abilities in Malayalam speaking typically developing children. *Language in India, 14*(2), 200–223. Last retrieved October 2017 from http://www.languageinindia.com/feb2014/somasekharareading.pdf.

Su, Y., & Samuels, S. J. (2009). Developmental changes in character-complexity and word-length effects when reading Chinese script. *Reading and Writing, 23*(9), 1085–1108. https://doi.org/10.1007/s11145-009-9197-3.

Tiwari, S. (2011). *Literacy development in the alphasyllabaries – implications for clinical practice*. Rapporteur's theme summary, language and literacy Cognititve development symposium, Bangalore. Last retrieved October 2017 from http://www.linguaakshara.org/yahoo_site_admin/assets/docs/Alphasyllabary_1_S_Tiwari.8094827.pdf

Tiwari, S., Nair, R., & Krishnan, G. (2011). A preliminary investigation of akshara knowledge in the Malayalam alphasyllabary: Extension of Nag's (2007) study. *Writing Systems Research, 3*(2), 145–151. https://doi.org/10.1093/wsr/wsr013.

Wijayathilake, M. A., & Parrila, R. (2013). Predictors of word reading skills in good and struggling readers in Sinhala. *Writing Systems Research, 6*(1), 120–131. https://doi.org/10.1080/17586801.2013.846844.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3.

Biliteracy Spelling Acquisition in Akshara and English

Pooja Nakamura, R. Malatesha Joshi, and Xuejun Ryan Ji

Abstract Spelling is one of the important literacy skills to be mastered. Most studies on the influence of first language orthography on spelling in a second language have been conducted primarily from one alphabetic language to another alphabetic language (e.g., English and Spanish) or from an alphabetic to a morphosyllabic language (e.g., English and Chinese). However, very few studies have been conducted on cross-linguistic influences from an akshara orthography – a unique orthography different from alphabetic, syllabic, alpha-syllabic, and morphosyllabic scripts – to an alphabetic language. The present study explored the influence of such an akshara orthography in spelling English words by administering phonological awareness, decoding, oral vocabulary knowledge, and spelling tasks to students in Grades 1–5 in low-income communities in South India. Results of quantile regression analyses showed that the orthography of the first language did influence spelling in English words. However, the contribution was different depending on the children's proficiency in English. Among students with higher proficiency in oral English, the influence of first language was less strong. Theoretical and educational implications are discussed.

Keywords Akshara · Biliteracy · Kannada · Low-income communities · Telugu · Spelling

Most extant literature on the influences of orthography on children's spelling development has been conducted with learners of English and other alphabetic languages. The relatively fewer studies that have focused on spelling acquisition in various

P. Nakamura
American Institute for Research, Washington, DC, USA

R. M. Joshi (✉)
Department of Teaching, Learning, and Culture, College of Education and Human Development, Texas A&M University, College Station, TX, USA
e-mail: mjoshi@tamu.edu

X. R. Ji
Department of Educational and Counselling Psychology and Special Education, Faculty of Education, The University of British Columbia, Vancouver, British Columbia, Canada

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_6

writing systems demonstrate that there are important orthographically-based differences that predict spelling skills in non-alphabetic languages (Ho, Yau, & Au, 2003; Joshi, Hoien, Xiwu-Feng, Chengappa, & Boulware-Gooden, 2006; Nag, 2011; Nag, Treiman, & Snowling, 2010; Shu, Anderson, & Wu, 2000). However, there is still a dearth of literature on the biliteracy acquisition of spelling skills in children learning akshara-alphabetic language pairs. This study aimed to address that gap by examining the relative contributions of phonological awareness, decoding, and oral vocabulary skills within and across languages in Telugu or Kannada (primary literacy) and English (secondary literacy) spelling acquisition.

While learning to read can be described as the process of converting print to spoken language, learning to write can be described as the process of converting spoken language to print. Although the underlying cognitive process in both involves matching sounds and symbols, the two processes are not cognitively symmetrical. Monolingual, English literacy acquisition studies underscore the fact that reading does not always necessitate a letter-by-letter process of matching symbols to their corresponding sounds, whereas spelling does require a letter-by-letter process of matching sounds to their corresponding symbols (Ehri, 1997, 2000). This task of having to select the correct grapheme from a set of two or more possible candidates adds a cognitive step in writing development compared to reading development (Geva, Wade-Wooley, & Shany, 1993). Furthermore, monolingual spelling acquisition studies tend to converge on the idea that the earliest stages of spelling acquisition are characterized by the straightforward application of the perceived phonological representation to the graphemic representation that is rendered, followed by more accurate spelling rules as the child's sight word repertoire increases (Ehri, 1997; Gentry, 1982; Henderson & Templeton, 1986; Read, 1975; Treiman, 1993, 2004; Treiman & Cassar, 1996). Not surprisingly, the impact of phonological awareness training – specifically, phonemic awareness training – on spelling ability has been clearly demonstrated (Ball & Blachman, 1991). As such, most monolingual frameworks of alphabetic literacy acquisition agree that the quality and integration of the phonological, orthographic and semantic representations of a lexical item contribute to early reading and spelling ability (Ehri, 2014; Perfetti & Hart, 2002).

Studies examining the cognitive processes underlying spelling acquisition in aksharas are far fewer. Even though many of the characteristics of uniqueness of the akshara orthography have been outlined throughout the volume, we would like to highlight some aspects here. The Indic writing system uses Devanagari script and the basic unit of writing is akshara. Even though the orthography is sometimes referred to as alphabetic, syllabic, and/or alpha-syllabic, the current trend is to classify it as a separate category (Daniels & Share, 2017; Share & Daniels, 2016). The earliest writing using akshara can be traced to the inscriptions of Asoka, circa 272–231 B.C.E and is arranged first by vowels followed by the basic consonants based on the articulation:– velars, palatals, retroflexes, dentals, and labials. This is the basic structure sometimes referred to as 'aksharamala.' In addition to the basic aksharas mentioned above, all aksharas also have a secondary form, which are smaller in size and can be placed either to the right, left, above, or below the primary consonant. Even though the akshara orthography is fairly transparent, decoding takes a long time to master due to the complex visuo-spatial structures. It may be possible to have 1000s of vowel-consonant combination forms, only about 400 clusters may appear in texts (Karanth, 2006; Nag, 2007). The sub-syllabic information may be encoded as 'matras' – vowel phonemic markers – which temporally follow a consonant, but may precede the consonant, or be placed above, below, or after the consonant in spatial representation. However, in Kannada and Telugu, they usually appear either to the right of or above the sound they follow. For more information on Kannada and Telugu, please refer to the previous publications by Nag and colleagues (Nag, 2007; Nag & Perfetti, 2014) and Nakamura, Joshi, and colleagues (Nakamura, Joshi, & Ji, 2017; Nakamura, Koda, & Joshi, 2013).

There is an emerging consensus that in the processes of acquisition of akshara knowledge, reading and writing are different from the acquisition of alphabetic knowledge in significant ways (Nag & Perfetti, 2014; Nag & Snowling, 2012; Share & Daniels, 2015). One example is that vowels and consonants do not have the same visual prominence leading to slower acquisition of the vowel diacritics (Nag & Snowling, 2012; Share & Daniels, 2015). Second, there is no resemblance between the diacritic form of vowels and consonants and their base forms, resulting in added challenges in the acquisition of the sound-symbol mapping (Share & Daniels, 2015). Third, there is a dual, yet asymmetric prominence of syllables and phonemes, leading to the role of phonological awareness in literacy acquisition being rather different between alphabetic and akshara writing stems (Nag, 2007; Nakamura, Joshi, & Ji, 2018; Reddy & Koda, 2013). Fourth, there is non-linearity in the placement of the graphemes which is not directly reflective of the placement of phonological units in speech, significantly impacting acquisition trajectories and timings of akshara scripts (Kandhadai & Sproat, 2010; Share & Daniels, 2015). All of these factors have also been implicated in contributing to the challenges in learning to spell Kannada (Nag et al., 2010). While all of these differences are rule-governed, the acquisition process is slower than the acquisition of alphabetic spelling due to these complexities and nuances (Nag, 2007; Nag & Snowling, 2012).

Studies on bilingual reading acquisition have consistently shown significant cross-linguistic effects (August & Shanahan, 2006; Koda, 2008; Koda & Reddy, 2008; van Gelderen Schoonen, Stoel, Glopper, & Hulstijn, 2007; Verhoeven, 2000). Although these effects are commonly referred to as "transfer" (Cummins, 1979, 1981), empirically, the mechanism of transfer can be traced to the sharing of resources that is common between the two languages (Koda, 2008). Metalinguistic awareness, especially phonological awareness and morphological awareness, has been identified as the shared candidate in numerous biliteracy studies (see Koda & Reddy, 2008 for an overview). This is not surprising given that the theoretical underpinning for sharing is the fact that universally, learning to read is governed by the mapping principle, which postulates that in all languages children need to become aware that graphemes carry phonological information (Perfetti, 2003). It is the mapping details – the precise phonological units that are graphed in each language – that constrain the degree to which transfer occurs (Reddy, 2008; Reddy & Koda, 2012).

The pattern of cross-linguistic influences in spelling are similar. Figueredo (2006) reviewed 27 studies of L1 influence on ESL spelling acquisition to find evidence for both negative transfer when there is a dissimilarity between the L1 and L2 linguistic and orthographic properties, as well as for facilitative resource sharing and cross-linguistic interdependence. Wang and Geva (2003) demonstrated positive and negative transfer in Chinese ESL students' spelling acquisition. The ESL learners' performance on pseudoword spelling was weaker than for L1 students; however, the ESL learners outperformed their L1 peers on a task of spelling legitimately and illegitimately spelled words. The authors hypothesized that this difference can be traced to the ESL students' positive transfer of visual orthographic processing of whole units in Chinese spelling. Li, McBride-Chang, Wong, and Shu (2012) provided further evidence of transfer, at the orthographic level, through significant relationships between Chinese and ESL spelling ability in 8 and 10-year-old Hong Kong Chinese children. These studies emphasize the role of cross-linguistic resources sharing, and consequently, transfer in bilingual spelling acquisition as well.

In another set of studies comparing L1 and L2 spellers, researchers showed that among the strongest predictors of L2 spelling in early grade children are L1 phonological awareness and L1 spelling skills (Sparks, Patton, Ganschow, Humbach, & Javorsky, 2008; Verhoeven, 2000). Verhoeven (2000) also demonstrated that L2 Dutch learners clearly lagged behind their monolingual Dutch counterparts in spelling acquisition in the early grades; however, the underlying predictors of phonological awareness, orthographic knowledge, and cipher knowledge were the same for both groups. The difference was that vocabulary played a much larger role in L2 reading comprehension for the L2 Dutch learners than for the Dutch-speaking participants. Jongejan, Verhoeven, and Siegel (2007) demonstrated that "lexical access" was more important than verbal working memory and syntactic awareness for ESL children compared to L1 English children. Taken together, these studies point to the notion that the orthographic and phonological properties of the L1, along with the semantic properties of the L2, impact L2 spelling acquisition.

Based on a combination of the orthographic properties of the akshara, research studies outlined above, and the transfer facilitation model theoretical framework (Koda, 2008), we hypothesized that L1 phonological skills and L2 oral language skills would explain significant variance in L2 akshara spelling acquisition. In order to begin addressing the paucity of cross-linguistic spelling acquisition studies, especially in akshara-alphabetic spelling acquisition, this chapter investigated the relative roles of L1 akshara (Kannada and Telugu) and L2 alphabetic (English) sub-skills in L2 English spelling acquisition. Specifically, we asked the following research questions:

1. What are the overall contributions of phonological awareness, decoding, and oral vocabulary knowledge explaining spelling acquisition within each language depending upon the level of spelling acquisition (L1 Kannada or Telugu, and L2 English)?

2. What are the overall contributions of L1 Kannada or Telugu phonological awareness and L1 Kannada or Telugu decoding to L2 English spelling?
3. What are the overall contributions of L1 Kannada or Telugu phonological awareness, L1 Kannada or Telugu decoding and L2 PA, OVK, and decoding for L2 English spelling?

Method

Participants

The present study analyzed reading and spelling data from 320 children from Grades 1–5 in two states of South India – Andhra Pradesh and Karnataka. The data were part of a larger study on literacy acquisition of children from very low-income communities in Karnataka state and Andhra Pradesh state in India. The mean age of the children was 5.87 (n = 23, SD = 0.81) in Grade 1, 7.07 (n = 68, SD = 0.98) in Grade 2, 8.44 (n = 68, SD = 1.56) in Grade 3, 9.28 (n = 74, SD = 1.32) in Grade 4, and 10.71 (n = 83, 1.77) in Grade 5. Participating students came from 13 different schools that were both urban and rural, and government-run or operated by non-governmental organizations (NGOs). About 15% of the sample reported speaking a mother tongue at home that was different from the medium of instruction in the respective schools. The mother tongues included Kannada, Telugu, Tamil, Urdu, Hindi, and Marathi.

Measures

Deletion Phonological deletion was measured in Kannada and Telugu in the L1, and in English as an L2. The test was conducted by asking participants to say a word out loud after removing a phonological unit from it. Given the dual syllable and phonemic level representation of phonology in aksharas, both syllabic deletion and phonemic deletion items were included. An example of a deletion task in Kannada/Telugu is to ask a child to say /naadu/ ನಾಡು and then to say /naadu/ ನಾಡು without the syllable /du/ ಡು for a syllable-deletion task and without the phoneme /u/ for a phoneme-deletion task. An example of a syllable deletion task in English is to ask the child to say ‘pencil’ and then to say pencil without saying ‘pen’ and, for the phoneme deletion task, to say the word ‘cat’ without /k/. The target deletion phonological unit occurred in the beginning, middle, or end of the word. There were 12 phoneme deletion items in both Kannada and Telugu and 16 and 15 syllable deletion items in Kannada and Telugu, respectively. Cronbach’s alpha coefficient was .96 for syllabic awareness in Kannada and Telugu, combined; and was .97 in

English. Cronbach's alpha coefficient was .98 for phonemic awareness in Kannada and Telugu, combined, and was .95 in English.

Oral Vocabulary Knowledge Based on the Peabody Picture Vocabulary test-Revised (PPVT-R; Dunn & Dunn, 1981), a receptive oral vocabulary knowledge test was constructed. Utilizing pictures drawn for the Indian context specifically for the study, each child was shown a card with four pictures and was asked to identify the pictures that matched the word that was spoken by the test administrator. After two rounds of piloting aimed at removing items that were deemed unfamiliar to children in low-income communities in India, there were 40 target words in Kannada, 49 in Telugu, and 25 in English. Words were all drawn from a word bank generated after an analysis of the government textbooks. One point was given if the correct picture was selected and no points were given if the wrong picture was selected, or if the child did not point to any picture. Cronbach's alphas were .90 in Kannada and Telugu, combined, and .91 in English.

Decoding The ability to read out isolated words with accuracy was measured through a researcher-created decoding instrument. Students were presented with each word printed in black ink on individual flashcards. They were instructed to sound out the word as clearly as possible as soon as they were shown the flashcard. The students were given 1 point if they sounded out the word correctly, 0.5 points if only 1 phonological unit was sounded out incorrectly or if there were any large, unnatural pauses before the word was fully decoded, and 0 points if more than 1 phonological unit was sounded out incorrectly, or if there were multiple stops during decoding. The word list consisted of both real words and pseudowords, including single akshara syllabic units (vowels and consonants), and more complex akshara syllabic units (consisting of multiple component phonemic units). The Cronbach's alpha for these tests were .90 in Kannada, .97 in Telugu and .98 in English.

Spelling Spelling was assessed by asking students to write down a list of words that were dictated to them in each language. The words list was developed using the same criteria for items as the decoding word list. Scoring was also the same, where students scored either 1, 0.5, or 0. Cronbach's alpha was .94 for Kannada and Telugu, combined and .91 for English.

The descriptive statistics of the literacy measures are shown in Table 1.

Analytic Procedure

Four conventional linear regression models were used to address the research questions as posed above:

Model1. Spelling (L1) ~ PA(L1) + Decoding (L1) + OVK (L1),

Table 1 Descriptive statistics of literacy skills measures

	M	SD	Range
Kannada/Telugu			
Spelling	0.29	0.24	[0.00, 0.97]
Deletion	0.55	0.29	[0.00, 1.00]
OVK	0.76	0.20	[0.05, 1.00]
Decoding	0.69	0.32	[0.00, 1.00]
Second language			
Spelling	0.08	0.13	[0.00, 0.67]
Deletion	0.40	0.31	[0.00, 1.00]
OVK	0.48	0.27	[0.00, 0.96]
Decoding	0.23	0.30	[0.00, 1.00]

Model2. Spelling (L2) ~ PA(L2) + Decoding (L2) + OVK (L2),
Model3. Spelling (L2) ~ PA(L1) + Decoding (L1),
Model4. Spelling (L2) ~ PA(L1) + Decoding (L1) + PA(L2) + Decoding (L2) + OVK (L2),

where L1 refers to Kannada/Telugu combined; L2, English; PA, phonological awareness; and OVK, oral vocabulary knowledge.

Due to the highly skewed distribution of response variables (skewness $_{spelling(L2)}$ = 2.43; Skewness $_{spelling(L1)}$ = 0.88) traditional linear regressions (i.e., conditional means model) were likely to lead to biased results in the present study. Specifically, the estimated association would have been weaker due to the influence of the skewness, possible floor effects, or the magnitude of the association may have been overestimated (Petscher & Logan, 2014). Therefore, we also conducted quantile regression analyses for each model on the 25th percentile, 50th percentile (median), 75th percentile, and 90th percentile of response variables (i.e., Spelling (L1/L2)). Moreover, we combined four grade levels groups into one group for data analysis. Because age varies across grade levels, age might also contribute to the variation in the L1 and L2 spelling performance. Therefore, we also included age as a covariate during modeling. However, the contribution of age was not significant enough to be recognized in detail in the results section. Accordingly, we retained the models without the covariate of age, and present results from both conventional regression and quantile regression for comparison. The conventional regression analysis assumes that there would be a single regression line that best fits the data to represent the association between dependent variable and independent variables. The quantile regression allowed us to examine the association between the dependent variable and independent variables based on the level of response variable. That is, this method would estimate multiple regression lines simultaneously. In our study, three regression lines were estimated based on the level of spelling skills in L1 and L2 at four percentile levels. As a result, we could obtain multiple estimated associations between spelling skill (L1 and L2) and literacy skills (i.e., phonological awareness, decoding, and oral vocabulary knowledge) based on the level of spelling skill (L1 and L2).

Results and Discussion

We found that the associations between spelling skills and other literacy skills (PA, decoding, and OVK) were overestimated if we only employed the conventional linear regression model. More specifically, the results (i.e., the overall *R*-squared) of our conventional linear regression model overpredicted as compared to the results of quantile regression results, especially at the 50th percentile level (i.e., median; see Tables 2, 3, 4, and 5). The results from both estimation methods should have been approximated to be the same, if the skewness of response variables was relatively minor. These results showed that it was necessary to incorporate the quantile regression into the analysis. Moreover, the results of overall R-squared with age as a covariate are also presented in the brackets in Tables 2, 3, 4, and 5. The differences were not significant, and, therefore, age was dropped from the modeling.

Overall Contribution of PA, Decoding, and OVK to Spelling Skills (L1 and L2) We found that within both L1 and L2, there was an increasing trend in associations (i.e., overall R-squared) between spelling skill and other literacy skills (PA, OVK, and decoding) as the spelling skill (L1 and L2) level increased. Namely, PA, OVK, and decoding skills had a higher explanatory power at the higher level of

Table 2 Model 1. Overall contribution of deletion, decoding and oral vocabulary knowledge to spelling skills in Kannada and Telugu (L2)

Model	Parameter	Estimate	SE	t-value	p	LB	UB
Linear regression – L1 Spell ~ L1 PA (Del) + L1 Decoding + L1 OVK							
$R^2 = 0.53$	Del	0.18	0.05	3.75	0.00	0.08	0.27
[0.53]	Decoding	0.40	0.04	9.12	0.00	0.31	0.49
	OVK	0.07	0.06	1.14	0.26	−0.05	0.19
	Intercept	−0.14	0.04	−3.80	0.00	−0.21	−0.07
Quantile regression – L1 Spell ~ L1 PA (Del) + L1 Decoding + L1 OVK							
QR-25	Del	0.04	0.04	1.17	0.24	−0.03	0.12
$R^2 = 0.29$	Decoding	0.23	0.03	7.61	0.00	0.17	0.29
[0.29]	OVK	0.09	0.02	3.67	0.00	0.04	0.13
	Intercept	−0.08	0.01	−14.51	0.00	−0.09	−0.07
QR-50	Del	0.11	0.04	2.75	0.01	0.03	0.20
$R^2 = 0.34$	Decoding	0.32	0.03	9.59	0.00	0.25	0.39
[0.34]	OVK	0.08	0.04	2.34	0.02	0.01	0.15
	Intercept	−0.09	0.02	−5.00	0.00	−0.13	−0.06
QR-75	Del	0.28	0.07	3.87	0.00	0.14	0.43
$R^2 = 0.38$	Decoding	0.46	0.06	8.02	0.00	0.34	0.57
[0.39]	OVK	−0.05	0.08	−0.60	0.55	−0.20	0.11
	Intercept	−0.03	0.05	−0.54	0.59	−0.13	0.07
QR-90	Del	0.30	0.04	6.85	0.00	0.22	0.39
$R^2 = 0.41$	Decoding	0.54	0.05	10.89	0.00	0.44	0.63
[0.41]	OVK	−0.07	0.04	−1.73	0.09	−0.16	0.01
	Intercept	0.02	0.02	1.10	0.27	−0.02	0.06

Table 3 Overall contribution of deletion, decoding and oral vocabulary knowledge to spelling skills in English (L2)

Model	Parameter	Estimate	SE	t-value	p	LB	UB
Linear regression – L2 Spell ~ L2 PA(Del) + L2 Decoding + L2 OVK							
$R^2 = 0.57$	Del	0.03	0.02	1.48	0.14	−0.01	0.07
[0.57]	Decoding	<u>0.27</u>	0.02	13.81	0.00	0.23	0.31
	OVK	<u>0.04</u>	0.02	1.99	0.05	0.00	0.09
	Intercept	−0.02	0.01	−1.75	0.08	−0.04	0.00
Quantile regression – L2 Spell ~ L2 PA(Del) + L2 Decoding + L2 OVK							
QR-25	Del	0.00	0.00	1.78	0.08	0.00	0.01
$R^2 = 0.19$	Decoding	<u>0.14</u>	0.02	8.20	0.00	0.11	0.18
[0.19]	OVK	0.00	0.00	0.74	0.46	−0.01	0.01
	Intercept	0.00	0.00	−1.68	0.10	−0.01	0.00
QR-50	Del	0.01	0.01	0.49	0.63	−0.02	0.03
$R^2 = 0.34$	Decoding	<u>0.24</u>	0.03	8.08	0.00	0.18	0.30
[0.34]	OVK	0.00	0.01	0.24	0.81	−0.01	0.02
	Intercept	0.00	0.00	−0.35	0.73	0.00	0.00
QR-75	Del	0.03	0.02	1.56	0.12	−0.01	0.07
$R^2 = 0.45$	Decoding	<u>0.34</u>	0.04	7.99	0.00	0.25	0.42
[0.46]	OVK	<u>0.04</u>	0.02	2.18	0.03	0.00	0.07
	Intercept	0.00	0.00	0.00	1.00	−0.01	0.01
QR-90	Del	0.02	0.02	0.81	0.42	−0.02	0.06
$R^2 = 0.55$	Decoding	<u>0.49</u>	0.06	8.22	0.00	0.37	0.60
[0.56]	OVK	0.04	0.02	1.51	0.13	−0.01	0.09
	Intercept	0.02	0.01	2.30	0.02	0.00	0.04

spelling skill (L1 and L2; see Tables 2 and 3). We also found that decoding skill in L1 and L2 had statistically significant explanatory power across all levels of spelling skill (L1 and L2), which indicated that mastery of decoding skill is a necessary pre-requisite for mastery of spelling skill in both Kannada/Telugu and English (see Tables 2 and 3). Unlike decoding skills in Kannada/Telugu, the L1 oral vocabulary knowledge (OVK) only retained explanatory power at the level of 25th and 50th percentile in L1 spelling skills, indicating that OVK did not play an important role for higher level spelling, since students might have mastered this skill already. In English, OVK only had significant explanatory power at the level of the 75th percentile of spelling skill ($b = 0.04$, $p = 0.03$, 95% CIs [0.00, 0.07]). After closer examination of this result, we concluded that the effect of OVK on the 75th level of spelling skills might be too small. Hence, we could not determine that OVK had explanatory power in spelling skill at all levels in English. Surprisingly, PA in L1 did not show statistically significant explanatory power at the level of the 25th percentile of L1 spelling skill, but showed significant explanatory power at other higher levels. This is possibly due to the floor effect of spelling words at the lower level. In contrast, PA in English did not have any statistically significant explanatory power in spelling skill at any levels. In conclusion, decoding skill in both Kannada/Telugu

Table 4 Overall contribution of deletion, decoding and oral vocabulary knowledge in Kannada/Telugu (L1) to spelling skills in English (L2)

Model	Parameter	Estimate	SE	t-value	p	LB	UB
Linear regression – L2 Spell ~ L1 PA(Del) + L1 Decoding							
$R^2 = 0.19$	Del	<u>0.08</u>	0.03	2.79	0.01	0.02	0.14
[0.19]	Decoding	<u>0.11</u>	0.03	3.89	0.00	0.05	0.17
	Intercept	−0.04	0.02	−2.82	0.01	−0.07	−0.01
Quantile regression – L2 Spell ~ L1 PA(Del) + L1 Decoding							
QR-25	Del	*NA*					
R^2 = NA	Decoding	*NA*					
	Intercept	*NA*					
QR-50	Del	0.01	0.01	0.36	0.72	−0.02	0.03
$R^2 = 0.10$	Decoding	<u>0.08</u>	0.01	6.36	0.00	0.06	0.11
[0.11]	Intercept	−0.01	0.00	−10.41	0.00	−0.02	−0.01
QR-75	Del	0.06	0.04	1.36	0.18	−0.03	0.15
$R^2 = 0.15$	Decoding	<u>0.10</u>	0.03	3.48	0.00	0.04	0.16
[0.16]	Intercept	−0.01	0.01	−2.36	0.02	−0.02	0.00
QR-90	Del	<u>0.29</u>	0.07	4.36	0.00	0.16	0.43
$R^2 = 0.24$	Decoding	0.10	0.06	1.90	0.06	0.00	0.21
[0.24]	Intercept	−0.01	0.01	−0.78	0.43	−0.04	0.02

and English played an important role in mastery of spelling skill for all levels of spellers.

Overall Contribution of Literacy Skills (L1) to Spelling Skill (L2) First, we built the model with only PA (L2) and decoding (L2) as independent variables, with spelling skill (L2) as dependent variable. The R-squared from conventional linear regression analysis was 0.19 indicating that around 19% of the variation in English spelling skill was explained by Kannada/Telugu phonological awareness and decoding skills. The results from quantile regression analysis further explicated the associations in details: 1) the association between literacy skills(L1) and spelling skill (L2) had an increasing trend along with the level of L2 spelling skill (QR-50: R-squared = 0.10, QR-75, R-squared = 0.15, QR-90, R-squared = 0.24). The model did not converge at the 25th percentile; therefore, we did not further consider the result.

Conventional linear regression indicated that both PA (L1) and decoding (P1) had statistically significant explanatory power in spelling skill (L2). Conversely, the results from quantile regression indicated that decoding skill (L1) was statistically significant at QR-50 and QR-75, but not at QR-90, indicating that high level L2 spellers did not transfer their L1 decoding skills. However, the PA (L1) showed the reverse patterns: It was only significant at QR-90. Thus, the high level spellers transferred their L1 PA skills, while median and median-high level spellers only transferred their L1 decoding skills when performing English spelling. This is in line with the transfer facilitation model (Koda, 2008) and previous studies on

Table 5 Overall contribution of deletion, decoding and oral vocabulary knowledge in L1 and L2 to spelling skills in English (L2)

Model	Parameter	Estimate	SE	t-value	p	LB	UB
Linear regression – L2 Spell ~ L1 PA(Del) + L1 Decoding + L2 PA(Del) + L2 Decoding + L2 OVK							
$R^2 = 0.57$	L1 Del	−0.03	0.02	−1.06	0.29	−0.08	0.02
[0.56]	L1 Decoding	0.02	0.02	0.84	0.40	−0.02	0.06
	L2 Del	0.03	0.02	1.58	0.12	−0.01	0.08
	L2 Decoding	0.27	0.02	13.67	0.00	0.23	0.31
	L2 OVK	0.05	0.02	1.93	0.06	0.00	0.09
	Intercept	−0.02	0.01	−1.46	0.15	−0.04	0.01
Quantile regression – L2 Spell ~ L1 PA(Del) + L1 Decoding + L2 PA(Del) + L2 Decoding + L2 OVK							
QR-25	L1 Del	0.00	0.00	−0.64	0.53	−0.01	0.01
$R^2 = 0.19$	L1 Decoding	0.01	0.00	2.05	0.04	0.00	0.01
[0.19]	L2 Del	0.00	0.00	1.12	0.27	0.00	0.01
	L2 Decoding	0.14	0.02	7.88	0.00	0.11	0.18
	L2 OVK	0.00	0.00	1.30	0.19	0.00	0.01
	Intercept	−0.01	0.00	−2.47	0.01	−0.01	0.00
QR-50	L1 Del	−0.01	0.01	−1.48	0.14	−0.03	0.00
$R^2 = 0.34$	L1 Decoding	0.02	0.01	2.02	0.04	0.00	0.03
[0.34]	L2 Del	0.02	0.01	1.15	0.25	−0.01	0.04
	L2 Decoding	0.22	0.03	7.51	0.00	0.16	0.28
	L2 OVK	0.01	0.01	1.19	0.24	−0.01	0.03
	Intercept	0.00	0.00	−2.41	0.02	−0.01	0.00
QR-75	L1 Del	−0.02	0.01	−1.52	0.13	−0.04	0.01
$R^2 = 0.46$	L1 Decoding	0.03	0.01	4.18	0.00	0.02	0.05
[0.46]	L2 Del	0.03	0.01	2.27	0.02	0.00	0.05
	L2 Decoding	0.35	0.04	8.32	0.00	0.27	0.43
	L2 OVK	0.00	0.01	0.42	0.67	−0.02	0.03
	Intercept	0.00	0.00	0.00	1.00	−0.01	0.01
QR-90	L1 Del	−0.03	0.03	−0.98	0.33	−0.10	0.03
$R^2 = 0.55$	L1 Decoding	0.05	0.03	1.75	0.08	−0.01	0.11
[0.56]	L2 Del	0.01	0.04	0.30	0.77	−0.07	0.09
	L2 Decoding	0.49	0.06	7.74	0.00	0.37	0.62
	L2 OVK	0.04	0.03	1.07	0.29	−0.03	0.10
	Intercept	0.00	0.01	0.27	0.78	−0.03	0.03

akshara-English biliteracy acquisition (Nakamura et al., 2014; Reddy & Koda, 2013), which showed that more proficient L2 readers are drawing on the available metalinguistic skills from their L1, such as phonological awareness.

We also included all L1 and L2 literacy skills measures as independent variables to investigate overall contributions in L2 (English) spelling skills. The findings were consistent with previous results, with a range of R-squared from 0.19 (QR-25) to 0.55 (QR-90; see Table 5). Moreover, English decoding skill was statistically significant at all levels of English spelling skills. The magnitude of explanatory power also increased from 0.14 to 0.49, indicating that English decoding skill played an important role in English spelling skills even for English as second language learners. The L1 decoding skill only showed significant results from QR-25 to QR-75; however, the magnitude of the coefficient was not large enough to be noticeable, indicating that the L1 decoding skills might serve as a scaffolding function for English learner. However, after becoming a proficient speller, perhaps English learners no longer need the L1 literacy skills as the scaffold for their English spelling skill development as shown in the non-significant results of L1 measures at QR-90 of English spelling in Table 5.

In conclusion, we found that English (L2) decoding skills were associated in an important way with developing English spelling skills among Kannada/Telugu ESL learners, even though we found that they appear to transfer their first language decoding skills when spelling English words. This result is in accordance with the biliteracy acquisition framework that suggests that L2 reading and spelling ability are predicted by a combination of transferable L1 script processing and orthographic ability as well as L2-specific skills (Koda, 2008). After becoming a proficient English speller, L1 decoding skills serve less as a scaffold, presumably since some threshold level of L1 decoding has been reached for the purposes of transfer. Moreover, we also found that the associations of phonological awareness, decoding, and oral vocabulary knowledge with spelling skill have an increasing effect regardless of L1 or L2. This might indicate that the importance of PA, decoding, and OVK instruction may only become apparent as the child's spelling mastery continues to increase in either language. At the earliest levels of learning to spell – whether it is an akshara language or an alphabetic L2 – it may be worthwhile to consider the role of more basic skills such as symbol knowledge, and possible gross motor skills needed for writing as well.

Discussion and Conclusions

Most of the studies hitherto have examined how the knowledge of first language can influence spelling in a second language. However, most of these studies are from one alphabetic language to another alphabetic language or from an alphabetic language to a morphosyllabic system such as English and Chinese. The present chapter is among the first to report how the knowledge of an akshara orthography might influence spelling English words. We administered phonological awareness, oral vocabulary, decoding, and spelling tasks in either Kannada or Telugu as L1 and English as L2. Results from quantile regression analyses showed that knowledge of L1 does influence spelling of English words; however, once students become proficient in English, especially at later grades, the influence of L1 to English spelling lessens, indicating that better spellers in English have reached a level of mastery where they do not have to depend on the sharable metalinguistic properties of the first language. An earlier study provides some evidence for this notion by showing that students who achieve a threshold of about 60% proficiency in decoding skills in the first language are able to perform significantly better in L2 decoding acquisition (Nakamura, de Hoop, & Holla, 2018). Further research is needed to unpack the extent to which such a threshold might exist in spelling biliteracy; yet this study sheds light on the fact that cross-linguistic resource-sharing occurs even in akshara-English spelling biliteracy acquisition.

Acknowledgments This research was conducted under the All Children Reading Grand Challenge initiative funded by the United States Agency for International Development, World Vision, and Australian Aid.

References

Ball, E. W., & Blachman, B. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*, 49–66.

Daniels, P. T., & Share, D. L. (2017). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading, 22*, 101–116. https://doi.org/10.1080/10888438.2017.1379082.

Dunn, L. M., & Dunn, L. M. (1981). *Manual for the peabody picture vocabulary test-revised*. Circle Pines, MN: American Guidance Service.

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 237–269). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders, 20*, 19–36. https://doi.org/10.1097/00011363-200020030-00005.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading, 18*, 5–21. https://doi.org/10.1080/10888438.2013.819356.

Figueredo, L. (2006). Using the known to chart the unknown: A review of first-language influence on the development of English-as-a-second-language spelling skill. *Reading and Writing: An Interdisciplinary Journal, 19*, 873–905.

Gentry, J. R. (1982). An analysis of developmental spelling in GNYS AT WRK. *The Reading Teacher, 36*, 192–200.

Geva, E., Wade-Wooley, L., & Shany, M. (1993). The concurrent development of spelling and decoding in two different orthographies. *Journal of Literacy Research, 25*, 383–406. https://doi.org/10.1080/10862969309547827.

Henderson, E. H., & Templeton, S. (1986). A developmental perspective of formal spelling instruction through alphabet, pattern, and meaning. *Elementary School Journal, 86*, 304–316.

Ho, C. S.-H., Yau, P. W.-Y., & Au, A. (2003). Development of orthographic knowledge and its relationship with reading and spelling among Chinese kindergarten and primary school children. In C. McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children* (pp. 51–71). London, UK: Praeger.

Jongejan, W., Verhoeven, L., & Siegel, L. S. (2007). Predictors of reading and spelling abilities in first- and second-language learners. *Journal of Educational Psychology, 99*, 835–851. https://doi.org/10.1037/0022-0663.99.4.835.

Joshi, R. M., Hoien, T., Xiwu-Feng, R., Chengappa, R., & Boulware-Gooden, R. (2006). Learning to spell by ear and by eye: A cross-linguistic comparison. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 569–577). Mahwah, NJ: Lawrence Erlbaum Associates.

Kandhadai, P., & Sproat, R. (2010). Impact of spatial ordering of graphemes in alphasyllabic scripts on phonemic awareness in Indic languages. *Writing Systems Research, 2*, 105–116. https://doi.org/10.1093/wsr/wsq009.

Karanth, P. (2006). The kagunita of Kannada–learning to read and write an Indian alphasyllabary. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 389–404). Mahwah, NJ: Lawrence Erlbaum Associates.

Koda, K. (2008). Impacts of prior literacy experience on second language learning to read. In K. Koda & A. M. Zehler (Eds.), *Learning to read across languages: Cross-linguistic relationships in first- and second language literacy development* (pp. 68–96). New York, NY: Routledge.

Koda, K., & Reddy, P. (2008). Cross-linguistic transfer in second language reading. *Language Teaching, 41*, 497–508.

Li, T., McBride-Chang, C., Wong, A., & Shu, H. (2012). Longitudinal predictors of spelling and reading comprehension in Chinese as an L1 and English as an L2 in Hong Kong Chinese children. *Journal of Educational Psychology., 104*, 286–301. https://doi.org/10.1037/a0026445.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*, 7–22.

Nag, S., & Perfetti, C. A. (2014). Reading and writing: Insights from the alphasyllabaries of South and Southeast Asia. *Writing Systems Research, 6*, 1–9. https://doi.org/10.1080/17586801.2014.883787.

Nag, S., & Snowling, M. J. (2012). Reading in an Alphasyllabary: Implications for a language universal theory of learning to read. *Scientific Studies of Reading, 16*, 404–423. https://doi.org/10.1080/10888438.2011.576352.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*, 41–52. https://doi.org/10.1093/wsr/wsq001.

Nakamura, P. R., Joshi, R. M., & Ji, X. (2018). Investigating the asymmetrical roles of syllabic and phonemic awareness in akshara processing. *Journal of Learning Disabilities, 51*, 499–506. https://doi.org/10.1177/0022219417718120.

Perfetti, C. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*, 3–24.

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsman (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam, the Netherlands: John Benjamins Publishing.

Petscher, Y., & Logan, J. A. (2014). Quantile regression in the study of developmental sciences. *Child Development, 85*, 861–881.

Read, C. (1975). *Children's categorization of speech sounds in English* (NCTE research report No. 17). Urbana, IL: National Council of Teachers of English.

Reddy, P., & Koda, K. (2013). Orthographic constraints on phonological awareness in biliteracy development. *Writing Systems Research, 5*, 110–130. https://doi.org/10.1080/17586801.2012.748639.

Share, D. L., & Daniels, P. T. (2015). Aksharas, alphasyllabaries, abugidas, alphabets, and orthographic depth: Reflections on Rimzhim, Katz, and Fowler (2014). *Writing Systems Research, 8*, 17–31. https://doi.org/10.1080/17586801.2015.1016395.

Shu, H., Anderson, R. C., & Wu, N. (2000). Phonetic awareness: Knowledge of orthography–phonology relationships in the character acquisition of Chinese children. *Journal of Educational Psychology, 92*, 56–62. https://doi.org/10.1037/0022-0663.92.1.56.

Sparks, R. L., Patton, J., Ganschow, L., Humbach, N., & Javorsky, J. (2008). Early first-language reading and spelling skills predict later second-language reading and spelling skills. *Journal of Educational Psychology, 100*, 162–174. https://doi.org/10.1037/0022-0663.100.1.162.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York, NY: Oxford University Press.

Treiman, R. (2004). Spelling and dialect: Comparison between speakers of African American Vernacular English and white speakers. *Psychonomic Bulletin & Review, 11*, 338–342. https://doi.org/10.3758/BF03196580.

Treiman, R., & Cassar, M. (1996). Effects of morphology on children's spelling of final consonant clusters. *Journal of Experimental Child Psychology, 63*, 141–170. https://doi.org/10.1006/jecp.1996.0045.

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4*, 313–330. https://doi.org/10.1207/S1532799XSSR0404_4.

Wang, M., & Geva, E. (2003). Spelling performance of Chinese children using English as a second language: Lexical and visual-orthographic processes. *Applied PsychoLinguistics, 24*, 1–25. https://doi.org/10.1017/S0142716403000018.

Akshara Processing in Telugu Depends on Syllabic and Phonemic Sensitivity: Preliminary Evidence from Normal Hearing and Hearing-Impaired Children

Vasanta Duggirala

Abstract The long-held view that most Brahmi-derived Indic scripts are prototypically alpha-syllabaries as well as a more recent proposal that these scripts are functionally alphabetic in terms of how they are read, have met with challenges in recent years. A consensus seems to be emerging about the idea that cognitive processes involved in beginning reading cannot be adequately understood without reference to sub-lexical processing that draws on language-specific syllable structures, and the extent of phonological information coded in the script. The research reported in this chapter deals with Telugu *akshara* processing abilities of 15 normal-hearing, and 15 hearing-impaired children. They were assessed using three specially designed tasks: akshara substitution, akshara deletion, and adding length marker to the vowel within one of the aksharas of target words. All three tasks made use of meaningful printed Telugu words. Both groups found deletion task more difficult than substitution task. All the children exhibited considerable phonemic sensitivity in the task requiring them to attach a vowel length marker to vowel in the first *akshara o*f each target word. However, they scored poorly in making lexical decisions. These results are discussed in the light of sonority principles, and characteristic features of graphic syllables associated with Telugu writing system. Implications for the assessment and literacy instruction are discussed briefly.

Keywords Akshara · Alphasyllabary · Brahmi-derived scripts · Sonority · Telugu

Introduction

Scripts associated with the languages of the world were traditionally classified as belonging to three major types: (1) Logographic (2) Syllabic and (3) Alphabetic, depending on the minimal linguistic unit deployed by a given script. Salomon (2000) argued that this classification was too simplistic and unsatisfactory, and that scripts are to be classified based on two different criteria, viz., mode of

V. Duggirala (✉)
Department of Linguistics, Osmania University, Hyderabad, Telangana, India
e-mail: vasantad@gmail.com

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*,
Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_7

representation by individual graphs within the system (e.g. Japanese Kana represents syllables, while Kanji represents whole words), and overall level of the script (e.g. Japanese is a mixed syllabo-logographic type). The Indic scripts are considered to be prototypes of alphasyllabaries (also referred to as syllabic alphabets or abugidas) because, they have the following main features: Graphic unit is mostly an open syllable of the type V, CV, CCV; The consonant graph has an inherent vowel such as /a/ or schwa; there are extra diacritic signs for vowels attached to consonant graphemes; vowels that do not follow consonant graphemes are represented by independent signs; and finally, vowel-less consonant graphemes are represented using a *halant* or *virama* sign. For more detailed descriptions of phonological features coded in Devanagari see Patel (2007) and Pandey (2007). More recently, drawing on their earlier research on skilled reading in Devanagari by adults, Rimzhim, Katz, & Fowler (2014) argued that Brahmi-derived orthographies of Indo-Aryan languages are predominantly and functionally alphabetic in terms of how they are read. In other words, for reading and writing in Devanagari, the relevant written unit, according to them is, the "letter" representing all phonemes except the inherent vowel or schwa.

Share (2014) argued against the 'culture' of 'alphabetism' or the belief that alphabetic writing systems are inherently superior to other kinds of non-alphabetic writing systems by stressing the point that the concept 'letter' has questionable applicability to many writing systems including Chinese and Brahmi-derived scripts. He also stated that what is important is to examine the match between language structure and writing system by looking at the complexity of syllable inventory of a language. More recently, Share and Daniels (2016) strongly disagreed with the view that Indic scripts should be considered as alphabetic functionally. In response to Rimzhim et al. (2014), Share and Daniels (2016) stated that the status of consonant and vowel graphic symbols represented in the Brahmi-derived scripts are not at all equal; that reduced consonant graphemes are not equivalent to vowel *matras*; that written signs in these scripts represent /Cə/ graphic syllables, and that the way consonants and vowel graphemes are spatially represented in Brahmi-derived scripts is not as linear as in most alphabetic scripts. Their main point was, akshara-based scripts are NOT fundamentally alphasyllabic, nor are they syllabic. They also opined that caution should be exercised by researchers in interpreting the results of phonemic processing tasks in Brahmi-derived scripts because skilled readers in these languages might be biliterate and biscriptal involving an alphabetic script (e.g. English), and therefore, their performance might reflect reciprocal influence of one script over another.

Considerable published research exists on various aspects of reading and spelling in Brahmi-derived scripts associated with various Indian languages. Within the limited space of this chapter, it is not possible to review all those studies, or even those pertaining to the four major Dravidian literary languages and scripts. Instead, select studies dealing with Telugu and Kannada scripts (with similar visuo-spatial organization of symbols) are briefly discussed here to provide a background to the research described in this chapter. Purushothama (1994) reported 118 words and 136 orthographic syllables that he had put-together keeping in mind various orthographic rules of Kannada language. He presented empirical evidence demonstrating that these materials successfully distinguish good versus poor III grade readers (N = 20). Other researchers interested in the process of reading acquisition have reported that both syllable and phoneme awareness does impact on word decoding abilities to different degrees at different points during elementary school years (For e.g. Vasanta, 2004; Nag, 2007; Reddy & Koda, 2013). Nag (2007) described a Literacy Acquisition Battery (LAB), and provided empirical evidence that children learning to read Kannada initially develop syllabic awareness, possibly due to the salience of orthographic syllable during literacy instruction offered in schools. She also stated that as akshara knowledge about orthographic rules increases, children do develop sensitivity to sub-syllabic structures. Nag, Treiman, and Snowling (2010) reported a study based on 60 primary school Kannada children classified at the outset of the study as less skilled (N = 29) and skilled spellers (N-31). Both groups of children performed poorly on words with complex syllable structure (containing geminates or clusters) or those involving secondary graphemes. In their concluding paragraph (Nag et al. 2010:49) commented: "Our analysis of errors at the intra-akshara level clearly demonstrate that children attempt to use phonemic information to build up aksharas even though their teaching has focused on the syllable level". In a large-scale more recent study, Nakamura, Joshi & Ryan Ji (personal communication, January 2017) investigated the nature of asymmetrical role played by syllabic and phonemic awareness in Akshara processing in relation to Telugu and Kannada, two languages associated with Brahmi-derived scripts. In this study involving 488 children from grades 1-V selected from 13 different schools in the States of Karnataka and erstwhile Andhra Pradesh in India, Nakamura et al. examined relative contribution of syllable and phoneme awareness in decoding Kannada and Telugu words using phoneme and syllable deletion tasks. The authors also treated oral vocabulary knowledge as another independent variable, but it did not turn out to be a strong predictor of decoding ability. Detailed statistical analysis of the results of this study revealed that while phonemic awareness plays a role in early grades, it makes lesser contribution to decoding ability as children progress through elementary years such that by grade V, syllable level representations incorporate (and overshadow) phoneme level representations.

It should be noted that in all these studies, there is inadequate discussion on the transparency / opacity of orthographic syllables within Brahmi-derived scripts and its possible impact on decoding words. Alvarez, Garcia-Saavedra, Luque, and Taft (2016) reported robust evidence for syllabic level processing in 6–8 year old Spanish speaking children by using a word-spotting task that permits manipulation of syllable boundaries. They talked about 'syllabic-bridge' hypothesis which foregrounds mapping between phonological syllables and orthographic units by offering considerable support in reading languages with clear syllable boundaries such as French. Their own study involving young Spanish readers offered evidence for the functionality of syllable as a processing unit in visual word recognition.

In the past, most researchers assumed that orthographic knowledge is a fully lexicalized knowledge that arises only after instruction in written language begins in schools. This assumption was questioned by Kaefer (2016) who argued that children are sensitive to statistical patterns in print and that they do extrapolate rules of speech and orthography implicitly by noting common patterns in the limited written language that they encounter. Further, it was pointed out that there are contexts when orthographic knowledge requires selective overriding of phonological knowledge and in those contexts, children must learn to overcome the conflict between orthography and phonology. One of the arguments being offered in this chapter is that there are system-internal contexts (e.g. presence of specific type of orthographic syllables in Telugu) as well as system-external contexts (e.g. severe to profound congenital hearing-impairment in the participants) that might force readers to rely more on orthographic knowledge than on phonological knowledge (see Harris & Moreno, 2004 for elaboration of this point).

Published literature on the acquisition of spellings by deaf children are very few, and even those are limited to English with alphabetic orthography (see for e.g. Padden, 1993; Olson, 2004). These studies have shown that deaf students still organize their written material in terms of syllables suggesting that syllabic structure can be relatively independent of the system devoted to speech. Olson (2004) pointed out that 65% of 3-letter words in English in the CELEX corpus are CVC syllables; that most syllable nuclei appear in the second position of these three-letter words; with knowledge about syllable nuclei, one can predict word initial and word final consonant clusters. This knowledge guides learners to grasp the idea that the interior consonants of s-cluster are more sonorous than the exterior consonants. While researchers have long demonstrated that sonority principles do make substantial predictions about sequencing of syllables in certain scripts (see Miller, 1994 for a detailed discussion), reading research in relation to Indic scripts has not paid much attention to the concept of sonority. The main aim of the research described in this chapter is to offer some insights about sonority profiles and graphic syllable structures and discuss their role in visual processing of Telugu script by normal-hearing, and hearing-impaired school children. The contents of the rest of this chapter are organized under the following main headings: Telugu script and Orthography (2.0); Assessment of Grapheme Awareness in Telugu (3.0); Empirical Findings (4.0); and Implications for Assessment and Instruction (5.0).

Telugu Script and Orthography

Telugu is one of the four major Dravidian literary languages (the other three are, Kannada, Tamil and Malayalam) spoken primarily in South India. The southern group of Indic scripts associated with these four languages date back to sixth to eighth century and are linked to Grantha, Kalinga and Kadamba scripts. The modern-day Kannada and Telugu scripts are descendants of Kadamba. These writing systems treat orthographic syllable as the main unit of representation, while at the same time, recognizing segments as units of analysis. Some of the Indic scripts that can be traced back to Brahmi writing system are discussed in detail in Daniels (2008).

Telugu script has as many as 56 graphemes and allographs: 16 graphemes corresponding to the vowels; 22 to stop consonants; five to nasals; one to alveolar flap; two to laterals; three to voiceless fricatives; two to frictionless continuants; one to glottal fricative; one representing the *visarga (*nasal before the consonants) and a few others to represent complex syllables (Sastry, 1972). Telugu writing system provides symbols for all these phonemes except for [æ] and [f]. The sound /f/ is represented as /ph/ and it occurs mostly in borrowed words. The *anuswara* represented by symbol [0] is used in place of /m, n/ before homorganic stops and affricates as well as for bilabial nasal /m/ when it occurs word finally. Vowel and consonant graphemes of Telugu are shown in Fig. 1.

Krishnamurti and Gwynn (1985) pointed out that Telugu writing system, even after excluding aspirated consonants is more complex with 12 graphemes for the primary vowels which would generate 12 × 22 or 264 CV sequences; 22 × 22 or 484 consonant plus consonant combinations, although only 140 of the latter occur in the language. All in all, more than 400 printed characters (12 + 264 + 140) are needed to represent most of the speech sounds and sound combinations in Telugu script.

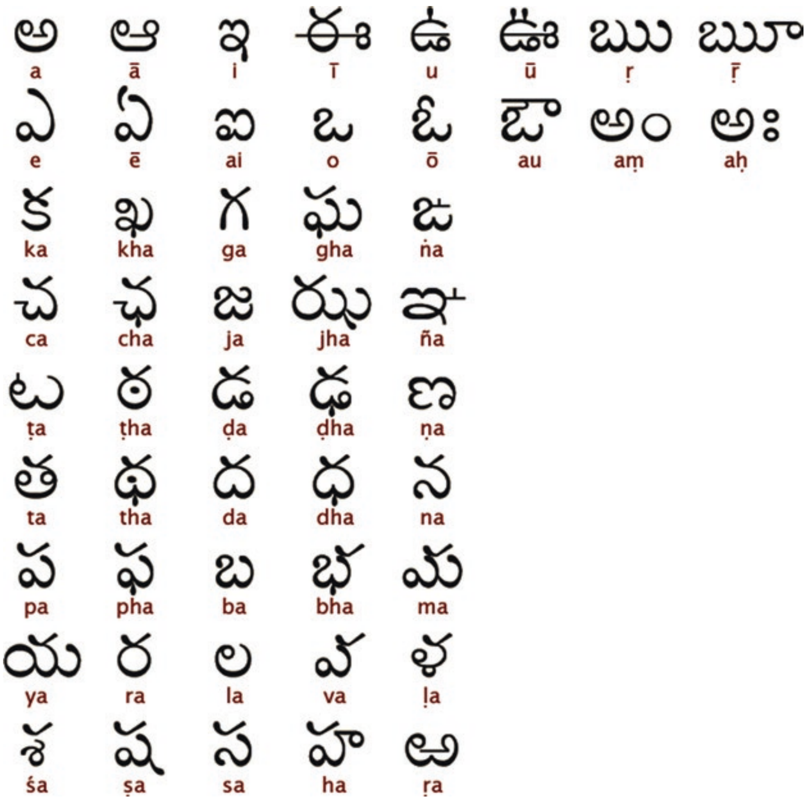


Fig. 1 Vowel and consonant graphemes of Telugu

Turning to graphic representation of two consonant clusters, in Telugu, those occurring at the beginning of a word are typically stop consonants followed by /y/, /w/, /r/, or /l/; these are mostly confined to loan words from Sanskrit. In the medial position of words, both geminates and consonant cluster graphemes occur abundantly in both native and non-native vocabulary. Aspirated stops and affricates rarely occur in clusters. Brahmi-derived Indic scripts including Telugu represent loan words from English differently in their orthographies. Besides such temporo-spatial non-linearity, there is also considerable variation in the way secondary forms of consonants are represented within Telugu script (see Hill, 1991 for examples). The point being stressed here is that the graphic syllable structure complexity varies in representation of native versus borrowed words. There are many words borrowed from Sanskrit and English that are assimilated into everyday Telugu. Children must learn to represent all these in academic writing.

Telugu Orthographic Syllables: Statistical Patterns

Linguistic Analysis of Telugu text corpus of over 3 million words was developed by the Central Institute of Indian Languages (CIIL) by Uma Maheshwar Rao (2006) who provided a list of 1000 most frequently used Telugu words. The following facts about orthographic syllables are pertinent to the topic under discussion:

- The vowel characters (graphemes) cover 43.16% of the entire corpus while sonorant consonants /n, r, l/ cover about 15%; another 15% is covered by /k, v, ṭ, m, and P/. The remaining 28 consonant characters have coverage of 23%. Aspirated consonants characters cover <2% of the corpus.
- Of the total 14,000 distinct orthographic syllables, only 500 syllables contributed to 94% of the total coverage in the corpus. The (graphic) syllable types that contribute to 95% coverage in this corpus are CV, CCV, CVm, V, CCVm, Vm, CCCV, and CCVC. Further analysis revealed that the most frequently occurring cluster type is CCV.
- Among the 2000 different consonant combinations that include both geminates and clusters, about 100 of them contributed to >90% of the total consonant sequences that included considerable number of homorganic nasal consonant clusters (e.g. ms, mv, mh, mḍ, mṭ, mg, mk, mb, mdh, mv, mbh, mgh ms, mc, etc) in which phonological and orthographic syllabification match (e.g. the word, *gampa* basket is syllabified phonologically and orthographically as *gam.pa*).
- The most commonly occurring geminate is /nn/ followed by /TT/, /kk/, /pp./, /ll/, /cc/, /ḷḷ/, /dd/ and /mm/. Biconsonantal clusters are more frequent than tri-consonantal clusters. There are very few four consonant clusters and those are mostly confined to loan words from English (making a minimal contribution of 0.03%).

We need to examine frequently occurring words to delineate the principles underlying syllable boundaries in a given language. It is a well-attested fact that sonority is one of the factors influencing syllabification. It is possible that sonority is one of the organizing principles underlying specific patterns of frequently occurring graphic syllables in Telugu.

Syllable Structures and Sonority Considerations

Sonority is an abstract phonological property that corresponds with perceived loudness of speech sounds and is influenced by the extent of jaw opening and closing. It is a language-independent construct that functions to organize segments within syllables (Ettlinger, Finn & Hudson Kam, 2012). Every language has co-occurrence restrictions specifying legal / illegal onsets and rimes. While many languages permit onsets, few require codas and no language prohibits onsets. The 'best' syllable (one that enhances production and perception) is the one that is made up of a single consonant (onset) and a vowel (nucleus) with no coda, that is, a syllable with CV structure. Syllables with complex onsets (CCV) or those with a coda (CVC) or those which lack an onset (V) are all considered to add a degree of complexity to the basic CV template. Observe English examples of words with different syllable structures listed in Table 1.

According to Clements and Keyser (1983), syllables can be ranked according to complexity (from simple to complex) as follows: (1). CV (2). V, CCV, CVC (3). VC, CCVC (4). VCC. Since optimal syllables are those which should enhance production and perception, it was predicted that simple syllables should occur more often in the languages of the world than complex syllables. This turned out to be generally true (see Joanisse, 1999). Vennemann (1988) argued that a hierarchy of CV syllable templates is inadequate to capture all linguistic facts, and that syllables should be ranked not only based on their structural properties (whether they have onsets, if so, whether those onsets are simple or complex etc.), but also according to the nature of the segments making up the syllable (that is, for example, whether the onset consonant is an obstruent or a sonorant).

Different classes of speech sounds differ from one another based on *sonority.* At least two factors determine the sonority of a sound; one, the degree of jaw opening or stricture, and two, its propensity for voicing. Thus, it was noted that stops, affricates and fricatives (which come under the cover-term, Obstruents) with greater degree of stricture are less sonorous than liquids and nasals. Within each category,

Table 1 Examples of words with different syllable shapes in English

English word	Syllable Shape	Description
See	CV:	Simple onset, no coda
Seat	CV:C	Simple onset, simple coda
Tree	CCV:	Complex onset, no coda
Street	CCCV:C	Complex onset, simple coda
Treats	CCV:CC	Complex onset, complex coda

voiced ones have greater sonority than their voiceless counterparts. Clements (1990) proposed a sonority hierarchy (SH): ONLGV that puts Obstruents (O) at the bottom end of the hierarchy and Nasals (N), Liquids (L), Glides (G), and Vowels (V) in that order towards the most sonorous end as shown below:

Natural Classes:	Stops < Fricatives/affricates < Nasals < Liquids < Glides < Vowels				
Abbreviation:	**O**	**N**	**L**	**G**	**V**
Examples	[t, d] [s, f, c, dz]	[m, n]	[l, r]	[y, w]	[a, i]
Sonority index	1	2	3	4	5

The Sonority Sequencing Principle (SSP) states that sonority must rise from the onset until nucleus, while codas must exhibit sonority fall from syllable peak (nucleus) to coda. That is what happens in permissible syllable shapes across languages.

The Minimal Sonority Distance (MSD) criterion restricts the distance between the onset consonants to two steps in English. Therefore cluster /bla/ with a sonority distance of 2 (3-1) is allowed, but a cluster such as /bda/ with a sonority distance between /b/ and /d/ being zero is banned. Sonority Hierarchy also correctly predicts a nasal plus sonorant over nasal plus stop onsets in English. It should be noted that the legality of cluster types varies across phonologies of the languages as does the legality of orthographic syllables due to language specific phonotactic, and graphotactic rules. Besides the 5-point ONLGV sonority hierarchy, there are other more phonetically grounded sonority scales, but the 5-point scale is the one that is applied in measuring sonority dispersion within the segments constituting a syllable using a metric called 'D' given below:

$$D = \sum_{i=1}^{m} \frac{1}{d_i^2}$$

Where d = the distance between sonority indices of each pair of segments; m = number of pairs of segments whose value is: n (n-1) /2 where n = # of segments. Clements (1990) argued that the value of 'D' varies according to the sum of inverse of the squared values of sonority distance between the members of each pair of segments within a demi-syllable (that is, onset + nucleus). Thus, for an onset cluster Obstruent-Vowel (OV / pa), the D = 0.06 (1/16); NV = 0.11 (1/9); LV = 0.25 (1/4) and so on. Sonority Dispersion values of 'D' for syllables with complex onsets in English are shown below:

Markedness	**cluster type Onset/Coda**	**'D' value**
Most natural	OLV/VLO	0.56
↑	OGV/VGO	1.17
	ONV/VNO	1.17
↓ Least natural	NGV/VGN	1.36
	NLV/VLN	1.36
	LGV/VGL	2.25

In most languages, preferred syllables are those with lowest ‘D’ value, that is, onset clusters in which Obstruents (O) are followed by Liquids (L) because sonority rises steadily and gradually. Therefore, they are considered as unmarked relative to the other types. The Minimum Sonority Distance (MSD) Principle favors Liquids in C2 position of a tautosyllabic cluster. Thus Obstruent-Liquid (OL) clusters such as /kr/ or /pl/ having an SD of 3–1 = 2 are predicted to be more common and natural than other types. On the other hand, according to the Sonority Dispersion Principle (SDP), in the demi-syllable, the segments should be maximally dispersed, and therefore OG clusters are considered more optimal. The question of which cluster type (OL) or (OG) is preferred in the C2 position cross-linguistically was taken up in a large-scale corpus data from 122 languages by Parker (2012) who observed neither of these two principles alone can account for all languages, but some notion of sonority differential between C1 and C2 is indeed crucial to account for preferred syllable types. These facts have implications for our understanding of acquisition of reading processes. For example, it has been demonstrated that in both bilinguals (Spanish-English) and English monolinguals, sonority influences phoneme segmentation performance (Yavas & Core, 2001; Yavas & Goldstein, 2006). For example, the English word, ‘de**sk**’ was found to be easier to segment compared to a word such as ‘mi**lk**’ because the distance between [l/] and [k] is greater than that between [s] and [k].

Telugu Orthographic Syllables and Sonority Dispersion

Of the total 1600 different syllables attested in the CIIL Telugu corpus reported by Uma Maheshwar Rao (2006), 635 were observed by this writer to be C1C2V type including about 100 closed syllables ending in a consonant /m/, that is of the shape C1C2Vm. They tend to occur in hetero-syllabic contexts, in the second syllable of two or three syllable words (e.g. *pa.**rva**.tam* ‘mountain’). The percentage of occurrence in the corpus, and the ‘D’ values for four different orthographic syllable shapes shown in Table 2 suggest that sonority considerations might have had some influence on the frequency of occurrence of orthographic syllables (O.S) in Telugu script.

In Telugu words with clusters, the phonological syllable (P.S) does not match the orthographic syllable (O.S); further, the first consonant of the conjunct appears in full form whereas the second one takes the secondary grapheme; they also vary in sonority distance (S.D) depending upon the segments involved. For example, the word, *gurtu* ‘sign’ is syllabified phonologically as *gur.tu*, orthographically it is

Table 2 Frequency of occurrence and ‘D’ values for medial two-consonant clusters in Telugu

O.S Type	Telugu Word	Gloss	Frequency (%)	‘D’ value
OLV	*gu. ḍlu*	Eggs	20.0	0.56
OGV	*ma. dhya*	Middle	12.6	1.17
NLVm	*aa. mlam*	Acid	1.5	1.36
LGV	*suu. ryu.ḍu*	Sun	1.8	2.25

represented as gu.rtu. Even from the point of view of sonority profile, *gurtu* with an S.D of 3.0 is easier to segment than *goorlu* 'nails' (with an S.D of zero). The word, *goorlu* also contains the plural suffix *–lu* also requiring morphological knowledge.

It should be noted that not all orthographic syllables obey Sonority principles in every language. The /S/ + obstruent clusters very common in English do not follow SSP fully. Literacy level also has an impact on the way we segment words (Ramachandra & Karanth, 2007). Drawing on the results of her work with Telugu-English biliterate adults, and language games children play, Sailaja (1999) pointed out that the performance on syllable transposition and phoneme deletion reflect influence of orthography which she argued impinges very heavily in the analysis of words, both in Telugu and English.

Assessment of Grapheme Awareness in Telugu

It was reported in Section "Telugu Orthographic Syllables: Statistical Patterns" that Telugu abounds in graphic syllables containing geminates and clusters. Further the agglutinating morphology associated with Telugu results in considerable morpho-phonemic changes when two words are combined. A clear-cut developmental trend was observed by this writer in the production and segmentation of compound nouns in children belonging to grades III to IX (see Vasanta & Sailaja, 1999 for details). The extent of morpho-phonemic changes during compounding of single words seemed to have affected their segmentation ability. In another study, children in grades IV and VI found it difficult to separate consonants and vowels in closed orthographic syllables in words with clusters of the type CVC1C2Vm such as: *paʈnam* 'town', *mukhyam* 'important', *padyam* 'poem', *mandram* 'low', *swargam* 'heaven', *gaatram* 'voice', *varʃam* 'rain', *nasyam* 'snuff', *muulyam* 'price, *gunɖram* 'round' although they were studying in Telugu medium schools in Hyderabad. The task required the participants to pay attention to the second syllable of each word, separate the consonants and the vowels in that syllable and write them down against each test word as shown in the example *paʈnam* 'town' below:

*Pa.**ʈnam→ʈa, na, am***

In this word, [−ʈa] appears in its primary form, [−na] in the secondary form and the final segment is represented with a *sunna* or zero following vowel grapheme /a/. Surprisingly, each of the ten children (average age: 11.8 years) scored zero on every single item in this task. They did however perform much better on words with open syllables of the type: CVVC1C2V (e.g. *waa.rta* 'news') or *koo.staa* 'coast' suggesting that the difficulty has little to do with the presence of cluster as such, and that instead, it may be related to the type of segments making up the cluster. In a subsequent large- scale investigation into linguistic awareness of normal-hearing (N = 15) and hearing-impaired children (N = 15) studying in Telugu medium schools conducted by this writer in Hyderabad during 2003–2004, it was observed that majority of hearing-impaired children were much better at segmenting words with two consonant clusters even in closed syllables compared to normal-hearing children (See Duggirala 2012). Some examples of test stimuli used with hearing-impaired children were: *raa.ʈnam* 'spinning wheel', *ce.kram* 'wheel', *po.tram* 'grinding stone', *waa.kyam* 'sentence', *sa.bdam* 'sound', *ra.ktam* 'blood' and *pa.kʃam* 'side'. Since these words had similar syllable structure (CVC1C2Vm/CVVC1C2Vm) as those presented to normal hearing children, and since sonority considerations are largely irrelevant for these children who have severe to profound hearing impairment, it was felt that their orthographic knowledge might have helped them succeed in this segmentation task. To probe this question further, the data relating only to orthographic awareness tasks given to both normal hearing and hearing-impaired children was reanalyzed for the purposes of the present paper. Specific tasks and the test materials designed to collect this data are described next (See appendix for the list of target items used in each task).

Syllable Substitution Task

This task consisted of 20 Telugu words with three types of syllable structures such as CVCV(CV), CVC1C1V (with geminates), and CVC1C2V (with clusters). In each word, one orthographic syllable was enclosed in a box. The task of the participant was to replace the boxed syllable with either one of two alternative syllables written against each word such that the newly created word is yet another meaningful word in Telugu. See the examples below:

paa. $\boxed{mu}$ *.lu* 'snakes'	*ʈa, sa*	*paaʈalu* 'songs'
na. $\boxed{kka}$ 'fox'	*ʈa , tta*	*natta* 'leech'

Instructions I will show you some Telugu words. In each word, one aksharam is put inside a box. I want you to look at two other aksharaas written against each word, select one of them, replace the one in the box and write the new meaningful word. For example, in the word, *baa.lu,* 'ball' -lu is inside the box. If you replace it with –tu, you get *baatu* 'duck'. But if you replace –lu with –da, we get, *baada* which is not a meaningful word. Seven practice items were used to make sure that the participants understood the task. Correct response was given a score of 1.0, and wrong response, a zero.

Syllable Deletion Task

This task also consisted of 20 target items. All of them had three syllables each with slightly varying syllable structures (all three open syllables CVCVCV; syllables with geminates, CVC1C1V; and syllable with nasal or non-nasal cluster CVC1C2V). The task consisted of reading the word, deciding which one of three syllables can be deleted to create a new meaningful word. See examples given below in which the to-be-deleted orthographic syllable is mentioned, but that was not done in the data-sheet.

Target word	Deleted segment	New word
kaa. ku. lu 'crows'	-- *ku*	*kaalu* 'leg'
gon. ga. ḍi 'blanket'	-- *ga*	*gaḍi* 'latch'

Instructions I will show you some words. Each word has three *aksharaas.* I want you to take out one *akshara* from each word and make up a new meaningful word and then write the new- word against the original word. For example, look at the word *ti.waa.ci* 'carpet'. If we remove 'ti' *akshara* from it, it will become *waaci* 'wrist watch'. There were six practice items. None of the children experienced any difficulty in getting the practice items right. There were one or two items in which when either of the two syllables are deleted, two different meaningful words could result. For example, from the word, *taa.ma.ra* 'lotus', if we remove *–taa,* we get *mara* 'machine'; if we remove *–ma,* we get *taara* 'star'. In such cases, either response was scored as correct with a score of 1.0; wrong responses were scores as zero.

Changing Vowel Component of the Akshara

Vowel length is phonemic in Telugu such that a mere change in the duration of a vowel changes the meaning (for e.g. *nadi* 'river'; *naadi* 'mine'). The idea of administering this phonemic processing task was to see if children from both the groups succeed in making the vowel in the first orthographic syllable long (sensitivity to phonemes) and decide whether the resulting new combination is a legal or illegal word. If they thought the result of this manipulation creates a new meaningful word, they were asked to put a tick mark against the new-word next to each target word. Otherwise they were asked to put X mark. There were 20 items in which 10 would result in legal words and 10 non-words. The items were randomized and presented on a data sheet. Two sample items are shown below:

$$\boldsymbol{me}.da \text{ 'neck'} \rightarrow \boldsymbol{mee}.da \text{ 'two – story building'} \left(\surd\right);$$

$$\boldsymbol{pu}.li \text{ 'tiger'} \rightarrow \boldsymbol{puu}.li \left(\mathrm{X}\right)$$

Instructions In each word printed below, I want you to change the first *akshara* so that it becomes long (by using *diirgham* ‘length marker on a vowel), write the new-word against each item and if you think the new-word is meaningful in Telugu, put a tick mark against it, if not put an X next to it. Two practice items had to be completed before providing answers to the test items. Each child’s written responses were scored separately and assigned 1.0 for correct placement of vowel length marker, and one for deciding whether the newly created word is meaningful or not. A zero score was given if they failed to give either of these responses.

Empirical Findings

All the 30 children 15 normal-hearing children (with an average age of 10.01 years) and 15 hearing- impaired children (with an average age of 14.09 years) were native speakers of Telugu. They were studying in Telugu medium schools at the time of data collection. The normal- hearing children were in elementary grades III to V, whereas the hearing-impaired children were enrolled in grades V-VIII. The latter were on the average 4 years older than normal hearing children. The four-frequency pure tone average hearing loss of the hearing-impaired group was 94.5 decibels. Details about participants, data collection procedures and comparison of their phonological versus orthographic awareness abilities are available in Vasanta (2007, 2014). Results based on orthographic syllable substitution and syllable deletion tasks alone are summarized in Table 3.

The results displayed in Table 3 above clearly demonstrate that normal-hearing children did much better on syllable substitution task than on syllable deletion task. They might have drawn primarily on their phonological skills. The fact that this difference is not so stark in the case of hearing-impaired children indicates that they might have relied primarily on orthographic cues. Of the 20 target items, only five words had clusters (−nṭ, −kr, ṭl, nḍ, mp). On these words and those with geminates such as −nnu, −llu, hearing-impaired children exhibited just chance level performance (scoring 9, 10 or 11 out of 20). Normal-hearing children scored 11 or 12 only on words, *gampa* ‘basket’ and *meṭlu* ‘steps’.

In the syllable deletion task, the performance difference between the two groups is 5%. Some of the test items on which normal-hearing children did poorly (at

Table 3 Mean scores (% correct) and Standard Deviation values on substitution and deletion tasks for both the groups

Task	Normal Hearing children (N = 15)	Hearing Impaired children (N = 15)
Orthographic syllable substitution	91.7	79.0
	(7.2)	(13.1)
Orthographic syllable deletion	73.3	68.3
	(14.2)	(12.2)

chance level) were: *kum(pa)ṭi* 'coal-stove' *panḍu(ga),* 'festival', *muggu(ru)* 'three people', *(gon)gaḍi* 'blanket', *we(nne)la* 'moonlight', and *dza(lle)ḍa* 'sieve' all with either geminates, clusters or homorganic nasals. Choosing which syllable to delete to make up another meaningful word was not indicated in the data-sheet. It thus appears that this task required both phonological and orthographic knowledge. If the visual input from the *aksharas* alone is influencing the performance, then hearing-impaired children who were on the average 4 years older and who were already in secondary schools would have done well on these items. The words that contributed to below chance level performance in this group were: *kumpaṭi, paawaḍa, molaka, waasana, munaga, mogali, gongaḍi, wennela, muppoddu, dzaabili* (see appendix for meanings of these words). They scored 10 or less out of a maximum of 20. Diminished or absent phonological clues, poor lexical knowledge combined with complex orthographic structure made this task more difficult for all the children.

In the third task, each child was required to make the first *akshara* long by attaching vowel length marker, write the newly generated form, and make a lexical decision as to whether the result is a meaningful word in Telugu or not. Table 4 displays the results on this task:

It can be seen from Table 4, that while normal-hearing children were highly successful in attaching long vowel *maatra* to the first syllable of each test word, the hearing-impaired children were not so successful as evident by the fact that the difference between the group scores is >25%. While both groups did well on changing the length of the inherent vowel, [a], in eight of the 20 items, they attached wrong length markers on the vowel [e] in the word *meḍa* 'neck' creating non-words. When appropriate length marker is attached to [e], it should become a meaningful word, *meeDa,* 'a two-story building'. Instead they made use of length marker for [o] or [u] etc. Due to their hearing loss, they might have failed to draw on phonological knowledge.

Hearing-impaired children also judged many non-words as words as evident in their overall score of less than 50%. Again, it is only hearing-impaired children who attached wrong vowel *maatras* (e.g. *pani* 'work' in this task becomes *paani,* a non-word in Telugu, but means 'water' in Hindi. Hearing-impaired children made it *pauni*, another non-word) or attached it to the wrong (second) syllable, as in example, for the test word, *kona* 'corner' their written response was ko**naa**; they also copied wrongly (e.g. *weḷḷu 'go'* should become *weeḷḷu* 'fingers' after the vowel in the first syllable becomes long. Some of the hearing-impaired children wrote *peḷḷu* a non-word*)*. On occasion, they simply copied the original word without attaching long vowel *maatra* (e.g. *pani – pani; tala-tala)*. Normal-hearing children never gave such responses. The higher standard deviation values on the lexical deci-

Table 4 Results based on adding length marker to the vowels

Task	Normal-Hearing children	Hearing-Impaired children
Adding vowel length marker	93.33	66.33
	(5.9)	(9.5)
Lexical decision	74.0	44.33
	(19.4)	(15.6)

sion task also suggest that performance on this task is more variable than the one involving manipulating a vowel *maatra*. This test has successfully differentiated the performance of normal-hearing and hearing-impaired children, who had differential phonological processing abilities.

Since vowel length is phonemic in Telugu, by simply changing the length of the vowel, one can generate orthographic neighbors. For example: *naaru* ‘sapling’ *niiru,* ‘mouth’ *nooru,* ‘water’ *nuuru* ‘hundred’. This task surely taps on lexical knowledge. Further, words with homorganic nasals will have very few neighbors whereas those with geminates will have many neighbors (e.g. *kaḷḷu* ‘eyes’ is associated with *paḷḷu, guḷḷu, waḷḷu, weeḷḷu, gooḷḷu, muḷḷu, kooḷḷu, dzooḷḷu, raaḷḷu,* and so on, all meaningful words in Telugu). The second orthographic syllable in these words contains plural marker; only those with short [a] in the first syllable are phonological neighbors. If we hold on to the first orthographic syllable and change the next one, we get neighbors such as *kannu, kappu, kallu, kaṭṭu, kaccu, kalam, kawwam, kannaam,* and so on, again all of them are meaningful in Telugu. Insights such as these guided the assessment and instructional material designed in the research described in this chapter.

Implications for Assessment and Instruction

The National Council for Educational Research and Training (NCERT) in its 2006 position paper on ‘Teaching of Indian Languages’ in schools expressed concern over majority of normal children with good oral language competence in at least two or three languages when they come to school are still exhibiting dismal level of performance in relation to comprehension and writing skills even in their native languages. The question of assessment of reading / spelling skills in all school children (not just those who are diagnosed as ‘dyslexics’) is a major challenge. Most of the assessment batteries in relation to Indic languages, and Brahmi-derived scripts including those described in this chapter were developed primarily for the purposes of answering specific questions of interest to the researchers. For the most part, they are limited to testing beginning reading skills. Also, there is limited corpus-based studies providing information on frequency of graphic syllables or grapho-tactic patterns in relation to Indian languages.

In this context, the recently published Dyslexia Assessment for Languages of India (DALI, 2015) is a welcome addition to the language-resources *that* teachers, teacher-educators and special education personnel should have access to. DALI is an attempt to provide a standardized screening and assessment battery in English, Kannada, Marathi and Hindi to identify school children at risk for dyslexia (see this volume). DALI manual also describes a Language Assessment Battery (LAB) that is modeled after an English phonological assessment battery developed by Uta Frith and her colleagues. DALI-LAB test formats are likely to be useful to school teachers in identifying children who are not reading and spelling at the expected grade levels. However, using translation equivalents of English test items with relatively simple syllable structures into Indian languages can pose several problems: (1) Familiar constructs such as 'word' may have different realizations in English as opposed to Indic languages / scripts. Languages belonging to the Dravidian family also have a number of geminate consonants that are not so common in English (2) translation might miss certain critical features of our languages /scripts (for e.g. Marathi has aspirated nasals, laterals and semi-vowels, whereas Hindi does not have those sounds; nasal vowels have greater functional load in Hindi than Marathi), (3) translation might also affect the item difficulty due to differences in use frequency (for example, the vowel grapheme, /e/ is more frequent in Hindi than Marathi, while grapheme corresponding to /o/ is equally infrequent in Hindi, Marathi and Telugu (Ghatage, 1964; Ramaswami, 1999).

The research described in this chapter suggests that we could use the following tasks involving compound nouns (CN) to improve orthographic awareness in Telugu (see Vasanta & Sailaja, 1998 for a list of 50 compound nouns in Telugu classified into five groups based on the complexity of morpho-phonemic changes during compounding). Fifty different target sentences can be designed to elicit data that taps on both phonological and orthographic awareness.

1	**2**	**3**	**4**	**5**
Identify C.N in a sentence written without word breaks	Generate a CN when given two simple nouns	Segmentation of compound noun into its parts	Count # of aksharas in each CN	Isolate vowel and consonant graphemes in the CN
6	**7**	**8**	**9**	**10**
Segment words with clusters	Delete one akshara to make a new-word	Rhyme recognition & generation for printed words	Generate as many ON and PN as possible	Read short written sentences

To illustrate with an example, the test sentence, *waaḷḷiddaru samudrapuṭalalu cuustuu kuurconnaaru* 'They sat watching the waves in the sea'. This sentence can be presented without word breaks. Children should be told to mark the boundaries of words. They can then divide the two individual components of the compound word *samudrapuṭalalu* as *samudram(u)* + *alalu* (notice the segment [ṭ] is an intrusive one that is present in the compound but not in the single words that make up that compound). They can then isolate the consonant and vowel graphs within each orthographic syllable (s + a + m + u + d + r + am and a + l + a + l + u); they can be asked to provide a rhyme for the word, *ala* 'wave' and write as many orthographic neighbors for the word *a.la* 'wave' as possible by substituting one orthographic syllable in the second position of the target word with another orthographic syllable (some e.g. *ara, anu, accu, appu, aggi, ammu, addu, aḍḍu, allu, akka, amma, anna, awwa)* and then substitute the first orthographic syllable with another orthographic syllable (e.g. *kala, wa;a, tala, nela, goola, sila, gela, wela, maala, soola)*. They can be encouraged to check their answers in the dictionary to make sure that all these 'neighbors' are indeed meaningful in Telugu. A computer program can be developed for this purpose. If we collect such data from school children or their teachers, we will automatically generate a huge data-base of words with different syllable structures and discover

Table 5 Orthographic awareness scores of hearing-impaired children before and after training

Participant's name	Orthographic awareness % correct scores	Orthographic awareness % correct scores
	Before training	After training
T.B.	78.33	91.66
P.S.R.	80.0	93.33
S.P.V.	76.66	93.33
T.J.S.	73.33	98.33
C.V.P.	80.0	90.0
B.M.	81.66	90.0
G. B.	80.0	96.66
J.B.	81.66	86.66
K.K.	63.33	86.66
G.P.	56.66	65.0

phonological and orthographic properties that govern them. One can also make use of Telugu compound verbs in designing assessment or instruction material.

Turning to the issue of literacy instruction, at primary grades we should include exercises in manipulating phonemes and syllables represented in the script. It is pertinent to cite one such effort undertaken by this writer with 10 of the 15 hearing-impaired children who were already assessed for both phonological awareness and orthographic awareness. Six weeks of training was given to these 10 hearing-impaired children using 60 different exercises in manipulating orthographic syllables in words made up of different syllable structures (Table 5).

The difference in performance before and after training was statistically significant ($P > 0.01$) with respect to orthographic awareness while this was not true with phonological awareness training (for more details and comments on educational implications see Vasanta, 2007, 2014).

Final Comments

The research described in this chapter offers support to the idea that new assessment tools for testing phonological and orthographic awareness in Brahmi-derived scripts at least with reference to Telugu and Kannada should take into consideration language/script-specific syllable structures, sonority profiles, and word frequencies. Reading research on Brahmi-derived scripts will also benefit from more corpus-based studies that can provide information about frequency of occurrence and types of syllable structures in Indic languages and scripts associated with them. The fact that in India more and more children are receiving literacy instruction in English, even before they master the script associated with their native language, more studies should be undertaken to examine mutual influence of different scripts on decoding and comprehension during reading in multilingual environments.

Acknowledgments I am grateful to all the children who provided the data discussed in this chapter and to their parents and teachers for extending cooperation during data-collection and training phase. This research was supported by a grant from the University Grants Commission (UGC), New Delhi under their Research Award scheme.

Appendix

Target Words and their glosses in Orthographic syllable substitution, syllable deletion, and change of vowel *maatra* tasks

1. In the first column, syllable printed in bold must be replaced with one of two alternate syllables
2. In the fifth column under Vowel change task, the first ten words result in new meaningful words even after attaching a length marker to the vowel portion of the akshara, the next ten do not. They were mixed and presented randomly.
3. The syllable division indicated by a dot in the target words refers to orthographic syllables.

Syllable Substitution		Syllable Deletion		Vowel Change	
Target word/O.S	Gloss	Target word	Gloss	Target word	Gloss
*gam.**pa*** *ṭa, ka*	Basket	*daa.ni.mma*	Pomegranate	*na.di*	River
*na.**kka*** ppa, tta	Fox	*mu.gg.uru*	Three people	*pu.ri*	Hay-stack
***caa.**ku* mee, gaa	Knife	*kaa.ku.lu*	Crows	*me.ḍa*	Neck
*gu.**ḍḍi*** rram, ra	Blind	*wee.shaa.lu*	Costume	*si.la*	Statue
*ba.**llu.**lu* ṭlu, ssu	Lizards	*mo.la.ka*	Sprout	*pa.ḍu*	Fall
*ce.**lli*** ppu, pi	Sister	*kum.pa.ṭi*	Coal-burner	*ka.ḷḷu*	Eyes
*ce.**kka*** ḍii, kram	Wood	*dii.paa.lu*	Lights	ta.ḍi	Wetness
***li.**pi* i, ṭoo	Script	*gu.ḍi.se*	Hut	*dza.ḍa*	Plait
***ce.**Tlu* me, ta	Trees	*mu.ppo.ddu*	3rd quarter of the day	*ko.na*	Corner
*paa.**mu.**lu* ṭa, sa	Snakes	*we.nne.la*	Moon light	*we.ḷḷu*	Go
gun. ḍu rem, num	Bald head	mu.na.ga	Drumstick	*ga.ḍi*	Latch
*paa. **ṭa*** bu, lu	Song	*gon.ga.ḍi*	Blanket	*pa.ni*	Work
*mo.**kka.**lu* nnu, gga	Plants	*taa.ma.ra*	Lotus	*ta.la*	Head
*ba.**rre*** lli, di	Buffalo	*dza.lle.ḍa*	Sieve	*ce.ḍu*	Bad
*waa.**dyam*** tu, ram	Musical instrument	*mo.ga.li*	A flower	wi.nu	Hear
*ree.**wu.**lu* ku, ga	River banks	*paa.wa.ḍaa*	Skirt	po.da	Bush
***mee.**ku* aa, Taa	Nail	*pan.ḍu.ga*	Festival	*li.pi*	Script
***dum.**palu* dzam,gam	Root vegitables	*ki.raa.yi*	Rent	*si.ga*	Hair-do
pe.nnu ta, dza	Pen	*waa.sa.na*	Fragrance	*mu.ni*	Saint
*co.**ṭṭa.**lu* lla, kkaa	Dents	*dzaa.bi.lli*	Moon	gu.ri	Aim

OS Orthographic Syllable

References

Alvarez, C. J., Garcia-Saavedra, G., Luque, J., & Taft, M. (2016). Syllabic parsing in children: A developmental study using visual word spotting in Spanish. *Journal of Child Language, 44*. https://doi.org/10.1017/S0305000916000040. Downloaded on 23rd Feb. 2016.

Clements, G. N. (1990). The role of sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology* (pp. 283–333). Cambridge, UK: Cambridge University Press.

Clements, G. N., & Keyser, J. (1983). *CV phonology: A generative theory of the syllable*. Cambridge, MA: The MIT press.

DALI. (2015). *Dyslexia assessment for languages of India*. Manesar, India: N.B.R.C.

Daniels, P. T. (2008). Writing systems of major and minor languages. In B. B. Kachru, Y. Kachru, & S. N. Sridhar (Eds.), *Language in South Asia* (pp. 285–308). New York, NY: Cambridge University Press.

Duggirala, V. (2012). *Linguistic awareness and hearing impairment in Telugu children: Perspectives on assessment and training*. Saarbrücken, Germany: Lambert Academic Publishers.

Ettlinger, M., Finn, A. S., & Hudson Kam, C. L. (2012). Effect of sonority on word segmentation: Evidence for the use of a phonological universal. *Cognitive Science, 36*(4), 655–673.

Ghatage, A. M. (Ed.). (1964). *Phonemic and morphemic frequencies in Hindi*. Pune, India: Deccan College.

Harris, M., & Moreno, C. (2004). Deaf children's use of phonological coding: Evidence from reading, spelling and working memory. *Journal of Deaf Studies and Deaf Education, 9*(3), 253–268.

Hill, E. C. (1991). *A primer of Telugu characters*. New Delhi, India: Manohar Publications.

Joanisse, M. F. (1999). *Exploring syllable structure in connectionist networks*. Paper presented at the 14th international congress of phonetic sciences, San Francisco, CA.

Kaefer, T. (2016). Integrating orthographic and phonological knowledge in early readers: Implicit and explicit knowledge. *Child Development Research, 2016*, 1. Retrieved from: https://doi.org/10.1155/2016/6036/29.

Krishnamurti, B., & Gwynn, J. P. L. (1985). *A grammar of modern Telugu*. New Delhi, India: Oxford University Press.

Miller, D. G. (1994). *Ancient scripts and phonological knowledge*. Amsterdam: John Benjamins.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*(1), 41–52.

Nakamura, P. R., Joshi, R. M., & Ryan Ji, X. (P.C.). *Investigating the asymmetrical roles of syllabic and phonemic awareness role in akshara processing*. Unpublished paper.

Olson, A. C. (2004). Orthographic structure and deaf spelling errors: Syllables, letter frequency, and speech. The Quart. *Journal of Experimental Psychology, 57*(3), 385–417.

Padden, C. (1993). Lessons to be learned from young deaf orthographer. *Linguistics and Education, 5*, 71–86.

Pandey, P. (2007). Phonology-orthography interface in Devanagari for Hindi. *Written Language and Literacy, 10*(2), 139–156.

Parker, S. G. (2012). Sonority distance and sonority dispersion- a typological survey. In S. G. Parker (Ed.), *The sonority controversy* (pp. 101–165). Berlin, Germany: De Gruyter Mouton.

Patel, P. G. (2007). Akshara as a linguistic unit in Brahmi scripts. In P. G. Patel, P. Pandey, & D. Rajgor (Eds.), *The Indic scripts: Paleographic and linguistic perspectives* (pp. 167–213). New Delhi, India: DK Print World.

Purushothama, G. (1994). *A framework for testing Kannada reading*. Mysore, India: CIIL.

Ramachandra, V., & Karanth, P. (2007). The role of literacy in the conceptualization of words: Data from Kannada speaking children and non-literate adults. *Reading and Writing: An Interdisciplinary Journal, 20*, 173–199.

Ramaswami, N. (1999). *Common linguistic features in Indian languages: Phonetics*. Mysore, India, CIIL.

Rao, U. M. (Ed.) (2006). *An analysis of the Telugu corpus: Some preliminary findings*. Unit-VII. Corpus linguistics: PG Diploma in Computer Applications in Indian Languages (PGDCAIL) 421 (pp. 125–161). Hyderabad, India: University of Hyderabad.

Reddy, P. P., & Koda, K. (2013). Orthographic constraints on phonological awareness in biliteracy development. *Writing Systems Research, 5*(1), 110–130.

Rimzhim, A., Katz, L., & Fowler, C. (2014). Brahmi derived orthographies are typologically Aksharik but functionally predominantly alphabetic. *Writing Systems Research, 6*(1), 41–53.

Sailaja, P. (1999). *Syllable structure of Telugu*. Paper presented at the International Conference of Phonetic Sciences 14 (ICPhS) held at San Francisco, CA. Retrieved from: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_0743.pdf

Salomon, R. (2000). Typological observations on the Indic script group and its relationship to other alphasyllabaries. *Studies in Linguistic Sciences, 30*(1), 87–03.

Sastry, J. V. (1972). Telugu phonetic reader. Mysore, India: CIIL.

Share, L. D. (2014). Alphabetism in reading science. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2014.00752.

Share, L. D., & Daniels, P. T. (2016). Aksharas, alphasyllabaries, abugidas, alphabets, and orthographic depth: Reflections on Rimzhim, Katz, and Fowler (2914). *Writing Systems Research, 8*(1), 17–31.

Vasanta, D. (2004). Processing phonological information in a semi-syllabic script: Developmental data from Telugu. *Reading and Writing: An Interdisciplinary Journal, 17*, 59–78.

Vasanta, D. (2007). Phonological awareness and literacy acquisition in Telugu speaking normal-hearing and hearing-impaired children. *Contemporary Education Dialogue, 4*(2), 217–237.

Vasanta, D. (2014). Phonological knowledge: Theoretical and clinical considerations In D. Vasanta (Ed.), Clinical applications of phonetics and phonology. *Monograph of the Indian speech and hearing association* 14:1 (pp. 1–47).

Vasanta, D., & Sailaja, P. (1998). Word awareness and word formation: A study of compound noun production and segmentation by Telugu children. In B. Vijayanarayana & C. Ramarao (Eds.), *Word formation in Indian languages* (pp. 35–51). Hyderabad, India: Book Links Corporation.

Vasanta, D., & Sailaja, P. (1999). Making sense of compound nouns: A study of word relatedness in Telugu. *Journal of Psycholinguistic Research, 28*(4), 331–346.

Vennemann, T. (1988). *Preference laws for syllable structure and the explanation of sound change*. Berlin, Germany: Mouton.

Yavas, M. S., & Core, C. W. (2001). Phonemic awareness of coda consonants and sonority in bilingual children. *Clinical Linguistics and Phonetics, 15*(1–2), 35–39.

Yavas, M. S., & Goldstein, B. (2006). Aspects of bilingual phonology: The case of Spanish-English bilingual children. In Z. Hua & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 265–285). Clevedon, UK: Multilingual Matters.

Vasanta Duggirala has obtained her doctoral degree in Audiology from the University of Memphis, Tennessee, U.S.A. in 1986. During 1987–2015, she was on the teaching faculty of the Department of Linguistics, Osmania University in Hyderabad, India. After her retirement, she served as an Emeritus Fellow of the University Grants Commission, New Delhi from 2015–2017 and worked on the project titled, *Language and Experience: A typology of Telugu Ideophones*. Her post-doctoral research projects and publications deal with different aspects of applied linguistics, special education, gender studies and translation.

Multiliteracy in Akshara and Alphabetic Orthographies: The Case of Punjabi, Hindi and English Learners in Primary Schools in Punjab

Seema Gautam, John Everatt, Amir Sadeghi, and Brigid McNeill

Abstract This chapter reports research on Grade 2–5 children learning to be literate in Punjabi, Hindi and English. Children (80–100 per grade) were assessed on reading comprehension, listening comprehension, non-word reading ability, phonological processing, orthographic knowledge and speed of processing in all three languages. The focus was on predictors of reading in the two akshara orthographies (Punjabi and Hindi) and the one alphabetic orthography (English). For both Punjabi and Hindi, reading comprehension was predicted by listening comprehension and decoding ability. Orthographic knowledge also predicted reading comprehension after controlling for word recognition. In English, along with listening comprehension and decoding, rapid naming and orthographic knowledge were also independent predictor of reading comprehension. Findings will be considered in terms of the development of reading across different orthographies (akshara versus alphabetic) within children learning to be literate in multiple languages.

Keywords Akshara versus Alphabetic · Listening comprehension · Multiliteracy: Punjabi, Hindi, English · Non-word reading · Orthographic knowledge · Phonological processing · Reading comprehension predictors · Speed of processing

S. Gautam (✉) · J. Everatt · B. McNeill
School of Teacher Education, College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand
e-mail: seema.gautam@canterbury.ac.nz; john.everatt@canterbury.ac.nz; brigid.mcneill@canterbury.ac.nz

A. Sadeghi
Department of English Language Teaching, Islamic Azad University, Damavand, Iran
e-mail: amir.sadeghi@canterbury.ac.nz

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_8

Introduction

This chapter discusses research conducted in the Punjabi region of India involving trilingual children. These learners were being taught reading and writing in Punjabi, as their mother tongue, but were also learning language and literacy skills in Hindi and English as additional languages. The Punjabi language and orthography form the main focus of discussion, due to Punjabi belonging to akshara family of orthographies (the focus of the current book) but also because of it being relatively under-researched. Features of Punjabi will then be contrasted with those of Hindi, which is also among the akshara family of orthographies (Punjabi and Hindi have both been influenced by the Brahmi script), and is the national/official language in India (Hindi is also discussed in other chapters in this book; hence, it is used for contrast here rather than for specific focus). Punjabi and Hindi will also be compared to English, the third language that the children studied in this research were learning to be literate in English as an alphabetic orthography provides a potentially interesting comparison with the two akshara orthographies.

The constitution of India recognises 22 major languages out of the many languages spoken in this multilingual country. Simultaneous acquisition of reading, and the ability to comprehend written text in various languages, therefore, is very common in schools in India, making it highly appropriate for the conducting of research that aims to increase our understanding of the multi-faceted nature of the process required for multiliteracy acquisition. Clearly, reading comprehension is a complex process and requires the acquisition of numerous skills (Koda, 2007). However, the Simple View of Reading (SVR) provides a relatively concise model that can be applied across different orthographies, and has been used to support explanations of variability in reading performance among bilingual/multilingual children (within languages and across language). It, therefore, provides a useful framework with which to consider the current research on multilingual children learning to read in three orthographies. The SVR argues that reading comprehension is the product of two components. The first is described as linguistic comprehension, and comprises those processes that support comprehension of a discourse presented verbally. The second can be considered as decoding or efficient word recognition. The latter component should be predicted by those processes that support the translation of written material into language, such as phonological and orthographic processes. In present research, the SVR was used as theoretical base to determine potential predictors of reading comprehension in two aksahara orthographies and one alphabetic orthography.

The chapter starts with an overview of the Indian language policy in education to understand the use of the three taught orthographies, before focusing on some of the features of Punjabi and how they contrast with one other akshara orthography and an alphabetic orthography learned by the same children. It will then provide an overview of the findings of a study to investigate predictors of reading levels in these three orthographies within the same group of primary school learners.

The Indian Language Policy in Education

India can be considered as one of the highest ranked countries in the world in terms of language diversity. There are 300–400 languages used in the country (which belong to five distinct language families), with 22 being acknowledged as official languages in the VIIIth schedule of Indian Constitution (see discussions in Hornberger & Vaish, 2009; Koul, 2005; Mohanty, 2010). Hindi and Punjabi are two of these 22 official languages, and Hindi and English have been accorded the status of 'official language' and 'co-official language' by the central government for use in most of its administrative communications. Punjabi is recognized as the official language of the state of Punjab, and therefore is used in most of its administrative correspondence. These national policies have influenced education too. The National Commission on Education 1964–66 (Hornberger & Vaish, 2009) recommended the 'three language policy' in education. According to this policy, every state has the right to use its regional language as an official language and language of instruction, alongside Hindi and English, which should also be part of education (Meganathan, 2011). This guiding principle is implemented within Punjab, where education in government primary schools is provided in the regional language (i.e., Punjabi), with children also learning Hindi and English as compulsory languages.

The Punjabi Language

Punjabi (also spelled Panjabi) is the official language of the Punjab region of India, and is written in the Gurmukhi script. Punjabi is a relatively new Indo-Aryan language. It is spoken by approximately 100 million people around the world, with approximately 33 million speakers being from the north Indian states of Punjab, Haryana and Himachal Pardesh (Bowden, 2012). As compared to other modern Indo-Aryan languages, Punjabi has some unique features. First, its lexicon is closely related to the early Vedic Sanskrit. Furthermore, due to Punjab's adjacency to the Middle East and Central Asia, and the influence of Islam in the region, it has absorbed a range of words and expressions from Arabic and Persian/Farsi (Mann, 2011). Punjabi also has the feature (unique among North Indian language) of including tones; however, this linguistic feature has yet to be fully examined and explained (Bhatia, 1994).

Punjabi has been considered to include seven basic types of syllables (see Nigam & Sen, 1975): (i) Vowel (V), such as ਆ /a/ (which means 'come' in English); (ii) Vowel-Consonant (VC), such as ਉਹ /ʊh/ ('that' in English); (iii) CV, such as ਗਾ /ga/ ('sing'); (iv) VCC, such as ਅੱਗ /əgg/ ('fire'); (v) CVC, such as ਗੀਤ /giṯ/ ('song'); (vi) CCVC, such as ਪ੍ਰੀਤ/priṯ/ ('love'); and (vii) CVCC, such as ਨੱਕ /nəkk/ ('nose'). The language includes both single and multiple-syllable words (Singh & Lehal, 2010), and the basic phonemes that comprise Punjabi can be grouped into 32 consonant sounds, 10 vowel sounds, and 2 semivowel sounds (these are discussed further in the next section on orthography), based on the place and manner of articulation.

Stress is also an important feature of Punjabi and varies from the initial to final syllable of the word in polysyllabic words; within a monosyllable word, stress occurs at the sentence rather than the word level (Singh, 2010).Similarly, Punjabi shows features consistent with vocal pitch (tone) differentiating words, although there is no representation of such features within the written script (Gill & Gleason, 1969). The discrepancies found in the existing research means that the existence, number and use of tones within Punjabi is an area of debate by linguistics (see Bowden, 2012; Malik, 1995).

The Punjabi Orthography

Punjabi is an Indo-Aryan language used in the Punjab states of both India and Pakistan. In India, Punjabi is written in the Gurmukhi script, whereas in Pakistan the Shahmukhi script is used to represent the language. The Shahmukhi alphabet is a version of the Urdu script, which is written from right-to-left, and is a form of the Persian-Arabic orthography. In contrast, the Gurmukhi script is derived from Brahmi, and uses a similar arrangement as Devanagari (Campbell, 1995). The Gurmukhi script is written from left-to-right, and is syllabic in nature; although it also shows features of a phonetic writing system, with a mainly consistent mapping between sounds and aksharas (see Bhatia, 1994). Punjabi aksharas follow the style of a horizontal bar drawn on top, with the body of the akshara written below. The length of the horizontal bar is longer than the size of the rest of the akshara, meaning that each akshara can be connected with previous and subsequent aksharas within a word.

According to a number of scholars (see Bowden, 2012; Mann, 2011), Punjabi can be considered to comprise 35 basic aksharas, which can be divided into vowels and consonant. However, several features of the orthography (including its akshara background) make this figure debatable. For example, consonant-vowel distinctions are complicated by the inclusion of ਧ and ਵ which are considered as semi-vowels in Punjabi. Additionally, the basic set of aksharas includes three (ੳ, ਅ, and ੲ) that are used as vowel carriers. The independent form of all vowels is always written with the help of one of these vowel carriers; and, of the three, only ਅ is used in isolation. Furthermore, Punjabi has some additional aksharas that are used to represent sounds within borrowed words. These (ਸ਼/ʃ/, ਖ਼/x/, ਗ਼/ɣ/, ਜ਼/z/, ਫ਼/f/, ਲ਼/ɭ/) occur with a small dot at the bottom left of the basic form, and are referred to as ‘pɛr bindi’, which means ‘with a dot underneath’. All these dotted characters are used only in adapted loan words (mainly from Persian, but also from Arabic). Hence, the number of basic aksharas can vary depending on what is included in the calculations.

The system for vowel representation in the Punjabi writing system is less straightforward than that for consonants. Consistent with the abugida orthography, even though vowel representation is necessary, vowel symbols are always linked to other forms. The 10 vowels forms in Punjabi may be oral or nasalized, and each has an

Table 1 Gurmukhi vowels in their independent and dependent forms

	Independent form	IPA	Dependent form (Laga matra)	Example with ਕ/k/
1.	ਅ/ə/	/ə/	No sign (schwa)	ਕ/kə/
2.	ਆ/a/	/a/	◌ਾ	ਕਾ/ka /
3.	ਇ/ɪ/	/ɪ/	ਿ◌◌	ਕਿ/kɪ/
4.	ਈ/i/	/i/	◌ੀ	ਕੀ/ki /
5.	ਉ/ʊ/	/ʊ/	◌ੁ	ਕੁ/kʊ/
6.	ਊ/u/	/u/	◌ੂ	ਕੂ/ku /
7.	ਏ/e/	/e/	◌ੇ	ਕੇ/ke/
8.	ਐ/ɛ/	/ɛ/	◌ੈ	ਕੈ/kɛ/
9.	ਓ/o/	/o/	◌ੋ	ਕੋ/ko/
10.	ਔ/aʊ/	/aʊ/	◌ੌ	ਕੌ/kaʊ/

independent and dependent form with the exception of the schwa /ə/, which only has an independent form. Table 1 shows the difference between independent and dependent forms of vowel characters that the child is required to learn, along with the rules to produce these forms correctly. Also, note that the independent forms of the vowels are written with the three vowel carriers referred to in the previous paragraph. Finally, the dependent forms of the vowels (which are diacritic marks) are written non-linearly above, below and to the left or right of the consonants with which it is pronounced – and, again, the child has to learn how to recognise and use these diacritic forms appropriately.

There are additional diacritics that the child needs to recognize and learn. For nasalization and gemination (i.e., the doubling of the sound of a consonant), Gurmukhi includes three symbols: the tippi and bindi for nasalization, and the addhak for gemination.The usage of tippi or bindi has been determined by orthographic conventions: bindi (◌ਂ)is used with kanna (◌ਾ), bihari (ੀ) or the independent forms of onkar (ਉ), dulankar (ਊ), lavan (ਏ),dulanvan (ਐ), hora(ਓ) and kanora (ਔ). For example, in the word ਕੰਨ /kənn/ (which means ear), a small symbol ੰ (tippi) on the consonant ਕ /k/ denotes the combination of the nasal sound with the consonant ਨ /n/. Similarly, in the word ਬਾਂਹ /banh/ (which means arm), the small dot (◌ਂ) on the vowel sign kanna (◌ਾ) represents that nasalisation of the sound. In contrast, addhak (◌ੱ)is used to double the length of consonant sounds (gemination): e.g., in the word ਨੱਕ /nəkk/ (which means nose), a symbol on the top of consonant ਨ/n/is used to double the length of the sound ਕ/k/.

In the Gurmukhi script, most clusters are bi-consonants and appear in the final and medial position of the words: for example, the /ʤd̪/ in ਮਜਦੂਰ /mʤd̪ur/ (which means 'workman') and the /lk/ in ਪਲਕ /pəlk/ (which means 'a moment'). However, Punjabi has phonotactic constraints, leading to consonant clusters in the initial position in a syllable/word occurring only with a few loan words from Sanskrit, and with three consonants (i.e., ਹ /h/, ਰ /r/, and ਵ /ʋ/) that form conjuncts when they adjoin with another consonant, which produces a change in their orthographic form.

In Punjabi, only consonant conjuncts form word-initial consonant clusters; such as ਨ੍ਹਾਤਾ /nhata/ (which means 'bathed'), ਪ੍ਰਕਾਸ਼ /prkaʃ/ (which means 'rise') and ਸ੍ਵਰਗ /sʊərg/ (which means 'heaven').

Hence, learners have a large range of orthographic symbols which they need to learn in order to read and write Punjabi. They, most likely, will need to recognise that written symbols can relate to syllables, but also isolated phonemes – depending on how the orthography is taught. They will also have to learn a relatively large number of rules for how written symbols vary in form and combine together across different contexts. Although, once the rules are learnt, the script is relatively transparent in its relationships between written symbols and language sounds, this complexity of orthographic representation may present challenges to the reader, and particularly the writer. Therefore, the skills required to acquire such an orthography may vary from those needed to support the development of an alphabetic orthography such as English (on which many models of reading acquisition have been developed), where the challenges (primarily the rules linking graphemes and phonemes) may be of a different form. In order to consider these differences, we will first consider the current context of research in which the child participants are required to learn Punjabi as well as Hindi (another akshara orthography) and English. Then data from a relatively large number of Punjabi-Hindi-English multilingual learners will be presented to assess possible differences in underlying skills needed to support the development of literacy skills in the three orthographies.

A Comparison with Hindi and English

There are many similarities between Hindi and Punjabi, mainly due to their similar origins. Hindi, like Punjabi, is an Indo-Aryan language, and uses a writing system that was based on Brahmi. However, Hindi is written in the Devanagari script (used for writing Sanskrit and some other Indian languages), which differs from the Gurmukhi script used for Punjabi. Hindi is generally considered to comprise 46 sounds (10 vocalic and 36 consonantal), although the exact number of sounds can vary depending on differing social and/or linguistic perspectives (see discussions in Agnihotri, 2013). Hindi has the same number (10) of basic vowel sounds as Punjabi, but it is considered to incorporate more syllable patterns: in Hindi, there are nine types of syllable pattern (basically the same as Punjabi, plus CCV and CCVCC – although, again, loan words can make such general statements controversial), with most words being two to three syllables in length (see Nigam & Sen, 1975). In general, first and second syllables are stressed in Hindi, though if a word has more than three syllables, the stress is typically on the penultimate (second last) syllable (see also Agnihotri, 2013; Bhatia, 2015; Koul, 2009). Stress can also be used to emphasise a constituent in a sentence (Nigam & Sen, 1975).

As with Punjabi, the basic written unit in the Hindi script is typically learnt as an akashara: a consonant with a vowel or a consonant with a vowel diacritic, or a vowel in full form. Hence, each aksharais a syllable. However, consistent with akshara orthographies, these written syllable units can also be broken down into forms that relate to basic phonemes. This mixture of alphabetic and syllabary features means that Hindi and Punjabi resemble alphasyllabic orthographies but belong to abugida family. Similar to Punjabi, characters are written below a horizontal bar, which helps to discriminate word boundaries. Also, similar to Punjabi, Hindi writing is nonlinear (see Gupta & Jamal, 2007; Vaid & Gupta, 2002) and includes loans from Arabic/Persian: a dot below an akshara represents a loan-sound.

As discussed above, the Punjabi orthography uses a range of diacritic marks. This is the same for Hindi. As with Punjabi, most vowels in Hindi have independent and dependent forms, with the schwa being the only vowel with no dependent form. Again, the dependent forms of vowels are placed non-linearly above or below or to the side of consonants as diacritical marks (in Hindi these diacritic marks are known as matra) and may be oral and nasalized. Bindi (also called anuswara) is a dot above a consonant (ं) and represents a nasal combination with the subsequent consonant. For example, in the word हंस/həns/ (swan), a small dot aboveह/hə/ is the nasal combination with the consonant स्/s/. Ardhachandra/chanderbindu (also referred to as anunasika: ँ) denotes the nasalization of the vowel. For example, in the word चाँद / ʧand/ (which means moon), a symbol above the vowel sign /a/ denotes the nasalised version of the vowel (see also Gill & Gleason, 1969; Nigam & Sen, 1975). Unlike Punjabi, Hindi does not have any sign for gemination; i.e., the doubling of a sound (see Bhatia, 2015, for further discussion). Hindi also includes word-initial, medial and final forms of consonant clusters; though word-medial consonant clusters are most common.

English also has some similarities with Punjabi and Hindi, given their inclusion in the Indo-European family of languages – though along different branches. However, English uses analphabetic orthography based on the Roman script and, therefore, shows very different orthographic features from the other two languages. English, Punjabi and Hindi are all written left-to-right, but word boundaries in English are indicated by spaces only and not the horizontal bar used in Punjabi and Hindi. Additionally, English uses 26 symbols to represent a larger number of consonant and vowel sounds. Hence the relationship between language sounds and individual letters in English is more complex than similar relationships in Punjabi and Hindi. As such, Punjabi and Hindi have been considered relatively shallow orthographic systems (Bhatia, 1994; Daniels & Bright, 1996; Pandey, 2007), where the akshara to sound correspondence is more consistent than for English, which is a relatively opaque orthography (Pasquarella, Chen, Gottardo, & Geva, 2015; Share, 2008).

Symbols in all three languages can be used to represent basic phonemes in the respective languages, but the basic unit of orthography within English is less likely to be a syllable, in contrast to Punjabi and Hindi. As discussed above, generally, the smallest unit taught in Punjabi and Hindi is a combination of at least one consonant and one vowel sound. Similarly, the use of diacritic marks in English is rare, unlike in Punjabi and Hindi. In particular, the dependent forms of vowels lead to a large number of diacritic marks been used in Punjabi and Hindi. The difference in use of diacritic marks for vowel sounds is a further indication of the importance placed on the syllable form, along with the use of the schwa /ə/ which is not included in the written form of Punjabi and Hindi. Usually, the pattern of consonants and vowels predict the occurrence of schwa /ə/, but generally, initial consonants are always followed by a schwa /ə/ if no other vowel occurs, and schwa /ə/ never occurs at the end of the word. The features of schwa and the dependent forms of vowels in both Punjabi and Hindi are clear variations from English. However, the use of vowel markers also differentiates Punjabi and Hindi to some extent: in Hindi, the independent form is employed for a vowel that does not immediately follow a consonant or consonant cluster, as in word-initial position, or when the second of a sequence of vowels, whereas in Punjabi, when a vowel is not preceded by a consonant, it is written with one of the three vowel carriers. Other markers are used in Punjabi and Hindi, and can be used to contrast these two orthographies from English: for example, the use of gemination marks and dots to represent sounds within loan words.

The Current Research

The present study investigatedpotential underlying cognitive-linguistic predictors of Punjabi reading (decoding and comprehension), and compared these withpredictors of Hindi and English as additional languages. The study assessed children in grades 2–5 who were studying in primary schools following the multilingual education policy discussed previously in this chapter. Measures were developed across the three langauges to determine if predictors in Punjabi were similar to those for Hindi and English (see Gautam, 2017, for detail on test development and procedures).

Participants

Children were selected from two primary schools who agreed to take part in the work (several schools known to the researchers were contacted and the two that agreed were included in the study). The two schools were private schools affiliated with the Punjab School Education Board (PSEB). They followed the same curriculum and used the same textbooks. Two to three classes/sections from each grade (i.e., from grades 2 to 5) were tested, with approximately 100 children (about half the children were boys and half were girls) per grade included in the sample. All children available on the days of testing were assessed given parental/child/teacher informed consent – though data from 43 children were excluded mainly because of missing one test day. Schools were selected to ensure that Punjabi was the first language for all participants, and that all children started to learn Punjabi, Hindi and English languages from their initial stage of formal (i.e., school-based) literacy learning.

Measures

Measures comprised those that assessed reading comprehension, pseudo-word (non-word) decoding, listening comprehension, phonological awareness, orthographic knowledge and speed of processing. Reading comprehension, pseudo-word reading and listening comprehension were included as measures of the basic features of the SVR: listening comprehension should include most processes considered as part of linguistic comprehension and pseudo-word decoding will provide an indication of basic processes in accessing novel words – a key feature of word reading acquisition. Given the focus of the current research on relatively young trilingual learners (children from year 2 of school), additional measures of early reading-related skills were also included. These comprised measures of orthographic and phonolgical processing that should support decoding, but may also be vital in explaining variation in early reading comprehension via word decoding. Both of these areas of processing have been found to show evidence of cross-orthography relationships (see discussions in Bérubé & Marinova-Todd, 2012; Melby-Lervåg & Lervåg, 2011; Sadeghi, Everatt, McNeill, & Rezaei, 2014). Word processing skills have also been found to be related to speed of processing, particularly when considering more transparent orthographies (Georgiou et al., 2012; Sadeghi & Everatt, 2015). Given that some argue for Punjabi and Hindi to be relatively regular in their akshara-sound relationships, then measures of rapid naming were also included to determine if they explained additional variance in reading over that explained by the phonolgoical or orthographic measures.

All measures were derived based on the literature related to predictors of reading, and all measures were peer reviewed by ten primary teachers from Punjab to ensure the appropriateness of material for the primary children (from grade 2 to 5) targeted in the research. The whole assessment battery was piloted and where necessary measures were amended prior to their administration within the current study. Piloting was performed on trilingual children (N = 40) in Punjab, India, in grades 2–5, who were similar in terms of background to those in the main study but who were independently sampled and tested.

Reading Comprehension

For each language, the test comprised four passages and 26 multiple choice questions. The four passages were designed to increase in length and grade level. The test was based on the Neale Analyses of Reading Ability (NARA; Neale, 1999) with the exception that participants were required to read passages silently (rather than read aloud as in the NARA) so as to allow the test to be administered to groups. Punjabi and Hindi materials were derived from textbooks used by the schools, and with reference to teacher-developed tests used in the schools. After each passage, the participants answered a series of multiple-choice questions about the passage. Questions were developed to focus on details from, or inferences about, a passage. Each question was followed by one correct response and three distracters, and the children were required to choose one of these responses. Passages were available while answering the comprehension questions and there was no requirement to read quickly. The maximum score for each language-version of the test was 26. The below provides an example of the sort of text and questions used in the English measure.

Text:
A black cat came to my house.
She put her kitten by the door.
Then she went away.
Now I have her baby for a pet.
Question:
What colour was the cat?
a. brown b. yellow c. white d. black

Decoding (Pseudo-Word Reading)

Pseudo-word reading was used to assess the ability to read/decode novel words. Pseudo-words were pronounceable akshara/letter strings (based on grapheme-phoneme correspondence rules of relevant orthography), but had no meaning in the given language (an English example would be 'unyen'). Thirty pseudo-words were developed for each of the Punjabi, Hindi and English language measures. For the English version, items were based onthe Woodcock-Johnson word-attack measure (Woodcock, McGrew, & Mather, 2001). For the Hindi version, items were based on words that were taken from elementary textbooks used in the two participant school and measures used in Gupta (2004) and Gupta and Jamal (2007) for bilingual children. For the Punjabi version, the elementary school textbooks were again consulted and the patterns of non-words derived based on the Hindi version of the measure. All items were arranged in terms of number of syllables per item: since one syllable words are rare in Hindi and Punjabi (Koul, 2009 – and see background to this chapter), the pseudo-words used comprised two syllables or more. All 30 items were presented on an A4 size paper, printed in bold font. A stopwatch was used to record the time spent by each participant on this task. Participants were told that they would be given some made-up words and that they should try to pronounce them correctly for the examiner. The total correctly read out of 30 was considered the child's score of each of the language versions.

Listening Comprehension

To assess the oral language skills of participants, three listening comprehension measures (one for each language) were developed. The English language version was based on the Clinical Evaluation of Language Fundamentals (CELF4; Semel, Wiig, & Secord, 2006), and Punjabi and Hindi tests followed the same format with materials developed from school textbooks and with reference to teacher-developed tests of language skills. In each case, the development procedure was similar to that used in the passage reading comprehension measures outlined above. Measures comprised six passages and a total of 39 yes/no (referential and inferential) comprehension questions. Passages were designed to increase in length and difficulty level over the course of the test. Each passage was articulated by a tester in the accent familiar to participants. Children were required to listen to each passage carefully, followed by the verbally presented questions about the passage. The children then checked 'yes/no' on their answer sheets to indicate their answer to the verbal questions. Participants did not see the content of the passages, but every passage was given a sequence number, and a title and coloured picture related to title was provided on the answer sheet to avoid confusion between passages. Answer sheets were collected after the test and the number of correct responses out of 39 was used as score. The below provides an example of the sort of text and questions used in the English measure

Passage:
Ram made a kite. Sham wanted to make a kite, too. So, Ram showed Sham how to make a kite. When they were done, they went up the hill to fly their kites.
Question:
Did Sham make the kite? YES/NO

Phonological Awareness

To measure phonological awareness, measures of sound deletion and sound substitution were developed in Punjabi, Hindi and English. English measures were developed based on those within the literature (see Sadeghi et al., 2014). Consistent with this previous work, the sound deletion task assessed the ability to delete a phoneme within the initial, middle or final position (5 items each) of a word. The task consisted of 15 items and varied in level of difficulty by increasing the number of phonemes per word. The assessor verbally presented the word and sound to the student and was asked to say the word without the sound (e.g., cup without the /k/ sound). The number of correct responses (out of 15) was used as the final score for this measure.

The sound substitution measure assessed the ability to substitute a sound within a word with another sound (e.g., in cap, replacing the /k/ sound with a /t/ sound). Again 15 items were developed that varied in their level of difficulty by increasing the number of phonemes within the word. Substitutions were positioned in the initial, medial or final part of the word (5 trials each). The number of correct responses out of 15 was used as the final score for this measure.

In the Punjabi and Hindi tasks, phonological units to be deleted or substituted were a combination of a consonant and a schwa (normally represented as Cə, with ‘C’ referring to the any consonant and the ‘ə’ syllable representing the schwa), or a vowel in their independent form, or a vowel diacritic (i.e., the vowel dependent form), or a consonant. Examples of deletion tasks are given below for these four types of phonological units.

(a) Combination of a consonant and a schwa; for example, deleting /mə/ from the word ਕਮਲ /kəməl/ (meaning ‘lotus’ in English) would leave /kəl/ (meaning ‘tomorrow/yesterday’).
(b) Vowels in their independent form; for example, deleting /ɪ/ from the word ਇਮਲੀ /ɪməli/ (meaning ‘tamarind’ in English) would leave /məli/ (meaning ‘to apply’).
(c) Vowel diacritics; for example, deleting /i/ from the word ਸੋਚੀ /sotʃi/ (meaning ‘planned’ in English) would leave /sotʃ/ (meaning ‘thinking’).
(d) Consonants; for example, deleting phoneme /m/ from the word ਕਸਮ /kəsəm/ (meaning ‘oath’ in English) would leave /kəs/ (meaning ‘tight’).

Orthographic Knowledge

Children’s knowledge of orthography was assessed via two matching tasks and two word-chain tasks. One task involved matching words (e.g., man-mat versus kite-kite, where the latter match but the former differ by one letter). A second comprised non-word matching (e.g., boal-boal versus floxy-ploxy, where the former pair match but the latter differ by one letter). A third task comprised random words with spaces between those words removed (e.g., booksitoamsomego), and required the child to mark the spaces with a line. The fourth task involved meaningful sentences with the spaces between words removed (e.g., I will read some stories), and required the participant to indicate the missing spaces. Tests and materials were based on similar measures in the literature (Gupta & Jamal, 2007; Sadeghi, Everatt, & McNeill, 2016), with the school textbooks again been used to ensure that items were appropriate for the children included in the study. Measures were developed in the three languages: Punjabi, Hindi and English. Tasks were timed and required the participant to complete as many items as they could within one minute.

For the word-matching task, the children were presented with 50 pairs of words: 25 were the same and 25 differed by either one letter for English or one character for Punjabi and Hindi. For the Punjabi and Hindi different pairs, the words were selected based on the four phonological units (explained earlier):a combination of a consonant and a schwa, as in ਤਨ /tən/ (which means ‘body’) and ਮਨ /mən/ (which means ‘mind’), producing ਤਨ-ਤਨ versus ਮਨ-ਤਨ; the use of vowels in their independent form, as in ਇਮਲੀ /ɪməli/ (which means ‘tamarind’) and ਅਮਲੀ /əməli/ (which means ‘addicted’), producing ਇਮਲੀ-ਇਮਲੀ versus ਇਮਲੀ-ਅਮਲੀ; the use of vowel diacritics, as in ਪਿਤਾ /pɪta/ (which means ‘father’) and ਪੀਤਾ /pita/ (which means ‘drink’), producing ਪੀਤਾ-ਪੀਤਾ versus ਪਿਤਾ-ਪੀਤਾ; and consonants, as in ਕਮਲ /kəməl/ (which means ‘lotus’) and ਕਮਰ /kəmər/ (which means ‘waist’), producing ਕਮਲ-ਕਮਰ versus ਕਮਰ-ਕਮਰ. The participants were required to underline the pairs that matched and not those that did not match. The same procedure was used in the non-word matching task, except that 50 pairs (25 identical and 25 one character different) of pronounceable, but meaningless letter/akshara strings were the items. For English, the different pairs differed by one letter, whereas for Punjabi and Hindi they differed by the same units as for the word items.

For the two word chain measures, the children’s task was to recognize word boundaries. In one measure, ten trials of random words were used with the spaces between those words removed. In the second measure, the string of words from which boundaries were removed made up a meaningful sentence: again, ten trials were used. The child was required to distinguish each word in the sentence by drawing a vertical line at the end of each word and before the next. The ten trials comprised a total of 55 words for each language-version of the measure; hence, 55 was the maximum score for each measure.

Speed of Verbal Processing

To assess participants’ ability to process verbal information accurately and quickly from their lexicon, measures of Rapid Automatic Naming of Objects were developed in Punjabi, Hindi and English, based on similar measures in the literature (Denckla & Rudel, 1976; Sadeghi et al., 2016). The task measured the speed with which the student named 50 line-drawings of familiar objects. Ten different objects were used, each repeated five times. Prior to administration, familiarity of the objects was checked. Naming errors were few, so the time taken by each participant was used as the score for this measure.

Procedures

The whole assessment battery was administrated in two sessions: one group and one individual session. Group testing was performed in the normal classrooms by three trained research assistants: two had experience of teaching young children and one had experience of collecting data for research projects. All research assistants were trained prior to testing. Classrooms were arranged so that children could not see each other’s work and talking was not allowed during tests. Individual sessions were conducted by the first author in a quiet room, free from distractions, within the children’s schools. Sessions included short breaks to avoid fatigue and were performed over several days to allow different languages to be the focus of assessment on different days to reduce confusion. Practice trials were given before each test to ensure understanding of the requirement of task.

Results

Full details of results can be found in Gautam (2017). For present discussion, the results of regression analyses investigating predictors of decoding and reading comprehension in Punjabi are the focus, and these are contrasted with similar analyses for Hindi and English measures. In each of the hierarchical regression analyses, either reading comprehension or pseudo-word decoding was used as the dependent variable. Control variables of age, gender and grade were entered first, followed by predictor variables in a set order – the set order was based on previous work (e.g., Sadeghi et al., 2016) and a Simple View of Reading perspective (Gough, & Tunmer, 1986; Hoover, & Gough, 1990).

For decoding, the phonological measures (sound deletion and sound substitution) were entered followed by the orthographic measures (word matching, non-word matching, random word chain and sentence word chain) and the speeded/rapid naming measure, with listening comprehension added as the final step. Listening comprehension was used as the last step to investigate possible general language influences on word processes (e.g., oral language vocabulary may support written word knowledge – see Tunmer & Chapman, 2012). Phonological and orthographic measures were entered in the reverse order to that presented in this chapter in additional analyses, but the current focus was to determine any additional variability explained by orthography over phonology (as in Sadeghi et al., 2016). Naming was always entered after the phonological and orthographic measures in order to assess additional influences of speed over language and writing processes (particularly from an extended simple view of reading perspective – see Joshi & Aaron, 2000 – and perspectives related to the relative transparency of Punjabi and Hindi).

For reading comprehension, listening comprehension was entered first followed by the decoding measure – these were to assess the basic perspective of the Simple View of Reading. Next, phonological measures (sound deletion and sound substitution) were entered followed by the orthographic measures (word and non-word matching, word and sentence chain), and finally rapid naming: this order was used for consistency with the analyses using decoding as the dependent variable.

The results of these regression analyses are presented in Tables 2 (Punjabi), 3 (Hindi) and 4 (English). Additional analyses, not reported here due to space limitations, were also performed; these varied the entry steps of measures (e.g., orthographic measures preceded phonological measures in additional analyses) and contrasted younger (grades 2 and 3) and older (grade 4 and 5) cohorts of students. These additional analyses are only mentioned when pertinent to the discussion, but can be found in Gautam (2017).

The results for Punjabi were generally consistent with a Simple View of Reading. Reading comprehension in Punjabi was predicted by measures of decoding and language understanding, with the former being predicted by measures of phonological and orthographic skills. However, additional variability in decoding was also predicted by language skills (particularly for the younger, grade 2 and 3, cohorts of students), which suggests an influence of wider oral language processes on

Table 2 Regression analyses investigating the predictors of Punjabi

	R^2	R^2 change	Significance	Final beta
Decoding				
Control – grade, gender and age	.419	.419	F = 94.7, p < .001	
Phonological awareness	.547	.127	F = 55.0, p < .001	Deletion = .310*
				Substitution = .216*
Orthographic knowledge	.578	.031	F = 7.2, p < .001	Word matching = .013
				Non-word matching = .010
				Random word chains = .194*
				Sentence word chains = −.051
Speed of Processing	.572	.005	F = 4.32, p = .04	Rapid naming objects = −.078
Listening comprehension	.581	.010	F = 9.6, p < .01	Listening comprehension = .130*
Comprehension				
Control – grade, gender and age	.475	.475	F = 119, p < .001	
Listening comprehension	.492	.017	F = 12.8, p < .001	Listening comprehension = .078
Decoding	.542	.050	F=42.7, p<.001	Pseudo-word reading = .244*
Phonological awareness	.543	.001	F < 1, p = .54	Deletion = −.039
				Substitution = .048
Orthographic knowledge	.581	.038	F = 8.7, p < .001	Word matching = .120*
				Non-word matching = .112*
				Random word chains = .092
				Sentence word chains = .058
Speed of Processing	.582	. 001	F = 1.11, p = .29	Rapid naming objects = −.041

*p < .01

word-level literacy skills (Tunmer & Chapman, 2012); and there was some influence of rapid naming on decoding, though this was only marginally significant. Furthermore, there was evidence for measures of orthographic knowledge to predict reading comprehension independently of decoding skills. The latter finding is consistent with studies of Persian speaking cohorts in Sadeghi et al. (2016). (The results for Punjabi are modelled in Fig. 1.)

The results for Hindi were similar to those for Punjabi. Again, Hindi reading comprehension was predicted by measures of language understanding and decoding, with the latter being associated with phonological and orthographic skills. Variability in decoding was also predicted by naming speed, and there was a marginal influence of language skills (represented by a dotted line in Fig. 2). As with Punjabi, there was evidence for measures of orthographic knowledge to predict reading comprehension independent of decoding levels (again consistent with Sadeghi et al., 2016, for Persian). There was also a marginal influence of phonological awareness in Hindi influencing Hindi reading comprehension. Although marginal (the effect was significant only if phonological measures were entered before orthographic measures and the individual beta values for deletion and

Table 3 Regression analyses investigating the predictors of Hindi

	R^2	R^2 change	Significance	Final beta
Decoding				
Control – grade, gender and age	.310	.310	F = 58.8, p < .001	
Phonological awareness	.513	.203	F = 81.3, p < .001	Deletion = .207*
				Substitution = .318*
Orthographic knowledge	.560	.047	F = 10.4, p < .001	Word matching = .161*
				Non-word matching = −.033
				Random word chains = .130*
				Sentence word chains = .053
Speed of Processing	.569	.009	F = 7.9, p < .01	Rapid naming objects = −.113*
Listening comprehension	.574	.005	F = 4.5, p = .04	Listening comprehension = .092
Comprehension				
Control – grade, gender and age	.421	.421	F = 95.2, p < .001	
Listening comprehension	.464	.043	F = 31.2, p < .001	Listening comprehension = .134*
Decoding	.578	.115	F = 106, p < .001	Pseudo-word reading = .244*
Phonological awareness	.592	.014	F = 6.5, p < .01	Deletion = .062
				Substitution = .089
Orthographic knowledge	.625	.041	F = 10.7, p < .001	Word matching = −.006
				Non-word matching = .112*
				Random word chains = .111*
				Sentence word chains = .141*
Speed of Processing	.635	.002	F = 2.06, p = .15	Rapid naming objects = −.056

*p < .01

substitution were non-significant), this may also be indicative of exposure to the oral language supporting comprehension levels.

For English reading, again both listening comprehension and decoding measures were predictors of reading comprehension, and decoding was predicted by phonological awareness and orthographic knowledge. Orthographic processing also was an independent predictor of variability in reading comprehension. Additionally, there was a larger influence of rapid object naming (speed of processing) on measures of English literacy compared to the other two languages, with speeded naming showing independent levels of prediction of both decoding and reading comprehension, potentially consistent with extended simple model perspectives (see discussions in Joshi & Aaron, 2000; and also Joshi, Tao, Aaron, & Quiroz, 2012). Finally, there was also a marginal influence of phonological awareness in English influencing English reading comprehension. Consistent with Hindi, the effect was significant only if phonological measures were entered before orthographic measures (and the individual beta values for deletion and substitution were non-significant), but this again may be indicative of exposure to the oral language supporting comprehension levels. (The results for English are modelled in Fig. 3.)

Table 4 Regression analyses investigating the predictors English

	R^2	R^2 change	Significance	Final beta
Decoding				
Control – grade, gender and age	.303	.303	F = 56.9, p < .001	
Phonological awareness	.535	.232	F = 97.7, p < .001	Deletion = .162*
				Substitution = .279*
Orthographic knowledge	.637	.102	F = 27.0, p < .001	Word matching = .074
				Non-word matching = −.051
				Random word chains = .164*
				Sentence word chains = .242*
Speed of Processing	.648	.011	F = 11.8, p < .001	Rapid naming objects = −.127*
Listening comprehension	.648	.000	F < 1, p = .857	Listening comprehension = .007
Comprehension				
Control – grade, gender and age	.434	.434	F = 100, p < .001	
Listening comprehension	.465	.032	F = 23.3, p < .001	Listening comprehension = .108*
Decoding	.556	.091	F = 79.9, p < .001	Pseudo-word reading = .136*
Phonological awareness	.563	.007	F = 3.0, p = .05	Deletion = .049
				Substitution = −.049
Orthographic knowledge	.609	.046	F = 11.3, p < .001	Word matching = .073
				Non-word matching = .057
				Random word chains = .148*
				Sentence word chains = .068
Speed of Processing	.631	.023	F = 11.8, p < .001	Rapid naming objects = −.188*

*p < .01

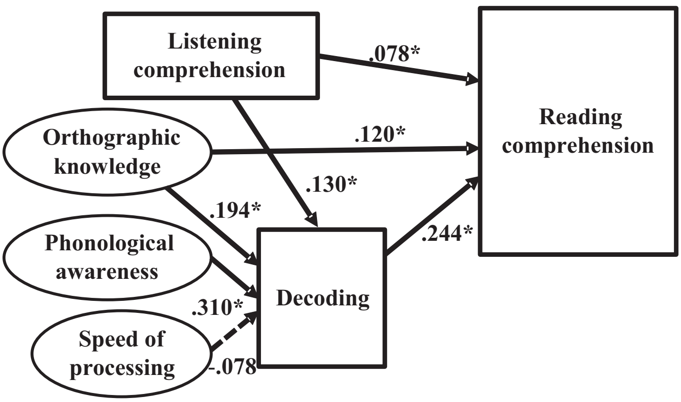


Fig. 1 Punjabi reading model

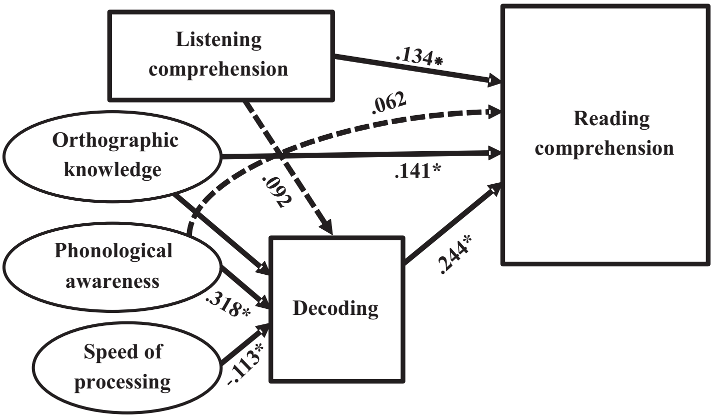


Fig. 2 Hindi reading model

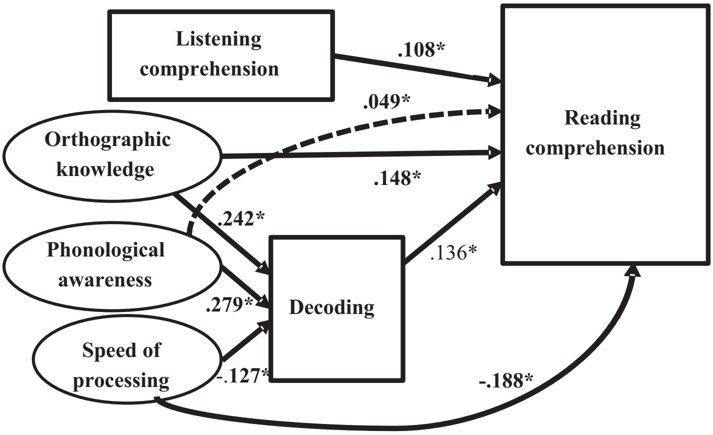


Fig. 3 English reading model

Discussion

Overall, the results argue for reading skills in Punjabi to be consistent with a simple (perhaps universal) model of reading in which both linguistic comprehension and word decoding are basic component processes in reading comprehension. Basic decoding processes are then supported by processes related to phonological awareness, orthographic processing and speeded naming. The current findings argue for this framework to be appropriate for describing Punjabi despite its akshara orthographic features, which means that in its representation of the language (at the level of syllable and phoneme) it varies somewhat from English (the language on which most models have been developed). Additionally, although the Punjabi orthography has been considered relatively transparent, its features (potentially associated with akshara, but also potentially the varying representations of vowels, the use of diacritic marks and the use of loan sounds) may make it more appropriate to consider it alongside other more mixed orthographies in terms of the underlying influences on reading acquisition. The finding that orthographic knowledge directly predicts Punjabi reading comprehension from an early grade, independently of the influence of decoding processes, argues for the orthography to be considered alongside orthographies that do not show a simple phonological decoding influence on reading acquisition.

This influence of orthographic processing on both decoding and comprehension has been found in other orthographies: see discussions in Sadeghi et al. (2016), on Persian, and Elbeheri, Everatt, Mahfoudhi, Al-Diyar, and Taibah (2011), on Arabic. In the case of Arabic, the orthographic influence has been discussed in terms of the complexity of the orthography. This may also be the case for Persian and Punjabi, both of which show influences from Arabic in their history. One possible factor that may associate Persian and Punjabi in terms of their potential complexity, is the use of diacritics to distinguish basic character forms and additional marks to represent sounds within loan words. This complexity may make orthographic knowledge even more important for learners. However, the fact that both Hindi and English also show similar influences of orthographic processing on both decoding and comprehension may be more consistent with views that suggest that orthographic knowledge is more important than some models may suggest (see Elbeheri et al., 2011; Torppa, Poikkeus, Laakso, Eklund, & Lyytinen, 2006). Further research comparing these different orthographies across different contexts of learning would be useful. The context of learner here is important since akshara orthographies can be taught as if syllabaries, and, hence, orthographic complexity may be represented in the number of syllables, or it may be taught from the perspective of individual phonemes, in which case complexity may be better seen in the variations in forms (e.g., dependent versus independent vowels) to represent the same sound.

The commonalities in predictors of Hindi and English, and their similarities with the predictors to the first language suggest an alternative explanation, though. These findings may be consistent with some level of inter-relationship between first and additional languages in terms of the underlying skills developed to support acquisition. Although controversial, cross-language influences indicative of transfer effects between languages have been found for phonological processing, morphological awareness and skills associated with linguistic comprehension (see discussions in Bialystok, McBride-Chang, & Luk, 2005; Bérubé & Marinova-Todd, 2012; Melby-Lervåg & Lervåg, 2011; Sadeghi et al., 2014; Sparks, Patton, Ganschow, & Humbach, 2012). Hence, it may be that the similarities in predictors apparent in the current data are more a factor of testing multilingual children learning to read and write in different orthographies at the same time. These similarities, therefore, may reduce if comparing children learning to reading and write only in an akshara orthography with those learning an alphabetic orthography.

Clearly, the three models developed from the data of the current study show a great deal of similarity. However, the English model showed larger influences of processes in speeded naming than the two akshara orthographies. This may be a more specific feature of English, and may relate to the need to access linguistic knowledge about words fluently to support decoding processes. Such fluency in linguistic access may be less of a feature of Punjabi and Hindi, where orthographic features and rules may provide the basis on which to decode words. However, the fact that both additional languages also show some aspects of phonological awareness supporting reading beyond the level of word decoding also argues for an influence of language experience. It may be that both influences (phonological and rapid naming) are related to exposure to the language and orthography among the children tested. Those with more exposure may be less influenced by basic processes in their understanding of text. Again, further research assessing the impact of levels of proficiency in the languages used by the children would be useful to distinguish potential alternative explanations (see also discussions in Sadeghi & Everatt, 2015).

Conclusion

The current work focussed on predictors of reading in the two akshara orthographies (Punjabi and Hindi) and the one alphabetic orthography (English). For all, reading comprehension was predicted by listening comprehension and decoding, consistent with a Simple View of Reading. However, orthographic knowledge was an independent predictor of reading comprehension indicating an important influence of this area of processing, particularly for orthographies that may be considered more complex. Commonalities between predictors, though, also argue for cross-language influences of basic processes within multilingual children, which may need to be considered in future studies within an Indian-linguistic context. Overall, the findings argue that the relative transparency, which has been associated with orthographies such as Punjabi, needs to be considered in the light of other challenges to reading acquisition, such as orthographic complexity; and it might be better to consider Punjabi as more akin to other orthographies (such as Persian) that do not show all the features of transparency that might be expected based on the correspondence between written symbols and language sounds.

References

Agnihotri, R. K. (2013). *Hindi: An essential grammar*. Abingdon, UK: Routledge.

Bérubé, D., & Marinova-Todd, S. H. (2012). The development of language and reading skills in the second and third languages of multilingual children in French Immersion. *International Journal of Multilingualism, 9*(3), 272–293.

Bhatia, T. K. (1994). *Punjabi: A cognitive-descriptive grammar*. Abingdon, UK: Routledge.

Bhatia, T. K. (2015). *Colloquial Hindi: The complete course for beginners* (2nd ed.). Abingdon, UK: Routledge.

Bialystok, E., McBride-Chang, C., & Luk, G. (2005). Bilingualism, language proficiency, and learning to read in two writing systems. *Journal of Educational Psychology, 97*(4), 580–590.

Bowden, A. L. (2012). *Punjabi Tonemics and the Gurmukhi script: A preliminary study.* MA thesis, Brigham Young University, USA.

Campbell, G.L. (1995). *Concise compendium of the world's languages*. Abingdon, UK: Routledge

Daniels, P. T., & Bright, W. (1996). *The world's writing systems*. New York: Oxford University Press.

Denckla, M. B., & Rudel, R. G. (1976). Rapid automatized naming (RAN): Dyslexia differentiated from other learning-disabilities. *Neuropsychologia, 14*(4), 471–479.

Elbeheri, G., Everatt, J., Mahfoudhi, A., Al-Diyar, M. A., & Taibah, N. (2011). Orthographic processing and reading comprehension among Arabic speaking mainstream and LD children. *Dyslexia, 17*(3), 123–142.

Gautam, S. (2017). *Predictors of Punjabi, Hindi and English reading comprehension among multilingual children in the Punjab region of India*. PhD thesis. University of Canterbury, New Zealand.

Georgiou, G. K., Papadopoulos, T. C., Fella, A., & Parrila, R. (2012). Rapid naming speed components and reading development in a consistent orthography. *Journal of Experimental Child Psychology, 112(1), 1–17*.

Gill, H. S., & Gleason, H. A. (1969). *A reference grammar of Punjabi*. Patiala, India: Department of Linguistics, Punjabi University.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6–10.

Gupta, A. (2004). Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 17*(1–2), 79–99.

Gupta, A., & Jamal, G. (2007). Reading strategies of bilingual normally progressing and dyslexic readers in Hindi and English. *Applied Psycholinguistics, 28*(1), 47–68.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*(2), 127–160.

Hornberger, N., & Vaish, V. (2009). Multilingual language policy and school linguistic practice: globalization and English language teaching in India, Singapore and South Africa. *Compare: A Journal of Comparative and International Education, 39*(3), 305–320.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology, 21*(2), 85–97.

Joshi, R. M., Tao, S., Aaron, P. G., & Quiroz, B. (2012). Cognitive component of componential model of reading applied to different orthographies. *Journal of Learning Disabilities, 45*(5), 480–486.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning 57*, 1–44.

Koul, O. N. (2005). *Language, education and communication*. Delhi, India: Indian Institute of Language Studies.

Koul, O. N. (2009). *Modern Hindi grammar*. Delhi, India: Indian Institute of Language Studies.

Malik, A. N. (1995). *The phonology and morphology of Panjabi*. New Delhi, India: Munshiram Manoharlal Publishers.

Mann, G. S. (2011). *An introduction to Punjabi: Grammar, conversation and literature*. Patiala, India: Punjabi University.

Meganathan, R. (2011). Language policy in education and the role of English in India: From library language to language of empowerment. In H. Coleman (Ed.), *Dreams and realities: Developing countries and the English language*. London: British Council.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading, 34*(1), 114–135.

Mohanty, A. K. (2010). Languages, inequality and marginalization: Implications of the double divide in Indian multilingualism. *International Journal of the Sociology of Language, 205*, 131–154.

Neale, M. D. (1999). *Neale analysis of reading ability* (3rd ed.). Camberwell, Australia: Australian Council for Educational Research.

Nigam, R. C., & Sen, S. (1975). *Grammatical sketches of Indian languages with comparative vocabulary and texts*. Delhi, India: National Government Publications.

Pandey, P. (2007). Phonology-orthography interface in Devanāgarī for Hindi. *Written Language and Literacy, 10*(2), 139–156.

Pasquarella, A., Chen, X., Gottardo, A., & Geva, E. (2015). Cross-language transfer of word reading accuracy and word reading fluency in Spanish-English and Chinese-English bilinguals: Script-universal and script-specific processes. *Journal of Educational Psychology, 107*(1), 96–110.

Sadeghi, A., & Everatt, J. (2015). Influence of language background on English reading comprehension skills: Cross-language transfer effects. In L. T. Wong & A. Dubey-Jhaveri (Eds.), *English language education in a global world: Practices, issues and challenges* (pp. 69–80). New York: Nova Science Publishers.

Sadeghi, A., Everatt, J., & McNeill, B. (2016). A simple model of Persian reading. *Writing Systems Research, 8*(1), 44–63.

Sadeghi, A., Everatt, J., McNeill, B., & Rezaei, A. (2014). Text processing in English-Persian bilingual children: A bilingual view on the Simple Model of Reading. *Educational and Child Psychology, 31*(2), 45–56.

Semel, E., Wiig, E. H., & Secord, W. A. (2006). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: The Psychological Corporation.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an 'outlier' orthography. *Psychological Bulletin, 134*(4), 584–615.

Singh, P. (2010). *Sidhantik Bhasha Vigyan*. Patila, India: Madaan Publications.

Singh, P., & Lehal, G. S. (2010). Corpus based statistical analysis of Punjabi syllables for preparation of Punjabi speech database. *International Journal of Intelligent Computing Research, 1*(3), 2042–4655.

Sparks, R. L., Patton, J., Ganschow, L., & Humbach, N. (2012). Do L1 reading achievement and L1 print exposure contribute to the prediction of L2 proficiency? *Language Learning, 62*(2), 473–505.

Torppa, M., Poikkeus, A., Laakso, M., Eklund, K., & Lyytinen, H. (2006). Predicting delayed letter knowledge development and its relation to Grade 1 reading achievement among children with and without familial risk for dyslexia. *Developmental Psychology, 42*(6), 1128–1142.

Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*(5), 453–466.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*(1–3), 679–690.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

Spelling and Reading Words in Bengali: The Role of Distributed Phonology

Shruti Sircar and Sonali Nag

Abstract Bengali orthography is only moderately transparent, and there are several striking instances of phonology-orthography and orthography-phonology inconsistency. In this chapter, we show that an index of advancing skills in Bengali word reading and spelling is knowledge of word level phonology. In particular, we show that in the early grades, an important task is to learn to read and spell words where phonological information is distributed across the word, neutralized or elided. While phonological neutralizations impact spelling more than reading, some forms of phonology-akshara mapping and non-linear arrangements in complex akshara impact both reading and spelling.

Keywords Bengali · Spelling · Reading · Consonant clusters · Distributed word level phonology · Inherent vowel · Phonology-akshara mapping · Phonological neutralizations · Orthographic representation

Introduction

An important contributive factor in spelling and reading acquisition is the relative consistency of mapping between phonology and orthography. In transparent orthographies such as Finnish and Turkish, children can spell and read most words once they know the basic symbol set and the sounds they map to. In contrast, the transition into spelling and reading does not run as much in parallel in the opaque orthographies, with cross-linguitic evidence building up from the 1990s to show that

S. Sircar (✉)
Department of Linguistics & Contemporary English, School of Language Sciences, English and Foreign Languages University, Hyderabad, Telangana, India
e-mail: shrutisircar@efluniversity.ac.in

S. Nag (✉)
Department of Education, University of Oxford, Oxford, UK

The Promise Foundation, Bangalore, India
e-mail: sonalinag@t-p-f.org; sonali.nag@education.ox.ac.uk

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_9

spelling development lags behind reading development (e.g. French: Geva, Wade-Woolley, & Shany, 1993; Portuguese: Pinheiro, 1995; Persian: Sprenger-Charolles, Siegel, & Bonnett, 1998; Kswahili: Alcock & Ngorosho, 2003; and Hebrew: Rahbari, Senechal, & Arab-Moghaddam, 2007; Ravid, 2011). This dissociation in development is because spelling, unlike reading, adds a new challenge with a phonological recoding process where speech sounds need to be mapped to symbols and each symbol has to be recalled anew. In the akshara orthographies not enough is known about whether spelling and reading development are in step or uneven. Single study reports suggest that spelling knowledge is slower to develop compared to reading, and the gap may be determined by the extent of transparency in an individual akshara-based language (e.g. Hindi: Vaid & Gupta, 2002; Kannada: Nag, 2014; Thai: Winskel & Iemwanthong, 2010). The pressure points appear to emerge because of the extensive symbol repertoires of these orthographies and the language-specific instances of inconsistency in phonology-akshara mapping. There is one further potential pressure point linked to the phonological information distributed across the word that may influence the phonological realization of a word. Although there are suggestions in the literature that such word level details matter for decoding (e.g. role of syllable weight: Pandey, 2014) their role on literacy acquisition is yet to be studied.

We turn to Bengali to study the relative patterns in spelling and reading focusing on inconsistencies and word level phonology. In Bengali, the mapping of orthography and phonology may be inconsistent and this inconsistency may occur either uni- or bi-directionally. Word level phonology may also contribute to rules of orthographic representation that are not immediately clear. In this chapter we begin with a description of Bengali phonology describing both the consistencies as well as the inconsistencies with the orthography. We also describe the underpinnings of word level phonology. Throughout, we use the Indo-Aryan sister language of Hindi to contrast the description of phonology, orthography, and the mapping principle, with a limited attention to some of the other Indic languages. These sections are followed by our analysis of spelling and reading development in a group of Bengali children in Grades 2, 3 and 4, which highlight that in both spelling and reading one needs to go beyond akshara-phonology mapping to actively engage with word-specific phonological rules. We will show that an index of advancing skills in word reading and spelling draws upon knowledge of word level phonology.

Bengali: Among the Five Largest Speaker Communities in the World

With nearly 230 million speakers, Bengali is one of the most spoken languages ranking fifth in the world. Bengali is the national and official language of Bangladesh, and in India, is the official language of the state of West Bengal and the co-official language of the states of Assam and Tripura, an official language of Sierra Leone and is spoken in Andaman and Nicobar Islands, especially the Neil and the Havelock Islands. Bengali is spoken by a significant immigrant population in the United Kingdom, Canada, USA, Malaysia, and Singapore. The Bengali orthography is also used to write several languages of India including Oxomiya (Assamese), Daphla, Garo, Hallam, Khasi, Manipuri, Mizo, Munda, Naga, Rian, and Santali.

Bengali has an elegant orthographic system: It is, on the one hand, organized around the alphasyllabic principle, and, therefore, sufficiently regular and predictable, and, on the other hand, wrought with complex orthographic conventions. This combination of regularity and complexity is what makes the Bengali orthography interesting for comparative research on literacy acquisition. In Bengali, the phonology-to-akshara and akshara-to-phonology mapping is not always balanced. There is typically a consistent one-to-one akshara-phonology mapping, but in a small number of akshara, there is one-to-two mapping, with one symbol linked with two phonological units. In addition, there are instances of mapping where one phonological unit is represented by either more than one akshara or by just one part of the akshara symbol block. Perhaps the starkest example of inconsistency in the orthography is with consonant clusters because several have unexpected pronunciations, including instances of one consonant in the cluster becoming a geminate in some contexts and the second consonant unvoiced in other contexts (further details in the next section). Making the phonology-to-orthography linkage in Bengali, therefore, depends on not just fine-grained knowledge about specific mapping features in the orthography, but more. We have argued elsewhere for the roles of lexical, morphological, and etymological factors in furthering orthographic knowledge (Sircar & Nag, 2014). In this chapter, we argue that akshara as well as word level phonology inform the linkages of orthography-to-phonology and phonology-to-orthography, and because these linkages are less evident, they are potential pressure points for a beginning learner, and mastery of these details is central to progress in literacy skills.

Bengali Phonology and Orthography

The Bengali symbol set evolved from the *Kutila lipi*, a descendant of Brahmi, which with some modification is also used for the Oxomiya, Manipuri, and Maithili languages. The Bengali orthography with 46 basic akshara (those with inherent vowels) is an alphasyllabary or abugida, and as in Sanskrit every independent consonant carries the inherent vowel /ɔ/ or /o/ [compared to /ə/ in Hindi], and in this, the architectural principle is similar to the Devanagari and all other Brahmi derived scripts. The systematic features of the orthography are as follows: It goes from left to right but quite unlike Roman, Arabic, Hebrew, and Cyrillic, symbols hang from a top line, using upstrokes as well as down strokes. The akshara are symbol blocks where diacritics are written in all four directions to indicate non-initial vowels and some consonants.

Phonologically, Bengali allows closed syllables, and has 16 canonical syllable patterns (Sarkar, 1986), with CV syllables constituting 54% of the whole language (Dan, 1992). Consonant clusters occur in only onset positions, with the exception of loan words, where coda clusters are permitted. Bengali allows a large set of consonant clusters in word medial positions, and a cluster may contain up to three consonants, e.g. স্ত্রী *stree* /stri:/ 'woman'.[1] The example of /stri:/ is also an excellent example of how one akshara captures multiple phonemes (in this instance three consonants and one vowel) and demonstrates how in Bengali some phoneme markers are clearly distinguishable segments in the symbol block but others are opaque and fused together (more details below).

Vowels and Diphthongs

Bengali has seven vowels and all of them have a nasal counterpart, changing the meaning of the word; for instance, the pronunciation of the honorific pronoun for third person is the nasalized /tãr/, clearly distinguishing the pronoun from the third person general /tar/. This contrast between oral and nasal vowels in Standard Bengali in India (Masica, 1991: 118) is not characteristic of Bangladeshi Standard Bengali, presumably due to an Eastern dialect influence within Bangladesh (Majumdar, 1997: 108). The Indo-Aryan distinction of short and long vowels (e.g. /i, i:/ and /ʊ, u:/) is merged in Bengali, and, therefore, the vowel length distinction does not hold anymore. However, the Bengali orthography continues to retain these distinctions, introducing systematic inconsistency between the spoken and written form. Bengali also has a unique low-high vowel distinction in /æ/ and /e/ which is captured in orthography in the grapheme এ.

Vowels have two forms, a full and a diacritical form (also called the primary and secondary forms). The full form appears in the word initial position, and after a CV akshara (with an inherent vowel as in বই *bai* /boi/ 'book' or any other vowel as in ভাই *bhai* /bʰai/ 'brother'). The diacritical form of a vowel is placed vertically above, below, or to either side of the consonant, and visual elements used in some diacritics can combine to make other unique diacritics. Taking the number of visual elements in a diacritic as a symbol parameter, Bengali vowels can be classified into: (1) vowels with a single element e.g. কা /ka/, কি /ki/, কে /ke/, কী /ki:/, কু /ku/, কূ /ku:/, (2) vowels with a two-part diacritic e.g. কো /ko/, কৈ /koi/ and (3) vowels with a three-part diacritic, as in the case of the diphthong marker /kou/ কৌ.

There is no consensus on the number of diphthongs in Bengali: It is not clear whether the resultant word is monosyllabic, or disyllabic with a hiatus in between the vowels as in খাই *khai* /kʰai/ 'eats' (habitual) as found in disyllabic words like টিয়া *Tia* /tia/ 'parrot.' Diphthongs are distinguished from the pure vowels (/i/, /u/, /e/, /o/,

[1] Bengali words are transcribed first as they are represented in orthography using ITRANS, followed by the phonetic transcription in IPA, and gloss. The inherent vowel of Bengali akshara pronounced as /ɔ/ or /o/ is denoted by [a].

/æ/, /ɔ/) by the presence of a non-syllabic semivowel (Chatterji, 1926). Old written Bengali shows evidence of a vowel-semivowel-vowel sequence as in সুইয়া *suiya* /ʃuija/ 'having slept,' লইয়া *loiya* /loija/ 'having brought,' and খাইয়া *khaiya* /kʰaija/ 'having eaten.' These were distinct because of the presence of a full form of the ই /i/ followed by a semivowel য় /j/. The count of diphthongs varies from 17 (Sarkar, 1986 as cited in Kar, 2010:17) to 31 (Hai, 1975). Most diphthongs are written as a combination of a diacritical form of the first vowel (or left unmarked if the vowel is inherent) and a full form for the second vowel, except for the two distinct diphthongal akshara, ঔ /ou/ and ঐ /oi/.

In Bengali the inherent vowel with a consonant akshara may be realized as /ɔ/ or /o/, or may be suppressed: The latter phenomenon is akin to the Hindi schwa deletion (Ohala, 1983; Pandey, 1990). The choice between /ɔ/ and /o/ is largely phonological, with /ɔ/ not occurring in the first syllable unless it is followed by a high vowel and /o/ occurring before and after a consonant cluster. Schwa deletion is subject to phonological and morphological conditions: The schwa cannot be deleted if it is in the first syllable, before and after a consonant, before a vowel (which results in a heavy syllable), and before a coda consonant. There is no specific orthographic (or other) rule to indicate to a reader how the inherent vowel of an akshara may be pronounced: For example, the realization of the inherent vowel in the akshara ত in তন *tana* 'body' is /tɔ/, in তটি *taTi* 'river' it is /to/ and in the word medial নাতনি *naatani* 'granddaughter' it is /t/ with a schwa deletion. Moreover, since the schwa deletion in Bengali is not indicated in orthography, it results in homographic words. For instance, শোল *shola* can be read as /ʃol/ 'a type of fish' or /ʃolo/ 'sixteen,' which only context can disambiguate. However, even while High varieties of Bengali continue to not indicate schwa deletion in orthography, low varieties may suggest closeness to phonology by writing /sholo/ with the /o/ vowel diacritic available in the symbol repertoire, শোলো (Dasgupta, 2003).

Bengali shows vowel height assimilation (Chatterji, 1926: 395–421; Dan, 1992: 57–63) where low vowel sounds gain height because of a neighboring vowel, e.g., প্যাঁচা *pyaancha* /pæ̃tʃa/ 'male owl' becomes পেঁচি *penchi* /pẽtʃi/ 'female owl' (æ> e), and নট naTa /nɔt/ 'male play actor' becomes নটী *naTi* /noti/ 'female play actor' (ɔ > o). We can expect that access to these fine-grained word level phonological rules would be essential to avoid reading words with the inherent vowel where a schwa drop is expected. In other words, orthography must be overridden by language-specific phonological processes, and we will examine these tensions in the reading and spelling of primary schoolers later in the chapter.

Consonants and Consonant Clusters

An obstruent-heavy language, Standard Bengali has 30 consonantal phonemes, of which 23 are plosives and affricates, each with a voiced and an unvoiced variant, and an unaspirated and aspirated variant. All consonants except /ŋ, ɽ, ɖh/ occur word initially as onsets.

Phonological neutralization or mergers have led to a discrepancy between Bengali phonology and its classical ancestry. Put differently, several sets of classical consonants have merged, resulting in loss of historical phonological distinctions. The voiceless sibilants শ, ষ and স have merged to the palatal /ʃ/, typically word initially and in intervocalic singletons. The /ʃ/ and /s/ distinction continues to be realized in medial clusters (e.g. /aste/ 'softly' vs. /aʃ1e/ 'to come') and word-finally (e.g. /bas/ 'enough' vs. /ba:ʃ/ 'bamboo'). Bangladeshi Standard Bengali, however, systematically maintains the contrast word-initially as well, e.g. /sirka/ 'vinegar' vs. /ʃira/ 'syrup' (Dasgupta, 2003: 360; Masica, 1991: 98). As in Hindi, Bengali has three rhotic akshara ড়, ঢ় and র (palatal, retroflex and dental respectively) – the first two have merged, though some speakers may maintain a marginal distinction between the two in formal registers, whereas Bangladeshi Standard Bengali and many Eastern dialects have only one rhotic /ɹ/ (Dasgupta, 2003: 359; Masica, 1991: 97). The retroflex and the dental nasals ণ and ন have neutralized to the dental, though Bengali orthography continues to mark these historical distinctions. This is in contrast to Hindi where all these phonemic distinctions are retained, though in fast speech they may get neutralized, as in the case of neutralization between the retroflex and palatal sibilant (Shapiro, 2007).

Ambiguity in Bengali spelling also is because of allography in the phoneme /j/: Word initially, it is written with য (as in যুব *juba* /dʒubo/), and as য় after a vowel (e.g. পায়রা *paayaraa* /paera/ 'pigeon' and realized as /e/) or a consonant with an inherent vowel (e.g. নয় *naya* /noi/ 'nine'). The /j/ in consonant clusters, however, appears in a diacritical form (e.g. প্যাঁচা *pyaancha* /pæ̃tʃa/ 'owl').

Turning to consonant clusters, Bengali has 62 consonant clusters out of which eleven occur in the word-initial position (for e.g., /pr, gr/) and 41 word medially. Though medial consonant clusters are common and may occur in any combination, onset consonant clusters are restricted and are in a distinct combination of rising sonority: Plosive obstruent + liquid/nasals /kr, kl, gr, gl, tr, d̪r, d̪hr, nr, pr, pl, br, bhr/ and nasal + liquid /mr, ml/ (Heimisdóttir, 2013; Kar, 2010). However, a coronal obstruent /s/ also occurs as a word initial cluster /sk, sk^h, st, st^h, sp, sn, str, spr, skr/ in words borrowed from Sanskrit. All sounds except /ŋ/, /h/ and /r/ can occur as geminates.

In Brahmi scripts, when two consonants meet without an intervening vowel, they coalesce and form a conjunct akshara (a CCV akshara). To demonstrate the pervasiveness of the coalesced akshara in Bengali, we continue our comparison with Hindi. The Hindi conjuncts are formed by writing them one after the other, one below the other, one above the other or by fusing them in a non transparent way (see Mathur & Nag, 2019). The first CCV form is the most frequently used while the second and the third are somewhat less common, and the fourth is rare. In Bengali, again by contrast, all four configurations are used frequently with the opaque fourth form being used for many clusters (e.g. ষ + ণ =ষ্ণ, /sha + na = shna/; ক + ত = ক্ত, /ka+ ta = kta/). It is possible that the various ways of coalescence add to the cognitive demands of learning the CCV akshara set in Bengali. Interestingly, in an attempt to reduce the burden of such learning on young learners, effort has been made in the two main

Bengali-speaking regions of South Asia (West Bengal in India and Bangladesh) to reduce the sheer variety in akshara with informal orthographic reform of specific consonant clusters. Modern Bengali textbooks are beginning to contain more and more transparent orthographic forms of consonant clusters in which each constituent consonant of a cluster is readily apparent (e.g. Kishalaya, 2004). However, this change is limited to school textbooks and is not reflected in the rest of the Bengali printed literature. A child eventually will have to learn to recognize both the new transparent and the old opaque forms, leading to a further increase in the already extensive Bengali symbol repertoire.

Moreover, conjunct consonants in Bengali with the second consonant /v/, /j/, /m/ have phonological values different from the ones represented in orthography (see Table 1). The choice of these akshara in spelling appears arbitrary because they have no discernible phonological cues suggesting that their use must be learned by rote. Unlike Sanskrit and Hindi, which have no silent consonants, Bengali also leaves some second consonants of clusters unpronounced. This phenomenon is especially for words borrowed from Sanskrit, where the spellings reflecting the historic pronunciation are retained but are pronounced differently now. The second consonant in a cluster also impacts the syllable differently depending on the position of the cluster in the word. Word initially, they influence the following vowel by nasalizing, raising, or lengthening them. Word medially, they induce a gemination of the primary consonant, a phenomenon similar to Hindi though without the consonant drop seen in Bengali; for e.g., *vishwa* 'world' /viʃvə/ > /viʃʃvə/; *mithya* 'lie' /mitja/ > /mit^{h}t^{h}ja/ (Pandey, 2007).

Another issue that confounds spelling in Bengali is how medial consonant clusters are represented: (a) with two full consonants, e.g. করতাল *karataala* /kɔrtal/ 'cymbal,' (b) with a conjunct, e.g., রিম্‌রিম্‌ *gandha* /gɔndho/ 'scent,' or (c) with a diacritic *hasanta* marker which indicates suppression of the inherent vowel in the first consonant of the cluster, e.g., ঝিমঝিম *jhimajhima* /dʒ^{h}imdʒ^{h}im/ 'tingling'. A final instance of inconsistency related to word level phonology is common, and

Table 1 Consonant clusters and their phonological realizations in Standard Bengali

Cluster	Pronunciation	Example (with unpronounced consonant in bold)	Effect on syllable
Word initial clusters with unpronounced consonant			
স্ম [sm]	/s/	স্মৃতি *s**m**riti* /srĩti/ 'memory'	Nasalizes following vowel
স্ব [sv]	/ʃ/	স্বাস / *s**v**aasa* ʃa:ʃ/ 'breath'	Lengthens following vowel
ত্য [ty]	/t/	ত্যাগ *t**y**aaga* /tæg/ 'sacrifice'	Raises the vowel from /a/ to /æ/
ক্ষ [ks]	/k$^{h/}$	ক্ষমা *k**sh**amaa* /k^{h}ɔ:ma/ 'pardon'	Lengthens following vowel
Word medial clusters with unpronounced consonant			
ত্ব [tv]	/tt/	তত্ব *tat**v**a* /tɔtto/ 'material'	Geminates primary consonant
দ্ম [dm]	/dd/	পদ্মা *pad**m**aa* /pɔdda/ 'lotus'	Geminates primary consonant
ক্ষ [ks]	/k^{h}k^{h}/	শিক্ষা *shik**sh**aa* /ʃik^{h}k^{h}a/ 'education'	Geminates primary consonant
হ্ন [hn]	/nn/	ছিহ্ন *chi**h**no* /tʃinno/ 'sign'	Geminates secondary consonant

this is when individual consonants in a cluster are not pronounced as would be implied by individual phonemic markers in the CCV akshara. For example, despite the cluster being an orthographic representation of /ʃ/ and /l/ (শ and ল = শ্ল) it is pronounced as /sl/. Thus, in the Bengali orthography, consonant clusters may be represented by different and sometimes irregular forms. Added to this, learning to read and spell is complicated by the sheer size of the full akshara set of CV and CCV combinations, which number to more than 350 (Sircar & Nag, 2014). The original claim by Nag (2007) that children take several years to learn the whole set of akshara symbols has been replicated in several languages including Bengali (e.g. Hindi: Vaid & Gupta, 2002; Malayalam: Tiwari, Nair, & Krishnan, 2011; and Bengali: Sircar & Nag, 2014). However, apart from the item-based learning of the symbol set, the ease with which words can be read and spelled depends on the extent of transparency (Nag, 2017), and perhaps the distributed phonological information in the word.

Instruction in Bengali akshara

In Bengali, orthography instruction happens in phases: Teaching of the simple akshara (with inherent vowel) is followed by akshara with a ligatured vowel, and finally children are taught the geminates and mixed clusters (Nag & Sircar, 2008). These symbol sets are taught through copy writing and recitation of singleton akshara and akshara embedded in word context as reported in other Indic orthographies (Kannada: Nag, 2013; Malayalam: Tiwari et al., 2011). However, unlike many other akshara orthographies, in Bengali learning, children are taught to spell out each component in an akshara, essentially making literacy instruction phoneme-based. Attention is drawn to consonant components in conjunct akshara through practice in words with these akshara. For example the cluster ন্দ /nd/ is taught using writing lists belonging to specific word families like ছন্দ *chhanda* /tʃʰɔndo/'verse,' বিন্দু *bindu* /bindu/ 'dot,' বন্দ banda /bɔndo/ 'close', and নিন্দে *ninde* /ninde/ 'criticism' (for comparison see Hindi instruction in Mathur & Nag chapter and Tamil instruction in Nag & Narayanan chapter).

Unlike in other akshara teaching practice (Kannada: Nag 2014; Gujarati: Patel, 2004; Sinhala: Wijayathilake, Parrila, Inoue & Nag, 2018), most of the consistent and frequent CCV akshara in Bengali are introduced by Grade 3. These are the transparent CCVs where the two consonants are placed on a vertical line or one above the other. A limited range of opaque CCV akshara where the two phonemic markers cannot be separated is also introduced. The infrequent and rare CCV akshara (e.g. ক্ত্র /ktr/, ত্ম্য /tmj/, ন্ট্র /ntr/ ম্ভ্র /mbhr/) are not taught explicitly but may incidentally feature in texts in the later grades (similar to the pattern of text encounters reported in Kannada, Patel, Bapi, & Nag, 2013). Though the general rules of ligaturing in CCVs are taught, not all CCVs are introduced, and the onus of learning these are on the learners on the basis of reading exposure mirroring the teaching practice in the Dravidian language, Kannada (Nag, 2014). Formal training in akshara is discontinued by Grade 3.

Taken together, the previous sections show that the Bengali orthography is only moderately transparent, and there are several striking instances of phonology-orthography and orthography-phonology inconsistency. The teaching of the orthography brings attention to the phonemic markers in the symbol blocks which could help learners disambiguate some but not all the inconsistencies in mapping triggered by word level phonological processes. We turn next to a survey of children's reading and spelling to examine whether these instances of inconsistencies and word level phonological processes impact performance, and, especially, based on the literature, whether spelling is impacted more than reading. We will particularly consider whether specific insights dependent on processes other than phonological recoding can describe advances in Bengali reading and spelling development.

Language-Specific Features Affect Bengali Reading and Spelling

We assessed Bengali children ranging in age from 7 to 10 years from Grades 2, 3, and 4. All participating children had Bengali as the main home and neighborhood language, and were learning to read and write Bengali at school. Ninety-four children were chosen from a pool of 2281 children from five schools across Kolkata (India) as average performers for their grade (scores between 0 and 1 standard deviation for a given grade). The scores were a composite of three tests: initial phoneme identification, word reading in context, and reading comprehension. The typically developing groups consisted of 37 children from Grade 2 ($M_{age} = 7;1$, $SD = 0.42$); 27 from Grade 3 ($M_{age} = 8;6$, $SD = 0.49$); and 30 from Grade 4 ($M_{age} = 9,4$, $SD = 0.45$). On the composite score, children in Grades 3 ($M_{accuracy} = 72.7\%$; $SD = 4.18$) and 4 ($M_{accuracy} = 77.5\%$; $SD = 2.60$) were higher than Grade 2 ($M_{accuracy} = 30.4\%$; $SD = 22.17$) but similar to each other, $U = 111.51$, $p < 0.01$. Given the substantial overlap in attainment levels in Grades 3-4 (a phenomenon often reported in Indian samples, e.g. Nag, Treiman & Snowling, 2010) we pooled the data and refer to them as the More Skilled (MS) Grade 3-4 group to contrast them with another group of children who we picked from the same grades (Grades 3-4) because their composite score was in the lowest 15%. This lower achieving group of 15 children ($M_{age} = 9;07$, $SD = 0.71$) was identified as the Lesser Skilled (LS) Grade 3-4 group. We were interested to see if the three groups (Grade 2, LS Grades 3-4, and MS Grades 3-4) would fall along a continuum of increasing attainments, or whether, for some aspects of skill development, the LS Grades 3-4 group would fall behind the younger Grade 2 group.

Children were assessed for reading accuracy on 15 words and 15 nonwords which had both CV and CCV akshara with the CCV either in a word initial or medial position. The words were sourced from Grade 2 and 3 textbooks, thus ensuring that all akshara in the assessment lists were from the taught set in school. The nonwords were derived by manipulating one akshara in a different word list also derived from akshara taught up to Grade 2. Orthographically, the word medial CCVs were either represented with two full consonants (i.e. with schwa suppression) or two half forms where the final CCV akshara was either transparent or opaque. Children were also assessed on 30 spellings on words with CV and CCV akshara (10 geminates or 10 mixed clusters). Of the words with the CCV akshara, some were irregular since the second consonant in the cluster was silent or obviated (a silent marker) by the geminating first consonant in the cluster (see Table 1). The words in the test also varied in word length with 10 each of monosyllabic, disyllabic, and trisyllabic words. Table 2 presents the accuracy scores on the three tasks for the three groups of children: those in Grade 2, and those who are lesser skilled and more skilled in Grades 3-4.

There was no difference between the younger readers and the lesser skilled older group on the spelling scores, although the more skilled older children were far ahead. Word reading was at ceiling for all groups perhaps because of our choice to limit the word list to those that contained words only taught up to Grade 2. The profile of scores may have been different if the list also contained words with low frequency, later encountered words. But scores show dispersal on the nonword reading task despite this list also comprising items with akshara taught by the end of Grade 2. Here, the LS Grade 3-4 group scored lower than the younger Grade 2 children even though they were better on word reading. This difference in pattern of scores is understandable because in reading words, children can identify individual phonemes and string them together and read them by mapping them onto a known lexical item. Such lexical cueing is unavailable in nonword reading. Nonword reading in Bengali, as in other languages, is therefore a better indicator of children's knowledge of akshara-phonology mapping. In addition, we found a moderate but significant correlation between nonword reading and performance on spelling (Grade 2: $r=0.414$, $p < 0.05$; MS G3-4 group: $r=0.386$, $p < 0.05$; however associations were not significant for the Low Skilled G3-4 group: $r=0.377$).

Table 2 Comparison of accuracy on reading and spelling across three groups

Measure	G 2 (37)	LS G 3-4 (15)	MS G 3-4 (57)	Mann-Whitney U and post hoc tests
Word reading	96.92 (*0.83*)	98.21 *(0.59)*	99.63 *(0.21)*	264.26 (p < 0.01)
				G2 < LS G3-4 < MS G3-4
Nonword reading	83.67 *(2.71)*	80.92 *(2.06)*	91.67 *(1.32*)	273.39 (p < 0.01)
				LS G3-4 < G2 < MS G3-4
Spelling	57.56 (*4.79*)	55.76 *(5.22)*	76.76 *(0.58)*	458.48 (p < 0.01)
				LS G3-4 = G2 < G3-G4

Note: G2 = the Grade 2 average performing group; LS G3-4 = the lesser skilled Grade 3-4 group; MS G3-4 = the MS Grade 3-4 group

Word Length and Akshara Type in Reading and Spelling

One would expect longer words (with more syllables) to be more difficult to read and spell than shorter words. Though polysyllabic words were relatively more difficult to spell across all groups ($M_{accuracy}$ was 30%, 32% and 58% in Grade 2, the LS Grade 3-4 and the MS Grade 3-4 groups, respectively), it was the phonetic density (the ratio of consonants to vowels) that explained the error patterns more than syllable length. For example, within the monosyllabic word set, the CVC nonwords were read more accurately than the words with CCVC (e.g., কেফ, *kefa* /kef/ was easier than গ্রিল *grila* /gril/). Or, put differently, longer items but with only CV akshara were easier than shorter items but with CCV akshara (e.g. বিছিন *bichina* /bitʃino/ was easier than রগ্না *ragnaa* /rɔgna/). This pattern replicates trends in other akshara orthographies on length and complexity in reading (Sinhala: Wijayathilake & Parrila, 2014) and in spelling (Kannada: Nag et al., 2010). We look at the specific points of difficulty in reading and spelling CCVs, with reference to transparency and visuo-spatial configuration, later in the chapter.

In spelling, CV akshara were easier to read (accuracy above 85% in all groups) than to spell (around 70% for Grade 2 and the LS Grade 3-4 group and 78% for the MS Grade 3-4 group). CV akshara were relatively less easy to spell because of the ambiguous nature of $C_{\varnothing}$ and C/o/ and the use of the long and short vowel digraph. They were also easier to read than mixed CCV akshara (76% and 79% in Grade 2 and the LS Grade 3-4 group respectively).[2] The difficulty in the acquisition of CCV clusters was more pronounced in spelling words with CCV akshara: Grade 2 and LS Grades 3-4 group accuracy was less than 50%, and the MS Grade 3-4 group was at about 73%. In the CCV clusters, geminates were less prone to error (accuracy rates: Grade 2: 73%, LS Grades 3-4 group: 69%; MS Grade 3-4 group: 89%) than mixed clusters (accuracy rates: Grade 2: 32%, LS Grades 3-4 group: 32%; MS Grade 3-4 group: 62%). Within mixed clusters, C/r/ clusters were less prone to error than /n/C clusters.

The Role of Vowel Length Neutralization and Vowel Harmony

Recall that in modern Bengali, phonology has neutralized the distinction between long and short vowels (i.e. /i, i:/ and /u, u:/ merged into /i/ and /u/, respectively). The trends in the data suggest that in word reading the neutralization was easily applied by a beginning reader. The application of neutralization was evident in the nonword লেরা lera which could be read as /lera/ or /læra/ because of ambiguity of representation. More than 80% of the children read the word as /læra/ where the high-mid vowel /e/ was replaced by a low vowel /æ/ in the presence of lower vowel /a/ because

[2] The nonword reading task has no item with geminates, therefore comparative scores on spelling and reading task are unavailable.

of vowel height assimilation in Bengali. In spelling, confusion due to neutralization (which does not give the cue for the use of a long vowel for e.g., ভীড় *bheeRa* /bʰir/ 'crowd') accounted for 6.8%, 18.8% and 2.8% of errors in Grade 2, the LS, and the MS Grade 3-4 groups respectively.

The Role of the Inherent Vowel and Mapping for Vowels /o/ and /ɔ/

Written Bengali does not distinguish between the inherent vowels /o/ and /ɔ/ and are written without a diacritic marker. Bengali nevertheless has another vowel full form of ও and its diacritic ো which represents /o/. On the face of it, it may appear that children have no phonological cues for choosing between the two representations for /o/ (the unmarked inherent vowel and the marked ো diacritic). But it is word level phonology that decides which to use when. The unmarked form is used usually before a consonant cluster, and very few children make an error (altogether 6 such errors were recorded and here, children used the marked form in words where আপত্তি *aapatti* /apotti/ 'objection' and থাপ্পড় *thaappoRa* /tʰa:poɹ/ 'slap' were erroneously spelt as আপোত্তি and থাপ্পোড়). A word which was contrary to this pattern রোদ্দুর *raddura* /roddur/ 'sunshine' was spelled wrongly by many children: 9 in Grade 2, 3 in the lesser skilled Grade 3-4 group and 1 in the more skilled Grade 3-4 group. The sound /o/ in the first syllable of a monosyllabic word is indicated with the marked form. Younger children appeared to be slower to pick up this rule and used the unmarked form instead, e.g. সোন *sona* /ʃon/ 'listen!' was spelt as সন (8 such errors).

The impact of the ambiguity around the inherent vowel was seen in reading word final and medial consonants, particularly in Grade 2 and the lesser skilled Grade 3-4 group. The difficulty was in deciding whether to drop or retain the inherent vowel in reading a word medial akshara (Sircar & Nag, 2014).[3] Interestingly, we find the word final consonant in nonwords also causing a similar confusion. Children read the last akshara in the nonword বিচিন *bichina* /bitʃin/ with the vowel /o/ as specified by orthography, though these errors were marginal, 2.2% in Grade 2 and 2.7% in the lesser skilled Grade 3-4 group. However, such errors in reading were not found in the word reading task. Neither did we find instances of children wrongly inserting a vowel diacritic ো in spelling words that end in akshara with a suppressed inherent vowel.

A rare but intriguing error was when পীরশা *pirashaa* /piroʃa/ was read as /poriʃa/ by 12 of the 109 children: Here, the first vowel shifts to the second syllable, and the first is reduced to the inherent vowel /ɔ/: $CV_\alpha CV_\beta > CV_\beta CV_\alpha$. Such vowel-vowel metathesis is rare in natural languages (McCarthy, 1985), but some languages shows

[3] In Sircar and Nag (2014) we showed that CVCaCV nonwords, such as দটকা *DoTaka* that could be read as /dɔtka/ or /dɔtoka/, are read as /dɔtka/ like a real word পটকা /pɔtka/ 'firecracker': More skilled children often rely on lexical and phonotactic information rather than akshara-phonology mapping rules.

similar vowel compensation.[4] Such vowel shifts were not seen in nonwords, where both the first and second syllable had inherent vowels or non-inherent vowels. Such shifts were also not seen in spelling.

Some children (altogether 10) inserted a vowel /o/ in between in a CCV. For example, they read তীপ্ন *tipna* /tipno/ as /tipon/. It is interesting that children tended to insert /o/ rather than /ɔ/ in word medial positions, respecting the phonological distribution of /o/ and /ɔ/. Such insertions were not seen in CCV with a non-inherent vowel; for instance, no child read টেল্পী *telpi* /telpi/ as /telip/ or রগ্না *ragnaa* /rɔgna/ as /rɔgon/.

Children rarely reduced vowels to the inherent vowel in reading words, clearly indicating that children recognized all vowel diacritics and mapped them onto their phonological values. However, in spelling, some Grade 2 children used an akshara with an inherent vowel for an akshara with a ligatured vowel diacritic in more cases than the older learners (error rate: 1.6%, 1.6% and 0.30% respectively in Grade 2, LS Grade 3-4 and MS Grade 3-4 groups). Finally, CCV akshara with the long vowel /a:/ were more likely to be reduced to the inherent vowel than vowels with /u/ and /i/.

Table 3 presents a summary of vowel diacritic errors.

A point to highlight is that the LS Grade 3-4 group was severely impacted by vowel length neutralizations in spelling (18.82%) and the suppression of the inherent vowel in reading (13.33%). In both cases, the phonology-orthography is insufficient for the task of decoding, and children needed to go beyond mapping to understanding of word level phonology, which is not always systematic and generalizable. In this they were poorer than the younger Grade 2 learners. Put differently, the younger but typically developing Grade 2 and the MS Grade 3-4 groups showed knowledge of word level phonology, and this supported their reading more than their spelling, while for the LS group, both spelling and reading vowels remained difficult.

Impact of Phonological Neutralization in Consonants

Orthography does not reflect the phonological neutralizations that some consonants have gone through. As predicted, homophonous consonants were one of the major sources of substitution in the spelling task where the wrong sibilant, rhotic or nasal was used in words across groups. In the 16 words providing scope of such phonological neutralizations, the pattern of errors was as follows: Grade 2: 22.63%; LS Grade 3-4: 18.75%; MS Grade 3-4: 10.96%). Most of the errors were made in the choice of the sibilants (স, শ ষ) and the rhotic (র and ড়). In reading, except for the rhotic /r/ being read as /ɹ/, phonological neutralizations showed little effect.

[4] Hawu shows a form of vowel metathesis which is more of a phonotactic repair strategy or a speech lapse.

Table 3 Percent errors in reading and spelling vowels

Type of errors	Example	Nonword reading			Word spelling		
		G 2	LS G3-4	MS G3-4	G 2	LS G3-4	MS G3-4
Vowel length (12 akshara in nonword reading;17 akshara in spelling)	ভীড় /b^{h}iɹ/ 'crowd' as ভিড় /b^{h}ɪɹ/	–	–	–	6.83%	18.82%	2.78%
Marked for unmarked for /o/ before or after CCV (8 akshara)	থাপ্পড় *thaappoRa* /t^{h}a:poɹ/ 'slap' as থাপ্পোড়	–	–	–	1.09%	0.83%	0
Inherent vowel for suppressed vowel in coda (7 akshara; 18 akshara)	বিচিন *bichina* /bitʃin/ as /bitʃino/	5.41%	13.33%	1%	–	–	–
Inherent for vowel diacritic (34 akshara with diacritics)	পুরুষ /puruʃ/ 'man' as পরুষ /poruʃ/	–	–	–	3.73%	2.74%	0.72%
Vowel shift (6 words)	পীরশা *pirashaa* /piroʃa/ *as* /poriʃa/	0.90%	2.22%	0.23%	–	–	–
Vowel insertion in clusters[a] (6 CCVs)	তীপ্ন *tipna* /tipno/ as /tipon/	1.35%	3.33%	1.16%	–	–	–

Note: G2 = the Grade 2 average performing group; LS G3-4 = the lesser skilled Grade 3-4 group; MS G3-4 = the MS Grade 3-4 group

[a]In spelling children split the consonant cluster: we report them in the CCV section

The CCVs

Reading and spelling of CCV akshara depend on knowledge beyond simple phonological-orthographical codes, and most often oral language indicates to a learner how specific clusters behave and how they influence the neighboring vowels and consonants. Some of the pressure points in gaining mastery are understanding of the visual layout of the consonants in a cluster, awareness (and representation) of the sonorous segment of a cluster, and knowledge of word level phonology, i.e., how a consonant in a cluster is pronounced and transcribed.

(i) *Retrieving phonology from a visuo-spatial configuration*: Unlike Hindi, Bengali arranges the consonants in a cluster in many ways. When the arrangement does not cue linearity, Grade 2 children impose a bottom-to-top linearity; treating the bottom as the first consonant, and the top as the second, particularly in nonwords where semantics does not come to their rescue. Therefore, the nonwords তেল্পী /telpi/, তীপ্ন /tipno/ and রগ্না /rɔgna/ get read as /tepli/, /tinpo/ and /rɔnga/, respectively, with lesser skilled group most affected by the lack of cues (40% errors) (compare: Grade 2 9.9% errors; More skilled Grade 3-4: no errors). What happens in spelling? We find that the lesser skilled group applies a simple linear rule (like in Hindi), i.e., left to right representation mirroring the phonological order, completely ignoring the various ways in which Bengali arranges consonants in a cluster. However, such a visual representation was not found in the MS group. Children figure out how specific clusters are represented in Bengali after some initial glitches.

(ii) *Dropping sonorous consonants in a CCV*: Children are known to drop the internal sonorous consonant in speech, and younger children in the study dropped them in reading a cluster and so did the lesser skilled group (Grade 2: 9.9%; LS: 8.8%) while the MS had few such omissions in the present study. However, interestingly, when nasals were dropped from a /pn/ or /gn/ cluster, the resulting sound was geminated (/tipno/ was read as /tippo/), quite in keeping with a language-specific trend of clusters with a silent consonant. In spelling, such omissions accounted for 33.6% errors in Grade 2 learners, 19% errors in LS, and 8.8% in the MS group (in this analysis words with silent consonants were not included). When the akshara block was complex with more than two or three ligatures, the omissions appeared to be higher. For instance, the /r/ in /bhr/ ভ্রুক্ষেপ *bhrukshepa* /bhrukhep/ ‘qualm’ was dropped more often than for others: This is an instance where the visuo-spatial organization may also be adding a cognitive challenge to the task of spelling. Fewer drops were seen in all three groups in geminates which are phonologically transparent: Grade 2: 4.05%; LS Grade 3-4: 3.33% and MS Grade 3-4: 1.92%).

(iii) *Clusters with an unvoiced consonant*: In the spelling task, unvoiced consonants in a cluster were routinely dropped (Grade 2: 54.05%; LS Grade 3-4: 52.22%, and MS Grade 3-4: 28.94%) since children used only the primary level of representation, i.e., the phonological-orthography mapping. In medial CCVs with /j/ as the secondary consonant, children produced a geminated cluster. For example, in তাচ্ছিল্য *taacchilya* /tatʃtʃ^{h}illo/ was spelled as তাচ্ছিল্ল, showing a tendency to assume that all phonological information is contained within the akshara rather than distributed akshara representing the phonology across the word. Spelling geminates with /Cy/ rather than /CC/ were not seen in other high frequency words in the test.

(iv) *A conjunct or two full akshara*: Two different ways of representation may result in a phonological consonant cluster as a conjunct or two full akshara. In the reading test, both forms were available in the target items. As described earlier, children had little difficulty in figuring out that both forms may lead to a CCV (except those read analogically as CVCVCV) In the spelling task, though no target CCV item was spelled with two full akshara, children were found to write select items /mɔnt^{h}on/, /roddur/ and /t^{h}appoɹ/ with two full akshara rather than a conjunct (G 2: 2.27%; LS G 3-4: 3.2%; and MS G 3-4: 0.98%). A medial CCV akshara usually gets torn apart and phonologically encoded as CVC.CV with a closed syllable followed by an open/closed syllable (Sircar & Nag, 2014). Children were more likely to write the cluster with two full consonants when it resulted in a CVC.CVC rather than CVC.CV. Table 4 summarizes the discussion on consonant errors so far.

Table 4 Percent errors in reading and spelling consonants

Type of errors	Examples	G 2	LS G3-4	MS G3-4	G 2	LS G3-4	MS G3-4
Phonological neutralization (5 nonwords; 16 words)	ভীড় /bʰiɹ/ 'crowd' as ভীর /bʰir/	3.24%	12%	1.7%	22.63%%	18.75%	10.96%
Omitting sonorous consonants in CCV (3 nonwords; 7 words)	রগ্না *ragna* /rɔgna/ as /rɔgga/ পরিশ্রান্ত *parishantra* /porisranto/ as /porisanto/	9.90%	8.8%	0.2%	33.6%	19%	8.8%
Omitting consonants in geminates (10 words)	আপত্তি *apatti* /apotti/ 'dissent' as আপতি /apoti/	–	–	–	4.05%	3.33%	1.93%
Silent consonant in CCV (5 words)	স্মরনিয় *smaraniya* /sɔronijo/ 'memorable' as সরনিয়	–	–	–	54.05%	52.22%	28.94%
Full akshara for CCV (25 CCVs)	মন্থন manthana /mɔnthon/ ‘ as মনথন	_	_	_	2.27%	3.2%	0.98%

Note: G2 = the Grade 2 average performing group; LS G3-4 = the lesser skilled Grade 3-4 group; MS G3-4 = the MS Grade 3-4 group

The varied akshara-phonology associations that are reported in many other akshara languages (Bright, 1996; Nag, 2013) interacts with spelling and reading acquisition in beginning learners of Bengali. While phonological neutralizations impact spelling more than reading, inconsistent mapping and visual complexity in CCV impact both reading and spelling. A point that becomes clear is that to circumvent inconsistencies in phonology-orthographic links, children need to move beyond the phoneme and syllable (which are sub-lexical and thus not affected by word boundaries) and understand phonological rules that apply at the level of the word or lexical unit. For instance, spelling CCVs with unvoiced segments requires more fine-tuned knowledge of phonology (nasalization, germination, or vowel length with reference to the position of the cluster in the word) which can cue the learner to the presence of an orthographic segment.

Going Beyond: Word Level Phonology

The intention in the present study was to explore language specific features, such as vowel and consonant neutralization, the inherent vowel, and how consonant clusters impact reading and spelling in different ways. In neither task is the mere knowledge of phonological-orthographic mapping sufficient.

Briefly, what we have found are the following:

(i) Reading and spelling performance increased rapidly in the first two years of schooling, and this result is in line with findings in other alphasyllabic orthographies. Grade 2 children achieved 96% accuracy for reading words with taught akshara and 84% accuracy for reading nonwords. A particular area of rapid mastery is knowledge about the unmarked inherent vowel. Familiarity pulls children out of the inherent vowel puzzle when in a word, but where the vowel is underspecified in the printed nonword, decoding becomes tricky. The way to read these items would be to read by analogy, or with the help of word level knowledge (position, vowel context, etc.): The inherent vowel is pronounced as /ɔ/, or /o/, or is suppressed. Therefore, children go beyond simple principles of mapping, and we argue that most of such word level phonological learning is derived from exposure and use of oral language. Those who are able to tune in to these phonological nuances succeed in reading, whereas the others fall back on taught linkages and overextend these akshara-sound correspondences to contexts that are, in fact, not permissible. In spelling, children additionally need to understand the language specific orthographic conventions within a word with respect to the inherent vowel which is governed by position (e.g., /o/ in a word initial syllable is represented with a diacritic, and /ɔ/ without) and type of akshara preceding or following them.

(ii) The assembly of spelling, particularly, is made complex by phonological neutralizations, where the phonological information is insufficient for the purpose. Because of phonological neutralizations of both consonants and vowels, the inventory of phonemes in Bengali has been whittled down to a smaller set though orthography stubbornly preserves all these once-distinct phonemes. Nevertheless, the orthography makes up for this spelling difficulty by simplifying the task of reading. The only way in which a speller can assemble the right spelling is by bypassing underspecified phonological information. Frequency of reading exposure and use of specific words in writing can play a contributive role here but may not be productive for low frequency words and nonword spelling.

(iii) A cluster, where a consonant is not voiced, is a test case of what children have learned about Bengali phonology. It is not only the phonological presence or absence of a sound in a word that needs to be noted and represented, but also how a segment changes other phonological features in the word: They may introduce new segments (like germination), or substitute one segment with another (nasalization, vowel lengthening). If these are used as cues, then children would infer the presence of an unvoiced segment, and be able to orthographically represent it. Since these rules pertain to only specific clusters and are not productive (i.e., they cannot apply to (recently) borrowed words), they are slower to be acquired.

In conclusion, we have shown that Bengali has its own distinctive orthography with homophonous consonants and vowels, complex graphemic expression of CCVs, inconsistent pronunciations, and ambiguous representation of the inherent vowel:

These may be learned only if one goes beyond the usual mapping that literacy instruction typically provides. Delving into distributed word level phonology is also clearly important for Bengali literacy development, and advancing towards mastery is dependent on substantial exposure to print and a language-rich environment

Acknowledgements This study was funded by a Seed Grant from The Promise Foundation to the first author and supported by Sarva Shiksha Abhiyan, Kolkata. We thank the Principals, teachers and children of the five participating children and their teachers, and the research assistants: Syed Saurov, Madhurima Chatterjee, Sulagna Ganguly, Jayeeta Mazumdar, and Dipannita Chakraborty for data collection and data coding.

References

Alcock, K. J., & Ngorosho, D. (2003). Learning to spell a regularly spelled language is not a trivial task-patterns of errors in Kiswahili. *Reading and Writing: An Interdisciplinary Journal, 16*, 635–666.

Bright, W. (1996). The Devanagari script. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 384–390). New York: Oxford University Press.

Chatterji, S. K. (1926). *The origin and development of the Bengali language* (Vol. 1). Calcutta, India: University of Calcutta.

Dan, M. (1992). *Some issues in metrical phonology of Bangla: The indigenous research tradition.* Ph.D. dissertation, University of Poona, Pune.

Dasgupta, P. (2003). Bangla. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 351–390). London: Routledge.

Geva, E., Wade-Woolley, L., & Shany, M. (1993). The concurrent development of spelling and decoding in different orthographies. *Journal of Reading Behavior, 25*, 383–406.

Hai, M. A. (1975). *Dhonibiggan o Bangla dhonitatto*. Dhaka, Bangladesh: Barnamichil.

Heimisdóttir, L., O. (2013). The phonetics of Bengali consonant clusters. In *Cornell working papers of phonetics and phonology*. http://conf.ling.cornell.edu/plab/paper/CWP3_2013_Heimisdottir1.pdf.

Kar, S. (2010). *Syllable structure of Bangla: Anoptimality theoretic approach*. Newcastle, UK: Cambridge Scholars.

Kishalaya, Book 1, 2. (2004). Paschim Banga Vidyalaya-Shiksha Adhikar, Kolkata.

Majumdar, A. (1997). *Bhasha Tatva*. Kolkata, India: Naya Prakas.

Masica, C. P. (1991). *The Indo-Aryan languages*. Cambridge, UK: Cambridge University Press.

Mathur, C., & Nag, S. (2019). Language-focused Instruction for literacy acquisition in *Akshara*-based languages: Pedagogical considerations and challenges. In M. Joshi, & C. McBride, (Eds). *Handbook of Akshara languages* (pp. 279–309). Cham, Switzerland: Springer.

McCarthy, J. J. (1985). *Formal problems in Semitic phonology and morphology*. New York: Garland.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22.

Nag, S. (2013). Low literacy attainments in school and approaches to diagnosis: An exploratory study. *Contemporary Education Dialogue, 10*(2), 197–221.

Nag, S. (2014). Akshara-phonology mappings: The common yet uncommon case of the consonant cluster. *Writing Systems Research, 6*(1), 105–119.

Nag, S. (2017). Learning to read alphasyllabaries. In K. Cain, D. Compton, & R. Parrilla (Eds.), *Theories of reading development*. Amsterdam: John Benjamins.

Nag, S., & Narayanan, B. (this volume). Orthographic knowledge, spelling and reading development in Tamil: The first three years. In M. Joshi & C. McBride (Eds.), *Handbook of Akshara languages*. Cham, Switzerland: Springer.

Nag, S., & Sircar, S. (2008). *Learning to read in Bengali: Report of a survey in five Kolkata primary schools*. Bangalore, India: The Promise Foundation.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*(1), 41–52.

Ohala, M. (1983). *Aspects of Hindi phonology*. New Delhi, India: Motilal Banarasidass.

Pandey, P. (1990). Hindi schwa deletion. *Lingua, 82*, 277–311.

Pandey, P. (2007). Phonology-orthography interface in Devanagari for Hindi. *Written Language and Literacy, 10*(2), 139–156.

Pandey, P. (2014). Akshara-to-sound rules for Hindi. *Writing Systems Research, 6*(1), 54–62.

Patel, J., Bapi, R. S., & Nag, S. (2013). *Akshara counts in child directed print: A pilot study with 101 texts.*

Patel, P. G. (2004). *Reading acquisition in India: Models of learning and dyslexia*. New Delhi, India: Sage.

Pinheiro, A. M. V. (1995). Reading and spelling development in Brazilian Portuguese. *Reading and Writing: An Interdisciplinary Journal, 7*(1), 111–138.

Rahbari, N., Sénéchal, M., & Arab-Moghaddam, N. (2007). The role of orthographic and phonological processing skills in reading and spelling of monolingual Persian children. *Reading and Writing: An interdisciplinary Journal, 20*, 511–533.

Ravid, D. (2011). *Spelling morphology: The psycholinguistics of Hebrew spelling*. Springer-Verlag/Sci-Tech/Trade.

Sarkar, P. (1986). *Aspects of Bengali syllables.* Paper presented at the National seminar on the syllable in phonetics and phonology, Hyderabad, India: Osmania University.

Shapiro, M. C. (2007). Hindi. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages* (pp. 250–285). London: Routledge.

Sircar, S., & Nag, S. (2014). Akshara-syllable mappings in Bengali: A language-specific skill for reading. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 409–425). Cambridge: Cambridge University Press.

Sprenger-Charolles, L., Siegel, L. S., & Bonnett, B. (1998). Reading and spelling acquisition in French: The role of phonological mediation and orthographic factors. *Journal of Experimental Child Psychology, 68*, 134–165.

Tiwari, S., Nair, R., & Krishnan, G. (2011). A preliminary investigation of akshara knowledge in the Malayalam Alphasyllabary: Extension of Nag's (2007) study. *Writing Systems Research, 3*(2), 145–151.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690.

Wijayathilake, M. A. D. K., & Parrila, R. (2014). Predictors of word reading skills in good and struggling readers in Sinhala. *Writing Systems Research, 6*, 120–131.

Wijayathilake, M. A. D. K., Parrila, R., Inoue, T., & Nag, S. (2018). Instruction matters to the development of phoneme awareness and its relationship to *Akshara* knowledge and word reading: Evidence from Sinhala. *Scientific Studies of Reading, 22*(5), 420–433.

Winskel, H., & Iemwanthong, K. (2010). Reading and spelling acquisition in Thai children. *Reading and Writing: An Interdisciplinary Journal, 23*(9), *1021–1053*.

Assamese Orthography: An Introduction and Some Applications for Literacy Development

Hemanga Dutta

Abstract This chapter outlines the basic characteristics of Assamese orthography from the northeastern part of India. Some of the inconsistencies in the orthography that should be taken into consideration while providing beginning literacy instruction are highlighted. The main purpose of this work lies in reducing the inconsistencies between the written and spoken synchronic Assamese keeping in mind the issues pertaining to learners. It is found that Assamese phonology adopts different repair strategies in order to incorporate borrowed syllables and segments into its native phonotactic domains. However, such patterns and scientific insights are absent in the orthography used in the language which causes difficulties on the part of the learners. The phonological and graphemic consistency is more visible in English loan words than in Sanskrit loan words which have been regularized in Assamese in the course of time.

Keywords Orthography · Literacy · Phonology · Repair strategies · Loan words

Introduction

There is a high degree of relationship between phonological processes and writing system as a representational unit (Pandey, 2014). Here I am trying to demonstrate the intrinsic relationship existing between phonology and orthography of the Assamese language, spoken mainly in the north eastern part of India. The paper draws many examples and illustrations from the language in order to indicate the exact nature of the relationship existing between Assamese sound system, speech patterns exhibited by the segments and the writing system which has been derived

I am grateful to the editors for their comments and suggestions on earlier version of this chapter.
Correspondence regarding this chapter can be sent to Dr. Hemanga Dutta (hemangadutta1@gmail.com).

H. Dutta (✉)
Department of Linguistics and Contemporary English, School of Language Sciences, The English and Foreign Languages University (EFLU), Hyderabad, Telangana, India
e-mail: hemangadutta1@gmail.com

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_10

from Brahmi script in the course of time. According to Pandey (2014), the relationship between akshara and sound in relation to Hindi can be based on an interdisciplinary approach that involves linguistic and cognitive studies. The main purpose of this work lies in reducing the inconsistencies between the written and spoken synchronic Assamese keeping in mind the issues pertaining to learners. The data used in this language are collected from various published descriptions of the language, mostly from Goswami (1966), Kakati (1941) and Dutta (2017).

The paper is divided into seven sections. Section "Assamese language" presents a general description of the Assamese orthography along with its phonemic inventories and phonotactic patterns. Section "Evolution of Assamese orthographic system" provides a glimpse of the evolution of Assamese script in the course of history through various phases. Section "Assamese phonology and orthography correlates: Some linguistic insights" shows the phonemic and graphemic correlations in Assamese with reference to the representation of vowel and consonant sounds in the language. Section "Nasal assimilation and Assamese orthography" highlights the nasal assimilation instances in Assamese where coda nasal consonants undergo assimilation in place to a following obstruent. In Assamese, place nasal assimilation occurs at word internal position, but across word boundaries nasals resist assimilation. Section "Gemination and orthography in Assamese" illustrates the inconsistencies found in the writing system of Assamese in the light of the data drawn from gemination. It is interesting to observe that the geminated patterns that Assamese phonology exhibits have got no orthographic correspondence. Section "Loan word phonology and orthography in Assamese" describes the orthographic representations of English loans in Assamese which has been phonologized in the course of time. The orthographic and phonemic consistency is more visible in English loans rather than Sanskrit loans in Assamese. The concluding section offers certain suggestions and reforms of the Assamese script keeping in consonance the need for a more economic and more easily learnable synchronic Assamese orthography.

Assamese Language

Assamese is an Indo Aryan language spoken in north eastern part of India. It is also the official language of Assam. Assamese is the eastern most Indic language and is surrounded by the languages belonging to Tibeto-Burman and Austro-Asiatic language families. To have a comprehensive understanding of the orthography phonology correlates in Assamese, it is important for us to look at the phonemic inventory of the language (both vowels and consonants) as shown below Dutta, 2017) (Tables 1 and 2):

Vowel Inventory

Table 1 Vowel phoneme inventory in Assamese

	Front	Central	Back
Close	i		ʊ
Close-mid	e		o
Open-mid	ɛ		ɔ
Open		a	ɒ

Consonant Inventory

Table 2 Consonant phoneme inventory in Assamese

Articulation place → Manner ↓	Bilabial		Alveolar		Post-Alveolar	Palatal	Velar		Glottal
Plosive	p	b	t	d			k	g	
	p^h	b^h	t^h	d^h			k^h	g^h	
Nasal		m		n				ŋ	
Fricative			s	z				x	h
Approximant		w	ɹ			j			
Lateral approximant				l					

According to Masica (1991), the Assamese phoneme inventory is unique in the Indic group of languages in its lack of a dental-retroflex distinction among the coronal stops. Historically, the dental and retroflex series merged into alveolar stops. This makes Assamese resemble non-Indic languages of Northeast India such as Austro-Asiatic and Sino-Tibetan (Moral, 1997). The only other language to have fronted retroflex stops into alveolars is the closely related eastern dialects of Bengali (although a contrast with dental stops remains in those dialects). Note that /r/ is normally realized as [ɹ], an approximant.

Assamese is unusual among Eastern Indo-Aryan languages for the presence of the /x/ (Dutta, 2003) although it shares this phonemic feature with its neighboring language Sylheti. The presence of /x/ in the phonemic inventory of Assamese can be attributed to the Tibeto-Burman influence. The voiceless velar fricative is absent in the West Goalpariya dialects[1] though it is found in lesser extent in East Goalpariya and Kamrupi, otherwise used extensively further east (Goswami & Tamuli, 2003).

The language has quite a few regional variations. Kakati (1941) identified two broad dialects which he named Eastern and Western dialects, of which the eastern dialect is homogeneous, and prevalent to the east of Guwahati, and the western dialect is heterogeneous. Western variety is divided into Goalpara dialect and Kamrupi dialect. Below is the linguistic division of the Assamese language and its varieties as per the classification of Kakati (1941) (Fig. 1).

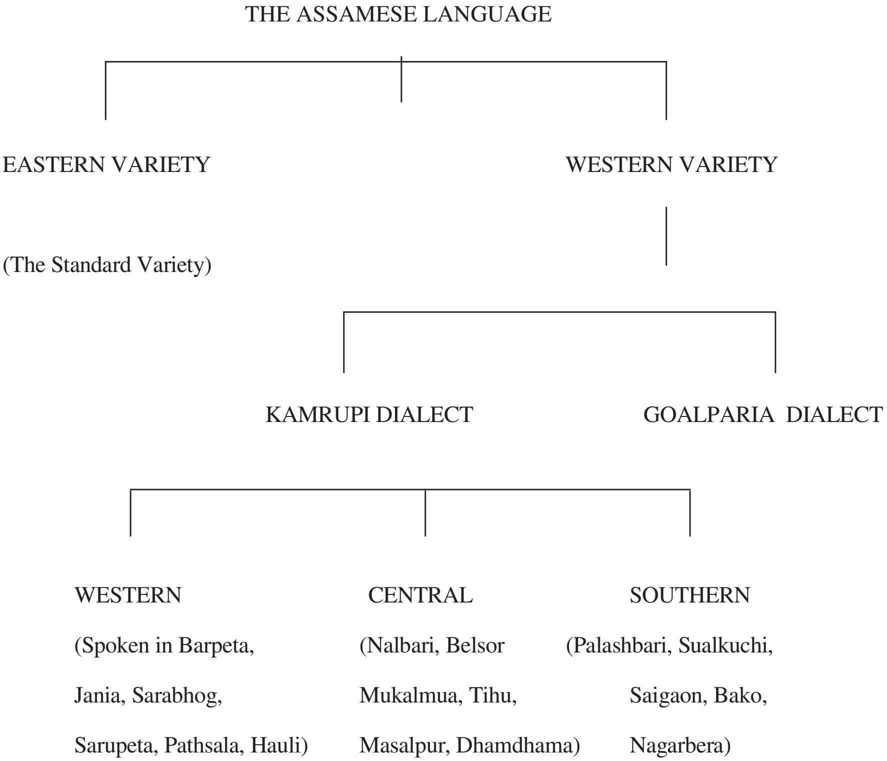


Fig. 1 Assamese and its different varieties

Evolution of Assamese Orthographic System

Assamese script is one the important scripts of Eastern India. The origin of Assamese script can be traced back to Gupta stage of Brahmi script. The evolution of this Assamese Script can be divided into three different stages: Old Assamese Script or the *Kamrupi* Script (fourth/fifth –thirteenth century); Medieval Assamese Script (thirteenth to early part of nineteenth century); and Modern Assamese Script from early part of nineteenth century.

There is evidence of birth of script in Assam in the fourth/fifth century itself. According to Goswami (1987), Assamese Script is a direct descendent of the Brahmi script in its original evolutionary stage, and Umachal script and the aksharas found in village script of Nagajari-Khanikar bear a close resemblance with Gupta script. The fourth/ fifth century inscription of Nagajari-Khanikaris considered as the oldest known proof of Assamese script. Scholars state that there is a similarity between the aksharas of this inscription of Nagajari-Khanikar with scripts of the pre-Gupta period.

After these two scripts, the influence of Brahmi and Gupta scripts was found in the seventh century Copper Plates of *Kamrupa* King Bhaskaravarman. *Dubi* Copper plate inscriptions of *Kamrupa* during Bhaskarvarman's rule indicate the use of ষ and স. Since mid nineth–tenth century onwards, aksharas of copper plates started to take almost modern forms, perhaps around early part of thirteenth century. *Kanai-Barashibowa Turaska Khyay* rock inscription is the latest example of Modern Assamese inscription. Modern Assamese script has almost exact similarity to the aksharas of this plate. Assamese script was found to be fully developed in rich Assamese literature from the later part of twelth century. During thirteenth century to seventeenth and eighteenth century, the vast Assamese literature was written mostly in manuscripts. During fourteenth to eighteenth century, rock inscriptions, aksharas written on copper plates show the proof of the growth of modern Assamese akshara. Further, Assamese Script gets its Modern form when the Missionaries started publishing printed books and magazines etc. in the early nineteenth century (Bhattacharjee, 2013).

Assamese Phonology and Orthography Correlates: Some Linguistic Insights

Speech sounds of a language always get their representation in written form in a systematic way although there is always discrepancy between spoken forms and written representations. Most of the phonemes in Assamese have one to one correspondence with graphemes. However, there is plenty of evidence in which the relationship between grapheme and phoneme appear to be quite arbitrary. To have an understanding of the issues involved in such discrepancy and its far reaching impact on learners we need to know in details about Assamese orthographic shapes representing vowels, consonants and consonant clusters of the language. I am describing the Assamese vowel alphabet chart and Assamese consonant alphabet chart below (Tables 3 and 4).

Since the number of aksharas used in Assamese script and phonemes in Assamese is not identical there is bound to be some mismatch between graphemes and phonemes Assamese by virtue of being a modern Indo Aryan language is bound to have

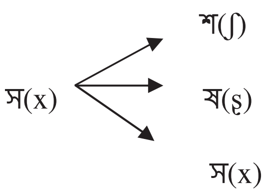
Table 3 Assamese vowel chart according to orthography

Assamse script	Phonetic form	Assamese script	Phonetic form
অ	ɔ	ঋ	ɹi
আ	a	এ	ɛ
ই	i	ঐ	ɔi
ঈ	i:	ও	o
উ	ʊ	ঔ	ou
ঊ	ʊ:		

Table 4 Assamese Consonant chart according to orthography

Assamese alphabet	Phonetic form	Assamese alphabet	Phonetic form
ক	k	ফ	p^h
খ	k^h	ব	b
গ	g	ভ	b^h
ঘ	g^h	ম	m
ঙ	ŋ	য	z
চ	s(s)	ৰ	ɹ
ছ	s(sh)	ল	l
জ	z	ৱ	w
ঝ	zh	শ	x (ʃ)
ঞ	ɳ	ষ	x (ʂ)
ট	ʈ	স	x
ঠ	ʈ^h	হ	h
ড	ɖ	ক্ষ	k^hj
ঢ	ɖ^h	ড়	ɹ(ɽ)
ণ	ɳ	ঢ়	ɹ(ɽ)
ত	t	য়	J
থ	t^h	ৎ	t
দ	d	ং	ŋ
ধ	d^h	ঃ	h
ন	n	ঁ	˜ (Nasalization)
প	p		

the influence of its source language Sanskrit and other sister languages such as Hindi and Bangla. Although Assamese does not have retroflex segments in its phonemic inventory, aksharas representing retroflex fricatives are very common in the orthographic system and the reason can be attributed to the influence of its source and neighboring languages where retroflexion is an eminent phenomenon. For instance, শ(ʃ),ষ(ʂ),স(x) are present in the Assamese Orthography, there is no one to one correspondence in terms of phonetic representation. Only স(x) represents all three শ(ʃ),ষ(ʂ),স(x) sounds.



In a similar fashion, the presence of voiceless affricate in Assamese orthography doesn't have a phonetic correspondence in the Assamese sound system. Assamese phonemic inventory allows only alveolar fricartive /s/ and velar fricative /x/ which are in contrastive distribution as the substitution of one sound with the other creates a semantic difference.

Consider the following examples:

/s/	sap	'pressure'	(চাপ)	/x/	xap	'snake'	(সাপ)
	sat	'a whip'	(চাট্)		xat	'seven'	(সাত)

However, in orthography the palatoalveolar /ʃ/ and retroflex /ʂ/ do exist. As far as the phonetic output is concerned they are either realized as /s/ or sometimes /x/.

	Orthographic form	phonetic form	
/ʃ/	ʃrɔm	srɔm	(শ্ৰম)
	ʃibɔ	xibɔ	(শিৱ)

The voiceless retroflex segment /ʂ/ is represented independently and also as homorganically with the following retroflex segment.

Consider the examples:

Assamese script	Orthographic form	Phonetic form	Gloss
শেষ	ʃeʂ	xex	end
বৰ্ষ	bɔrxɔ	bɔrʂɔ	year
কৃষ্ণ	kriʂɳɔ	krisnɔ	name of a god
কষ্ট	kɔʃʈɔ	kɔstɔ	pain

There are certain graphemes which are pronounced as consonant clusters but which are represented as single grapheme, following the practice in Sanskrit such as gnan 'knowledge". Assamese has a grapheme which is a combination of /k^h / and /j/. This is derived from the Devanagari which is an amalgamate of the segments <kʃ> as in the words /kʃəʈi/. The phonetic correspondence of the grapheme in synchronic Assamese is quite regular as the consonant clusters, rather than a single sound unit.

The correlates between Assamese phonology and orthography can be addressed in relation to various phonological processes such as nasal assimilation, gemination, loan word phonology.

Nasal Assimilation and Assamese Orthography

In this section, I am going to discuss the instance of nasal assimilation in Assamese where coda nasal consonants undergo assimilation in place to a following obstruent. Here, nasal assimilate in terms of feature [place] at word internal level but across word boundaries nasals resist assimilation. It can be represented in the following schemata:

Homorganic Nasals in Word Internal Position

Words	Gloss	Assamese script
ɒŋkɒ	(sum)	(অংক or অঙ্ক)
ɒŋgɒ	(organ)	(অংগ or অঙ্গ)
xɒŋgit	(music)	(সংগীত or সঙ্গীত)
xɒŋg^{h}ɒ	(club)	(সংঘ or সঙ্ঘ)
xɒŋkɒɹ	(the name of Lord Siva)	(শংকৰ or শঙ্কৰ)
kɔmbɔl	(blanket)	(কম্বল)
bɔndɔna	(worship)	(বন্দনা)

But across word boundaries nasals resist assimilation, as shown in the following words:

bɒn git – bɒn git	*bɒŋgit	(folk songs)	(বনগীত)(*বঙগীত)
g^{h}ɒn kola – g^{h}ɒn kola	*g^{h}ɒŋkola	(dark black)	(ঘনকলা)(*ঘঙকলা)
g^{h}amxɒɹa – g^{h}amxɒɹa	*g^{h}aŋxɒɹa	(to shed sweat)	(ঘামসৰা)(*ঘাঙসৰা)
gangua – gangua	*gaŋgua	(to sing)	(গানগোৱা)(*গাংগোৱা)

In Assamese, nasal assimilation takes place at the lexical level, but it is blocked at the post lexical level. From the data it is also possible to draw a conclusion that post nasal voicing is optional in Assamese. As far as the orthographic representation with regards to nasal assimilation is concerned it is evident that Assamese orthography has two ways of representing the homorganic nasals, one by means of an archiphonemic device of anuswara that occurs independently (when n turns in to ŋ) and the other by representing the nasal as a part of the conjunct akshara. The device of archiphonemic device of anuswara is applicable in all the cases of nasal assimilation in Devanagari script whereas in Assamese it is confined to the representation of velar nasal.

Gemination and Orthography in Assamese

Gemination is a phonological process which mainly refers to the doubling of the consonants. There is a common pattern observed in Sanskrit and most of the Indo Aryan languages that word internal obstruents are geminated when they precede the liquids and approximants /j/, /r/, /l/ and /w/. Since Assamese is a modern Indo-Aryan language it is bound to have words borrowed from Sanskrit but with the passage of time these words have developed independent phonotactic patterns with the usage of different repair strategies, but unfortunately the orthography still retains the original Sanskrit pattern.

Let’s see the differences between the Assamese words and their source counterparts.

Here are the obstruent geminates followed by /r/.

Sanskrit source words	Gloss	Assamese counterparts	
put.trə	son	‘put.trɔ’	(পুত্ৰ)
səmud.drə	ocean	‘xɔ.mud.drɔ’	(সমুদ্ৰ)
vik.krəm	aggression	‘bik.krɔm’	(বিক্ৰম)
as.srəm	holy place	‘as.srɔm’	(আশ্ৰম)

If we see the Assamese orthographic pattern, it is evident that Assamese script displays the non geminated form. Similar patterns are exhibited by Devanagari script too.

Obstruent geminates followed by /j/.

Sanskrit source words	Gloss	Assamese words	Orthography
sət.tjə	‘truth’	‘xɔit.tɔ’	(সত্য)
nepətt^{h}jə	‘background’	‘ne.poit.t^{h}ə’	(নেপথ্য)
vak.kjə	‘sentence’	‘baik.kə’	(বাক্য)
mrit.tjuː	‘death’	‘mrit.tu’	(মৃত্যু)
pəd.djə	‘poetry’	‘pɔid.dɔ’	(পদ্য)
gəd.djə	‘prose’	‘gɔid.dɔ’	(গদ্য)
nrit.tjə	‘dance’	‘nrit.tɔ’	(নৃত্য)
vəid.djə	‘healer’	‘boiddɔ’	(বৈদ্য)

The above examples bear close affinities between Devanagari and Assamese script in terms of the representation of the geminated patterns in non geminated forms. In the surface form of Assamese, the liquids /j/ gets dropped and a process of metathesis occurs in the form of an insertion of a vowel before the geminates. In other words it can be argued that the [high or [+ATR] element of /j/ gets metathesised and surfaces as a high front vowel forming a diphthong with the vowel in the preceding syllable (Dutta, 2018). In addition, the schwas are replaced by full rounded vowel although it does not have any orthographic representation in Assamese script. Consider the example:

kabbjə (sanskrit root) ----- kaibbɔ (Assamese) ‘epic’ (Assamese Orthography কাব্য)
(Phonetic Output কাইব্ব)

For explaining this process of metathesis the following the constraints need to be invoked within Otimality theory (Prince and Smolensy 1993):

*CCG: Ban geminates before glides.
MAX C: The segment in the input must have a corresponding segment in the output.
Linearity: The linear order of segments in the output be faithful to the linear order of segments in the input (Table 5).

Table 5 OT tableau for deriving the optimal candidate 'kaibbɔ'

/kabbjə/	*CCG	MAX C	LINEARITY
a)kabbjə	*!		
b)kabbɔ		*!	
c)☞kaibbɔ			*

Here from the table, we can say that to get the optimal candidate /kaibbɔ/ in Assamese the f constraint has to be ranked in the following fashion:

*CCG>> MAX-C >> LINEARITY

It is interesting to observe that although in spoken phonetic form, Assamese phonology exhibits such patterns it has no equivalent orthographic correspondence which might lead to the spelling errors on the part of the new learners of Assamese. Hence, I think the Assamese learners should be encouraged to write 'কাইব্ব' (kaibbɔ) rather than কাব্য (kabbjɔ). If there is no one to one match between phoneme and grapheme it leads to confusion among the learners and stands as an impediment to the whole process of literacy acquisition.

Obstruent geminates followed by /w/.

Sanskrit root words	Gloss	Assamese words	Ass. Ortho.	Phonetic output
təttwə	'theory'	tɔttɔ	(তত্ত্ব)	(তত্ত)
məhəttwə	'significance'	mɔhɔttɔ	(মহত্ত্ব)	(মহত্ত)
sətittwə	'chastity'	xotittɔ	(সতীত্ত্ব)	(সতীত্ত)
əsswə	'horse'	ɔssɔ	(অশ্ব)	(অষ্চ)

However, it is seen that this metathesis does not occur in the case of obstruent geminates followed by /w/. We need one more markedness constraint to address the issue of obstruent geminates followed by approximant 'w'. In such cases 'w' gets dropped in the surface form and we don't get any process of metathesis operating here. Actually, what happens is a consequence of the fact that metathesis is not possible because Assamese does not allow / ɔw / or /Vu/ diphthongs. Therefore, *V[back, round] glides are not licensed in Assamese and ranked high in the constraint hierarchy. Also, complex onsets of the type *Cw are disallowed and ranked high. The only option left is to link part of the features of the glide with the preceding vowel i.e., [+round] which will avoid a Max violation of not parsing the glide in 'mɔhɔttɔ' (Table 6).

Consider the following word for the sake of illustration:

məhəttwə (Sanskrit root) ------- mɔhɔttɔ (Assamese) 'significance'

Here, we can conclude our analysis with the constraint ranking in shown below:

Table 6 OT tableau for deriving the optimal candidate

‘mɔhɔttɔ’/məhəttwə/	*V[back,round]glides	*Cw	Max Round >> Max W	Linearity
mɔhɔttwɔ		*!	*	
mɔhɔwttɔ	*!		*	*
☞mɔhɔttɔ				*

*V[back, round]glides, *Cw and *Long VG Round >> Max Round>>Max W and Linearity.

However, the phenomenon of dropping of ‘w’ is not manifested, yet Assamese orthography maintains the tradition of Brahmi script. It can also cause issues of confusion among the learners due to the mismatch between Assamese phoneme and akshara.

This phenomenon of gemination shows different repair strategies Assamese phonology adopts in order to incorporate borrowed syllables and segments in to its native phonotactic domains. Such kind of analysis can give us insights in to the specific segmental patterns that loanwords exhibit. The patterns which the languages display are not random but very systematic. However, such systemic patterns and scientific insights are absent in the orthographic mechanism representing the process of morpheme internal gemination process attested with obstruents followed by glides and approximants in Assamese phonology. Gemination of consonants preceeding approximants and liquids on account of a phonological process need to be represented as the process is quite predictable and the representation of double consonants in some cases will prove to be cumbersome as in the instances /পুত্ৰ/ instead of /পুত্ৰ/ (son) and /বাক্ক্য/ instead of /বাক্য/ (sentence). However, as we have discussed in this morpheme internal gemination process the liquids /j/ gets dropped in the surface form and a process of metathesis occurs in the form of an insertion of a vowel before the geminates. Hence the phonetic output should have correspondence with the writing. The ideal writing should be / বাইক্ক/ ‘sentence’. Otherwise the constraint rankings within OT grammar will not suffice to the representation in writing. In a similar process however, this metathesis does not occur in the case of obstruent geminates followed by /w/. Hence, məhəttwə ‘significance’ in Sanskrit becomes ‘mɔhɔttɔ’ in spoken Assamese, but is not manifested in writing. Assamese orthography still abides by the principle of Devanagari script. Hence for the sake of convenience of the learners and readers in Assamese this word should be written as /মহত্ত/. The similar form should be used for other words too with such pattern i.e., /সতীত্ত/ ‘chastity’ /সাত্তিক/ ‘pious’ and this writing pattern will be more consistent with the phonological patterns of the language. Assamese orthography seems to be very much indebted to the Devanagari script patterns as far as the representation of the morpheme internal gemination process is concerned. I think that educators and policy makers should make an attempt to revise Assamese orthographic shapes to enhance the correspondence between graphemes and phonemes. In other words, it should be our constant endeavor to regularize the Assamese spelling system for the convenience of the learners.

Loanword Phonology and Orthography in Assamese

Loan word phonology is a very fascinating process in the realm of segmental change and segmental behavior and it is to be noted here that the patterns displayed by the loan words are always systematic and not random. In the process of globalization, lots of English words have become the part and parcel of Assamese vocabulary as a result English sound system has been modified in accordance with the principles and rules of Assamese phonotactic mechanism

For instance, in English /m/, /n/, /l/, /r/ etc. serve as syllabic consonants. Since Assamese phonology does not allow syllabic consonants hence a vowel is epenthesized between the consonants as a repair strategy (Dutta, 2016). Some examples are provided below (Table 7).

The reason behind the occurrence of such forms can be brought in to scrutiny by taking in to our consideration some constraints in Optimality Theory (OT).

*Peak [+consonantal]: Avoid consonants which act as peak or syllabic consonants in a syllable.

DEP [IO]: output segments must have input correspondents.

*OR: words can't end with an obstruent followed by a sonorant.

ALIGN R: The right edge of a grammatical word coincides with the right edge of a syllable.

Contiguity: Avoid epenthetic segment.

Consider the following tableau in OT model (Table 8).

Here the optimal candidate is (a) tebul as it violates only one constraint contiguity. The candidate (b) teibl vilotes *Peak [+consonantal] which is the highest ranked constraint in Assamese phonology. In the same way, the candidate (c) teibl violates *OR constraint and the candidate (d) violates both DEP-IO and ALIGN R constraint.

To get this optimal candidate tebul the ranking of the constraints can be categorized in the following way:

*Peak [+consonantal]>> *OR>> DEP-IO>> ALIGN R>> Contiguity

This phonological process is very much supported by the orthographic system in Assamese.

tebul: টেবুল here, ু refers to the epenthetic vowel sound /u/.

However, in Assamese orthography the grapheme representing the low mid rounded vowel /ɔ / is missing in the word internal and final position.

pɛsɔn: পেচন here, in this example, there is no grapheme in Assamese to represent the low mid rounded vowel.

What we have seen from the orthographic representation of the English loanwords in Assamese, it is clear that there is relatively a higher degree of correspondence between the phonemes and graphemes. There is no space for arbitrary writing mechanism to represent the English loanwords. The graphemic and phonological consistency is more visible in the English loanwords than in the Sanskrit loanwords.

Table 7 English loan words in Assamese

Phonetic form in English	English orthographic representation	Phonetic form in Assamese	Assamese orthographic representation
/teibl/	<table>	/tebul/	টেবুল
/botl/	<bottle>	/bɔtɔl/	বটল
/kotn/	<cotton>	/kɔtɔn/	কটন
/pædl/	<paddle>	/pɛdɛl/	পেডেল
/sadn/	<sudden>	/sadɛn/	চাদেন
/tækl/	<tackle>	/tɛkul/	টেকুল
/ga:gl/	<gargle>	/gargul/	গাৰগুল
/kazn/	<cousin>	/kazin/	কাজিন
/chænl/	<channel>	/senel/	চেনেল
/hospitl/	<hospital>	/hospital/	হস্পিতাল
/neison/	<nation>	/nɛsɔn/	নেচন/নেশ্ব্যন
/pæʃn/	<passion>	/pɛsɔn/	পেচন
/babl/	<babble>	/babul/	বাবুল

Table 8 OT tableau for deriving the optimal candidate ‘tebul’

Teibl	*Peak	*OR	DEP-IO	ALIGN R	Contiguity
☞ (a)tebul					*
(b) teibl	*!				
(c) teibl		*!			
(d) teblu			*	*!	

Any reforms pertaining to the existing orthographic system of a languages is bound to be based on the understanding of the basic tenets of the writing system and their correlation with the linguistic and nonlinguistic aspects of cognition. It is always desirable that the relation between spelling and speech in Assamese is consistent and learner friendly, not arbitrary. Interestingly loanwords derived from English are regularized in the course of time and they have become part and parcel of Assamese vocabulary and the English sounds and phonotactic patterns of these words have undergone transformation in accordance with the principles of Assamese phonological patterns. The orthography representing these words in Assamese are very much consistent with the phonological principles.

Conclusion

This chapter presents the intrinsic relationship between speech and writing, which are two distinct output mechanisms of human language with reference to synchronic Assamese speech sounds and Assamese script which has evolved from Brahmi script in the course of time through various stages in history. As Coulmas (1997) has pointed out that speech is intrinsic to language learning with a locus in the phonetic module and the art of learning writing relies on both linguistic and non linguistic aspects. Here, I have pointed out the mismatch between Assamese aksharas and Assamese speech sounds with reference to different phonological processes such as nasal assimilation, gemination and loanwords. Hence the main crux of the chapter lies in displaying the relationship between the phonology of Assamese and its orthography and how this relationship leads to suggestions for reforms in the orthography. As shown here, it is clear that learner based needs and confusion faced by the learners are very much associated with this relationship. Hence, these aksharas that represent palatal and retroflex sounds could be dropped from the orthography of Assamese to make the system more systematic and easy to learn keeping in consonance the practical needs and expectations of the learners and these sounds can be replaced by alveolar sounds which are available in the phonemic inventory of Assamese.

References

Bhattacharjee, S. (2013). *Axomiya lipitatto adhyayan* (A study of Assamese orthography). Chandra Prakashan, Guwahati, India

Coulmas, F. (Ed.). (1997). *The Blackwell encyclopedia of writing system*. Oxford, UK: Blackwell.

Dutta, B. (2003). Non-standard forms of Assamese: Their socio-cultural role. In M. Miri (Ed.), *Linguistic situation in northeast India* (2nd ed., pp. 101–110). New Delhi, India: Concept Publishing.

Dutta, H. (2016). *English loans in Assamese: A phonological account*. The IAFOR International conference on language learning, Dubai official conference proceeding.

Dutta, H. (2017). *Phonological strength relations and the patterning of segmental speech sounds with reference to Indian languages*. Linguistic Editions: 109. Lincom GmbH, Germany.

Dutta, H. (2018). Gemination and segmental patterns with reference to Sanskrit and Assamese: an OT account. *Lingua, 203*, 102–110.

Goswami, G. C. (1966). *Assamese phonology*. Pune, India: Deccan College Press.

Goswami, G. C., & Tamuli, J. (2003). Asamiya. In G. Cardona & D. Jain (Eds.), *The Indo-Aryan languages*. New York: Routledge.

Goswami, U. (1987). *Axomiya Lipi* (Assamese script). Guwahati, India: Chandra Publications.

Kakati, B. (1941). *Assamese: Its formation and development*. Guwahati, India: Lawyers' Book Stall Publications.

Masica, C. (1991). *The Indo Aryan languages* (pp. 36–38). Cambridge/New York: Cambridge University Press.

Moral, D. (1997). North East India as a linguistic area (PDF). *Mon-Khmer Studies, 27*, 43–53.

Pandey, P. K. (2014). Akshara –to-sound rules for Hindi. *Writing systems Research, 6*(1), 54–72.

Princel, A., & Smolensky, A. (1993). *Optimality theory: Constraint and interation in generative grammar*. Malden, MA: Ms, Rutgers University. Later published as in 2004 from Blackwell Publishing.

Reading and Writing Sinhala

Marasinghe A. D. K. Wijaythilake and Rauno Parrila

Abstract Sinhala is one of the two official languages in Sri Lanka spoken by about 74% of the population. It belongs to the Indo-Aryan family of languages and is written with a distinct, highly cursive akshara script. The extensive Sinhala akshara set (600 plus) is mostly consistent from akshara to sounds, and the basic literacy rates in Sri Lanka are high. Advanced literacy skills, however, require extended study due to strong diglossia: Spoken Sinhala has been open to influences from Dravidian (mainly Tamil) and European (Portuguese, Dutch, and English) languages, but Literary Sinhala has mostly maintained the classic vocabulary, spelling, and grammar. As a result, spelling and writing are complicated by significant differences between the two forms. In this chapter, we will first describe both forms of Sinhala and then review the limited existing research on Sinhala literacy.

Keywords Sinhala · Phonology · Orthography · Diglossia · Reading development · Writing development

Original research presented in this chapter is from the PhD dissertation the first author completed under the supervision of the second author. We would like to thank Dr. Dileep Chandralal, Dr. Sonali Nag, Dr. David Share, Dr. Manjula Waniganayake and the editors for their helpful comments on earlier versions of this chapter. Correspondence regarding this chapter can be sent to Dr. Parrila at rauno.parrila@mq.edu.au

M. A. D. K. Wijaythilake
Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada
e-mail: wijayath@ualberta.ca

R. Parrila (✉)
Department of Educational Studies, Macquarie University, NSW, Sydney, Australia
e-mail: rauno.parrila@mq.edu.au

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*,
Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_11

Introduction

Sinhala (or Sinhalese in older texts) is the language of Sinhala people and one of the two official languages in Sri Lanka (the other official language is Tamil). Currently, Sinhala is spoken by about 74% of Sri Lanka's 21 million people (Central Bank of Sri Lanka, 2016). Sinhala is the southernmost Indo-Aryan language and linguistically related to Northern Indian languages, such as Hindi, Marathi, and Punjabi (De Silva, 1979; Disanayaka, 1991; Gair, 2006). Its closest relative is Dhivedi of Maldive Islands, and both languages have been isolated from their relatives in Northern India for over two millennia. Spoken Sinhala has been exposed to other language families of the region, such as Dravidian and Malayo-Polynesian, as well as to colonial languages. For example, Spoken Sinhala shows a heavy influence of Dravidian languages on its phonology (Elizarenkova, 1972; Gair, 1985), and includes numerous lexical borrowings from Malaylam, Portuguese, Dutch, and English (Chandralal, 2010; Hettiaratchi, 1965).

Sinhala is written with an extensive akshara script that originates from Brahmi and belongs to the group of South Indic Dravidian scripts (Dharmadasa, 1967; Fernando, 1949, 1950; Gunawardhana, 1918; Nicholas, 1949; Wijayaratne, 1956). Sinhala script started to appear in inscriptions during the third and second centuries B.C.E., although the visual appearance of the symbols has changed considerably since then (Gair, 1982). The forms of Sinhala akshara are highly cursive and show an obvious influence of the early Grantha script of South India (Fernando, 1949, 1950; Jayarajan, 2006; see https://en.wikipedia.org/wiki/Grantha_alphabet for examples). The classical Literary Sinhala has remained relatively unchanged since the 1300s, resulting in both strong diglossia and in emergence of hybrid texts that combine classical writing and transcription of oral language. Despite the differences between the Spoken and Literary Sinhala, the two are not mutually unintelligible, and the basic literacy rates in Sri Lanka are high: According to the most recent UNESCO estimates, 92.6% of Sri Lankans are literate (UNESCO, 2015).

The remainder of this chapter is divided into two main parts. In the first part, we will first describe Sinhala in more detail to help the reader to understand the learning task a literacy learner faces. We will then discuss the differences between Spoken and Literary Sinhala. The second part of the chapter focuses on what we know about reading and writing acquisition and instruction in Sinhala. Finally, we conclude with a discussion of what we see as the most significant gaps in the existing research for it to guide literacy instruction and policy development.

Sinhala Language

Phonology

Sinhala phonology includes about 40 consonant segments and 20 vowel sounds, of which 24 consonants and 14 vowels (seven vowel qualities each with two lengths) are common today (Gair & Paolillo, 1997). Table 1 shows the phonetic characteristics of the vowels.

Manipulating the vowel length changes the meaning and the pronunciation of a word, and all vowels can be in both long and short forms. Sinhala has a three-way contrast between /ə/, /a/, /aa/, but there is no separate symbol for /ə/ in Sinhala orthography. However, the distribution of /ə/ and /a/ are largely predictable (Gair, 1998). Long /əː/ occurs only in loanwords like ‘sir’ /səːr/ and ‘shirt’ /ʃəːʈ/. The sound of the vowels /æ/ and /æː/ (similar to the sounds of the English words - hat and bad, respectively) are distinctive characteristics of Sinhala as these sounds are not present in Indo-Aryan or Dravidian languages (Gunasekara, 1981/1999; Ramanayake, 2006).

While many of the consonants are similar to neighboring languages, there is a set of four “half nasals” (Chandrallal, 2010) or pre-nasalized voiced stops — ^{m}b, ^{n}d̪, ^{n}ɖ, and ^{ŋ}g — specific to Sinhala (Karunatillake, 2004). These half nasals are sometimes treated as independent phonemes (e.g., Gair & Paolillo, 1997) and sometimes as consonant clusters with an extra short allophone (e.g., Chandrallal, 2010). They contrast with nasal + voiced stop consonant clusters mb, nd, nɖ and ng (Gair, 1970), and are treated as independent phonemes in Table 2 that shows the Spoken Sinhala consonant classification.

Fricative /f/ is bilabial and occurs only in borrowed words of English origin, such as /foʈo/. In Spoken Sinhala, speakers regularly substitute it with the bilabial stop /p/ that is native to Sinhala (Chandralal, 2010). Similarly, the palatal fricative /ʃ/ mainly occurs in borrowed Sanskrit words and is sometimes substituted by /s/. In Spoken Sinhala, /s/ is commonly replaced by /h/ combined with a vowel change. Thus, Literary Sinhala ‘eye’ /æsə/ and ‘rubbish’ /kasələ/ become /æhæ/ and /kahalə/, respectively, in Spoken Sinhala.

Table 1 Spoken Sinhala – vowel classification

	Front		Central		Back	
	Short	long	Short	long	Short	long
High	i	iː			u	uː
High-mid	e	eː				
Mid			ə	əː	o	oː
Low-mid	æ	æː				
Low			a	aː		

Table 2 Spoken Sinhala – consonant classification

	Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Stops - Voiceless	p	t̪		ʈ		k	
Voiced	b	d̪		ɖ		g	
Affricates - Voiceless					tʃ		
Voiced					dʒ		
Pre-nasalized voiced stop	ᵐb	ⁿ̪d̪		ⁿɖ		ᵑg	
Nasal	m		n		ɲ	ŋ	
Trill			r				
Lateral			l				
Fricatives	f	s			ʃ		h
Glides/semivowels	w				y		

Similar to the vowels, most consonants can be both long and short, with the long forms limited to the medial position. The distinction between single and double consonants is critical for distinguishing many common words, as in /atə/ and /attə/ (hand and branch) or /malə/ and /mallə/ (flower and bag). The phonetic length of consonants is indicated most of the time in writing by doubling the consonant symbol. Consonant clusters are common in Sinhala and can occur in initial and medial positions. In the initial position, the first consonant is pronounced short but in the medial position the first consonant is usually pronounced long (but orthographically represented by only one consonant letter, creating one of the few opaque features to the Sinhala orthography). Clusters longer than two consonants are also possible.

Syllabification and permissible syllable structure depend on whether the word is of local origin (*nishpanna*), borrowed from other languages in their (near) original form (*thathsama*), or originates from another language, but modified to be incorporated to Sinhala (*thadbhava*) (Weerasinghe, Wasala & Gamage, 2005). However, syllabification of almost all of the words borrowed from languages other than Sanskrit are consistent with the syllabification rules for words in the nishpanna category (Weerasinghe et al., 2005). There are four legal syllable structures – V, VC, CV and CVC – for nishpanna words. This can also be represented as (C)V(C) (Disanayaka, 1991). Syllabic structures of borrowed Sanskrit words can be more complex: (C)(C)(C)V(C)(C)(C). Syllabification of these words will be altered according to the ease of pronunciation. While oral syllabification is relatively uncomplicated, the written syllable does not always match the oral syllable as we will see below.

Finally, Sinhala suprasegmental features are relatively simple. Sinhala stress is weak and its placement is difficult to specify. In modern Spoken Sinhala, the first syllable is usually stressed. If the word is long and includes a long vowel, the syllable with the long vowel can be stressed. Sinhala uses also phrasal stress and allows stress to fall on any other words but verbs in a phrase (Wasala & Gamage, 2004–2007). The function of phrasal stress position is to emphasize a specific meaning of the sentence (e.g., "OUR father" versus "our FATHER"). Similarly, pitch is irrelevant for distinguishing between words, but is used to distinguishing utterance types.

Orthography

Sinhala is written from left-to-right with its own distinct, highly cursive semi-syllabic script. The basic orthographic unit is an akshara representing either a vowel (if there is no onset) or a consonant-vowel sequence. Each Sinhala akshara stands for a distinct sound and there is a high degree of regularity in akshara to phonology correspondences. In this sense, Sinhala is highly transparent for reading as the pronunciation of an akshara is almost always clear from its written form. The same is not necessarily true for writing as most phonological syllables can be represented by either a *śuddha* akshara or by a *miśra* akshara with only one usually considered correct. The śuddha akshara set is a subset of the miśra akshara set that contains all the akshara necessary to write classical Literary Sinhala. The current Spoken Sinhala can be represented fully by the śuddha akshara, but Literary Sinhala retains reference to special Sanskrit and Pali sounds, such as aspirates that have disappeared from Spoken Sinhala over time (Gair & Paolillo, 1997; Paolillo, 1997), captured by the miśra akshara.

Each Sinhala vowel has a distinct akshara (primary form) and one or more secondary forms consisting of diacritic signs. The primary form of the vowel is used only when the vowel is in the word-initial position (Gair, 1996). The secondary forms are attached to the consonants in all syllables starting with a consonant(s). Each consonant also has a distinct grapheme (primary form) that implies an inherent vowel /a/. When a consonant-vowel syllable with a vowel other than /a/ is written, the secondary sign of the vowel is attached to the consonant base and the /a/ is dropped from the pronunciation of the syllable. Similarly, some consonant clusters are formed by combining the independent written forms of consonants, sometimes in such a manner that a new visually complex and distinct akshara is created (see below for examples).

Table 3 shows the primary akshara and the secondary diacritic signs for Sinhala vowels, including two common dipthongs /ai/ and /au/ and the syllabic r. In turn, Table 4 shows Sinhala consonants (40). Note that where two akshara are present for a single consonant sound, the second usually represents the aspirated version that is no longer phonemically present but is preserved in writing.

Table 3 Sinhala vowel primary aksharas and secondary diacritics used in combination with a consonant

Vowel	Primary Symbol		Secondary diacritic	
	Short	Long	Short	Long
a	අ	ආ		ා
i	ඉ	ඊ	ි	ී
u	උ	ඌ	ු or ු	ූ or ූ
e	එ	ඒ	ෙ	ේ
ə				
o	ඔ	ඕ	ො	ෝ
æ	ඇ	ඈ	ැ	ෑ
ai		ඓ		ෛ
au		ඖ		ෞ
ru	ඍ	ඎ	ෘ	ෲ

Table 4 Sinhala consonants; all are pronounced with an inherent vowel not shown on the table

		Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Stops	Voiceless	/p/ ප	/t̪/ ත		/ʈ/ ට		/k/ ක	
	Aspirated	/p^{h}/ ඵ	/t̪^{h}/ ථ		/ʈ^{h}/ ඨ		/k^{h}/ /ඛ	
	Voiced	/b/ බ	/d̪/ ද		/ɖ/ ඩ		/g/ ග	
	Aspirated	/b^{h}/ භ	/d̪^{h}/ ධ		/ɖ^{h}/ ඪ		/g^{h}/ ඝ	
Affricates	Voiceless					/tʃ/ ච		
	Aspirated					/tʃ^{h}/ ඡ		
	Voiced					/dʒ/ ජ		
	Aspirated					/dʒ^{h}/ ඣ		
Pre-nasalized	Voiced stops	/^{m}b/ ඹ	/^{n}d̪/ ඳ		/^{n}ɖ/ ඬ		/^{ŋ}g/ ඟ	
Nasal		/m/ ම		/n/ න	/ɳ/ ණ	/ɲ/ ඥ or ඤ	/ŋ/ ඞ or ං	
Trills				/r/ ර				
Lateral				/l/ ල	/ɭ/ ළ			
Fricatives		/f/ ෆ	/s/ ස			/ʃ/ ශ /ʃ/ ෂ		/h/ හ or ඃ
Glides/semi-vowels		/w/ ව				/y/ ය		

Note: /n/ න is identified as dental in some sources

The Sinhala akshara is best understood as a symbol block that is quite unlike syllables formed from discrete letters in alphabetic languages. First, consonants and vowels are not treated equally. When writing CV and some CCV/CCCV akshara, the akshara is built up with the primary consonant symbol as the base and with all subsequent sounds in the symbol block written in the secondary form and ligatured to this base. The ligaturing position of the secondary forms can be to the right, left, above, below, and even around the base consonant. For example, the syllables /ka/, /ka:/, /ke/, /ku/, /ko/, and /ki/ in Sinhala are shown below with the base consonant in bold:

Example 1:

ක	**කා**	**කෙ**	**කු**	**කො**	**කි**
/ka/	/ka:/	/ke/	/ku/	/ko/	/ki/

Note the variable placements of the different diacritics: /a:/ is placed after the consonant, /e/ before the consonant, /u/ under the consonant, /o/ around the consonant, and /i/ above the consonant base that stays the same in all instances.

Two additional diacritics – called ‘hal lakuna’ or ‘hal kirima’ — are used for representing the pure consonant form. Consonants imply the inherent vowel ‘a’ unless marked for its absence by adding hal lakuna (Gair, 1996). For example, you add ් to the /ka/ consonant ක to make it ක් /k/. Figure 1 shows the two visual forms hal lakuna can take depending on the shape of the base consonant, with /p/ on the left, and /b/ on the right.

Clusters with two consonants (CCV) are the most common form and clusters with more than two consonants, while possible, are rare in Sinhala. Conjunct consonants can appear at the beginning, middle, or end of a word and more than one cluster is possible in a word. There are two types of conjuncts in Sinhala script, *combining* and *touching*. Example 2 shows two common combining conjuncts:

Fig. 1 The two different hal lakuna shapes. (From https://en.wikipedia.org/wiki/File:Sinhala-hal-kiriima.svg)

Example 2:

ප්	+	රා	=	ප්‍රා
/p/	+	/ra:/	=	/pra:/
ක්	+	ර	=	ක්‍ර
/k/	+	/ra/	=	/kra/

Example 3 shows two examples of touching akshara

Example 3:

ක්	+	ෂ	=	ක‍්ෂ
/k/	+	/ Ša/	=	/kŠa/
න්	+	ද	=	න‍්ද
/n/	+	/d̪a/	=	/nd̪a/

In each case, the visual form of the resulting akshara changes, and often substantially in the case of combined conjunct akshara, and the conjunct is never visually a pure combination of the two akshara. Furthermore, very frequent combinations are often written in one stroke, like /nd̪a/ above. When this is the case, the first consonant is not marked with hal lakuna. Forming the conjunct consonants is rule-governed and covered by instruction in Grade 4. However, children encounter frequently used conjunct consonants in their reading materials well before receiving formal reading instruction on them in school.

Moreover, as the Sinhala orthography signifies distinctions that no longer are phonemically recognized, attention to context is required for choosing the appropriate akshara in writing. For example, there are three sets of akshara in Sinhala in which the usage is determined by etymology: න and ණ for spoken /na/, ල and ළ for /la/, and ස, ෂ and ශ for /sa/. These akshara pairs sound the same in modern usage but differ in shape and the meaning of a word can change depending on the akshara in use.

Diglossia and Morphology

All Sinhala dialects are mutually intelligible as prominent differences remain restricted to the lexicon while phonological and morphological differences are less prominent between dialects (Chandralal, 2010; De Silva, 1979). The most important varietal distinction in Sinhala is diglossia: There are two major functional varieties in Sinhala, the Spoken and the Literary (De Silva, 1967; Gair, 1968, 1986; Geiger, 1938 Gunasekara, 1981/1999). Literary Sinhala dates back to the earliest language inscriptions, Sinhala Prakrit, of the third or second century BCE. According to De Silva (1967), current diglossia results from traditionalists resisting changes to Literary Sinhala, whereas Spoken Sinhala has undergone significant changes (see also Chandrallal, 2010). As a result, there now is a sharp distinction between the two varieties that differ in their form, structure, typical use, and functions.

Literary Sinhala is considered the 'higher' variety and it is generally used for all literary texts and published materials (Dharmadasa, 1967; Weerasinghe et al., 2005). Literary Sinhala is no one's first language and children have to learn Literary Sinhala in school. Spoken Sinhala is used by everyone in all societal levels as the language of communication in everyday life (Gair, 1968, 1982, also discusses Formal Spoken Sinhala as a separate variety used in formal situations, such as lectures, public speeches, sermons, and TV and radio news broadcasts). Differences between Literary and Spoken Sinhala exist at all levels of language structure, including morphology.

Sinhala *verbs* inflect for person, number, gender, tense, volition, and mood generally through suffixing but sometimes through vowel changes in the root (e.g., fronting the vowel). Both Literary and Spoken Sinhala inflect for two tenses - past tense and non-past tense. The non-past form can refer either to the present or the future. In Spoken Sinhala, time adverbials are used to express the future tense without changing the verb form (the verb form stays in indicative form). Thus:

Example 4:

a. Spoken Sinhala:

Mamə (I)/Eyaa (he/she)/Api (we)/Eyaala (they) hetə gedərə yanəwa – tomorrow home go

b. Literary Sinhala:

With time adverbial	*With krudantha suffix*
Ohu (he) hetə gedərə yayi	Ohu gedərə yanneya
Æyə (she) hetə gedərə yayi	Æyə gedərə yanniya
Api hetə gedərə yamu	Api gedərə yannemu
Owuhu hetə gedərə yati	Owuhu gedərə yannoya

Spoken Sinhala sentences differ only in the pronoun, but in Literary Sinhala the verb is also inflected for person and number. Further, Literary Sinhala allows use of *krudantha* forms of the verbs to express future tense by adding suffix –nne to the verb root prior to person and number suffixes, as shown in the right column above.

The active-passive morphological distinction is an important characteristic of verb morphology in Sinhala. Active or causative verbs are used to express volitional (conscious choice or decision) actions and passive forms (including a special causative-passive form made by adding first a causative suffix and then a passive suffix) are used to express non-volitional actions or events. Active form is usually identical to the verb stem while passive and causative forms are derived forms. Active forms, however, can be made to indicate additional moods (e.g., imperative, hortative, optative) by adding appropriate suffixes. Finally, non-verbal sentences (“this book new”) are common in Spoken Sinhala, but not allowed in Literary Sinhala.

In both Literary and Spoken Sinhala, *nouns* inflect for case, number (singular and plural) and definiteness (definite and indefinite) and the inflection is ruled by the natural gender which is the animate/inanimate distinction (Gair, 1967; Karunatillake, 1987). Animate nouns inflect for six cases (nominative/direct, accusative, dative, instrumental/ablative, locative/genitive, and vocative). Inanimate nouns inflect for four cases: nominative, dative, instrumental, and locative. Case distinctions are expressed by suffixes that are not always identical for Spoken and Literary Sinhala (see Table 5 for examples).

In Literary Sinhala, animate nouns are further divided into feminine and masculine, leading to a three-way gender system. In Spoken Sinhala, this three-way gender system is seen only in very formal speech when the speaker is more careful to use –ak alternate for the feminine indefinite nouns. For all other purposes, the spoken variety has only a two-way gender system in noun inflection. Both animate and inanimate common nouns are inflected further to mark for number. Table 5 shows

Table 5 Examples of the six cases for animate nouns

		Singular				Plural	
		Masculine		Feminine			
Case	Variety	Definite	Indefinite	Definite	Indefinite		
Nominative / direct	Literary	මිනිසා /minisa:/ (the man)	මිනිසෙක් /minisek/ (a man)	කෙල්ල /kellə/ (the girl)	කෙල්ලක් /kellak/ (a girl)	මිනිස්සු /minissu/ (men)	ගැහැණු /gæhænu/ (women)
	Spoken	/miniha/	/minihek/	/kellə/	/kellek/kellak/	/minissu/	/ gæ:nu/
Accusative	Literary	මිනිසා /minisa:/ (the man)	මිනිසකු /minisəku/ (the man)	කෙල්ල /kellə/ (the girl)	කෙල්ලක /kelləkə/ (a girl)	මිනිසුන් /minisun/ (the man)	ගැහැණුන් /gæhænun/ (women)
	Spoken	/minihawə/	/minihekwə/	/kelləwə/	/kellekwə/	/minissunwə/	/ gæ:nunwə/
Dative	Literary	මිනිසාට /minisa:ʈə/ (to the man)	මිනිසකුට /minisəkuʈə/ (to a man)	කෙල්ලට /kelləʈə/	කෙල්ලකට /kelləkəʈə/	මිනිසුන්ට /minisunʈə/	ගැහැණුන්ට /gæhænuntə/
	Spoken	/minihaʈə	/minihekuʈə/	/kelləʈə/	/kellekuʈə/	/minissunʈə/	/gæ:nuntə/
Instrumental/ ablative	Literary	මිනිසාගෙන් /minisa:gen/ (from the man)	මිනිසකුගෙන් /minisəkugen/ (from a man)	කෙල්ලගෙන් /kelləgen/	කෙල්ලකගෙන් /kelləkəgen/	මිනිසුන්ගෙන් /minisungen/	ගැහැණුන්ගෙන් /gæhænungen/
	Spoken	/minihagen/	/minihekugen/	/kelləgen/	/kellekgen/	/minissungen/	/gæ:nungen/
Locative/ genitive	Literary	මිනිසාගේ /minisa:ge:/ (the man's)	මිනිසකුගේ/ /minisəkuge:/ (a man's)	කෙල්ලගේ /kelləge:/	කෙල්ලකගේ /kelləkəge:/	මිනිසුන්ගේ /minisunge:/	ගැහැණුන්ගේ /gæhænunge:/
	Spoken	/minihage/	/minihekuge/	/kelləge/	/kellekge/	/minissunge/	/gæ:nunge/
Vocative	Literary	මිනිස /minisə/		කෙල්ල /kellə/		මිනිසුනේ/නි /minisune:/ni/	ගැහැණුනේ/නි /gæhænune:/ni/
	Spoken	/miniho/		/kelle/keli:/		/minissune/	/gæ:nune/

examples of noun inflections for a masculine and a feminine animate common noun separately for Literary and Spoken Sinhala.

As is clear from Table 5, case marking for animate nouns differs widely between Literary and Spoken Sinhala. Sinhala pronouns, which are in first, second, and third person, and mark for number, show similar differences between the two forms. However, for inanimate nouns, the differences are less noticeable. For example, /potə/ 'the book' has separate literary and spoken forms only in genitive definite form, whereas all the other forms are the same. A large number of nouns, both animate and inanimate, appear in either singular or plural forms in both varieties without any suffix. For example, /amma/ 'mother' and /thaththa/ 'father' are singular form kinship nouns that are identical to their noun stems. The plural forms of many animate nouns, such as /nari/ 'foxes' and /koti/ 'tigers,' are also identical with their noun stems. The respective examples of inanimate singular nouns are /ratə/ 'country' and /paarə/ 'road' and of inanimate plural nouns are /gas/ 'trees' and /putu/ 'chairs,' all of which are identical to their stems.

Finally, Sinhala allows for compounding, and there are a number of compound nouns made of three or more words. While shorter compound nouns are common in Spoken Sinhala, the use of longer compounds is mostly limited to Literary Sinhala. A compound is considered grammatically as a single word but often writers keep a space between the two components of a short compound; longer compounds are regularly written with spaces between the components. There is a fair amount of confusion among writers as to when to use spaces between compounds, and there is little consensus among grammarians and language users regarding the maintenance of the internal boundaries (Chandralal, 2010).

Learning to Read and Write Sinhala

Sri Lanka has provided a free primary, secondary, and tertiary education since 1940s. Compulsory education lasts 9 years from age 5 to 13 and practically all children in Sri Lanka attend primary education (UNESCO, 2015, estimates access to primary education at 98.8%). As a result, the average adult literacy rate in 2015 was reported at 93.2% (male literacy rate was 94.1% while female literacy rate was 92.4%; Central Bank of Sri Lanka Annual Report, 2016). Both youth (98.77%) and adult (92.6%) literacy rate estimates are higher than regional averages (UNESCO, 2015), and basic literacy problems are mostly limited to older adults. About 85% of adolescents complete secondary education, but access to tertiary education is limited and only about 21% of young adults access tertiary education (24% for females; 17% males). University access is limited to about 20,000 students per year, a number that equals roughly 15% of those who pass the General Certificate of Education Advanced Level examinations and are thus eligible for universities.

Despite the well-attended basic education system that performs better than most in the region, learning to read and write Sinhala is a prolonged process. Several of the features of Sinhala reviewed above – such as large akshara inventory, their visual complexity and nonlinear placement of diacritics, presence of phonemic contrasts in writing that no longer exist in speech, large differences between Spoken and Literary Sinhala – has created a literacy learning environment where basic word reading skills are relatively easy to acquire, whereas reading comprehension and writing skills seem to be more elusive for students. Further, the difference of being able to read and being able to write can be particularly pronounced, as can the differences between levels of writing skills. As research on literacy acquisition in Sinhala is sparse, we will complement the existing studies with Government statistics and with observations from Sri Lankan teacher informants. Where possible, we draw on yet unpublished data from our ongoing project that examines reading development and predictors of it across the first six grades in three relatively well-functioning schools in Sri Lanka.

Learning to Recognize Akshara and Words

According to Chandralal (2010), contemporary Literary Sinhala can be written with about 54 independent symbols combined with 18 diacritics, and 38 symbols are sufficient to represent the Spoken Sinhala. Altogether, the Sinhala writing system includes well over 600 akshara; our analysis of all language arts textbooks from grades 1 to 6 identified over 400 unique akshara that children are expected to learn (Wijaythilake & Parrila, unpublished data). This extensive symbol set creates a learning situation quite different from learning alphabetic orthographies: Compared to 20–30 symbols of most alphabetic orthographies, learning the 400–600 symbols in Sinhala is an extended process and seldom mastered by the end of the primary school education (see e.g., Das, Kumar, Bapi, Padakannaya, & Singh, 2010; Nag, 2007; Nag & Snowling, 2011; Nag, Treiman & Snowling, 2010; Tiwari, Nair & Krishnan, 2011, for similar examples from other akshara orthographies, and Daniels & Share, 2017, for discussion). To illustrate this in Sinhala, we recently asked children from grades 1 to 6 to name 80 akshara taken from their textbooks varying in frequency, type (V, Ca, CV, C, CCa, CCV), visual complexity, and few other characteristics assumed to impact akshara learning. Table 6 shows the performance levels across the grades.

Table 6 Akshara recognition across grades

Akshara Recognition (max = 80)				
Grade	Mean	SD	Minimum	Maximum
1	35.60	10.67	16	67
2	56.42	10.26	30	70
3	63.28	7.98	37	77
4	69.02	7.48	44	77
5	74.10	4.50	60	80
6	72.86	4.23	60	80

As the akshara used were taken from Language Arts textbooks across the grades, the older children were exposed to more of them than the younger children and the variability among the older children was reduced; the most mistakes older children made were with infrequent vowel akshara that are not needed to represent modern Spoken Sinhala. However, it is clear from Table 6 that children had difficulties recognizing less frequently presented akshara; similarly, akshara frequency has been shown to account for unique variance in akshara knowledge (Nag, Snowling, Quinlan, & Hulme, 2014) and word reading (Nag, 2007; Nag et al., 2010; Sircar & Nag, 2013; Tiwari et al., 2011; Vasanta, 2004) in other akshara orthographies. Akshara learning continues well beyond the elementary school and even Grade 6 Sinhala students have not fully mastered all types of akshara.

Further, given the consistency of the akshara-phonology relationships, we expected that being able to read words accurately would be closely associated with knowing the akshara in those words. Akshara recognition in this sample correlated .81, .77, and .60 with both word recognition and nonword recognition tests across grades 1–3, and the word and nonword recognition tasks correlated .94, .88, and .72, respectively, across grades 1–3 (the slight declines in these values reflect reduction in variability). Moreover, the lexical status of the items (word or nonword) did not matter for *accuracy* when roughly comparable (in terms of akshara frequency) materials were used – nonwords were read as accurately as words in all grades. However, words were read slightly faster than nonwords from Grade 2 onwards, indicating that at least some of the common words were likely recognized as whole units.

One possible reason for high dependency of word/nonword reading on akshara knowledge is that in Sri Lankan schools, reading and writing instruction follows a fixed sequence (also reflected in the Language Arts text books we analyzed) where the approach to teaching reading for the first 4 years is at the akshara level and not at the level of individual markers of phonemes. Grade 1 Sinhala Language Arts books cover the primary vowels, CV akshara with an inherent vowel, and CV akshara with a vowel other than the inherent vowel. In Grades 2 and 3, students are taught all the secondary forms of consonants. In Grade 4, consonant clusters are introduced formally for the first time. Only in Grade 5 are students explicitly taught to decompose CV and CCV akshara into phonemic markers, thus opening the door to a more combinatorial understanding of Sinhala akshara formation. From Grade 6 onwards, the practice of using clusters and ligaturing rules continues. Thus, the emphasis on the recall of whole akshara rather than its parts and ligaturing rules is a likely culprit for why children can for the most part only decipher whole akshara they already know. More research is needed to examine if the instruction in phonemes, diacritic marks, and ligaturing rules in Grade 5 increases children's ability to decipher hitherto unseen akshara.

The fixed sequence of instruction is naturally perturbed by the real world reading materials and vocabulary children are exposed to. Children's books and story books have a mixture of all akshara types and children's vocabulary contains a sizable number of words with complex akshara. For example, in Sinhala, the CCV akshara /kka/ and /mma/ are common in early text books in two-akshara words like /akka/

(sister) and /amma/ (mother), and children typically learn these akshara alongside the akshara with inherent vowels due to frequent exposure. Children who read frequently should therefore gain exposure to a more extensive akshara set, and we would expect this alone to improve their reading skills – in our data, time spent reading at home indeed correlated significantly with akshara knowledge even after controlling for grade, gender and phonological processing skills at both syllable and phoneme levels. We were not able to locate other studies examining home literacy practices in Sinhala, and the direction of effect needs to be established in future studies.

One obvious additional complexity with all akshara orthographies is the visual complexity of the symbols (Daniels & Share, 2017). Adding diacritic markers to akshara adds both phonemic complexity and visual complexity to the base consonants that themselves can already be visually complex (see Table 4). The more diacritic markers are attached to the base consonant the more complex the akshara becomes. Example 5 shows first a consonant with an inherent vowel (no additional diacritics), followed by the same consonant with a different vowel, and then with three examples of consonant clusters.

Example 5:

(a) /k/ with inherent vowel:	ක	/ka/
(b) /k/ with /i/:	කි	/ki/
(c) /kr/ with inherent vowel:	ක්‍ර	/kra/
(d) /kr/ with /i/:	ක්‍රි	/kri/
(e) /kr/ with /e/:	ක්‍රේ	/kre/

The extent to which the diacritic markers that are stacked together retain their individual visual identity can differ. In Sinhala, the individual consonants in a cluster can be partially or completely identifiable, as in Example 5, making them visually easier to locate. However, their visual identity can be substantially different from their primary form, as in Example 3 above, essentially requiring the children to learn a new akshara. Regardless, the resulting symbol is visually complex. There is preliminary evidence from other akshara orthographies that visual complexity affects akshara learning (e.g., Nag et al., 2014), and we know that words with consonant clusters (that *per se* are visually complex) are more difficult to recognize in Sinhala than words without them (Wijaythilake & Parrila, 2014). We are currently examining the item level characteristics that impact akshara recognition rates, and the preliminary data suggest that visual complexity (as rated by individuals not familiar with an akshara orthography) indeed affects the learning rate independently of the presence of consonant clusters.

The placement of secondary forms around the base consonants can also introduce nonlinearity to the left-to-right flow of symbols and occasionally a departure from the expected sequence of mappings between the akshara components and the phonemes as phoneme markers may not be located in a place that is consistent with their position in the sound sequence of the spoken syllable (see Example 1 above). Jayawardena and Winskel (2016) suggested that because the onset consonant can occur up to the third position in a written syllable, Sinhala readers have a heightened attentional response to the first three graphemic markers as opposed to only the first. Out of sequence orthography-to-phonology mappings have been shown to slow down akshara recognition in Kannada readers (Nag et al., 2010), and readers seem to make fewer errors with diacritic markers when the spatial order of the akshara in the script is consistent with the phonological order (Kandhadai & Sproat, 2010; Vaid & Gupta, 2002). Phoneme-grapheme sequence irregularity has a negative impact on gaining akshara competence for Kannada readers (Karanth, Mathew & Kurien, 2004; Padakannaya & Chaitra, 2002). Our count of all independent akshara in grades 1–6 Sinhala reading text books identified 332 unique akshara (out of 411) that were consonants with other vowels (CV) or consonant clusters (CCV). More than half of them included off-the-line phoneme markers and broke the phonology-orthography sequence match.

No existing studies have examined how phonology-orthography sequence irregularities affect Sinhala reading. Our preliminary data indicate that grapheme-phoneme sequence mismatching and the number of breaks from linearity of visual presentation are robust predictors of how accurately an akshara is recognized. Given that these mismatches also lead to phonological syllables being different from orthographic syllables (e.g., /am-ma/ versus a-mma in writing) and the importance of syllable awareness for learning akshara orthographies (Nag, 2007; Prakash, Rekha, Nigam, & Karanth, 1993), this is clearly an area of interest for future research. In sum, there are very few studies that have examined the influence of various akshara-specific factors on akshara knowledge and reading development and more studies are needed.

In terms of word reading, most languages written with akshara have a predominant CVCV(CV) word structure. For example, there are only a few one syllable words in Sinhala and most words consist of two, three, or four syllables. Words longer than four syllables are also present. Researchers examining word recognition in other akshara orthographies have reported that word length affects reading latencies and that longer words are more vulnerable to errors than shorter words (Karanth et al., 2004; Nag et al., 2010; Padakannaya & Chaitra, 2002). Further, only a few studies have examined the effect of word frequency on word recognition, and the results are mixed (Karanth et al., 2004; Padakannaya & Chaitra, 2002); one possible interpretation of these results is that significant differences are seen between high and low frequency words only when the words have complex akshara, such as consonant clusters. If this is the case, future research clearly needs to examine whether the word frequency effect is separable from various akshara effects, including their frequency. To conclude, we hope that even on the basis of this very brief and incomplete review of akshara recognition and word reading, it is apparent that the kinds of questions reading research in akshara orthographies needs to answer are often very different from the questions imported from alphabetic orthographies.

Learning to Write Sinhala

Although Sri Lankan students' school enrolment and retention rates are high and the basic literacy rates are above regional averages, Sinhala students' writing performance seems to be a cause of continuous concern. For example, recent national exam results for Grades 4 and 5 (Department of Examination, Sri Lanka, 2015) identified writing as the main area of concern, similar to two earlier reports. In both grades, about 6% of all students did not write anything at all, and the majority of students who did write a response failed to score a point in essay writing. Both spelling and grammatical errors were common in both grades (see also Jhingran, 2011, and Sandyanganie, Jeewandara, & Perera, 2016) and akshara with diacritic markers and consonant clusters were particularly difficult. According to our teacher informants, correct spelling is particularly difficult for children when they encounter the three pairs of akshara (see above) that sound the same in Spoken Sinhala and whose usage is based on the etymology.

Some additional differences between Spoken and Literary Sinhala seem to be problematic as well. The major difference between the two varieties is the absence of subject-verb agreement in the spoken form. As a result, subject-verb agreement was not kept by 15% of Grade 5 students in the national exams (Department of Examination, Sri Lanka, 2015), and the teacher informants suggested that this problem is common throughout elementary school. The use of colloquial language in writing was another problem noted in the national exams, although the Literary and Spoken varieties are supposed to be mutually exclusive domains (De Silva, 1976). When De Silva (1976) analyzed children's free writing samples from Tamil, Kannada, and Sinhala Grade 4 and 5 classrooms, he noted first that in about the same number of writing samples, Sinhala students made twice the errors of Tamil students and 25% more than Kannada students. Second, a higher proportion of errors were grammar errors, including use of colloquial grammar (25% in Sinhala compared to 7% in Tamil and 15% in Kannada); inflection errors were common in both Tamil and Sinhala, and spelling errors made up about half of all errors in all three orthographies. The somewhat poorer writing performance of Sinhala children was perhaps more striking if we keep in mind that their basic literacy levels were (and still are; Census of India, 2011; UNESCO, 2015) expected to be higher than those of Tamil speaking or Kannada speaking children in India.

Our teacher informants indicated that many of the common spelling errors involve the spelling of long vowels and aspirated consonants. Although long vowels are marked in writing using the appropriate diacritic marks, long vowels are rarely pronounced in full in colloquial Sinhala. Aspirated consonants are met in the books and written in Literary Sinhala when words are of Sanskrit or Pali origin, but replaced by their unaspirated counter parts in Spoken Sinhala. Lack of phonological familiarity is likely to lead to problems in marking long vowels and using the right akshara for aspirated sounds in writing (see Nag et al., 2010, for similar difficulties in Kannada).

Finally, writing problems may not be limited to children who are learning to write. De Silva (1976) surveyed social expectations of literacy in Tamil, Kannada and Sinhala adults. The respondents had all completed their secondary school education and ranged from the first-year university students to professionals holding doctoral degrees; in De Silva's words, "they had all gone through the machinery for acquiring sufficient literacy as required by their communities." The survey first showed an interesting split between Sinhalese speaking respondents in their definition of literacy where 30% (termed 'purists' by De Silva) agreed with the statement "Literacy is the ability to read and write correctly" but 50% chose the statement "Literacy is the mere ability to read and write." Perhaps more interestingly, when De Silva analyzed the quality of the written responses to the follow up questions, he found that the purists made on average one grammatical and one spelling error per sentence produced, whereas the less purist (but no less educated) respondents made on average three grammar errors and one spelling error per sentence. For the sake of comparison, similarly educated Kannada respondents made similar numbers of spelling errors but only .5 grammar errors per sentence, and Tamil respondents made 1 grammatical error and .3 spelling errors per sentence written. While the writing samples examined were not extensive (190 sentences in Sinhala), these results clearly indicated that writing grammatically correct Sinhala remained a lifelong learning task, likely due to a variety of differences that separate the Spoken (and familiar) Sinhala from its Literary counterpart. Given that those differences are no lesser now and that students' writing performance remains a concern (e.g., Ministry of Education, Sri Lanka, 2016), we suspect that De Silva's conclusions would still stand.

In sum, learning to write Literary Sinhala is a demanding task and a continuous source of concern in Sri Lanka. Somewhat surprisingly, we were not able to locate any studies on Sinhala spelling and writing instruction. There clearly is a need for comparative studies examining different approaches to writing instruction, as well for studies examining what constitutes sufficient functional writing skills (as opposed to correct writing).

Conclusion

We hope that the above review of the characteristics of Sinhala (together with the other chapters in this volume) convinces even the most skeptical reader that the learning demands that akshara orthographies impose on the learners are both similar and different from those that learners of alphabetic or morphographic orthographies face (see also Daniels & Share, 2017; Nag, 2017). In terms of similarities, reading is a language activity in all orthographies, and when the written language is as different from the everyday oral language as in Sinhala, the learners face additional challenges. These challenges, however, are not limited to akshara orhographies as the growing literature on diglossia in Arabic, for example, attests (see e.g., Abu-Rabia & Taha, 2006; Saiegh-Haddad, 2003). Similarly, the visual complexity of the orthographic symbols is arguably shared with many morphographic writing systems, such as Chinese or Japanese kanji. What may be specific to akshara orthographies is the mapping of oral language to written symbols on multiple phonological grain-sizes (Nag, 2017) with little morphological support – most of the 600+ akshara in Sinhala, for example, do not map directly to morphemes (Chandralal, 2010). While learning syllable-to-symbol mappings may *per se* be relatively straightforward (see e.g., Inoue, Georgiou, Muroya, Maekawa & Parrila, 2017, for Japanese kana learning), complexities abound when syllable and phoneme mappings are combined. For example, vowels and consonants in akshara orthographies are not on an equal structural footing (Share & Daniels, 2016) due to the fact that the overwhelming majority of vowels are not full-sized, and the consonants frequently represent more than just the consonant sound. Even when the orthographic symbol, the akshara, is fully consistent and follows all the rules of ligaturing, the learner has to pay simultaneous attention to syllable and phoneme level cues, some of which negate the pronunciation of a component sound (the inherent vowel) rather than elicit it, to decipher the final pronunciation. This, we argue, is a different process, and currently very poorly understood process, than anything we know from alphabetic orthographies, including the English deciphering process that at times requires simultaneous attention to phonological, morphological, and etymological cues.

The ability to segment visually and phonologically complex orthographic symbols that you may have never seen before into their separate phonemic parts is the major challenge in learning to read and write in akshara orthographies (Nag, 2014; Nag et al., 2010). However, it is also clear from Table 6 that the size of the symbol set is a challenge in itself in Sinhala. Nag (2007) distinguished between *extensive* orthographies (those with large symbol inventories, such as akshara orthographies and Chinese) and *contained* orthographies that use more limited symbol inventories, such as alphabets. As learning the symbol set in an extensive Sinhala orthography is a demanding process by itself, it is likely that the cognitive skills employed during this process are at least partly different from those needed to learn a contained orthography. Elsewhere, we have called this the *orthographic breadth hypothesis* (Inoue et al., 2017) and noted that very few studies have investigated how all the additional complexities extensive orthographies bring to the learning task affect the learning process and the skills supporting it. We suspect that it is these complexities, together with diglossia, that account for poorer than expected reading comprehension levels in Sinhalese children (Sandyanganie et al., 2016). To simplify, they have to manage akshara recognition AND word recognition before they can allocate resources to comprehending of what is being read.

We would like to end this chapter – perhaps not surprisingly – with a call for research. While the number of research studies conducted in akshara orthographies is steadily rising, as is evidenced by this volume as well, we need more and larger studies to guide educational policy and literacy instruction. More scientific studies are needed in all aspects of literacy acquisition and development in Sinhala and other akshara orthographies focusing on akshara knowledge, word recognition, phonological and other cognitive skills, reading comprehension, vocabulary, and visual processing skills (Nag, 2007; Nag & Snowling, 2011, 2012; Nakamura, Koda, & Joshi, 2014). While we wait for the results to arrive, we urgently need well-controlled educational experiments. The scarcity of experimental studies allows educational innovations to proceed in a scientific vacuum: To give an example, the most recent changes in the Sri Lankan language arts curriculum increased the practice of akshara writing in elementary school (likely a good thing), and moved the explicit instruction on phoneme markers and ligaturing rules from Grade 5 to Grade 6, a theoretically counterintuitive (see e.g., Nag, 2017) and likely a poor decision given the preliminary evidence from Bengali that moving this instruction to earlier grades benefits reading and writing acquisition (Nag, 2014; Sircar & Nag, 2013). However, in the absence of well-controlled large scale educational experiments, it is impossible to build anything resembling an "evidence-based" policy and practice at the moment.

Our call for more research in akshara orthographies also concerns the progress of reading science. While comparisons of cognitive-linguistic and reading processes across alphabetic languages have produced interesting results and guided policy and instructional practice (at least in many European countries), it is time to broaden the pool of languages and scripts that are typically considered to include akshara orthographies. Nag (2017) reviews evidence suggesting that learning to read in akshara orthographies builds on many of the same cognitive–linguistic processes (phonological awareness, rapid naming, phonological memory) needed to learn an alphabetic orthography, but her review also suggests that additional processes are needed (for example, visual processes), and that the importance of syllable versus phoneme awareness may be flipped in akshara orthographies compared to alphabetic orthographies. Needless to say, many other processes that can affect literacy learning remain unstudied, and likely undiscovered. Further, the mechanisms of effect are poorly understood, leaving open the possibility that statistically similar relationships are based on different effective mechanisms. In sum, the bar for claiming universality of any theory of reading or writing has to be set higher, and testing the universality has to include studies in akshara orthographies.

References

Abu-Rabia, S., & Taha, H. (2006). Reading in Arabic orthography: Characteristics, research findings, and assessment. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 321–338). Mahwah, NJ: Lawrence Erlbaum Associates.

Census of India. (2011). *Provisional population totals-Karnataka-Data sheet*. Retrieved from http://www.censusindia.gov.in/2011-prov-results/census2011_PPT_paper1.html

Central Bank of Sri Lanka. (2016). *Sri Lanka socio-economic data*. Colombo, Sri Lanka: Central Bank of Sri Lanka.

Chandralal, D. (2010). *Sinhala*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Daniels, P. T., & Share, D. L. (2017). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading, 22*, 101–116.

Das, T., Kumar, U., Bapi, R. S., Padakannaya, P., & Singh, N. C. (2010). Neural representation of an alphasyllabary – The story of Devanagari. *Current Science, 97*, 1033–1038.

Department of Examination, Sri Lanka. (2015). *Reviewing of performance at grade 5 scholarship examination*. Retrieved from http://www.doenets.lk/exam/docs/comm/Grade%2005%20-%20 2015%20Symposium.pdf

De Silva, M. W. S. (1967). Effects of purism on the evolution of written language. *Linguistics, 36*, 5–17.

De Silva, M. W. S. (1976). *Diglossia and literacy*. Mysore, India: Central Institute of Indian Languages.

De Silva, M. W. S. (1979). *Sinhalese and other island languages in South Asia*. Tübingen, Germany: Gunther Narr Verla.

Dharmadasa, K. N. O. (1967). *Spoken and written Sinhalese: A contrastive study*. (Unpublished Master of Philosophy dissertation). University of York, UK.

Disanayaka, J. B. (1991). *The structure of spoken Sinhala*. Maharagama, Sri Lanka: National Institute of Education.

Elizarenkova, T. (1972). Influence of Dravidian phonological system on Sinhalese. *International Journal of Dravidian Linguistics, 1*, 126–137.

Fernando, P. E. E. (1949). Palaeographical development of the Brahmi script in Ceylon from 3rd century B.C. to 7th century A. D. *University of Ceylon Review, 7*, 282–301.

Fernando, P. E. E. (1950). Development of the Sinhalese script from 8th century A. D. to 15th century A. D. *University of Ceylon Review, 8*, 222–243.

Gair, J. W. (1967). Colloquial Sinhala inflectional categories and parts of speech. *Indian Linguistics, 27*, 31–45.

Gair, J. W. (1968). Sinhalese diglossia. *Anthropological Linguistics, 10*, 1–15.

Gair, J. W. (1970). *Colloquial Sinhalese clause structures*. The Hague, The Netherlands: Mouton.

Gair, J. W. (1982). Sinhala, an Indo-Aryan isolate. *South Asian Review, 6*, 51–64.

Gair, J. W. (1985). How Dravidianized was Sinhala phonology? Some conclusions and cautions. In V. Z. Acson & R. L. Leed (Eds.), *Festschrift for Gordon H. Fairbanks, Oceanic Linguistics Special Publication* (pp. 37–55). Retrieved from http://www.jstor.org/stable/pdf/20006708.pdf

Gair, J. W. (1986). Sinhala diglossia revisited, or diglossia dies hard. In B. Krishnamurti, C. P. Masica, & A. K. Sinha (Eds.), *South Asian languages: Structure, convergence and diglossia* (pp. 322–336). Delhi, India: Motilal Banarsidass (Reprinted in Gair, 1998a, pp. 224–236).

Gair, J. W. (1996). Sinhala writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 408–412). New York, NY/Oxford, UK: Oxford University Press.

Gair, J. W. (1998). Selections from the verb in Sinhala, with some preliminary remarks on Dravidianization. In J. W. Gair & B. C. Lust (Eds.), *Studies in South Asian linguistics, Sinhala and other South Asian languages* (pp. 202–209). New York, NY/Oxford, UK: Oxford University Press.

Gair, J. W. (2006). Sinhala. In *Concise encyclopedia of languages of the world* (Vol. 1, pp. 964–968). Oxford, UK: Elsevier Ltd.

Gair, J. W., & Paolillo, J. C. (1997). *Sinhala*. Munich, Germany: Lincom Europa.

Geiger, W. (1938). *A grammar of the Sinhalese language*. Colombo, Sri Lanka: Royal Asiatic Society.

Gunasekara, A. M. (1981/1999). *A comprehensive grammar of the Sinhalese language*. New Delhi, India: Asian Educational Services.

Gunawardhana, W. F. (1918). *The origin of the Sinhalese language*. Colombo, Sri Lanka: W. E. Bastian & Co.

Hettiaratchi, D. E. (1965). Influence of Portuguese on the Sinhala language. *Journal of the Ceylon Branch of the Royal Asiatic Society, 9*, 229–238.

Inoue, T., Georgiou, G., Muroya, N., Maekawa, H., & Parrila, R. (2017). Cognitive predictors of literacy acquisition in syllabic Hiragana and morphographic Kanji. *Reading & Writing, 30*, 1335–1360.

Jayarajan, P. M. (2006). History of the evolution of the Sinhala alphabet (2nd ed.). Michigan: Colombo Apothecaries' Company, Ltd.

Jayawardena, R., & Winskel, H. (2016). Assessing the modified receptive field (MRF) theory: Evidence from Sinhalese-English bilinguals. *Acta Psychologica, 171*, 65–71.

Jhingran, D. (2011). *Reading failure in early primary grades: A serious challenge to equity in primary education* [PowerPoint slides]. Retrieved from www.create-rpc.org/pdf_documents/Delhi2011D.Jhingran.pdf

Kandhadai, P., & Sproat, R. (2010). Impact of spatial ordering of graphemes in alphasyllabic scripts on phonemic awareness in Indic languages. *Writing Systems Research, 2*, 105–116.

Karanth, P., Mathew, A., & Kurien, P. (2004). Orthography and reading speed: Data from native readers of Kannada. *Reading & Writing, 17*, 101–120.

Karunatillake, W. S. (1987). Category of gender in Sinhala. In M. Hiran & F. Jayasoriya (Eds.), *Gate Mudaliyar W. F. Gunawardhana conmemoration volume*. Colombo, Sri Lanka: Sri Lanka.

Karunatillake, W. S. (2004). *An introduction to spoken Sinhala* (5th ed.). Colombo, Sri Lanka: M.D. Gunasena & Ltd.

Ministry of Education, Sri Lanka (2016). *Annual performance report.* Retrieved from http://www.moe.gov.lk/english/images/publications/2018/Anual_performance_report/anual_performance_report_e.pdf

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*, 7–22.

Nag, S. (2014). Akshara-phonology mappings: The common yet uncommon case of the consonant cluster. *Writing Systems Research, 6*, 105–119.

Nag, S. (2017). Learning to read alphasyllabaries. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 78–95). Amsterdam, The Netherlands: Benjamins.

Nag, S., & Snowling, M. J. (2011). Cognitive profiles of poor readers of Kannada. *Reading and Writing, 24*, 657–676.

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a language universal theory of learning to read. *Scientific Studies of Reading, 16*, 404–423.

Nag, S., Snowling, M. J., Quinlan, P., & Hulme, C. (2014). Child and symbol factors in learning to read a visually complex writing. *Scientific Studies in Reading, 18*, 309–324.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*, 41–52.

Nakamura, P. R., Koda, K., & Joshi, R. M. (2014). Biliteracy acquisition in Kannada and English: A developmental study. *Writing Systems Research, 6*, 132–147.

Nicholas, C. W. (1949). Palaeographical development of the Brahmi script in Ceylon from the 3rd century B. C. to the 7th century A. C. *UCR, 7*, 60–65.

Padakannaya, P., & Chaitra, R. (2002). Effect of word frequency and lexicality on reading speed. *The Second International Conference on Neurology, Language and Cognition*. Institute for Communicative and Cognitive Neurosciences, Cochin, India.

Paolillo, J. C. (1997). Sinhala diglossic variation: Continuous or discrete? *Language in Society, 26*, 269–296.

Prakash, P., Rekha, D., Nigam, R., & Karanth, P. (1993). Phonological awareness, orthography and literacy. In R. Scholes (Ed.), *Literacy and language analysis* (pp. 55–70). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ramanayake, R. (2006). Characteristics of Sinhala pronunciation [PowerPoint slides]. Retrieved from https://www.w3.org/2006/10/SSML/slides/Ruvini/Presentation_on_Sinhala.pdf

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied PsychoLinguistics, 24*, 431–451.

Sandyanganie, M. S. V., Jeewandara, K. C., & Perera, H. (2016). Prevalence and correlates of reading and spelling difficulty in 10 year old children in a semi-urban population in Sri Lanka. *Sri Lanka Journal of Child Health, 45*, 193–198.

Share, D. L., & Daniels, P. T. (2016). Aksharas, alphasyllabaries, abujidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler. *Writing Systems Research, 8*, 17–31.

Sircar, S., & Nag, S. (2013). Akshara-syllable mappings in Bengali: A language-specific skill for reading. In H. Winskel & P. Padakannaya (Eds.), *South and south-east Asian psycholinguistics* (pp. 202–211). Cambridge, UK: Cambridge University Press.

Tiwari, S., Nair, R., & Krishnan, G. (2011). A preliminary investigation of akshara knowledge in the Malayalam Alphasyllabary: Extension of Nag's (2007) study. *Writing Systems Research, 3*, 145–151.

UNESCO Institute for Statistics. (2015). *Adult literacy rate, population 15+ years, both sexes (%).* Retrieve from http://data.worldbank.org/indicator/SE.ADT.LITR.ZS?locations=LK

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690.

Vasanta, D. (2004). Processing phonological information in a semi-syllabic script: Developmental data fromTelugu. *Reading & Writing, 17*, 59–78.

Wasala, A. & Gamage, K. (2004–2007). *Research report on phonetics and phonology of Sinhala*. Retrieved March 26, 2015 from https://www.researchgate.net/publication/266296671_Research_Report_on_Phonetics_and_Phonology_of_Sinhala

Weerasinghe, R., Wasala, A., & Gamage, K. (2005). A rule based syllabification algorithm for Sinhala. In R. Dale, K. F. Wong, J. Su, & O. Y. Kwong (Eds.), *Natural language processing – IJCNLP 2005. Lecture notes in computer science* (Vol. 3651, pp. 438–449). Berlin, Heidelberg: Springer.

Wijayaratne, D. J. (1956). *History of the Sinhalese noun*. Colombo, Sri Lanka: University of Ceylon Press.

Wijaythilake, M. A. D. K., & Parrila, R. (2014). Predictors of word reading in good and struggling readers in Sinhala. *Writing Systems Reseach, 6*, 120–131.

Learning to Read and Write in Thai

Heather Winskel and Theeraporn Ratitamkul

Abstract Thai has a distinctive alphabetic script that shares some common characteristics with Indic writing systems, due to common origins. It has syllabic characteristics as it has inherent vowels for some consonants. Furthermore, it has a non-linear configuration in that consonants are written in a linear order, but vowels can be written above, below, or to either side of the consonant as full letters or diacritics, and commonly combine across the syllable to produce a single vowel or diphthong. Notably, Thai is also a tonal language and does not normally have interword spaces. Hence, when one reads in Thai, words have to be segmented using cues other than spaces. There is a high level of consistency of mapping between phonemes and graphemes but there are multi-grapheme to phoneme correspondences. Consequently, spelling development lags behind reading in Thai. The particular challenges this distinctive orthography poses to beginning readers and writers of Thai is discussed. First, the characteristics of the Thai language and its orthography are outlined. Secondly, relevant background literature is reviewed prior to examining some research conducted on learning to read and write in Thai. Finally, some future research directions on reading and spelling in Thai are presented.

Keywords Children · Interword spaces · Orthography · Reading · Spelling · Thai · Tonal language · Writing

---

H. Winskel (✉)
Psychology, School of Health and Human Sciences, Southern Cross University, Coffs Harbour, NSW, Australia
e-mail: heather.winskel@scu.edu.au

T. Ratitamkul
Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand
e-mail: Theeraporn.R@chula.ac.th

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_12

Introduction

Many of the scripts of South and Southeast Asia share features and characteristics as they are historically-related and descendants of the ancient Brahmi script. They have been termed 'alphasyllabaries' as they have hybrid characteristics of both alphabetic and syllabic scripts. The term 'alphasyllabary' has also been used to refer to writing systems in which vowels are denoted by symbols that are subsidiary to consonants and that do not necessarily adhere to a linear temporal order of pronunciation (Bright 2000; Daniels & Bright, 1996). A common feature of many Indian scripts is *akshara*, which is basically an orthographic unit that represents sound at the level of the syllable, as well as marking constituent phonemes (see Bhuvaneshwari & Padakannaya, 2014; Joshi, 2014; Sircar & Nag, 2014). These historically-related scripts are very diverse in terms of their characteristics and the languages that they map onto.

Thai shares some common characteristics with Indic writing systems, due to these common origins. It has syllabic characteristics as it has inherent vowels for some consonants. Furthermore, it has a non-linear configuration in that consonants are written in a linear order, but vowels can be written above, below, or to either side of the consonant as full letters or diacritics, and commonly combine across the syllable to produce a single vowel or diphthong. Thai does not have ligatures connecting adjacent consonants (a characteristic of scripts such as Devanagari), similar in this respect to scripts such as Tamil. Notably, Thai is also a tonal language, which is a characteristic it shares with other regional languages (e.g., Chinese, Burmese and Vietnamese). Interestingly, this characteristic is also shared with Punjabi, which has three tones: high falling, low rising and level. Punjabi can be written either in Gurmukhi (ਗੁਰਮੁਖੀ) script or in a version of the Urdu script known as Shahmukhi (شاہ مکھی). Furthermore, Thai script does not normally have interword spaces. Hence when one reads in Thai, words have to be segmented using cues other than spaces. Thai shares this feature with regional neighbours, including Lao, Khmer and Myanmar, as well as Chinese and Japanese. Notably, Thai also represents lexical tone orthographically, which is a shared characteristic with the scripts of Lao, Vietnamese and Myanmar.

The particular challenges this distinctive orthography poses to beginning readers and writers of Thai are discussed below. First, the characteristics of the Thai language and its orthography are outlined. Secondly, relevant background literature will be reviewed prior to examining some research conducted on learning to read and write in Thai. Finally, in the concluding paragraph, future research directions will be proposed.

Characteristics of the Thai Language and Its Orthography

Thai Language

Thai is a predominantly monosyllabic, analytic (or isolating) language. That is, the word forms do not change, similar in this respect to many other languages in the Southeast Asian region. Thai, consequently, does not have complicated inflectional morphology; instead affixation, compounding, elaboration and concatenation (strings of verbs juxtaposed around a head verb to form complex verb phrases) are used. Aspect, not tense, is the primary inflectional category for verbs (Matisoff, 1983). Thai verbs have no inflection for time or number. Instead, context, added time expressions or preverbs generally specify the tense. Subject-Verb-Object contributes the most favoured word order. Thai also has a noun classifier system. When a verb is reduplicated, the action indicated by the verb is generally intensified. Thai is a topic-prominent language, which has "a surface coding for the topic, but not necessarily for the subject" (Li & Thompson, 1976, p. 466) whereas English is a subject-focused language and requires overt subjects. Another characteristic of Thai, similar to other languages in the region, is that many word forms have multiple functions, which change depending on their position in a clause or sentence (e.g., /paj0[1]/ 'go', /ma:0/ 'come', or /da:j2/ 'can'). Thai shares many of these features with other regional neighbours such as Chinese due to areal diffusion.

Thai Orthography

Thai has 44 basic consonants plus 4 archaic consonants (ฤ, ฤๅ, ฦ, ฦๅ, which are rarely used) for 21 consonant sounds. Thai initial consonants are listed in Table 1. Thai is predominantly monosyllabic but also has polysyllabic words, which have been borrowed mainly from Khmer, Pali, or Sanskrit. Spoken Thai has clearly defined syllable boundaries (Hudak, 1990). Possible syllable structures are $C_1(C_2)V$, $C_1(C_2)V$:, $C_1(C_2)VC_3$, and $C_1(C_2)V{:}C_3$ (e.g., vowel (V and V: indicates a long vowel), vowel-consonant (VC)) (Tumtavitikul, 1998).

Consonants are written in a linear order, but vowels are nonlinear in their configuration. In addition, there is a high level of consistency of mapping between phonemes and graphemes but there are multi-grapheme to phoneme correspondences. For example, there are multi-grapheme to phoneme correspondences for some consonants (e.g. for /p^h/, /t^h/, /k^h/, /s/, /$tɕ^h$/, refer to Table 1). In addition, there is a change in grapheme-phoneme correspondences of consonants when they occur in final position. Thai has only nine final consonant phonemes, /p/, /t/, /k/, /ʔ/, /m/, /n/, /ŋ/, /j/ and /w/, and other consonants occurring in final position have to take one of these sounds (e.g., อาหาร < ʔa:ha:**r** > 'food' becomes /ʔa:ha:**n**/). In addition, Thai is predominantly a monosyllabic language with words with very similar spellings; words often vary by just one letter or the tone of the syllable, which adds to the challenges of learning this particular orthography.

There are 15 initial consonant clusters, which all start with stop consonants and are followed by laterals, trills, or bilabials; กร /kr/, กล /kl/, กว /kw/, ขร /k^hr/, ขล /k^hl/, ขว /k^hw/, คร /k^hr/, คล /k^hl/, คว /k^hw/, ตร /tr/, ปร /pr/, ปล /pl/, ผล or พล /p^hl/, พร /p^hr/. Thai does

[1] Tones are marked in the Thai examples cited in this paper as follows; 0 = mid, 1 = low, 2 = falling, 3 = high, 4 = rising. This system is based on the system that was developed at the Linguistics Research Unit (LRU) of Chulalongkorn University (Luksaneeyanawin, 1993). IPA transcription is used for the transcription of all other Thai text.

Table 1 Thai initial consonants (IPA symbols are in parentheses)

Place of articulation / Manner of articulation		Bilabial	Labio-dental	Alveolar	Palatal	Velar	Glottal
Stop	Voiceless aspirated	ผ พ ภ (pʰ)		ฐ ฑ ฒ ถ ท ธ (tʰ)	ฉ ช ฌ (tɕʰ)	ข ฃ ค ฅ ฆ (kʰ)	
	Voiceless unaspirated	ป (p)		ฏ ต (t)	จ (tɕ)	ก (k)	อ (ʔ)
	Voiced	บ (b)		ฎ ด (d)			
Nasal		ม (m)		ณ น (n)		ง (ŋ)	
Fricative			ฝ ฟ (f)	ซ ศ ษ ส (s)			ห ฮ (h)
Lateral				ล ฬ (l)			
Trill				ร (r)			
Approximant		ว (w)			ญ ย (j)		

Table 2 Thai vowel expressions classified in terms of vowel combination types

	Vowel expressions	Example
(1) Single vowel expressions		
(i) Preceding the consonant vowels:	เ_ /e:/, แ_ /ɛ:/, โ_ /o:/, ไ_ /aj/, ใ_ /aj/	แปรง /prɛ:ŋ0/
(ii) Following the consonant vowels:	_ะ /a/, _า /a:/, _อ /ɔ:/, _◌ำ /am/	คอ /k^{h}ɔ:0/
(iii) Upper vowels occurring above the consonant:	_อั /a/, _◌ิ /i/, _◌ี /i:/, _◌ึ /ɯ/, _◌ื /ɯ:/	สี /si:4/
(iv) Vowels occurring below the consonant:	_◌ุ /u/, _◌ู /u:/	งู /ŋu:0/
(2) Simple vowel combinations (combination of 2 vowel letters)		
(i) Preceding vowel plus following vowel:	เ_ะ /e/, แ_ะ, /ɛ/, เ_อ /ɤ:/, โ_ะ /o/, เ_า /aw/	โปะ /poʔ1/
(ii) Vowel above plus following vowel:	_◌ัว /ua/	วัว /wua0/
(3) Complex vowel combinations (combination of more than 2 vowel letters)		
(i) Preceding plus two following vowels:	เ_อะ /ɤ/, เ_าะ /ɔ/	เงาะ /ŋɔʔ3/
(ii) Preceding, upper plus following vowels	เ_◌ีย /ia/, เ_◌ือ /ɯa/	เรือ /rɯa0/
(iii) Preceding, upper plus two following vowels:	เ_◌ียะ /iaʔ/, เ_◌ือะ /ɯaʔ/	เอียะ /ʔiaʔ1/
(iv) Upper plus two following vowels:	_◌ัวะ /uaʔ/	ลัวะ /luaʔ3/

have irregularities, which include silent consonants and vowels that are not pronounced (e.g., the final consonant ‘ย’ in อาทิตย์ ‘week’ is not expressed and in this case is explicitly marked by the diacritic ◌์) (for more details see Winskel, 2014, Winskel & Iemwanthong, 2010).

As stated previously, vowels vary in position and orthographic expression (see Table 2). Vowel length is linguistically significant, and Thai has both short and long vowels. There are also five commonly used vowels (i.e., เ_ /e:/, แ_ /ɛ:/, โ_ /o:/, ไ_ /aj/, ใ_ /aj/) that precede the consonant orthographically and yet operate after the consonant in speech (e.g., แฟน < ɛ:fn > is spoken as /fɛ:n0/ ‘boy/girlfriend’). Other vowels are written and spoken in a corresponding order as occurs in English (e.g., ฝาก < fa:k > is spoken as /fa:k1/ ‘deposit’). In addition, vowels are not always explicitly specified. Instead, inherent vowels can occur with the consonant; an *a* is usually found in words of Sanskrit, Pali, or Khmer origin, whereas an *o* is found in native Thai words (e.g., ตลาด /t[a]la:t1/ ‘market’ or ลม /l[o]m0/ ‘wind’, the inherent vowels are in parentheses). Thai vowels are complex graphemically in their expression, but have, in general, consistent grapheme to phoneme mapping.

Tone forms an integral element of the syllable in Thai, and serves an essential function in distinguishing meanings of syllables and words with identical phonological structure. There are five tones in Thai conceptualised as high, mid, falling, rising, and low, but only four tone markers; *ma:j3 ʔe:k1* ◌่, *ma:j3 tho:0* ◌้, *ma:j3 tri:0* ◌๊ and *ma:j3 tɕ*^{h}*at1ta1wa:0* ◌๋, which occur above the initial consonant in the syllable or word. The tone determination of a syllable is complex, as it is influenced by a combination of the class of initial consonant, the type of syllable (open or closed), the final consonant in the syllable, the length of the vowel and the tone marker. There are three classes of consonants (11 high, 9 middle, 24 low), which reflects old voicing distinctions that have been neutralized in modern Thai (Gandour, 2013; Hudak, 1990). These particular classifications of the initial consonant (high, middle, low) in conjunction with the other factors contribute to phonological tone realisation in written Thai. In addition, there are two silent orthographic class-change consonants, ห (high class) or อ (middle class), which are used to differentially change the class of consonant to a high or middle class expression with a corresponding change in tone, for example, นาว /na:w0/ with 'ห' becomes หนาว /na:w4/ 'cold' or ยาก /ja:k2/ with 'อ' becomes อยาก /ja:k1/ 'want'.

The Thai Education System and the Teaching of Reading and Writing

Thai children typically begin school in Grade 1 when they are 6 or 7 years of age. Prior to that, the majority of children attend Kindergarten where they are taught the letters of the alphabet; the consonants and single vowel letters are taught first. Initially, they are taught how to read real words or syllables composed of a consonant and simple vowel (e.g. มา /ma:0/ *come*) without final consonants; then, easy words with final consonants are introduced. At the next stage, both monosyllabic and bisyllabic words are typically presented to the children. Gradually, children are taught to read words with more complex combinations of vowels. Simple sentences are subsequently introduced to the children. When children begin to learn to read Thai, they are given spaced text, but then unspaced text is introduced in Grade 2 (around 7 or 8 years of age).

According to the Basic Education Core Curriculum 2008 (Bureau of Academic Affairs and Educational Standards, 2008), children in Grade 1 should be able to read at least 600 basic, everyday words, which are words with and without tone markers, with both regular and irregular final consonants, and with initial consonant clusters and silent orthographic class-change consonants. They learn to read and spell by practicing reading sets of similar words. By the third grade, children should master at least 2,600 words, including words with silent graphemes and those with more irregular spellings. They are taught words with orthographic irregularities as well as spelling rules, for example, rules concerning consonant classes and tones. Teaching reading and spelling is further elaborated in Grade 4 and children learn more about the principles of the Thai language, characteristics of syllables in Thai, and types of syllables.

Relevant Background Literature

The psycholinguistic grain size theory (Ziegler & Goswami, 2005) asserts that the way the phonology of a language is represented by its orthography affects the psycholinguistic grain size that is primarily accessed by children beginning to learn to read. In regular or transparent alphabetic orthographies (e.g., German, Indonesian or Finnish), acquisition occurs relatively fast and children appear to be able to access small grain size units, grapheme-phoneme correspondences, whereas in other less consistent orthographies such as English, which is notoriously irregular, children can take several years to gain a similar level of competence, and appear to use a mixed or flexible grain size strategy (Goswami, 2000, 2003; Goswami, Ziegler, Dalton, & Schneider, 2003; Seymour, Aro, & Erskine, 2003).

Research on European alphabetic orthographies has found that phoneme awareness is a robust predictor of reading and spelling (Bradley & Bryant, 1983; Byrne & Fielding-Barnsley, 1993; Caravolas & Bruck, 1993; Caravolas, Hulme, & Snowling, 2001; Cardoso-Martins, 1995; Goswami, 2002; Hulme et al., 2002; Maclean, Bryant, & Bradley, 1987). However, research on reading in Asian languages has drawn attention to the syllable as an important processing unit and de-emphasised the degree of prominence of the phoneme. It appears that due to the hybrid characteristics of these scripts, children may be less aware of phonemes than readers of European languages (e.g., Nag, Caravolas, & Snow, 2011; Vaid & Gupta, 2002; Vaid & Padakannaya, 2004). For example, the syllable is the optimal unit for children acquiring Kannada, although more proficient readers and spellers can also manipulate phonemes (Padakannaya, Rekha, Vaid, & Joshi, 2002).

As stated previously, a common characteristic in Brahmi-derived scripts is that vowels can be written prior to the consonant but follow it in speech (e.g., the written word “odg” would be spoken as /dog/). In an innovative study, Vaid and Gupta (2002) used this feature to investigate the relative prominence of the phoneme in relation to the syllable when reading and writing Devanagari. In Devanagari, the short /i/ vowel produces an incongruency between the spoken and written order. For example, the word /t**i**lak/ is written with the vowel placed before the /t/, resulting in what could be read as /**i**tlak/. There are also more severely non-aligned vowels where the vowel is phonologically realized in the subsequent syllable. For example, medial /i/ occurs in the second syllable of a word, in which case the vowel is actually written to the left of the consonant in the first syllable (e.g., /masj**i**d/ is written as <m**i**sj[a]d>). The researchers used a rapid naming task to compare responses to the aligned and non-aligned words. They found that the aligned control words were named significantly faster than both types of non-aligned words, in particular the more severely non-aligned words. Moreover, children incorrectly placed the vowel /i/ after the first consonant rather than before the second consonant. For example, they read /mas**i**jad/ for /masj**i**d/. This reveals that children tend to favour the spatial placement rather than temporal order of the vowels. Moreover, when writing the words, Hindi speakers initially wrote the consonant and then the preceding non-aligned vowel. Vaid and Gupta (2002) argued that this indicates that the initial short vowel /i/ is segmented at the phoneme rather than syllable level. These results were interpreted as supporting a phonemic and syllabic level of processing in Hindi readers/writers.

In contrast to these results on Devanagari, a somewhat different pattern of results was found for Thai; a greater emphasis on segmenting at the syllable level was found (Winskel, 2009). Similar to Devanagari, Thai has words with non-aligned vowels whereby vowels can precede the consonant in writing but follow it in speech. These types of vowels can operate within the syllable (e.g., แมว < ɛːmw > is spoken as /mɛːw0/ 'cat') or across the following syllable in a more severely non-aligned example (e.g., the word แมลง /ɛːmlŋ/ 'insect' is spoken as /m[a]lɛːŋ0/ where [a] represents an inherent vowel). In this study, eye movements of adults reading words with and without non-aligned vowels in sentences were recorded using the EyeLink II tracking system. The participants read sentences with pairs of aligned and non-aligned vowel words matched for length and frequency embedded in same sentence frames. Similar to Vaid and Gupta (2002), rapid naming data was also collected from adults and children aged 6;6–8;6 years old. The children also were asked to spell comparable words. Results indicated that there was a processing cost associated with reading and spelling the more severely non-aligned words where the vowel operates across the syllable, but not in general for the non-aligned words where the vowel operates within the syllable in both adults and children. In fact, results for the non-aligned words (where the vowel operates within the syllable) were not significantly different from the aligned control words. Furthermore, in contrast to what occurred in Hindi writers, Thais typically adhered to the written order and wrote the vowel first, and then the consonant, in the non-aligned vowel words. This disparity in results for the Thai and Devanagari readers, is likely due to orthography-specific differences. Thai has five commonly-used vowels that precede the consonant but phonologically occur after the consonant, whereas in Devanagari there is only one short vowel /i/ that occurs prior to the consonant (Vaid & Gupta, 2002). Thai vowels also commonly combine with other vowels and diacritics across the syllable to produce a single vowel or diphthong. This characteristic does not occur in Devanagari. Another consideration is that the representation of lexical tone occurs at the syllable level in Thai. These orthography-specific characteristics appear to underlie the processing differences observed in these related writing systems, with segmentation at the syllable level playing a more prominent role in Thai than Devanagari.

In a more recent study, the development of reading and spelling of Thai words and nonwords was investigated in children from Grade(s) 1–3 (from 7 years to approximately 9 years) (Winskel & Iemwanthong, 2010). Non-words were created by changing the first consonant of the corresponding monosyllabic word, for example the word เดือน /dɯan0/ 'month' became the nonword เกือน /kɯan0/. In Grade 1 after 4 months of school, the children achieved 42% correct for reading words and 24% for reading the corresponding non-words, and 32% correct for word spelling and only 17% for non-word spelling. In grades 2 and 3, children scored about 86% correct for reading and spelling words. However, notably, they still lagged behind in non-word reading and spelling with only approximately 60% correct. These results are more comparable with results found for children learning to read relatively irregular alphabetic orthographies such as English, rather than the more regular orthographies such as German.

This lexicality effect, that is, the better performance on reading and spelling of words than non-words, suggests that a larger grain size (lexical or syllabic) is being utilised when reading and spelling Thai. Furthermore, there were more lexical than phonological errors for both reading words (lexical: 59%, phonological 41%) and non-words (lexical: 53%, phonological: 47%) in this study; such results are in accord with this view. These effects were more pronounced in the Grade 1 children as they had 61% lexical errors and 39% phonological errors for word reading, with 56% lexical errors and 44% phonological errors for non-word reading. The relatively high proportion of lexical errors, in conjunction with the relatively poor performance on non-word reading, suggests that Thai children, particularly the youngest Grade 1 children, were predominantly using a larger grain size (syllabic or lexical) to decode both words and non-words.

The degree of balance between the mapping of phonemes and graphemes in different orthographies can also affect reading and spelling development. When there is consistent mapping between phonemes and graphemes but there are multi-grapheme to phoneme correspondences, as occurs in Thai, then spelling development generally lags behind reading (e.g., French, Hebrew, Portuguese, and Persian; Geva, Wade-Woolley, & Shany, 1993; Pinheiro, 1995; Rahbari, Sénéchal, & Arab-Moghaddam, 2007; Sprenger-Charolles, Siegel, & Bonnett, 1998). Congruent with these findings, the youngest Grade 1 Thai children lagged substantially behind in word spelling in comparison with word reading ability, but by Grade 2 this disparity had largely been overcome (Winskel, 2009; Winskel & Iemwanthong, 2010).

In this study, the reading and spelling errors of the children were also analysed. The youngest Grade 1 children made a high proportion of unrelated and initial consonant errors, which is similar to what has been found for children learning to read other alphabetic orthographies. This reflects their poorly developed phonological and orthographic skills, as well as their inability at this stage to fully decode or encode words and non-words. Other errors found were related to vowel length and to visually similar vowels, a result which concurs with results found in Hindi (Gupta, 2004). Children also tended to make errors on words with consonant clusters. When reading, they either tended to not pronounce the second consonant in the cluster or segment the word incorrectly, or, when spelling, they tended to drop the second consonant in the cluster. Similar difficulties with consonant clusters have been found in other studies (Sprenger-Charolles & Siegel, 1997; Stuart, 2005; Treiman, 1997; Treiman & Weatherston, 1992). Specific to Thai, the younger children had difficulty with the orthographic class-change clusters, in which the first consonant of the cluster, ห or อ, is silent. Children tended to pronounce the silent consonant and segment the word inappropriately when reading, (e.g. เหงา /ŋaw4/ was pronounced as เห–งา /he:4-ŋa:0/) and when spelling, children often omitted the silent class-change consonants.

In Thai, vowels are relatively more consistent than consonants in terms of phoneme-grapheme correspondences. In line with this, it was found that fewer errors were made spelling vowels than consonants (Winskel, 2010). In addition, spoken language affected the spelling errors made by the children. In central Thai dialect, the /r/ is often pronounced as /l/, which was reflected in the children's spelling errors. This also concurs with what has been found in other orthographies (Alcock & Ngorosho, 2003; Joshi, Hoien, Xiwu-Feng, Chengappa, Boulware-Gooden, 2005; Treiman, 1993). The most persistent errors made by the older children involved the incorrect selection of homophonous consonants and incorrect expression of lexical tone. The older children typically made minus-one errors (Sprenger-Charolles, Siegel, & Bonnett, 1998; Fernandes, Ventura, Querido, & Morais, 2008), with correct vowel expression but with the incorrect consonant and/or tone. The tone rules for Thai are complex and irregular, and consequently, errors were still being made by the older children.

Research indicates that children are more likely to spell a word correctly if it is regular in its spelling (Romani, Olson, & Di Betta, 2005) and "the greater number of rules, exceptions, and sources of inconsistency, the more slow and difficult the learning process is likely to be" (Caravolas, 2006, p. 499). In English, when word spellings do not correspond to the spoken language (e.g., *yacht* or *knee*), spelling development is relatively delayed for children as they need to specifically learn how to spell these types of irregular words (Treiman & Kessler, 2005). In a follow-up study, we further investigated the spelling strategies used by Thai children from grades 1–3 (38 Grade 1 children (mean age 6;8), 35 Grade 2 children (mean age 7;7) and 35 Grade 3 children (mean age 8;9)) when spelling words. In this study, children were given familiar words (as judged by their teachers and the fact that they occurred in the children's readers) to spell with the following characteristics: (1) regular words with a one-to-one correspondence between phonemes and graphemes, (2) words with inherent vowels where the vowel is not orthographically represented, (3) words with silent class-change consonants, which change the class of the consonant and the tone of the syllable, (4) words with final consonants that change due to a restricted number of legal final consonants, and (5) words with silent graphemes where there is a high degree of arbitrariness. More detailed information about the specific characteristics of the words presented to the children in this study and the errors that children made when spelling these words is given below.

1. Regular words with a one-to-one correspondence between phonemes and graphemes acted as a comparison with the other word spellings.
2. Inherent vowel words. In these types of words, the vowel is not orthographically represented (e.g., ฝน <fn> /f[o]n4/ 'rain' or สนุก <snuk> /s[a]nuk1/ 'fun', the inherent vowel is represented by the letter in brackets). These types of spellings are a core characteristic of these types of scripts. However, there are alternative ways of spelling these types of words orthographically using explicit vowels or diacritics in Thai such as _ะ /a/ or _ั /a/. For example, สนุก <snuk> /s[a]nuk1/ 'fun' could be written as สะนุก <sanuk> and ฝน <fn> /f[o]n4/ 'rain' could be written as โฝ็น <ofn> /fon/ or โฝน < o:fn> /fo:n4/ with a longer vowel, although this usage would not be common. We found that these types of errors were infrequent (only two children from Grade 1 made this type of error) and by the end of

Grade 1, accuracy rate was comparably high for both regular words (91%) and inherent vowel words (88%). These results reflect the fact that inherent vowels are a common and core feature of this script.

3. Silent class-change consonants. These are two silent orthographic consonants (in bold) which influence tone realisation of the syllable or word (e.g., หมอ /mɔ:4/ ‘doctor’ and อยาก /ja:k1/ ‘want’). Children may omit these silent-class change consonants (e.g., for หมอ to write มอ without the ‘ห’). We found that the youngest Grade 1 children had a tendency to initially omit these silent class-change consonants (9% of these word spellings), which declined in the Grade 2 (1.6%) and Grade 3 (1.6%). Accuracy was relatively high for class-change consonant words by Grade 2 (91%) indicating that children had largely learned this rule.
4. Irregular final consonants. There are only a limited number of legal final consonant phonemes, /p/, /t/, /k/, /m/, /n/, /ŋ/, /j/, /w/, and /ʔ/ and other consonants occurring in final position map onto one of these sounds (e.g., ควร <k^{h}uar> ‘should’ is read as /k^{h}uan0/ and อาจจะ <?a:tçtça> ‘maybe’ is read as /?a:t1tça?1/). Children may substitute the phonological form for the more restricted legal form (e.g., to write ควน <k^{h}uan> for ควร <k^{h}uar> ‘should’). These types of errors did occur in Grade 1 children (7.9% of the words), and to a lesser extent in Grade 2 (1.9%) and Grade 3 (0.4%) children. From this error data, we can infer that the younger Grade 1 children had a tendency to adopt this strategy, whereas by Grade 3, children had largely learned this final consonant spelling rule.
5. Silent graphemes. Silent graphemes (in bold) are added at the end of some words (e.g., สัตว์ <sat**w** > ‘animal’ is spoken as /sat1/ and อาทิตย์ <ʔa:thit**j** > ‘week’ is spoken as /ʔa:0thit3/, the silent graphemes are in bold). These types of irregularities have a high degree of arbitrariness and may need to be specifically learned. Grade 1 children had a tendency to omit the silent graphemes (e.g., for สัตว์ <satw> ‘animal’ to write สัต without the ว์, in this example occurring with a suppression marker) whereas Grade 2 children had a tendency to either omit or substitute incorrect graphemes, and Grade 3 children either substituted an incorrect grapheme or were correct in their spellings. Not surprisingly, it takes time to learn these word-specific spellings. It was not until Grade 3 that accuracy was relatively high for spelling these types of words (88%).

It can be seen that children appear to learn the more common or perceptually transparent forms earlier than words with spelling forms that apply to relatively small sets of words or individual words. In terms of acquisition, the inherent vowel words, which are a core feature of Thai (and other related scripts), were acquired relatively early and prior to the class-change consonant spellings, which were in turn acquired prior to the irregular final consonants, and the silent grapheme irregularities were the last acquired. Not surprisingly, the words with more arbitrary silent graphemes proved most problematic to the Thai children. This order of acquisition also roughly corresponds to the order of teaching recommended in the Basic Education Core Curriculum document (Bureau of Academic Affairs and Educational Standards, 2008).

Difficulties and Challenges of Learning to Read and Write Thai

All writing systems and orthographies pose distinct challenges to beginning readers and writers and influence the development of reading and spelling related skills (Abu-Rabia & Taha, 2004). Thai has a distinctive orthography, which presents particular challenges to the young learner. It has visually complex graphemic expressions, particularly in relation to the complex non-linear vowel combinations and for the expression of tone. This requires the memorisation of a complex combination of vowels and diacritics. In addition, Thai is primarily a monosyllabic language with words with very similar spellings, which often vary in just one letter or the tone of the syllable, which contributes to the challenges of learning this particular orthography. As Thai is more consistent in terms of spelling-to-sound (feedback) than sound-to-spelling (feedforward) correspondences, spelling development generally lags behind reading. The most persistent errors found in older children were related to the homophonous consonants and the complex expression of tone. Based on research findings, it appears that both larger grain size, lexico-syllabic as well as phonemic grain size are important when learning to read Thai. This reflects the hybrid characteristics of the orthography; it has alphabetic and syllabic characteristics. In order to become a skilled reader of Thai it appears to be beneficial to use both larger and smaller phonological units. Future research needs to investigate this further.

Additional research on Thai needs to focus on reading words in sentence contexts and for comprehension purposes. Moreover, there are rich opportunities for comparisons of reading and spelling development with other Brahmi-derived scripts with typologically similar and distinctive characteristics. A neglected area of research is on the acquisition of lexical tone when learning to read and write. There are intriguing prospects for further cross-linguistic research on Thai and other Brahmi-derived scripts in the South and Southeast Asian region. In particular, many children in this region are learning to read and write in more than one script so again it is important to conduct research on learning to read in closely related scripts and scripts that are quite diverse.

References

Abu-Rabia, S., & Taha, H. (2004). Reading and spelling error analysis of native Arabic dyslexic readers. *Reading and Writing: An interdisciplinary Journal, 17*, 651–689.

Alcock, K. J., & Ngorosho, D. (2003). Learning to spell a regularly spelled language is not a trivial task-patterns of errors in Kiswahili. *Reading and Writing: An Interdisciplinary Journal, 16*, 635–666.

Bradley, L., & Bryant, P. E. (1983). Categorising sounds and learning to read: a causal connection. *Nature, 310*, 419–421.

Bright, W. (2000). A matter of typology: Alphasyllabaries and abugidas. *Studies in the Linguistic Sciences, 30*(1), 63–71.

Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*, 104–111.

Caravolas, M. (2006). Learning to spell in different languages: How orthographic variables might affect early literacy. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 497–511). Mahwah: Lawrence Erlbaum Associates.

Caravolas, M., & Bruck, M. (1993). The effect of oral and written language input on children's phonological awareness: A cross-linguistic study. *Journal of Experimental Child Psychology, 55*, 1–30.

Caravolas, M., Hulme, C., & Snowling, M. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*, 751–774.

Cardoso-Martins, C. (1995). Sensitivity to rhymes, syllables, and phonemes in literacy acquisition in Portuguese. *Reading Research Quarterly, 30*, 808–828.

Daniels, P. T., & Bright, W. (Eds.). (1996). *The world's writing systems*. New York, NY: Oxford University Press.

Fernandes, S., Ventura, P., Querido, L., & Morais, J. (2008). Reading and spelling acquisition in European Portuguese: A preliminary study. *Reading and Writing: An interdisciplinary Journal, 21*, 805–821.

Gandour, J. T. (2013). A functional deficit in the sensorimotor interface component as revealed by oral reading in Thai conduction aphasia. *Journal of Neurolinguistics, 26*, 337–347.

Geva, E., Wade-Woolley, L., & Shany, M. (1993). The concurrent development of spelling and decoding in two different orthographies. *Journal of Reading Behaviour, 25*, 383–406.

Goswami, U. (2000). Phonological representations, reading development and dyslexia: Towards a crosslinguistic theoretical framework. *Dyslexia, 6*, 133–151.

Goswami, U. (2002). In the beginning was the rhyme? A reflection on Hulme, Hatcher, Nation, Brown, Adams and Stuart (2002). *Journal of Experimental Child Psychology, 82*, 47–57.

Goswami, U. (2003). Why the theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences, 7*, 534–540.

Goswami, U., Ziegler, J. C., Dalton, L., & Schneider, W. (2003). Nonword reading across orthographies: How flexible is the choice of reading units? *Applied PsychoLinguistics, 24*, 235–247.

Gupta, A. (2004). Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 17*, 79–99.

Hudak, T. J. (1990). Thai. In B. Comrie (Ed.), *The major languages of East and South-East Asia* (pp. 757–773). London, UK: Routledge.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82*, 2–28.

Joshi, R. M. (2014). Literacy in Kannada, an alphasyllabic orthography. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 184–191). Cambridge, UK: Cambridge University Press.

Joshi, R. M., Hoien, T., Xiwu-Feng, R., Chengappa, R., & Boulware-Gooden, R. (2005). Learning to spell by ear and by eye: A cross-linguistic comparison. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 551–568). Mahwah, NJ: Lawrence Erlbaum Associates.

Bhuvaneshwari, B., & Padakannaya, P. (2014). Reading in Tamil: A more alphabetic and less syllabic *akshara*-based orthography. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 192–201). Cambridge, UK: Cambridge University Press.

Li, C. N., & Thompson, S. A. (1976). Subject and topic: a new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457–489). New York: Academic Press.

Luksaneeyanawin, S. (1993, March 17–21). Speech computing and speech technology in Thailand. *Proceedings of the Symposium on Natural Language Processing in Thailand* (pp. 276–321). Chulalongkorn University.

Maclean, M., Bryant, P. E., & Bradley, L. (1987). Rhymes, nursery rhymes and reading in early childhood. *Merrill-Palmer Quarterly, 33*, 255–282.

Matisoff, J. A. (1983). Linguistic diversity and language contact. In J. McKinnon & W. Bhruksasri (Eds.), *Highlanders of Thailand*. Oxford: Oxford University Press.

Nag, S., Caravolas, M., & Snowling, M. J. (2011). Beyond alphabetic processes: literacy and its acquisition in the alphasyllabic languages. *Reading and Writing: An Interdisciplinary Journal, 24*(6), 615–622.

Pinheiro, A. M. V. (1995). Reading and spelling development in Brazillian Portuguese. *Reading and Writing: An Interdisciplinary Journal, 7*, 111–138.

Padakannaya, P., Rekha, D., Vaid, J., & Joshi, M. (2002). *Simultaneous acquisition of literacy skills in English and Kannada. A longitudinal study.* (Poster presented at the 13th World Congress of the International Association of Applied Linguistics, Singapore).

Rahbari, N., Sénéchal, M., & Arab-Moghaddam, N. (2007). The role of orthographic and phonological processing skills in reading and spelling of monolingual Persian children. *Reading and Writing: An interdisciplinary Journal, 20*, 511–533.

Romani, C., Olson, A., & Di Betta, A. M. (2005). Spelling impairments. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 431–447). Blackwell Publishers.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.

Sircar, S., & Nag, S. (2014). Akshara syllable mappings in Bengali: A language specific skill for reading. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian Psycholinguistics* (pp. 202–211). Cambridge, UK: Cambridge University Press.

Sprenger-Charolles, L., & Siegel, L. S. (1997). A longitudinal study of the effects of syllabic structure on the development of reading and spelling skills in French. *Applied PsychoLinguistics, 18*, 485–505.

Sprenger-Charolles, L., Siegel, L. S., & Bonnett, B. (1998). Reading and spelling acquisition in French: The role of phonological mediation and orthographic factors. *Journal of Experimental Child Psychology, 68*, 134–165.

Stuart, M. (2005). Phonemic analysis and reading development: some current issues. *Journal of Research in Reading, 28*, 39–49.

Treiman, R. (1993). *Beginning to spell: A study of first grade children*. Oxford, UK: Oxford University Press.

Treiman, R. (1997). Beginning to spell in English). In C. Hulme & R. M. Joshi (Eds.), *Reading and spelling: development and disorders* (pp. 371–393). Mahwah, NJ: Lawrence Erlbaum Associates.

Treiman, R., & Kassar, B. (2005). Writing systems and spelling development. In M. J. Snowling & C. H. Hulme (Eds.), *The science of reading: A handbook* (pp. 120–134). Blackwell Publishers.

Treiman, R., & Weatherston, P. (1992). Effects of linguistic structure on children's ability to isolate initial consonants. *Journal of Educational Psychology, 84*, 174–181.

Tumtavitikul, A. (1998). The metrical structure of Thai in a non-linear perspective. In Warotamasikkhadit, U. & Panakul, (Ed.), *Papers from the Fourth Annual Meeting of the Southeast Asian Linguistics Society* (pp. 53–72). Arizona State University, Program for Southeast Asian Studies.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690.

Vaid, J., & Padakannaya, P. (2004). Reading and writing in semi-syllabic scripts: An introduction. *Reading and Writing: An Interdisciplinary Journal, 17*, 1–6.

Winskel, H. (2009). Reading in Thai: The case of misaligned vowels. *Reading and Writing: An Interdisciplinary Journal, 22*, 1–24.

Winskel, H. (2010). Spelling development in Thai children. (Special issue on Reading Development and Impairment in Asian Languages Guest Editor Prof. Greg Simpson). *Journal of Cognitive Science, 11*, 7–35.

Winskel, H. (2014). Learning to read and write in Thai. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 171–178). Cambridge: Cambridge University Press.

Winskel, H., & Iemwanthong, K. (2010). Reading and spelling acquisition in Thai children. *Reading and Writing: An Interdisciplinary Journal, 23*, 1021–1053.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3–29.

Part III
Assessment

The Assessment of Emergent and Early Literacy Skills in the Akshara Languages

Shaher Banu Vagh and Sonali Nag

Abstract The assessment of language and literacy skills in the akshara languages pose distinct challenges that are related to the specificities of the orthography, contextual variability and population diversity, which in turn tend to be compounded by generally low levels of achievement. In this book chapter, we discuss an assessment framework linked to a language and literacy framework for the early years for akshara languages. Using assessments for Hindi and Kannada as cross-linguistic illustrative examples, the discussion focuses on psychometric rigor and sensitivity to contextual factors that include demographic diversity (e.g., socioeconomic factors, home language(s), and access to literacy resources), classroom practices and orthography.

Keywords Akshara languages · Assessment · Emergent literacy · India · Oral language proficiency · Reading comprehension · Writing

Assessments are used to address a wide range of important questions. Is the second grader reading at grade level? Is the literacy program effective? Is the child able to read with comprehension? To effectively address questions such as these we suggest that there are at least three key issues to consider. First, characteristics of the language of literacy instruction are important because they define the pace of literacy acquisition and the skill-set that is useful to assess. Unique characteristics may be related to the spoken and written forms of the language (e.g., its morphology, syntax and orthography-specific features). Second, the socio-cultural, socio-economic and

S. B. Vagh (✉)
ASER Centre, New Delhi, India

Pratham, New Delhi, India
e-mail: shahervagh@gmail.com; shaherbanu.vagh@pratham.org

S. Nag
Department of Education, University of Oxford, Oxford, UK

The Promise Foundation, Bangalore, India
e-mail: sonalinag@t-p-f.org; sonali.nag@education.ox.ac.uk

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*,
Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_13

linguistic status of young learners is central because they each determine the opportunities for learning and may explain performance on an assessment. In addition, children's performance on a given task may be reflective of the extent of exposure to the culture of testing and familiarity with assessment formats rather than simply the skills that underpin literacy. Finally, assessors must consider the psychometric appropriateness of individual tools by which skills and areas of knowledge are assessed especially given the varied purposes that motivate assessment. Since contextual realities are such that many young learners may be deprived of opportunities to encounter, learn and consolidate their emergent and early literacy skills, it is clear that a key consideration in use of testing tools is the fairness and sensitivity of assessments to demographic diversity (e.g., socioeconomic level, home language(s)), socio-contextual factors (e.g., access to literacy resources) and classroom practices (e.g., the task demand of various types of reading comprehension questions). In this chapter, we discuss these three concerns to address the assessment of emergent and early literacy skills in the akshara languages.

The chapter is structured around an assessment framework that gives weight to skills and knowledge linked to not just literacy but also the foundations of literacy in broader language skills. This framework lends itself to a component by component analysis within the written language and spoken language domains and across literacy development (Nag, 2017a); the framework also fits well with our theoretical account of literacy acquisition as crucially dependent on language. The next sections introduce the akshara writing system and the foundational skills that contribute to literacy acquisition followed by the outlining of an assessment framework. The ensuing sections focus on issues related to psychometric rigor, test utility for research versus practice, appropriateness of test adaptations and challenges related to monitoring growth in language and literacy skills over time. A key consideration is the sensitivity of assessments to demographic diversity. We conclude with suggested guidelines for test development for research and practice. Throughout, we use assessments for Hindi and Kannada—belonging to the Indo-Aryan and Dravidian families respectively—as cross-linguistic illustrative examples.

The Orthography, Phonology and Morpho-syntax of Akshara Languages

Our writing is focused on the symbol block called *akshara*. The akshara orthographies are most commonly termed an alphasyllabary (Salomon, 2000), with other names being abugida (Daniels, 2009) and the akshara writing system (Nag, 2017b). Akshara are versatile units and run into several hundred, hence these orthographies are also labeled 'extensive' (Nag, 2007). Individual phonemes in the akshara writing system can have either a primary or a secondary form as the written representation based on position in a word and other orthotactic rules. Akshara can be simple or complex depending on the amount of encoded phonological information. This is because the akshara unit can represent an open syllable, the body and coda of a closed syllable and a phoneme. In some akshara orthographies (but not all), the akshara unit also encodes information from across the syllable boundary. Transparency of the akshara orthography further contributes to what will be easier to read and spell. A distinct feature of these orthographies is that the phonemic units or markers can be visually decomposed for the majority of the akshara with some exceptions, e.g., the high frequency inherent vowel /a/ (also called the 'schwa') and the low frequency conjoint consonants like the /ksha/ and /jna/ in Hindi (similar clusters not found in Kannada). These exceptions introduce some opacity in an otherwise transparent system of phonology-to-orthography mapping. Additionally, akshara tend to behave differently when represented as a singleton akshara versus when embedded within a word. In the context of bi-syllabic and multi-syllabic words, the post-vocalic consonant visually maps on to the next akshara. This process known as re-syllabification accounts for a wide range of akshara for consonant clusters and can contribute to an additional layer of opacity. This is because the schwa suppression in the consonant akshara is left unmarked or the representation is across the syllable boundary or there is a mismatched phonology-orthography mappings (e.g., the first two are common in Hindi and the second two in Kannada). As with all writing systems, the mastery of these ambiguities in sound-to-symbol linkages clearly take longer (reading: Nag, 2014a; Purushothama, 1990; Vaid & Gupta, 2002; spelling: Nag, Treiman, & Snowling, 2010). At the printed level, the visual complexity of the akshara is attributable to the non-linear representation of markers within the symbol block. Typically, non-linearity results from the placement of vowel or consonant markers to the left, right, top, or bottom of a consonant. Visual complexity may also stem from the density of visual features in the akshara block, and whether these are disconnected from other segments in the symbol block (Kannada akshara: Nag, Snowling, Quinlan, & Hulme, 2014a). Taken together these multiple parameters of phonological and visual complexity of the akshara have implications for learning to read and spell well past primary school (for review see Nag, 2017b). Added to this, when there is limited availability of children's literature or the occurrence of such complexities is low in child-directed print that too presents an additional challenge to the literacy acquisition task.

Prior to learning about the written language, there are the foundation skills related to spoken language. At the level of single words, while all languages allow for word formation based on permitted combinations called morphotactics, languages differ in the range and variety of word formation processes allowed. One important process of word formation is to modify words by the addition of inflections where the new word carries new grammatical information (e.g., the plural marker –s added to words like mat, hat and bat and the past tense marker –ed to words like sway, stray and spray). The akshara based languages vary in the complexity of inflection use but are typically never as sparse as English. Among them, the Dravidian languages (Kannada, Malayalam, Tamil and Telugu) are strongly agglutinating in nature with a high density of inflections suffixed to a word (for reviews see Krishnamurti, 2003). In contrast, Hindi would be considered to be morphologically less dense. An example of grammatical information for the noun 'at night' in Kannada and Hindi is *raatriyalli* (suffix *-alli*) and *raat ko* (postposition

/*ko*/). Apart from inflectional morphology, languages also have the facility to generate new words through derivation. Here, the new word form can either belong to a different part-of-speech (e.g., magic, magician, magical in English; *pratham* + *ik* = *praathamik* in Hindi and *prathama* + *ika* = *praathamika* in Kannada, primary, here the suffix *–ik/−ika* forms adjectives from nouns) or remain in the word class but with a change of meaning (Hindi: *khet* and *khet* + *ii*, farm and agriculture; *chamak* and *chamakdaar,* bright and sparkling). Examples of prefixes that change word meaning in Hindi and Kannada are *bahu-* (multi-) as in *bahu-bhaashii and bahu-bhaashika* (multi-lingual), and *mano-* (of the heart and mind) as in *mano-vigyaan* and *mano-vigyaana* (psych-ology, the science of heart and mind). Another example of word formation is compounding (e.g., moonlight, daydream and nevertheless) which is also common in most akshara-based languages. Finally, there are words formed through reduplication (Hindi: *chalte chalte*, Kannada: *hoogutta hoogutta*, while walking;) and echo-compounding (Hindi: *khaana-vaana*, Kannada: *thindi-gindi,* food and related matters); both these linguistic phenomena are found in several akshara-based languages and are considered as defining features of the South Asian linguistic space.

At the sentential level the ordering of words is rule-governed. Both Hindi and Kannada are Subject-Object-Verb (SOV) languages, although there is a relatively free word order. Given the agglutinating nature of akshara languages, morpho-syntactic knowledge gains prominence in language development. There is a small body of evidence to suggest that knowledge of inflectional morphology, derivational morphology, morpho-syntax and vocabulary are important not just for language comprehension but also reading comprehension (e.g., Nag & Snowling, 2011a; Vagh & Sharma, 2018).

While the unique and distinct grapho-phonological and morpho-syntactic specificities of the akshara languages define skills and knowledge that must be considered when planning an assessment, it is important to note that the attainments in each of these areas are firmly constrained by the context. Examples of contextual variability include bilingualism, multi-lingualism, diglossia, and dialect varieties of the school language. Depressed socioeconomic conditions, low access to books at home and at school and in general low resources and low family and community engagement in literacy activities signal potential disadvantage on a literacy-related assessment task. Turning to instruction in the classroom, children's learning histories may differ in several ways. Instruction may differ in the extent to which the textbook is privileged over a range of other language sources and the mechanics of reading and writing is privileged over comprehension and making inferences (see Vagh, Nag & Banerji, 2017). There is growing evidence to show that these multiple factors individually and cumulatively contribute to low levels of achievement in the akshara languages (e.g., India: Bhattacharjea, Wadhwa, & Banerji, 2011; Sharma, 1997; Singh, 2014; Vagh, 2009; Nepal: Pinto, 2010; Sri Lanka: Aturupane, Glewwe, & Wisniewski, 2013; for multi-country review see Nag, Snowling, & Asfaha, 2016).

Some ensuing implications of ecological and psycholinguistic constraints for the assessment of language and literacy skills are the high prevalence of zero scores and an inability to capture individual differences at the bottom of the attainments spectrum. Floor effects mask children's true abilities when for instance, their bi- or multi-linguistic knowledge base is unaccounted for or failure is linked to testing format rather than skill, proficiency or knowledge. Moreover, while floor effects or low levels of attainment arising from classroom instructional practices provide information about the performance of a school or education system they become inaccurate representations of individual abilities. Similar constraints also function at the high end of the attainments spectrum; assessments may turn out to be too easy and register ceiling effects because individual differences in children's mature skills and knowledge bases are not captured. In summary, patterns of attainments are driven by the learner's experiential ecology of home, community and school and this cannot be easily disentangled from the distinct learning demands posed by the specificities of a language or orthography.

What Early Skills Contribute to Literacy Acquisition in the Akshara Languages?

Oral Language Proficiency Word knowledge and the grammatical rules of the spoken language help children effectively communicate and successfully comprehend the written word. The concurrent and predictive role of language proficiency to literacy development, well into middle school, is widely acknowledged for both first and second language learners (Kannada: Nag & Snowling, 2012; Gujarati: Patel, 2004; Hindi: Vagh & Sharma, 2018). Aspects of oral language that have been shown to contribute to literacy development in the akshara languages include vocabulary knowledge, syntactic knowledge (the grammar rules that determine word order and sentential structure to communicate the message) and inflection knowledge (markers in words to index different grammatical information including case, tense, person, number and gender). Given the multiple languages spoken in many communities, it becomes imperative to better understand the variability in children's skills in the home, community and school languages to better address their literacy and broader academic needs. But more importantly, oral language is the first and continuous support for literacy acquisition; low proficiency in oral language implies limited supply of linguistic cues to resolve the ambiguities in reading and writing that the child will inevitably encounter.

Phonological Processing Skills Phonological processing refers to the ability to identify and manipulate sound units in a language. This is an important skill for decoding words in the akshara languages (for review see Nag, 2017b). Aligning with the nature of the akshara orthography, phonemic and syllabic awareness skills appear to be salient in the reading acquisition process well into middle school. While syllable processing has consistently emerged as a predictor of individual differences in reading attainments, the role of phonemic awareness remains somewhat unclear. Fluent readers tend to demonstrate increased proficiency in phonemic processing skills and these skills tend to correspond with the acquisition of complex akshara (Kannada: Nag & Snowling, 2012). But there is also research to suggest that despite increasing proficiency in phonemic level skills across the primary grades, its influence on decoding appears to be subsumed by syllabic level skills (Kannada and Telugu: Nakamura, Joshi, & Ji, n.d.). It is unclear whether differences in these two studies can be attributable to different reading outcomes, i.e. rate and accuracy of reading words in connected text (Nag & Snowling, 2012) versus the accurate decoding of words in list form (Nakamura et al., n.d.) or to differences in the specificity with which phonological awareness skills have been operationalized. While, Nag and Snowling (2012) sample a wider range of phonological manipulations (a beginning and final deletion task and a substitution task at the phonemic and syllabic levels), Nakamura et al. employ a single task, the deletion of the final syllable and phoneme. Moreover, the absence of unique contribution of phonemic level skills to reading accuracy in a multiple regression model as shown in Nakamura et al. cannot be interpreted to indicate that the skill does not matter to reading acquisition. It is entirely possible that its influence could be indirect or mediated. Although the mechanisms of influence of different levels of phonological processing are as yet unclear, it appears that analytical skills that allow readers to access the phonological units in words and akshara accrue benefits for accurate and efficient reading and spelling. For the purposes of assessment, insights of the child's mastery with phonological processing at each level can be expected to provide useful information.

Concepts About Print Exposure to print helps children develop rudimentary understandings about the uses and functions of print and how the spoken and written word are connected (Teale & Sulzby, 1986). This exposure may be through experiences that are either direct (e.g., reading labels and flashcards, shared storybook reading) or indirect (e.g., watching others engage with print when reading the newspaper and filling out printed forms). Children with greater conceptual understandings about print tend to demonstrate better proficiency in listening and reading comprehension (Hindi: Vagh & Sharma, 2018). Children's concepts about print provide a useful index of early and frequent exposure to print (Nepal: Pinto, 2010), which arguably is critical when we consider that the majority of children learning to read and write in the akshara languages may come from home and community environments characterized by low print or low literacy engagement with the purposes and functions of literacy. The language of print exposure is an area that is yet to receive systematic research attention though surveys with low-income households suggests that availability of child-oriented print and parent engagement with children in print-related activities is extremely low (India: Bhattacharjea, Wadhwa, & Banerji, 2011; Nag et al., 2014a; Vagh, 2009; Nepal: Pinto, 2010). In parallel, surveys with middle-income households suggest that print in some homes is multi-scriptal while in others is focused on either the dominant language of the region or the minority language of the home (India: Kalia & Reese, 2009; Nag, Snowling & Mirkovic, 2017). One implication for assessment is the need to contextualize test items as closely as possible to children's experiences; for example, in exceptionally low print contexts, the wall calendar (Bhattacharjea, Wadhwa & Banerji, 2011; Nag et al., 2014b) and religious texts (Vagh, 2009) may be sources of print exposure providing potential material for assessment of conceptual knowledge about print.

Orthographic Knowledge Orthographic knowledge is an important foundation skill for literacy acquisition in all written languages. Learning to decode and spell begins with the learning of the basic sound-symbol units. For the akshara languages these are the vowels (V), consonants with the inherent vowel (Ca) and the consonants with ligatured vowels (CV). While these akshara tend to be easily learned, the acquisition of the complex akshara (CCV, CCCV) extends well into fourth grade (Nag, 2007). Emergent akshara recognition is a correlate of phonological skills (Hindi: Vagh & Sharma, 2018; Kannada: Nag et al., 2014a). In later grades, increasing akshara knowledge is associated with better visual word recognition, speed of processing as assessed on rapid automatised naming tasks, visual memory, vocabulary and spelling (Kannada: Nag & Snowling, 2011b, 2012; Nag, Treiman, & Snowling, 2010; Purushothama, 1990; Hindi: Gupta, 2004; Vagh & Sharma, 2018). Visual complexity attributable to the visual features of the akshara and phonological complexity attributable to the number of sound units encoded in the akshara tend to contribute to the slower acquisition of akshara knowledge. Arguably then, children who have more exposure to print and greater exposure to the spoken language will come to the task of literacy learning with more cognitive reserves; they will show greater sensitivity to visual features in the symbol set and greater ease with manipulating the sounds in the language.

Emergent Writing Emergent writing refers to the early explorations around print in the form of markings, drawings, scribbles, symbol-like forms, invented spellings and conventional spellings. These word productions represent children's understanding that writing conveys meaning and that writing is a form of communication. Increased opportunities to engage in emergent writing activities help to refine children's understandings of how print encodes and conveys meaning. Opportunities to support emergent writing differ widely both at the level of instruction and apprenticeship as well as access to and type of writing tools; from this perspective the experiences for emergent writing are no different from the experiences for drawing in various contexts (for drawing, see Jolley, 2010). Thus, children's markings or scribbles evolve developmentally over time culminating in the acquisition of conventional spelling; with a growing body of evidence to show that emergent writing is far from random but reflective of patterns in the language and orthography and a form of statistical learning (Treiman & Kessler, 2013). Children's learning environments that value all forms of early explorations with print tend to be conducive to the early emergence of writing abilities and children who demonstrate early engagement in emergent forms of writing tend to demonstrate increased performance in listening and reading comprehension in the primary school years (Vagh & Sharma, 2018).

In summary, a variety of emergent and early skills contribute to literacy acquisition. The array of skills relate to the linguistic system (lexicon, syntax and phonology), the mechanics of writing (transcription) and the orthographic system (symbols and mapping principles). However, these skills and knowledge are not discrete operators. On the contrary, successful reading and writing depends on interplay between and interdependence amongst these skills and knowledge. Poor readers in one domain may show difficulties in several other domains, but also strength in one domain can mitigate the losses associated with deficits in another domain. Hence, there is clearly a case for using assessments to develop *a profile of strengths and difficulties* in foundation skills and from this make a judgment about children's learning.

An Assessment Framework

Assessments evaluate a learner's ability in some measurable way and provide a profile of relative strengths across component skills and knowledge bases. Assessments typically represent a small sampling of a domain of interest and help us to draw inferences about an individual or a group's level of ability or skill in the larger domain. For example, a picture naming task with a limited set of items is used as an estimate of a child's vocabulary repertoire while performance on a listening comprehension task serves to index broader language proficiency.

Well-designed assessments of language and literacy must have content (or test items) that account for language- and orthography-specific characteristics and thereby provide a representative sample of the underlying domain. In parallel, the format and content of assessments must be sensitive to local and contextual factors such as children's home and community language. In the context of diglossia for example, ways to demonstrate sensitivity could be to avoid words and sentences that are particularly vulnerable to variation between the ordinary spoken variety and the more complex formal variety. But such selectivity may lead to a narrow and thus irrelevant test in which case the preferred way to show sensitivity would be through explicitly worked out scoring schemes and interpretation guidelines when diglossia effects are noted. Also important to note is whether classroom instructional practices have acquainted children with certain formats of assessments. For instance, absence of creative writing opportunities restricts the legitimacy of narrative writing assessments; similarly, when the sole focus is on verbatim responses from supplied texts then it is difficult for children to respond to inferential questions. In other words, unfamiliar test formats not only underestimate children's 'true' ability but also bring into question the validity of the assessments. Sensitivity in test construction and test use reduces the chances of unfair penalties on test performance. Hence, integral to our framework of assessment is the reporting of the test development process and the psychometric properties related to the operationalizing of a construct. Psychometric evaluations help us to understand whether an assessment provides a reliable profile of children's skills and knowledge and whether test inferences are defensible and valid. They also indicate adequacy for use with specific populations of interest. Classical test theory, generalizability theory and item response theory are the dominant approaches to evaluating the psychometric properties of an assessment. The use of the latter two approaches, despite their many merits, is however extremely limited in assessments both in the akshara languages as well as internationally (Nag, 2017a). Our intent in this chapter is not to compare and contrast these approaches but rather to highlight the need for rigor in evaluating and in the reporting of an assessment's technical quality, including the validity associated with the interpretation of test scores and their use.

In keeping with the psychometric approach, reporting about individual tests should at the least include information about item level analysis (item difficulty and where appropriate differential item functioning), consistency in measurement (e.g., internal consistency or test-retest consistency from a classical test theory perspective), stringent assessments of interrater reliability (e.g., Cohen's Kappa estimates that accounts for chance versus mere estimates of correlations between raters) and validation evidence. The latter is an accumulation of evidence based on but not restricted to the relevance and representativeness of test content and concurrent and predictive information as support for specific test score interpretation (more below). A recent review of assessment tools for emergent and early literacy found transparent reporting of these parameters to be poor in the few akshara languages in which tools are currently available namely Bengali, Gujarati, Hindi, Kannada, Malayalam, Odia, Tamil, Telugu, Thai, and Sinhala (Nag, 2017a).

A fundamental issue related to the validity of assessments is its intended purpose. Teachers, other practitioners, researchers and policy makers tend to use assessments for varied purpose. For instance, practitioners are likely to use assessments for summative, formative or diagnostic reasons, researchers may seek to understand the extent of variability in specific skills or knowledge and their concurrent or predictive role in relation to specific language or literacy outcomes and policy makers may want to monitor and evaluate entire school systems. While there is overlap in the uses of tests for teaching, research and policy making, the larger point is that these intended purposes of assessments have implications for test design, administration, test evaluation and subsequent reporting of scores. While some assessments may be suitable for more than a single purpose, it is often the case that assessments developed for a specific use are not appropriate for alternative uses. For instance, an assessment developed to evaluate group level performance may neither be specific nor sensitive enough for use as a diagnostic tool or to make pedagogical decisions about individual students. Additionally, an assessment on which a substantial number of children fail or succeed, i.e. one which captures little variability, provides useful information to teachers, practitioners and even policy makers about presence or absence of mastery and effectiveness of instructional programs but is of little value to researchers who for instance are examining individual differences and pathways of influence of early skills on later development. In the ensuing sections, our discussion focuses on assessments from a research standpoint. Assessments at scale to evaluate extensive school systems, clinical assessments for diagnostic purposes and assessments by teachers to guide and inform classroom practice are other important purposes of testing, but these are beyond the scope of the current chapter.

Taken together, our assessment framework for language and literacy acquisition has a strong psychometric perspective and accounts for the specificities of the orthography and the psycholinguistic characteristics of the language such that the sampled skill set is critical to concurrent associations with and predictive of future literacy outcomes. This framework, drawing from research in the akshara languages and mindful of contextual constraints, helps determine the domain-specific skills to be assessed and has the potential to enhance teaching-learning processes and in turn inform assessment design for further research. Current research in the akshara languages, as discussed, suggests that the skills and knowledge associated with reading comprehension and writing ability, the litmus tests of literacy acquisition, are language proficiency and grasp of its morphology and syntax, concepts about print and emergent writing, phonological processing skills at the syllable and phoneme level, the acquisition of simple and complex akshara, reading accuracy and fluency, spelling ability and narrative generation skills. In the next few sections we discuss some of the ways these skills and knowledge are assessed and highlight contextual as well psycho-linguistic challenges associated with their measurement. While a discussion of every assessment format is beyond the scope of this chapter, we focus on some of the most common formats including assessments from our work in Hindi (Vagh, 2009; Vagh & Sharma, 2018) and Kannada (Nag and colleagues) to illustrate and highlight some promising assessments for research and practice.

Oral Language A wide spectrum of measures is available to assess oral language proficiency. This section will cover a selection of oral language tasks at the word level (picture-vocabulary, word definition) and at the sentence level (sentence repetition and listening comprehension). These assessments, while broadly indicative of children's linguistic knowledge, vary substantially in the specific area of oral language. As a consequence, the ensuing inferences from each task are either limited to or inclusive of children's lexical, morphological and syntactic knowledge.

Picture-Vocabulary Tasks Vocabulary refers to the repertoire of words that the child understands (receptive vocabulary) and the words in a child's lexicon that she or he can actively use in oral or written communication (expressive vocabulary). Typically, receptive vocabularies tend to be larger than expressive vocabularies. The former may be assessed via picture identification tasks (identifying the picture that matches the target word from multiple choices) and the latter via picture naming tasks (providing a label for the target picture). Picture identification and picture naming tasks are the most commonly used assessments of oral language ability (for international trends see Nag, 2017a). This is perhaps because picture-vocabulary tasks are relatively easy to administer and score compared to other oral language measures. However, picture-vocabulary measures provide only a slice of oral language restricted primarily to those nouns and verbs that can be visually represented in a simple and clear illustration. In addition, a test of even 30–50 items provides only a small sampling of a very extensive domain bringing into question the representativeness of the assessment. Another considerable challenge is assessing children growing up in bilingual or multilingual contexts. Restricting assessments to just the language of instruction not only underestimates children's conceptual knowledge base but also undermines effective strategies to foster word knowledge (Amritavalli, 2007; Snow, 2017). An ever present concern with picture-vocabulary tests is whether oral language assessments are biased against speakers of non-school languages or dialects.

Two separate longitudinal studies of emergent and early Hindi literacy skills in contexts of significant diglossia and dialect varieties illustrate innovations to avoid such a bias (Allahabad city in North India: Vagh & Sharma, 2018; Mumbai city in West India: Vagh, 2009). In these studies credit was given for vernacular and non-standard variants of picture labels prevalent in children's home and community language environments. For instance, 30% of children provided a vernacular variant such as /chova/ for the target label /keel/ (nail) and 9% provided variants such as /fulva/ for the word /fool/ (flower). Such contextualizing of the scoring scheme to the language(s) of the region permits for a more accurate estimate of children's lexical knowledge although it then limits generalizability to contexts where other variants are prevalent. Despite this, we recommend that in line with the general principal of test fairness there should be explicit guidance notes for responses in non-standard languages.

Other design related challenges for picture-vocabulary tasks are the selection of distractor items and the clarity of the pictures chosen to represent the target word and accompanying distractors. Phonologically similar sounding or semantically similar distractor items can help increase item difficulty but may in parallel cause confusion, leading to inconsistencies in measurement. Hence, pilot tests and item analyses sensitive to these issues are imperative. In the absence of word frequency lists or normative data, the selection of highly discriminating target words, particularly for longitudinal research in the akshara language is another considerable challenge. Rather than intuitively compiled graded lists, word frequency based on a corpus of children's literature can better inform the item selection process for vocabulary measures. Similarly, rather than judging age of acquisition amongst monolingual speakers, developing norms representative of multilingual contexts can help provide meaningful age- and grade-appropriate benchmarks.

Word Definition Tasks Tasks that require children to define select words are a measure of vocabulary depth in contrast to vocabulary breadth that is captured by picture-vocabulary tasks. Word definition tasks require children to be able to deduce the salient properties of the target word and present the information in a format that resonates with school-like tasks. Unlike the other assessments of oral language, definitional ability is regarded as a formal, decontextualized language skill as it lacks a shared temporal and physical context between narrator and listener (Kurland & Snow, 1997; Snow, 1991). The decontextualized nature of the task and the requirement for coherent production at the sentential level increases the cognitive and linguistic demand on this task compared to the picture-naming task. Coding schemes help evaluate children's responses and herein lies the challenge associated with this task because examiners have to rate oral responses in real-time or be accurate in recording running responses to be coded off-line at a later time. Thus, establishing inter-rater reliability in scoring children's responses is essential. Box 1 presents two coding schemes from three studies (Kannada and Hindi: Nag, 2008, 2014b; Hindi: Vagh & Sharma, 2018).

Box 1 Scoring Schemes for the Word Definition Task

Word Definition Task (Nag, 2008, 2014b)

The child had to explain the meaning of words representing actions, qualities, states, time, place, and result.

- Score 3: Accurate definitions, synonyms, and translated equivalents
- Score 2: Sentential use of the word and descriptions
- Score 1: Repeating of the word with an inflection or using idiomatic phrases

Noun Definition Task (Vagh & Sharma, 2018)

The child was asked the label of the pictured noun and was then asked to tell the examiner as much as they knew about it.

- Score 2: Target picture is described with four or more attributes relevant to the target noun including synonyms. E.g. A rat is an animal. A rat lives in a burrow. A rat has two ears. A rat runs fast
- Score 1: Target picture is described with one to three relevant attributes including synonyms or if the child relates a personal experience connected to the target word

There are two approaches to item selection for the word definition task. In the first approach, words are typically high frequency because the goal is defined as an assessment of definitional ability, which can be hampered if the target word is not well-known or understood by the child. Examples of high frequency words are mouse (/chuha/), cat (/billi/) and drum (/dholak/) for Hindi learners in Grade 1–2 (from Vagh & Sharma, 2018). A second approach to item selection is to cover a wide range of early and late acquired words. Examples of late acquired words are *garajna* and *gudugitu* (to roar, thunder); *parinaam* and *parinaama* (result) in Hindi and Kannada respectively, for use in primary and middle school (Nag, 2008, 2014b). The versatility of the graded task across age bands and proficiency levels makes such tests particularly attractive in longitudinal research programs.

Most concerns outlined in the section above on picture-vocabulary tasks remain relevant with word definition tasks. To mitigate the influence of dialects/diglossia, a pictorial representation of the target word can be used (Vagh & Sharma, 2018), and the coding scheme should not penalize for productions that reflect diglossia or dialect effects (Nag, 2008, 2014b; Vagh & Sharma, 2018). Another variation to the assessment format worth consideration is to offer children the option to respond in their home languages in addition to the instructional language in order to arrive at a more comprehensive profile of language proficiency. However, to the best of our knowledge, such an innovation in the design of word definition tasks is as yet untested with akshara languages.

Sentence repetition tasks are an innovative way to assess language processing mechanisms where children repeat verbatim a sentence spoken aloud by the examiner; a relatively simple and straightforward task. Primarily used in clinical settings and contexts of second language learning, the task presents a promising approach for assessing oral language because of its wide sampling of the language domain. Importantly, "the entire language system is recruited during sentence repetition while being sensitive to language-specific characteristics" (p. 19, Nag, Snowling, & Mirkovic, 2017). Areas that may be captured through the task include children's knowledge of content-function words, preferred word order (e.g. the subject-verb-object order or the subject-object-verb order) and inflections in the language.

Most sentence repetition tasks cover a range of sentence structures and use a word by word coding Scheme. A refinement to this word-level scheme to suit the morphologically rich and agglutinating akshara languages is to score for word-internal details. This includes assessment of accuracy in children's production of the word stem and all inflections that modify the meaning of the word to capture the intended event semantics such as who-did-what-to-whom. Assessment of inflections brings focus on morpho-syntax including children's assessment of case markers, person-number-gender markers (also called PNG markers), tense markers and markers for passive structures. A further area of interest is whether children's production retain the coherence of the 'message' in the just heard sentence or whether there is a collapse in meaning. A summary of such a scheme is given in Box 2, for a first application of this scheme see Nag et al. (2017).

Box 2 Scoring Scheme for the Sentence Repetition Task (Nag et al., 2017)

The child had to repeat verbatim a sentence spoken aloud by the examiner, which was scored at three levels:

Sentence level coding. Measured word-order accuracy and the length of the produced utterance. Word order was scored as correct when the exact order of the sentence elements was preserved and length was unchanged. When the repetition was a phrase or a shorter syntactically complete sentence that preserved event semantics, it was coded as a shortened utterance.

Word- and affix-level coding. Content words (nouns, verbs) with specific grammatical suffixes were scored for accuracy on the root, inflection, and the boundary (see online-only supplementary material for examples). Dialect variations for roots of words (e.g., dropping of the glottal fricative /h/ in word initial position; thus *aakidaLu* for *haakidaLu*) were accepted and all other substitutions were coded as a semantic change. Because inflections on the semantic substitutions and the boundary could be accurate for the new root–inflection pairing, semantic substitutions were included in the analyses of accuracy on boundaries and inflections.

As evident from the above description, while sentence repetition tasks are simple to administer and make low demands on testing time, its complexity is in the design of appropriate sentences and in the coding and analysis of errors. Sentence level factors that affect performance relate to word frequency, syntactical complexity (e.g., active and passive sentences), frequency of sentence type and plausibility of the events encoded in the sentence. In the field of bilingual and multilingual research as well in second language teaching open questions remain about the challenges for non-native speakers and dialect speakers on the processing of grammatical and phonological systems. Arguably, tasks such as sentence repetition that allow a broad sampling of the underlying complexities of the language system may provide valuable new information.

Listening Comprehension Tasks Listening comprehension is the ability to extract meaning from spoken language in the form of a discourse, an oral narrative or an aural presentation of an informational text. Listening comprehension tasks are particularly suitable for use with beginning and struggling readers and represent a dynamic approach to assessing oral language ability that is beyond the level of single word meanings and sentences. Multiple psycholinguistic skills underpin the ability to make meaning from spoken language. These include but are not limited to lexical knowledge, morphology, syntax and ability to hold and process information in working memory.

Key issues for development and administration of listening comprehension assessments are choice of text–topic and its grade-appropriateness, and inter-rater reliability among examiners to account for the reading of the text and scoring of children's responses. A text that requires specialized knowledge (word and world knowledge) is likely to impose greater cognitive demands for meaning extraction than a text that is on a common or familiar topic. Deeply embedded in choice of topics for listening comprehension tasks is the possibility of undue bias against learners with limited opportunities for exposure to certain types of world knowledge. Similarly vocabulary, grammatical structure, length of text and number of events in the text that a child has to keep track of in order to comprehend the narrative are other parameters to consider. In addition, a substantial challenge specific to all akshara languages is the absence of information about word frequency or commonly upheld parameters for leveling texts.

The format of questions eliciting comprehension of just-heard material can be deployed in a multiple-choice format using written, verbal or pictorial options. Pictorial formats help reduce the cognitive load particularly for young learners and reduce rating errors in scoring oral responses in real-time. However, the challenges related to diglossia and bi- and multilingual contexts (as discussed above) persist. In addition, it is important to consider issues of accent and pronunciation when examiners are not from the same language backgrounds as the assessed children. In summary when developing and interpreting listening comprehension tasks it is vital to be sensitive to text properties such as topic and grade appropriateness and training of examiners in reading texts and scoring responses.

Phonological Processing Tasks Phonological processing refers to the ability to isolate and manipulate the sounds of a language. The units of sound may be phonemes or syllables or the parsing of syllables into body and coda (or, if for languages like English, then into onset-rime). These skills are assessed in a variety of ways. For instance, and roughly in order of task demand and explicit display of phonological processing skill, the tasks cover blending of units, parsing or segmentation of units, identifying or generating rhyming words, identifying words that begin and end with the target sound, and substitution or deletion of the first, final or middle unit. Nonword repetition and decoding of nonwords are also often considered as broadly tapping the same underlying phonological domain as the processing tasks described above. Each of these task manipulations lend themselves to much experimental innovation in terms of item complexity and have allowed for teasing out challenges to learning to read in the akshara orthographies (e.g., Nag & Snowling, 2012).

Typically in research studies, one or a small selection of the above manipulations are used to index children's phonological awareness skills. This is understandable given that testing time is limited. However, there is a need for more research with comprehensive phonological processing batteries to critically examine whether the specificities of phonological skills assessed by individual manipulations are similar and whether the tasks are interchangeable. Another open question is whether the pathways of influence on reading development converge or differ when the skills have been elicited via different manipulations. It may turn out that multiple manipulations across multiple levels of phonology are the best option for indexing a meaningful profile of children's phonological abilities. Such research will also be helpful in resolving inconsistencies across research studies and for better defining the role of phonological awareness skills in learning to read in the akshara languages.

Within a psychometric perspective, the reporting of inter-rater reliability and test reliability is important, especially where raters have to make on-line judgments about children's performance on a response by response basis. Our experience has been that training for phoneme level tasks requires more time and practice and even though the akshara is not always a syllable representation, assessors may misunderstand syllable level tasks as demanding akshara-by-akshara manipulation. Inferences about phonological awareness have to be considered in light of the required manipulations and type of akshara. For instance, segmenting the phonemes of just simple akshara vs. a set of simple and complex akshara. Finally, as with all forms of oral language assessment, item selection has to be sensitive to the issue of diglossia, dialects and second language learners because any of these can influence the enunciation of certain phonemes placing some children at greater risk for poor performance (e.g. the use of 's' for 'sh' or 'r' for 'dh' in the regional dialects of Bihar in North India, and the dropping of the glottal /h/ in the regional dialects of Karnataka in South India).

Print-Based Skills

Emergent Literacy Tasks Concepts about print (CAP) refer to children's understanding of print conventions, its functions and uses. These for instance include understandings of print directionality, of book orientation, that print conveys meaning and that akshara strung together form words (e.g. see Table 1). One of the most widely used format for assessment of emerging print concepts is an adaptation of Marie Clay's work (2000) wherein the assessor whilst reading a book observes the child's book handling skills and asks questions such as "where should I begin reading?". Other assessment formats of emergent print concept include asking children to identify environmental print (e.g., pointing to the start point of a word on a calendar or a prominent name board), straighten akshara flashcards to face the right way up, copy visually simple to more complex features in akshara and distinguish (not name) akshara from amongst other symbols. Assessments of print concepts tend to differ in the number and nature of items which makes comparison across research studies less straightforward. As with all domains covered in this chapter, decisions on number of items and particular sub-skills are likely to be driven by issues of testing time, especially when the task is administered along with a battery of other assessments.

Table 1 Concepts About Print tasks (Nag et al., 2013; Vagh, 2009; Vagh & Sharma, 2018). Each prompt is assigned a score of 1 for a correct response

	Nag et al. (2013)	Vagh (2009)	Vagh and Sharma (2018)
Concept	Prompts	Prompts	Prompts
Parts of a book and book handling skills	Front of the book, back of the book, book title, open the book to where the story begins, open the book to the page just read	Front of the book, back of the book, book title, open the book to where the story begins	Front of the book
Print conveys meaning	Show the picture, show the words, meaning of word 'xxx'	Show the picture, show the words, where to start reading	Where to start reading
Print directionality	Show how the finger should move, return sweep to left	Which way to read, return sweep to left, line order, left page before right	Which way to read, return sweep to left line, left page before right
Word correspondence	Finger pointing	Finger pointing	Finger pointing
Difference between akshara and word	Show one word, count number of words on the page, show any one akshara, show first akshara of a word, show last akshara of a word, count number of akshara in the book title	First akshara of a word, last akshara of a word, number of akshara in a word, number of "matra", number of words	
Punctuation marks	Meaning of question mark, period and comma	Meaning of question mark, period, comma and quotation marks	

Our current work with the Concept about Print task suggests that children tend to reach ceiling on many of the test items (e.g., front of the book, where to start reading) by the end of the first year of instruction when the school is average to well functioning, yet the few items that continue to capture variation (e.g., mapping words with finger) show predictive validity and relate to children's later reading and listening comprehension attainments (Vagh & Sharma, 2018). Concepts about print and other emergent literacy tasks are also remarkably responsive to time-bound shared book reading interventions especially in print starved classrooms (Nag et al., 2014b). Both these preliminary trends suggest that there is a need for greater attention to the development of concepts about print and emergent literacy tasks in settings where akshara languages are in use. Finally, emergent literacy tasks have a unique motivational role in assessment situations in the early grades. We find that children are put at ease when participating in a book reading task or a copying task; these tasks do not seem to produce the same anxiety as other school-like tasks and this helps set a positive tone for the testing session.

Akshara Knowledge and Decoding Tasks For understanding individual differences across the primary school years, the research is clear that assessments of symbol knowledge need to encompass the extensive variety of akshara ranging from simple to complex (see Nag, 2017b). Assessments that focus on the primary forms of vowels (V) and consonants with the inherent vowel (Ca), which form the primary set (the '*varnamala*' in Hindi or '*varnamale*' in Kannada) and consonant-vowel pairs (CV) (the *barakhadi* or *aksharamala* matrix) are important but limiting when children move beyond the basics. Typically, these akshara—V, Ca, CV and CCV—tend to be assessed as independent singleton symbols but knowledge of akshara can also be assessed in the context of words either in list form or in connected text or nonwords. Assessments of akshara as isolated units versus assessments of akshara in the context of words tap different yet overlapping skills. The singleton approach and the reading of nonwords primarily targets akshara-phonology mappings while the reading of words is likely to also implicate lexical-semantic information including the use of contextual cues to decode unfamiliar words. All approaches – singleton akshara, nonwords, words in list form and words in connected text—are useful to assess as these allow researchers and practitioners to better identify sources of individual differences as well as children's areas of struggle. There is rich cross-linguistic evidence for the associations between symbol knowledge and literacy acquisition, particularly for beginning readers as well as for the ability to string akshara together to decode words in list form and connected text.

A significant challenge in developing assessments of symbol knowledge and decoding is the absence of frequency data for the different types of akshara and for words based on a corpus of children's literature. Informal age of acquisition corpora and frequency analysis on small child-directed print corpora has been rare in the akshara literature but where used have allowed for improved experimental manipulation (e.g., Nag, 2014a). An additional challenge is the absence of parameters for leveling reading materials that account for the linguistic properties of reading materials. See Box 3 for suggested guidelines. The absence of

Box 3 Relevant Text Features for the Akshara Languages: Suggested Guidelines

- Word frequency: Establishing grade appropriate word frequency and akshara frequency lists based on a corpus of children's literature, including the writing of children from different grades attending different types of schools.
- Curricular demands: Analyzing state-prescribed textbooks to understand the text characteristics of the prescribed curriculum and mandated pace of instruction (see rubric proposed by Vagh & Nag, 2014).
- Suggested metrics for reporting text characteristics that account for syntactic and semantic difficulty of the written materials for the akshara languages:

  - Average number of akshara per sentence to provide the syntactic complexity of a text. The akshara is a more relevant metric than words as it accounts for the morphologically rich and agglutinating akshara languages.
  - Average number of complex akshara per sentence.
  - Proportion of high frequency words.
  - Proportion of complex akshara.
  - Proportion of instances of re-syllabification.
  - Type-token ratios for akshara and words, which help account for repetitions. This is based on the premise that repetitions in a text (such as in early readers) make a text easier to read.
  - Overall length of text captured by number of akshara, words and sentences.

Further research is needed to address how best to synthesize text characteristics to index the grade-appropriate level of a text. We propose:

- Backward mapping of the metrics to state-mandated curricular as well as to children's reading performance at different grades to better understand alignment of text to grade.
- Analysis of children's reading errors can provide additional insights to the appropriateness of the metrics and the mapping process.
- Qualitative considerations of the topic or content's suitability for the desired population
- Consideration of issues such as font size, spacing, page layout, which are as yet understudied for the akshara languages.

evidence-based benchmarks to evaluate text level or its readability hinders comparability and generalizations across research studies as researchers may have employed different parameters to develop reading passages, thus creating passage level confounds. A common way around this problem has been to align text to state-mandated curricula. However, a limitation of this approach, particularly for research, is the absence of information about the appropriateness of the pace of instruction mandated in state curricula.

Turning next to the assessment of decoding, two common indices employed in assessment are accuracy and fluency. While accuracy is the correct decoding of words (or nonwords), fluency relates to the ability to accurately decode at an appropriate pace. Fluency is taken as indicative of a level of automaticity with sounding and blending of the akshara in words and is seen as allowing cognitive resources to be devoted to meaning making rather than to decoding (e.g., Wijayathilake & Parrila, 2014).

An important issue in the area of assessment of the component skills of reading relates to the utility of reading fluency as an index of reading comprehension for the akshara languages. Are fluency and accuracy measures adequate indicators of reading comprehension? If yes, then what constitutes as appropriate fluency rates, i.e. grade-appropriate norms? Evidence, albeit limited, suggests that accuracy explains reading comprehension above and beyond fluency and the joint contribution of accuracy and fluency (number of words read correctly in a minute) explains 58% of the variation in reading comprehension, which although substantial still leaves a little less than half of the variation unexplained (Hindi: Vagh & Biancarosa, 2011). This suggests that for the akshara languages, while fluency and accuracy are important, by themselves they are not sufficient indices of good comprehension. An additional open question is the efficacy of words as a metric of fluency for strongly agglutinative languages such as Kannada, Malayalam, Tamil and Telugu, where length of words are likely to be longer than for less agglutinative languages such as Hindi while morpheme counts are similar.

Reading Comprehension Tasks Reading comprehension is the ability to make meaning from the reading of a text. Reading for understanding is the result of a complex interplay between varied language-based and print-related skills including comprehension strategies such as looking back in the text or skimming ahead to enhance understanding. Reading comprehension, therefore is a multi-dimensional construct. Reading for meaning takes on many forms and purposes some of which are authentic practices carried out in the context of everyday life (e.g. following a recipe, reading a prescription), while others are more academic or school-based (e.g. reading to prepare a research paper, a book report or to respond to questions based on the text). The skills demonstrated in the latter context is what assessments of reading comprehension often aspire to capture.

Comprehension assessments usually are in the form of responses to questions based on a passage that is either read aloud or silently by the child. Questions can elicit factual information, inferences or a synthesis of the read text. The assessment format can be multiple-choice in verbal or pictorial format, can elicit an oral or written answer or can require children to provide a retelling. Alternatively, comprehension assessments can require children to provide the missing word in a sentence or the missing words (every n^{th} word) in a passage referred to as Cloze tasks.

The selection of the reading material and the framing of questions is central to the development of reading comprehension assessments. As discussed for listening comprehension, the topic of the text, vocabulary demands (familiar versus rare words), syntactic complexity and length of the text all have a bearing on the assessment of comprehension ability. See Box 3 for some recommendations on the reporting of text characteristics. Despite the simplicity of the assessment approach and strong alignment with classroom practices, evidence from other language contexts suggests that not all comprehension measures can be used interchangeably. Work with some popular assessments of English reading comprehension, for example, indicates for example that measures vary in the extent to which they capture decoding skills as against comprehension skills (Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 2017). We are not aware of similar direct comparisons between different formats of comprehension assessment in the akshara writing system. Linked to the issue of differences between measures of the same construct are the types of questions (e.g., factual, inferential) on a reading assessment and the format of capturing children's responses. The scoring of short answers in written format arguably includes an additional dimension of writing ability to the assessment of reading comprehension. Taken together, we wish to emphasize that assessments that go by the name of reading comprehension test are not all equivalent and it is prudent to carefully analyze task demand. A final point is related to test administration and reliability: adequate training of raters and the reporting of inter-rater reliability is essential irrespective of oral or written format of capturing children's responses to the comprehension questions.

Component Skills for Writing Assessment of writing abilities for the youngest learners evaluate emergent writing skills such as scribbles, drawings or invented spelling in response to a prompt as well as their ability to transcribe akshara, typically V, C, and CV units. Prompts that elicit children's emergent writings, for instance may be a read-aloud of a short story that the child is asked to draw and write about (Vagh & Sharma, 2018). Scoring schemes account for drawings, scribbles (non-symbols), independent akshara and string of akshara. On the other hand, assessments of akshara formation capture children's sound-symbol mappings and transcription ability. The stringing of words together to spell words captures children's ability to accurately transcribe the sounds of the spoken language, which involves knowledge of the symbol set i.e. sound-symbol mappings, rules of akshara formation including the mechanics of forming akshara. Spelling, for the akshara languages, tends to be more challenging than reading given the presence of a wide variety of complex akshara and mismatches in phonology-orthography mappings due to re-syllabification. Some akshara segments are also more prone to error given the minor differences in visuo-spatial orientation (e.g. the ligatured short and long vowel sounds /i/ and /ii/ in Hindi) and orthographic-phonological sequence (e.g. the mis-sequence in CCV arrangements in Kannada with the V ligatured to the first rather than the final consonant in a cluster).

Dictation, a test format common in most classrooms and familiar to most children, enables the assessment of young learners' spelling attainments. By including words on a dictation list that capture diverse aspects of akshara-phonology mapping (e.g., Nag, 2014a, 2014b; Nag et al., 2010), researchers and teachers can gain insights about the pace of their acquisition as well as individual children's strengths and gaps in knowledge. A further method of item selection is items selected for their morphological complexity such as manipulating the complexity of affixes in words and morphophonological changes due to sandhi rules although we are not aware of systematic assessment of this aspect of spelling knowledge in any of the akshara based languages. However, a considerable challenge is the effects of diglossia on spelling development. For instance, speakers of regional dialects in Bihar tend to articulate and spell the Hindi word 'shaadi' (marriage) as 'saadi' as they tend not to distinguish between the phonemes 'sh' and 's'. An open question is how best to account for these factors so that items on an assessment are not unduly biased against speakers of certain regional dialects?

Writing skills can also be assessed in the context of sentences and narratives. Just as with reading words in connected texts, sentence construction and writing longer narratives too draw upon multiple psycho-linguistic processes beyond those implicated/involved in/elicited by word-level dictation. For young learners, pictorial prompts are useful to elicit a sentence, a description or a narrative (e.g., picture prompt of a tiger with a cold clutching a handkerchief and warmly wrapped in a woolen scarf in Nag, 2013). The assumption here is that visual cues help provide a context for an otherwise decontextualized task and help emergent and beginning writers to generate ideas and as a consequence longer writing samples. Often, the prompt is just a target word, a written or oral prompt, or the retelling of a familiar story or a set of words (such as connectors, inflections, transitional words or phrases) that children are asked to use in the construction of their narratives.

In addition to spelling, scoring rubrics for sentence construction and narrative writing assessments evaluate syntactical knowledge, narrative organization, event sequencing, the extent to which settings and characters are developed, use of punctuations and length of narratives) along with at times a composite global indicator for writing ability. While these assessments are crucial to understand a desirable goal of literacy instruction, i.e. the ability to communicate in writing for varied purposes, two considerations are of critical importance. The first relates to handwriting and ambiguities in akshara formation. Poor handwriting can make it difficult for raters to decipher the written text, even if they are just singleton akshara. This makes the task of evaluating other attributes challenging. In addition, consistent application of the scoring rubric and agreements on the judgments of quality made by raters are considerable challenges. Hence, adequate training of raters and the reporting on inter-rater reliability for these measures is a minimum requirement.

The second issue relates to the limitations imposed by classroom instruction as writing ability is closely linked to opportunities offered within classrooms. Typically, in many akshara language classrooms, opportunities for creative writing and responding to open-ended writing prompts are typically uncommon. Children may perform poorly on measures that demand these skills simply due to lack of exposure. Hence, when assessing narrative writing abilities of children, two careful considerations have to be made. First, if there is little opportunity to practice narrative writing skills beyond the mechanics of writing and if there is little familiarity with the format of assessment then researchers have to consider how best to leverage or introduce appropriate supports. Examples of support could be in the form of modeling and practicing specific task demands and providing adequate number of trial sessions. Second, threats to the validity of the desired inference have to be well thought out. For instance, if the interest is to understand pace of acquisition of certain skills (e.g., use of low frequency or late acquired inflections) then lack of opportunity or usage undermines the inference of interest. But on the other hand, questions about quality of instruction can be validly addressed by ensuring that there is rich language input and an explicit inclusion of later acquired material in a selective and thoughtful manner (see Chapter on "Language-focused Instruction for Literacy Acquisition in *Akshara*-based Languages: Pedagogical Considerations and Challenges", this volume).

Discussion

Measurement undergirds all evidence-based research. Robust measures that are designed to evaluate specific theoretical accounts advance our understanding of the literacy acquisition process. However, with the prevalence of a wide variety of researcher developed tasks in the akshara languages, clarity with the test development process (e.g., selection of words, details about text features) becomes an imperative. Such transparency allows for greater comparability across research studies, promotes the generalizability of research evidence and mitigates confounds attributable to text and other differences in linguistic material. In addition, seemingly similar measures of the same construct can be tapping into very different skills, which also makes it difficult to resolve inconsistencies across research studies in learning to read in the akshara languages, e.g., the role of phonemic awareness (Nag & Snowling, 2012; Nakamura et al., n.d.); visuo-spatial arrangements (Vaid & Gupta, 2002; Winskel & Iemwanthong, 2010); word complexity (e.g. Husain, Vasishth & Srinivasan, 2015; Kandhadai & Sproat, 2010). Hence, research that specifically addresses comparability across diverse assessment approaches is essential. In the absence of such work, the use of multiple approaches such as different test formats as for oral language measures and varied set of manipulations as for phonological awareness tasks is preferable in arriving at a more comprehensive and accurate profile of the construct.

A constraint of literacy acquisition research in the akshara languages is the lack of norms for skills and knowledge such as vocabulary knowledge or oral reading fluency. The implementation of norming studies that can provide data representative of a wide segment of the population is unquestionably fraught with many challenges given the diversity in children's home language and socio-economic backgrounds. However, norm-referenced assessments help us understand the distribution of skills within a population. Undoubtedly, within a context of low quality schooling, norms are likely to be depressed and will not be an accurate representation of young learners' 'true' potential. Even so, norms help provide realistic benchmarks for the current status of skills and knowledge in the population of interest and further the ability to compare and generalize across the evidence base.

To conclude we present below a summary of the issues and recommendations discussed in this chapter.

(a) Diglossia is ubiquitous in the Indian context. Assessments of language and literacy need to be sensitive to variations attributable to regional dialects.
(b) Assessments of print concepts and book handling skills show promise for children from print impoverished home and community environments.
(c) Symbol knowledge assessments need to account for the diverse and extensive range of akshara beyond the basics of vowels, consonants with the inherent vowel and consonant with vowel ligatures.
(d) A review and analysis of a corpus of children's literature in the akshara languages is needed to understand frequency of occurrence of different akshara types and of words in print.
(e) For measures based on reading passages such as listening comprehension and reading comprehension, comprehensive reporting of text features related to semantics and syntax will allow for better comparability across research studies.
(f) Use of multiple formats in the assessment of skills will contribute to a more robust skill profile.
(g) Current reporting about test development and test properties is poor. Integral to a robust framework of assessment is the reporting of the test development process and the psychometric properties related to the operationalizing of a construct.

In the current chapter, our attempt has been to present an assessment framework for language and literacy acquisition in the akshara languages that is grounded in the psycholinguistic properties of the language and contextual realities of children's linguistic and socio-economic backgrounds including considerations of schooling opportunities. A limitation is that we haven't discussed issues related to cross-linguistic comparisons, which given its breadth, is beyond the scope of this chapter. Demonstration of psychometric robustness is an imperative. With the evolving landscape in akshara research, test formats and approaches will undoubtedly gain in specificity. Considerations of test appropriateness in relation to the inference must though remain paramount.

References

Amritavalli, R. (2007). *English in deprived circumstances: Maximizing learner autonomy*. New Delhi, India: Foundation Books and Cambridge University Press.

Aturupane, H., Glewwe, P., & Wisniewski, S. (2013). The impact of school quality, socioeconomic factors, and child health on students' academic performance: Evidence from Sri Lankan primary schools. *Education Economics, 21*(1), 2–37.

Bhattacharjea, S., Wadhwa, W., & Banerji, R. (2011). *Inside primary schools. A study of teaching and learning in rural India*. Mumbai, India: Pratham.

Clay, M. M. (2000). *Concepts about print: What have children learned about the way we print language*. Heinemann.

Daniels, P. T. (2009). Two notes on terminology. *Written Language and Literacy, 12*, 275–279.

Gupta, A. (2004). Reading difficulties of Hindi speaking children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 17*, 79–99.

Husain, S., Vasishth, S., & Srinivasan, N. (2015). Integration and prediction difficulty in Hindi sentence comprehension: Evidence from an eyetracking corpus. *Journal of Eye Movement Research, 1*(1):1, 1–9.

Jolley, R. (2010). *Children and pictures. Drawing and understanding*. Chichester, UK: Wiley-Blackwell.

Kalia, V., & Reese, E. (2009). Relations between Indian children's home literacy environment and their English oral language and literacy skills. *Scientific Studies of Reading, 13*(2), 122–145.

Kandhadai, P., & Sproat, R. (2010). Impact of spatial ordering of graphemes in alphasyllabic scripts on phonemic awareness in Indic languages. *Writing Systems Research, 2*(2), 105–116.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281–300.

Krishnamurti, B. (2003). *The Dravidian languages*. Cambridge: Cambridge University Press.

Kurland, B. F., & Snow, C. E. (1997). Longitudinal measurement of growth in definitional skill. *Journal of Child Language, 24*(3), 603–625.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*, 7–22.

Nag, S. (2008). *Kannada vocabulary test*. Bangalore, India: The Promise Foundation.

Nag, S. (2013). Low literacy attainments in school and approaches to diagnosis: An exploratory study. *Contemporary Education Dialogue, 10*(2), 197–221.

Nag, S. (2014a). Akshara-phonology mappings: The common yet uncommon case of the consonant cluster. *Writing Systems Research, 6*, 105–119.

Nag, S. (2014b). *Expressive vocabulary test-Hindi*. Bangalore, India: The Promise Foundation.

Nag, S. (2017a). Assessment of literacy and foundational learning in developing countries: Final Report. Health and Education Advice and Resource Team, xiii, 94p.

Nag, S. (2017b). Learning to read alphasyllabaries. In K. Cain, D. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 75–98). John Benjamins.

Nag, S., Chiat, S., Torgerson C., & Snowling, M. J. (2013). *Literacy, foundation learning and assessment in developing countries. Report of a rigorous review*. Department of International Development, UK.

Nag, S., & Snowling, M. J. (2011a, June). Reading comprehension, decoding and oral language. *The EFLU Journal, English and Foreign Languages University, 2*, 75–93.

Nag, S., & Snowling, M. J. (2011b, July). Cognitive profiles of poor readers of Kannada. *Reading and Writing: An Interdisciplinary Journal, 24*, 657–676.

Nag, S., Snowling, M. J., & Asfaha, Y. (2016). Classroom literacy practices in low- and middle-income countries: An interpretative synthesis of ethnographic studies. *Oxford Education Review, 42*(1), 36–54.

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a language-universal theory of learning to read. *Scientific Studies of Reading, 16*, 404–423.

Nag, S., Snowling, M. J., & Mirkovic, J. (2017). The role of language production mechanisms in children's sentence repetition: Evidence from an inflectionally rich language. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716417000200.

Nag, S., Snowling, M. J., Quinlan, P., & Hulme, C. (2014a). Child and symbol factors in learning to read a visually complex writing system. *Scientific Studies of Reading, 18*, 1–16.

Nag, S., Ramkumar, S., Miranda, R., Sutar, L., Krishna, M., Aravind, S., Kala, B., & Arulmani, G. (October, 2014b). Home and school learning environment: Field notes from eight urban, public-funded nursery and primary schools. Working paper. Bangalore: The Promise Foundation.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an Alphasyllabary: The case of Kannada. *Writing Systems Research, 2*, 41–52.

Nakamura, P. R., Joshi, R. M., & Ji, X. R. (n.d.). *Investigating the assymetrical roles of syllabic and phonemic awareness role in akshara processing*. Retrieved from http://www.linguaakshara.org/yahoo_site_admin/assets/docs/Nakamura_et_al_Final.58220626.pdf.

Nation, K., & Snowling, M. (2017). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*(3), 359–370.

Patel, P. G. (2004). *Reading acquisition in India: Models of learning and dyslexia*. New Delhi: Sage Publishers.

Pinto, C. (2010). *Literacy boost Kailali, Nepal year 1 report*. Save the Children.

Purushothama, G. (1990). Reading vowels in Kannada script. *Journal of Learning Disabilities, 23*(3), 198–200.

Salomon, R. (2000). Typological observations on the Indic script group and its relationship to other alphasyllabaries. *Studies on the Linguistic Sciences, 30*(1), 88–103.

Sharma, R. (1997). Dynamics of learning three R's in Madhya Pradesh. *Economic and Political Weekly, 32*(17), 891–901.

Singh, A. (2014). Test score gaps between private and government sector students at school entry age in India. *Oxford Review of Education, 40*(1), 30–49.

Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education, 6*(1), 5–10.

Snow, C. E. (2017). The role of vocabulary versus knowledge in children's language learning: A fifty-year perspective. *Infancia y Aprendizaje, 40*(1), 1–18.

Teale, W. H., & Sulzby, E. (1986). Introduction: Emergent literacy as a perspective for examining how children become writers and readers. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading* (pp. vii–xxv). New Jersey: Ablex Publishing Corporation.

Treiman, R., & Kessler, B. (2013). Learning to use an alphabetic writing system. *Language Learning and Development, 9*(4), 317–330.

Vagh, S. B. (2009). Learning at home and at school: A longitudinal study of Hindi language and emergent literacy skills of young children from low-income families in India. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 70*(11-A), 4183.

Vagh, S. B., & Biancarosa, G. (2011, July). *Early literacy in Hindi: The role of oral reading fluency*. Poster presented at the 12th International Congress for the Study of Child Language, Montreal, Canada.

Vagh, S. B., & Nag, S. (2014). *A template for the analysis of primary school language textbooks in Indian languages*. Retrieved from http://www.linguaakshara.org/yahoo_site_admin/assets/docs/Textbook_Analysis_Template_Vagh_and_Nag.58235348.pdf.

Vagh, S. B., & Sharma, D. (2018). How can localized assessments enhance our understanding of learning at the 'bottom of the pyramid'? Insights from Pratham's collaborative, longitudinal action research program. In D. A. Wagner, S. Wolf, & R. B. Boruch (Eds.), *Learning at the Bottom of the Pyramid: Science, measurement and policy in low-income countries* (pp. 121–140). Paris, France: UNESCO-IIEP.

Vagh, S. B., Nag, S., & Banerji, R. (2017). India: The policy and practice of early literacy acquisition in the akshara languages. In N. Kucirkova, C. E. Snow, V. Grover, & C. McBride (Eds.), *The Routledge international handbook of early literacy education: A contemporary guide to literacy teaching and interventions in a global context* (pp. 233–246). New York: Routledge.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690.

Wijayathilake, M. A. D. K., & Parrila, R. (2014). Predictors of word reading skills in good and struggling readers in Sinhala. *Writing Systems Research, 6*(1), 120–131.

Winskel, H., & Iemwanthong, K. (2010). Reading and spelling acquisition in Thai children. *Reading and Writing, 23*(9), 1021–1053.

The Role of Phonological Processing and Oral Language in the Acquisition of Reading Skills in Devanagari

Nandini Chatterjee Singh and Sumathi T. A

Abstract We examined the relative contributions of phonological processing and oral language on word and non-word reading abilities in Devanagari in 230 children from grades 1 to grade 5. Phonological skills were assessed using tasks of rhyming and syllable replacement while oral language was assessed using semantic and verbal fluency. Phonological awareness made independent contributions to word reading across all grades. In addition, across grades 1 and 2, akshara recognition significantly explained word reading while in grades 3–5, oral language fluency was uniquely associated with word reading performance. Our study provides new insights on reading acquisition in Devanagari and adds to a growing body of literature on reading in akshara orthographies.

Keywords Akshara · Devanagari · Fluency · India · Oral · Orthography · Phonological

Introduction

In recent years, there has been an upsurge of studies that investigate word recognition in akshara-based orthographies prevalent across South and Southeast Asia (Nag & Perfetti, 2014; Share & Daniels, 2016). Findings from these studies have suggested that the processing of akshara-based orthographies, which have origins in Brahmi, are neither syllabic nor phonemic but unique in many ways (Share & Daniels, 2016). A vast variety of Indic languages use the akshara based writing system and past research focused on reading in such systems has primarily explored how the structure of the writing system influences reading in akshara based

N. C. Singh (✉)
Language, Literacy and Music Laboratory, National Brain Research Centre, Gurgaon, Haryana, India
e-mail: nandini@nbrc.ac.in

Sumathi T. A
R & D Engineer(Project), Language, Literacy and Music Laboratory, National Brain Research Centre, Gurgaon, Haryana, India

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_14

languages. However, relatively little research has focused on the role of oral language, in learning to read in akshara based orthographies. It is well known that the successful acquisition of reading in any language is jointly explained by spoken language and its writing system (Perfetti, 2003). Here we examine how the relative contributions of features of oral language and writing systems modulate word recognition in Devanagari, an akshara-based orthography.

The writing system that a language uses affects acquisition of literacy because each system is based on a different set of symbolic relations and requires different cognitive skills (Coulmas, 1989). Devanagari is an akshara based orthography, reported to be accessed by a population of over 200 million Devanagari readers (Census 2011). Devanagari is used to write several languages in North India including Hindi, Sanskrit, Marathi, Nepali, Bihari, and Konkani. The basic symbolic unit in Devanagari is an akshara (Bright, 2000) which can be a vowel, or a consonant with either an inherent vowel or a diacritic representing the vowel, or a combination of two or three consonants with a vowel (Karanth, 2006). Basic consonant aksharas represent CV syllables with an inherent /ʌ/ or schwa. Word-initial vowels are written using basic vowel aksharas, while others are typically represented by diacritics written to the right, left, top or bottom of the preceding consonant. Geminates and conjunct consonants are represented by ligaturing a secondary form of one consonant onto the basic akshara of the other. For example, the basic aksharas द –/d̪ʌ/ and न – /nʌ/ combine with vowel diacritics ू and ी (/u:/ and /i:/ respectively) to produce दून (/d̪u:n/, meaning *valley*) and नदी (/nʌd̪i:/, *river*), while the word-initial /u:/ is written using the akshara ऊ in ऊन (/u:n/, *wool*). The secondary form न् representing न is ligatured onto द in नन्द (/nʌnd̪/, *joy*), and the secondary form ◌ी (standing for र, /rʌ/) is ligatured onto श (/ʃʌ/) in दर्शन (/d̪ʌrʃʌn/, *view*) (Rao & Singh, 2015). The most prominent feature of Devanagari is its visuospatially complex layout, arising from the nonlinear arrangement of vowels as well as consonant ligatures (see examples above). Similar to other Indic writing systems, successful phonological retrieval in Devanagari requires accurate deciphering of visuospatial details.

Psycholinguistic studies investigating the nature of phonological processing in Devanagari have been scarce, and the sole study on segmentation in Devanagari by native Hindi speakers has suggested a partly phonemic and partly syllabic level of segmentation (Vaid & Gupta, 2002). Children who begin to read in Devanagari are exposed to aksharas as whole units and do not receive explicit instruction in phonemic segmentation but still learn to read successfully. Assessment of the orthography of a writing system typically is done through tasks of phonological processing and letter identification while spoken language is assessed using measures of picture naming and fluency. In the present study, we used measures of rhyming and syllable replacement to assess phonological processing and semantic and verbal fluency, to ascertain how features of oral language and phonological processing influence word recognition in Devanagari in children between 5 and 10 years of age.

The role of phonological awareness in learning to read alphabetic languages is now well established, and it is known that phonological awareness proceeds from large units to smaller units —namely from awareness of syllables to phonemes (Carroll, Snowling, Hulme, & Stevenson, 2003; Treiman & Breaux, 1982). However, in order to develop robust phonological skills, early readers of alphabetic languages need to acquire letter knowledge and phoneme awareness (Byrne, 1998; Share, 1995). Only then will they be able to efficiently map the units of print to those of sound. Since phonemic information is embedded and often not explicitly taught in the classroom, an alternative view has suggested that phoneme awareness is a consequence rather than a precursor of learning to read (Castles & Coltheart, 2004; Goswami & Bryant, 1990; Morais, Carey, Alegria, & Bertelson, 1979).

However, the precise role of phonological awareness in reading akshara-writing systems continues to emerge and is a subject of active interest and research. The structure of the akshara orthographies is such that they primarily encode sounds at the syllable level. The sub-syllabic information is encoded in matras or diacritics and does not have a uniform representation. The matras are 'vowel phonemic markers' and in spatial representation may be located below or after, above or below the consonant. Further, when represented as diacritics, the 'matras' do not clearly resemble their original representations, and this adds to the symbolic inventory that needs to be learned during the process of reading acquisition. Studies that have investigated these complex ortho-phonological properties of aksharas have shown that both syllable and phoneme awareness are predictive of word decoding (Joshi, 2014; Nag, 2007; Reddy & Koda, 2013; Sircar & Nag, 2013; Tiwari, Nair, & Krishnan, 2011), but that syllabic awareness is acquired more rapidly as compared to sub-syllabic information (Nag, 2007; Vasanta, 2004). Interestingly, studies with monolingual Oriya (another akshara orthography) speakers have indicated that only syllable awareness was predictive of word reading (Mishra & Stainthorp, 2007). However Reddy and Koda (2013), who found that both syllabic and phoneme awareness predicted Kannada decoding, also reported that phoneme-level awareness did not contribute unique variance to decoding when introduced after syllable-level awareness in a step-wise regression model of Kannada reading. These studies seem to suggest that while syllabic and phonemic awareness influence word decoding in Devanagari, the role of syllabic awareness is more important, especially during early reading acquisition. These studies also suggest that the precise nature of the syllabic and phonemic relationship in akshara decoding may vary across Indic languages. Thus the two primary challenges while learning to read akshara orthographies are (1) the large inventory of symbols/syllabographs that are required to be learned (Nag, 2007) estimate this to be as high as 400 in Kannada and (2) the complex visuo-spatial representation of the script. At the same time, a distinct advantage in akshara orthographies include the consistency of akshara-sound mapping and the fact that the basic sound unit is a syllable that is easily identifiable. The acquisition of a large symbol set in addition to the complex visuo-spatial nature of the writing system prolongs the process of word reading (Nag, 2007) and prompts one to speculate that relative contributions of phonological and orthographic processes might be age- and instruction-dependent.

Finally, it is universally acknowledged that the successful acquisition of reading in any language is jointly explained by spoken language and its writing system (Perfetti, 2003). Thus, a third factor that will also impact reading skill and later reading comprehension is vocabulary or oral language skill (Ouellette, 2006; Ouellette & Beers, 2010; Perfetti & Hart, 2002; Scarbourgh, 2001; Sénéchal, Ouelette, &

Rodney, 2006). Past studies on decoding in akshara orthographies have been unable to establish a clear pattern for oral language in reading acquisition. According to the Simple View of Reading (SVR), the ability to decode words into constituent sounds and language comprehension are two necessary skills to achieve reading comprehension (Gough & Tunmer, 1986; Hoover & Gough, 1990). Past research has shown that each of these competencies makes independent contributions to reading comprehension (Kendeou, van den Broek, White, & Lynch, 2009). While in the early grades the relative contribution of decoding skills is higher, language comprehension skills become increasingly important after decoding is proficient (Byrne, Freebody, & Gates, 1992; Florit & Cain, 2011; Hoover & Gough, 1990; Tilstra, McMaster, van den Broek, Kendeou, & Rapp, 2009). We therefore propose that an investigation of word recognition in an akshara system in young children should include an assessment of the relative roles of vocabulary and phonological processing. In light of the consistency of akshara-sound consistency, coupled with the visuo-spatial complexity of the akshara orthography, we hypothesized that for beginning readers, akshara recognition should take precedence over phonological processing and oral fluency in explaining word recognition. However, as children achieve akshara recognition skills, we speculated that oral vocabulary skills and phonological processes would independently contribute to word recognition.

Methods

Participants

Devanagari is used as a writing system for multiple Indian languages (e.g., Hindi, Marathi, Konkani, Bihari and Nepali), and data for this study were collected from young learners of both Hindi and Marathi. The forms of Devnagari used by Hindi and Marathi are identical in every way except for the addition of one akshara /ळ/ used in the Marathi language. The data presented in this study were part of a larger scale survey of literacy skills conducted in primary and middle schools in the cities and neighboring districts of the National Capital Region (New Delhi and its surrounding areas) and the city of Allahabad located in the state of Uttar Pradesh in the northern part of India. The Marathi data were collected from Mumbai and Pune in the western part of India. Hindi is the dominant language in areas in and around New Delhi while Marathi is the dominant language of Maharashtra to which both Mumbai and Pune belong. Consent for the study was obtained from parents of the children as per the norms prescribed by the Human Ethics Committee of the National Brain Research Centre.

Data were collected from 28 schools. In order to reflect the profile of the socio-economic status (SES) in these regions, the survey sample comprised children from a mix of schools: 20.2% from upper SES, 70% from middle and 9.8% from lower SES. Schools in the Delhi region were under a common curriculum, namely, the Central Board for Secondary Education (Department of Elementary Education and Literacy & Department of Secondary and Higher Education, 2005) while schools in

Table 1 Demographic information by language

	Hindi	Marathi
Number of schools	18	9
Number of participants	130	100
Gender information	64 F, 66 M	55 F, 45 M
Grade 1–2	39	36
Grade 3–5	91	64

the Mumbai and Pune regions were under another common curriculum, namely, the State Board Curriculum (Department of Elementary Education and Literacy & Department of Secondary and Higher Education, 2005) However, to a large extent, schools followed a similar reading program. In these schools, formal literacy instruction for Hindi and Marathi began in grade 1, between the ages of 5 and 6 years. As per the education policy, many of the schools were also required to introduce English in grade 1. Data were collected from children across grades 1–5 (Grades 1–2 n = 75, M_{age} = 7.1, SD = 0.8; Grades 3–5 n = 155, M_{age} = 9.6, SD = 1.1). A detailed description of the sample is provided in Table 1. The analysis presented here describes the children who displayed no reading difficulties and formed a chronological group from each grade. An exhaustive battery of tests was administered for every child participating in the survey and is described in the next section.

Assessments

General Cognitive Ability

This was assessed using Raven's Coloured Progressive Matrices. General ability is reported in percentile scores based on Indian norms (Raven, 2004).

The battery of behavioral tests may be broadly divided into three groups, namely, reading (akshara, word list 1, word list 2, and non-word), phonological processing (rhyme and syllable replacement), vocabulary (verbal fluency and semantic fluency) and picture naming. While the phonological processing, vocabulary, and picture naming tests were common across all the children, the reading tasks were varied. Grades 1 and 2 read completed akshara reading and word list 1, while grades 3 to 5 completed word list 2 and non-word reading.

Literacy Skills

Akshara Knowledge

Children read 10 aksharas comprising no specific phoneme markers. The list of aksharas in Hindi and Marathi was identical, and included 3 vowels and 7 consonants. Cronbach's alpha was .68.

Word Reading 1(Grade 1–2)

Each child was asked to read a list of 25 words arranged in increasing order of difficulty. The list contained 3 syllabographs monosyllabic words, 13 bisyllabic words, 6 trisyllabic words, and 4 polysyllabic words. Correct reading of an item was given an accuracy score of 1. Cronbach's alpha was 0.90.

Word Reading 2 (Grade 3–5)

Each child was asked to read aloud a list of 50 words and all words were arranged in order of increasing difficulty. Cronbach's alpha was 0.90.

Non Word Reading (Grade 3–5)

Each child was asked to read aloud a list of 30 words. All items had been constructed by replacing one or more akshara and/or the consonant or vowel component or a real word in the respective language. The non word list was arranged in order of increasing difficulty of decoding. Cronbach's alpha was 0.82.

Phonological Awareness Tasks

Rhyming

A total of 12 sets of triplet of words were presented in each language (e.g., Hindi: नाम(nam)-काम(kam)-नील(nil)); Marathi: गोल(gol)-मोल(mol)-तोड़(tod)). Participants were required to listen 3 set of words carefully, and then say the two words that end with the same sound. Cronbach's alpha was 0.81.

Syllable Replacement

In the syllable replacement task, participants were instructed to replace the syllable at the beginning or end of a given word in Hindi or Marathi. The task consisted of 10 items in each language. Cronbach's alpha was 0.77.

Picture Naming

This test was common across the grade 1–5. The objective of the picture naming task was to assess the speed of retrieval and production of the phonology (sound) of whole words. The Picture Naming Tasks in both Hindi and Marathi made use of line drawings of five common objects (shoe, flower, house, chair, and key) where the participant had to identify and name the objects shown on the card, using only the task language.

Fluency

This test was common across grades 1–2 and grades 3–5. The Fluency Test had two sub-tests, namely, Verbal Fluency Test and Semantic Fluency Test.

Participants were given 1 min to produce as many unique words as possible within a semantic category (animals and vegetables) or starting with a given akshara (verbal fluency). Fluency tasks have emerged as important tools to assess both lexical access ability as well as executive control ability. These tests are designed to assess both verbal (sound-based) and semantic (meaning-based) long term memory. These tests also provide the information on the vocabulary of the participants. The sound and category fluency task differ in subtle but important ways in task demands. While the category fluency task may be compared to a shopping list, that can exploit existing links between related concepts (e.g., between the category label and the category members and among associated category members) the sound (verbal) fluency task, requires words with a same initial sound to be retrieved. Thus participants must consciously inhibit semantically or associatively related words and must instead employ novel retrieval strategies (e.g., Luo et al., 2010; Katzev et al., 2013).

Procedure

Each child was assessed individually for approximately 90 min (the testing battery included tests of oral language, picture naming, reading, phonological processing, dictation, and reading comprehension (some of which are not reported here) by a clinical psychologist and four research assistants who were native speakers of Hindi and/or Marathi and were trained to administer the tests.

Results

Only children who completed all 9 tasks were included in the analysis. All children from grades 3–5 were at ceiling level for the akshara identification task, and this task has not been included in the analysis for grades 3–5. Table 1 describes data by language and grade.

Based on the tests administered, data children from grades 1–2 (mean age = 7.1, SD = 0.8)) were combined into group 1 and those from grades 3–5 (mean age = 9.6,

Table 2 Descriptives of picture naming, phonological processing, and reading measures

	Grade 1–2, N = 75		Grade 3–5, N = 155		t-test
Measure	*Mean*	*SD*	*Mean*	*SD*	*P value*
CPM	106.7	11.1	109.2	9.8	NA
Picture naming (time in seconds)	61.4	17.4	52.0	10.2	<0.001
Semantic fluency	14.6	4.7	18.7	5.3	<0.001
Verbal fluency	9.8	4.0	14.1	5.2	<0.001
Rhyme (12)	10.1	1.9	11.1	1.3	<0.001
Syllable replacement (10)	8.4	1.6	9.3	1.2	<0.001
Letter Reading (10)	9.3	1.1	NA	NA	NA
Word Reading (25) (50)	21.4	4.4	44.0	6.5	NA
Non word Reading (30)	NA	NA	24.9	4.8	NA

Note. NA-Not applicable.

SD = 1.1) were combined to form group 2. We describe in Table 2 means and standard deviations of groups 1 and 2, respectively, on all tasks. The data reported here are for only those participants who showed reading accuracies above 60 percent in their grade. The picture naming scores are reported in terms of time taken to complete the task in seconds.

Bivariate Correlation Analysis

In order to ascertain sub-skills which might theoretically play a causal role in determining reading performance in word and non-word reading, we first examined correlations between various behavioral skills and reading performance. Bivariate correlations across all measures were performed and revealed several significant trends shown in Tables 3, 4, and Fig. 1. Fluency skills of semantic and verbal fluency were all highly correlated. Phonological skills of syllable replacement and rhyming also showed a strong significant correlation. Correlations between measures of fluency and phonological processing suggested that while fluency and phonological processing both explained word reading, phonological processing might be a stronger correlate as compared to fluency.

In terms of phonological skills, rhyming and syllable replacement both explained Hindi word reading and non-word reading almost equally. To understand this phenomenon better, hierarchical regression analyses were computed.

The correlational analyses clearly established that akshara reading, phonological processing, and fluency were moderately associated with reading and decoding outcomes in Devanagari. In order to test the relative contributions of each sub-skill, we conducted stepwise regression analyses. Since the bivariate correlational analyses showed strong correlations among semantic and verbal fluency and syllable awareness and rhyming, composite measures for fluency and phonological awareness (PA) were obtained. Fluency was an average measure of semantic and verbal fluency

Table 3 Correlations among measures for group 1 (Grade 1–2, mean age M 7.1, SD = 0.8)

	Picture naming	Semantic fluency	Verbal fluency	Rhyme	Syllable replacement	Letter reading	Word reading
Picture naming	1						
Semantic fluency	−.26*	1					
Verbal fluency	−.13	.38**	1				
Rhyme	.01	.09	.25*	1			
Syllable replacement	.02	.08	.32**	.32**	1		
Letter reading	−.19	.23*	.30**	.05	.37**	1	
Word reading	−.27*	.42**	.35**	.23*	.29*	.67**	1

Note. *p < 0.05, **p < 0.01.

Table 4 Correlation measures for group 2 (Grade 3–5, mean age = 9.6, SD = 1.1)

	Picture naming	Semantic fluency	Verbal fluency	Rhyme	Syllable replacement	Word reading	Non word reading
Picture naming	1						
Semantic fluency	−.26**	1					
Verbal fluency	−.19*	.31**	1				
Rhyme	−.11	.24**	.10	1			
Syllable replacement	.10	.16*	.14	.23**	1		
Word reading	−.11	.17*	.25**	.26**	.30**	1	
Non word reading	−.12	.16*	.18*	.29**	.25**	.82**	1

Note. *p < 0.05, **p < 0.01.

scores and PA was an average measure of rhyme and syllable replacement scores. Composite measures were used for fluency and phonological awareness (PA) as unique correlates for word reading.

In all stepwise regression analyses, non-verbal intelligence was entered as a covariate at the first step. Table 5 describes the hierarchical regression models for grades 1–2. It was found that irrespective of the order of entry, letter reading, fluency and phonological awareness (PA) were significant correlates of word reading.

Table 6 describes the hierarchical regression models for grades 3–5. Irrespective of the order of entry, both phonological awareness (PA) and fluency were significant predictors of Devanagari decoding. The same effects were seen with non-word reading too as reflected in Table 7. It may therefore be proposed that learners' awareness of the syllabic akshara is significantly related to the decoding ability in Devanagari.

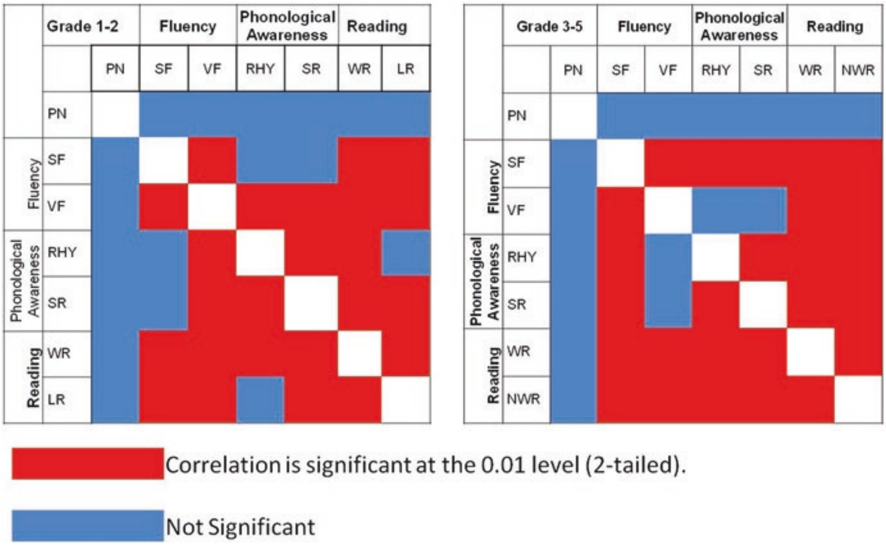


Fig. 1 Correlations for Grades 1–2 and Grades 3–5
Note. PN Picture Naming, *SF* Semantic Fluency, *VF* Verbal Fluency, *LR* Letter Reading, *WR* Word Reading, *NWR* Non Word Reading

Table 5 Stepwise regression models predicting word reading for grade 1–2

	ΔR^2	ΔF	p
Model 1			
CPM	.00	.31	.58
Letter reading	.45	59.71	.00
Fluency	.07	10.88	.00
Phonological awareness	.02	2.42	.12
Model 2			
CPM	.00	.31	.58
Letter reading	.45	59.71	.00
Phonological awareness	.03	4.32	.04
Fluency	.06	8.73	.00
Model 3			
CPM	.00	.31	.58
Fluency	.22	20.29	.00
Phonological awareness	.04	3.97	.05
Letter reading	.28	42.88	.00
Model 4			
CPM	.00	.31	.58
Phonological awareness	.10	7.92	.01
Fluency	.16	15.57	.00
Letter reading	.28	42.88	.00

Table 6 Stepwise regression models predicting word reading for grade 3–5

	ΔR^2	ΔF	P
Model 1			
CPM	.01	.95	.33
Fluency	.06	10.29	.00
Phonological awareness	.09	15.84	.00
Model 2			
CPM	.01	.95	.33
Phonological awareness	.12	21.23	.00
Fluency	.03	5.30	.02

Table 7 Hierarchical regression analysis predicting non word reading for grade 3–5

	ΔR^2	ΔF	P
Model 1			
CPM	.01	1.14	.29
Fluency	.04	6.96	.01
PA	.09	15.57	.00
Model 2			
CPM	.01	1.14	.29
PA	.12	19.95	.00
Fluency	.02	2.99	.09

It is noteworthy to mention that fluency also contributed significantly to word recognition in grades 3–5.

Discussion

This study investigated the predictors for reading among 5- to 11-year-old children, divided into two groups based on grade levels, learning to read in Devanagari, an akshara-based writing system. Hindi and Marathi both use the Devanagari script, which is a consistent orthography yet at the same time requires an extensive set of visuo-spatially complex symbols and syllabographs to be mastered. However given the visuo-spatial complexity of the symbols which form a large set, learning to read in Devanagari is quite challenging. Our first finding was that picture naming, phonological skills, and fluency showed distinct age effects. Since identical tasks were used across groups, we were able to estimate age effects and not surprisingly we found significantly increased performance for rhyming and syllable replacements across groups 1 and 2. Similarly, vocabulary was significantly larger in children belonging to group 2 as compared to group 1.

The primary objective of this study was to delineate the roles of phonological processes, and oral fluency in explaining reading performance in Devanagari. The most consistent finding of our study was the finding that oral fluency or vocabulary emerged as the most consistent independent correlate for word reading in Devanagari. Bivariate correlational analyses followed by hierarchical regression analysis indicated that oral fluency emerged as an important independent predictor for word reading skills across groups 1 and 2. However for group 1, as seen in Table 5, the variance in word reading was primarily determined by akshara knowledge but in group 2, it was determined by phonological processing. Given the transparent nature of the akshara-sound mapping, the finding that phonological awareness was a strong predictor for reading was not surprising. The finding that PA explained both word and non-word reading almost equally confirms its role in decoding in consistent orthographies.

With regard to fluency, a noteworthy finding was its role in explaining the variance in word reading but little change in the prediction of non-word reading performance. The role of oral language fluency in reading fluency has been well established (NICHD, 2005). The results of a longitudinal study wherein children were followed from age 3 through third grade, suggested that oral language broadly plays both a direct and an indirect role in word recognition and during the transition to school and serves as an excellent foundation for early reading skill (NICHD, 2005).

In summary, our results corroborate findings from other recent akshara reading studies particularly from Kannada which indicate an important role for phonological information encoded at the syllabic level in akshara systems (Nag & Snowling, 2012, Nakamura, Koda, & Joshi, 2013). However our second finding of a role for oral language is noteworthy though not surprising. In recent years a large body of work has revived an interest in the role of oral language in reading acquisition. A recent review by Nag et al. (2014) commissioned to address issues pertaining to foundation learning and literacy, from early childhood to Grade 8 (approximately 3–13 years) concluded that children with poor oral language were at a high risk for reading and education failure that included both numerical and scientific thinking. Critical to recognizing words is the ability to connect graphemic units to phonological segments. However, decoding the pronunciation of a word may not yield lexical understanding, and, therefore, vocabulary knowledge is also an important skill in learning to read.

Finally we evaluate our findings in the terms of psycholinguistic grain size theory proposed by Ziegler and Goswami (2005). One of the primary advantages of the akshara based writing system has been the availability of orthographic units or writing symbols at the level of the akshara. The syllable provides a concrete/tangible platform for the early reader who can readily map a spoken syllable to an orthographic unit. This process is facilitated further by the transparency of the mapping. The sound-symbol mapping in Devanagari is consistent to a large extent. This is beneficial during early learning because if akshara identification is systematic and thorough, children begin to appreciate early on the fun and joy of learning to read. However as pointed out by other studies (Nag, 2007; Nag, Treiman, & Snowling, 2010), it is the mastery of the low-frequency symbols that sets the pace of development in reading and spelling.

A limitation of this study was the absence of a specific task on phonemic awareness. Past research on akshara orthography has suggested that only syllabic awareness (Mishra & Stainthorp, 2007) is important in akshara decoding, while others have demonstrated that phonemic awareness correlates with reading and spelling scores (Nag, 2007); others have suggested that the role of phonemic awareness reduces with age. Unfortunately this study was unable to resolve this debate in the context of Devanagari.

A second limitation of this study was the absence of a task on orthographic processing. The acquisition of decoding skills is necessary not only for processing novel words in reading but also to generate orthographic strings in spelling. Thus in order to achieve skilled visual reading, children are also required to achieve visual processing and discrimination abilities. While children's orthographic knowledge and letter knowledge are causal factors in subsequent reading development in English (e.g., Badian, 1994; Lonigan et al., 2000), phonological processing abilities have emerged as the best predictors for reading in English. However, visual skills have emerged as stronger and more reliable predictors for reading in Chinese when compared to English (e.g., Huang and Hanley, 1994, 1995; Ho and Bryant, 1997; Siok and Fletcher, 2001; Mcbride-Chang and Ho, 2005; Luo et al., 2013). Given the visual complexity of the aksharas of Devnagari, it is crucial that that visual processing be examined in greater detail in Devnagari. We hope future research will attempt to include tasks on orthographic processing to provide a more holistic picture of reading in Devnagari. Finally, the data used in this study were cross-sectional and could therefore not provide causal links between reading and other related cognitive skills. Despite these limitations, this study adds to a growing body of literature on the akshara orthographies (Nag & Perfetti, 2014; Share & Daniels, 2016).

Acknowledgements The authors would like to acknowledge M. Joshi for valuable comments on the manuscript, Geet Oberoi (Orkids Centre for Learning Disabilities), Bhoomika R. Kar (University of Allahabad), Kate Currawala (Maharashtra Dyslexia Association) and Kshipra Vaidya (Shanta Vaidya Memorial Foundation) for data collection and Cognitive Science Initiative of the Department of Science and Technology, India, for financial support.

References

Badian, N. A. (1994). Preschool prediction: Orthographic and phonological skills, and reading. *Annals of Dyslexia, 44*(1), 1–25.

Bright, W. A. (2000). Matter of typology: Alphasyllabaries and Abugidas. *Studies in the Linguistic Sciences, 30*, 63–71.

Byrne, B., Freebody, P., & Gates, A. (1992). Longitudinal data on the relations of word-reading strategies to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly, 27*(2), 140–151. https://doi.org/10.2307/747683.

Byrne, B. J. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle*. East Sussex, UK: Psychology Press.

Carroll, J., Snowling, M. J., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in pre-school children. *Developmental Psychology, 39*, 913–923.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77–111.

Coulmas, F. (1989). *The writing systems of the world*. Oxford, UK: B. Blackwell.

Department of Elementary Education and Literacy, Department of Secondary and Higher Education. (2005). *Annual report*. New Delhi, India: Government of India.

Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review, 23*, 553–576. https://doi.org/10.1007/s10648-011-9175-6.

Goswami, U., & Bryant, P. E. (1990). *Phonological skills and learning to read*. London, UK: Erlbaum.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading ability. *Remedial and Special Education, 7*(1), 6–10. https://doi.org/10.1177/074193258600700104.

Ho, C. S. H., & Bryant, P. (1997). Development of phonological awareness of Chinese children in Hong Kong. *Journal of Psycholinguistic Research, 26*(1), 109–126.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*(2), 127–160. https://doi.org/10.1007/BF00401799.

Huang, H. S., & Hanley, J. R. (1994). An overview of research on phonological awareness and reading ability. *Journal of National Taipei Teachers College, 7*, 115–168.

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54*(1), 73–98.

Joshi, R. M. (2014). Literacy in Kannada, an alphasyllabic orthography. In *South and Southeast Asian psycholinguistics* (pp. 202–211). Cambridge, UK: Cambridge University Press.

Karanth, P. (2006). The Kagunita of Kannada—Learning to read and write an Indian alphasyllabary. In *Handbook of orthography and literacy* (pp. 389–404).

Katzev, M., Tüscher, O., Hennig, J., Weiller, C., & Kaller, C. P. (2013). Revisiting the functional specialization of left inferior frontal gyrus in phonological and semantic fluency: The crucial role of task demands and individual ability. *Journal of Neuroscience, 33*(18), 7837–7845.

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765–778. https://doi.org/10.1037/a0015956.

Mishra, R., & Stainthorp, R. (2007). The relationship between phonological awareness and word reading accuracy in Oriya and English: A study of Oriya-speaking fifth-graders. *Journal of Research in Reading, 30*, 23–37. https://doi.org/10.1111/j.1467-9817.2006.00326.x.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596.

Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition, 114*(1), 29–41.

Luo, Y. C., Chen, X., Deacon, S. H., Zhang, J., & Yin, L. (2013). The role of visual processing in learning to read Chinese characters. *Scientific Studies of Reading, 17*(1), 22–40.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323–331.

McBride-Chang, C., & Suk-Han Ho, C. (2005). Predictors of beginning reading in Chinese and English: A 2-year longitudinal study of Chinese kindergartners. *Scientific Studies of Reading, 9*(2), 117–144.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22.

Nag, S., Chiat, S., Torgerson, C., & Snowling, M. J. (2014). *Literacy, foundation learning and assessment in developing countries*. DFID publication retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/305150/Literacy-foundation-learning-assessment.pdf.

Nag, S., & Perfetti, C. A. (2014). Reading and writing: Insights from alphasyllabaries of south. *Writing Systems Research*, 1–9. https://doi.org/10.1080/17586801.2014.883787.

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a universal theory of learning to read. *Scientific Studies of Reading, 16*(5), 402–423. https://doi.org/10.1080/10888438.2011.576352.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 1*(2), 1–12.

Nakamura, P. R., Koda, K., & Joshi, R. M. (2013). Biliteracy acquisition in Kannada and English: A developmental study. *Writing Systems Research, 6*(1), 132–147.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*(2), 428.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554–566. https://doi.org/10.1037/0022-0663.98.3.554.

Ouellette, G. P., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading & Writing: An Interdisciplinary Journal, 23*, 189–208. https://doi.org/10.1007/s11145-008-9159-1.

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*, 3–24.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam, The Netherlands/Philadelphia, PA: John Benjamins Publishing Company.

Rao, C., & Singh, N. C. (2015). Visual complexity modulates reading in the brain. *Brain and Language, 141*, 50–61.

Raven, J. (2004). *Progressive matrices and Crichton vocabulary scale*. London, UK: Pearson.

Reddy, P., & Koda, K. (2013). Orthographic constraints on phonological awareness in biliteracy development. *Writing Systems Research, 5*, 110–130. https://doi.org/10.1080/17586801.2012.748639.

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis) abilities:Evidence, theory, and practice. In S. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). New York, NY: Guilford Press.

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In S. B. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 173–182). New York, NY: Guilford Press.

Share, D., & Daniels, P. (2016). Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). *Writing Systems, 8*, 17–31. https://doi.org/10.1080/17586801.2015.1016395.

Share, D. L. (1995). Phonological recoding and self teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology, 37*(6), 886.

Sircar, S., & Nag, S. (2013). Akshara-syllable mappings in Bengali: A language-specific skill for reading. In *South and Southeast Asian* (pp. 202–211). Cambridge, UK: Cambridge University Press.

Tilstra, J., McMaster, K., van den Broek, P., Kendeou, P., & Rapp, D. (2009). Simple but complex: Components of the simple view of reading across grade levels. *Journal of Research in Reading, 32*, 383–401. https://doi.org/10.1111/j.1467-9817.2009.01401.x.

Tiwari, S., Nair, R., & Krishnan, G. (2011). A preliminary investigation of akshara knowledge in the Malayalam alphasyllabary: Extension of Nag's (2007) study. *Writing Systems Research, 3*(2), 145–151.

Treiman, R., & Breaux, A. M. (1982). Common phoneme and overall similarity relations among spoken syllables: Their use by children and adults. *Journal of Psycholinguistic Research, 11*, 569–598.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*(1–3), 679–690. https://doi.org/10.1006/brln.2001.2556.

Vasanta, D. (2004). Processing phonological information in a semi-syllabic script:Developmental data from Telugu. *Reading and Writing: An Interdisciplinary Journal, 17*, 59–78. https://doi.org/10.1023/B:READ.0000013830.55257.3a.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3.

Part IV
Instruction

Language-Focused Instruction for Literacy Acquisition in *Akshara*-based Languages: Pedagogical Considerations and Challenges

Chandrika Mathur and Sonali Nag

Abstract In this chapter we explore the role of language- and meaning-focused instruction for supporting literacy acquisition in *akshara*-based languages, drawing upon examples from Hindi and the Devanagari script. Instruction of lower- and higher-level literacy skills typically targeted in pre- and primary school years are examined with a particular focus on the phonological, orthographic and morpho-syntactic domains. The role of affective-motivational aspects in learning and a strong cultural embedding of teaching-learning materials and pedagogical strategies are highlighted, with implications for native speakers and second language learners, in contexts of diglossia, and with demographically diverse groups. Such a multi-component curriculum is dependent on the perceptiveness and sensibilities of teachers, their attitudes towards teaching and learning, specific pedagogical skills and concrete knowledge-bases. We discuss the challenges this holds.

Keywords Akshara knowledge · Cultural embedding · Devanagari · First language · Hindi · Literacy instruction · Morpho-syntax · Pedagogy · Reading · Second language · Teaching-learning materials (TLMs) · Writing · Teacher training

The UN-led Jomtein Summit (World Declaration on Education for All, 1990) was a spur for the widespread emergence of pre-schools across South and Southeast Asia. With this the age of formal instruction in the languages of the region moved to around the age of three in an ever growing number of communities. Instruction in languages that use the *akshara* writing system thus usually commences at an age when several events take place simultaneously in a child's life: moving away from

C. Mathur
Rishi Valley Institute of Teacher Education, Rishi Valley Education Centre, Krishnamurti Foundation India (KFI), Chittoor District, Andhra Pradesh, India
e-mail: chandrika.rvs@gmail.com

S. Nag (✉)
Department of Education, University of Oxford, Oxford, UK

The Promise Foundation, Bangalore, India
e-mail: sonalinag@t-p-f.org; sonali.nag@education.ox.ac.uk

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_15

the familiarity of the home, entering the unfamiliar environment of a pre-school/school, and the introduction to formal instruction. This transition could turn out to be a trying experience for the child both emotionally and cognitively, or it could become an opportunity for growth and development beyond the confines of the family. The pedagogical challenges and considerations in literacy acquisition explored in this chapter are rooted in this complex range of issues as well as the new possibilities that arise in this transition.

Children across South and Southeast Asia may come to the task of literacy learning either in their first language or in a new language to which they may have little or no prior exposure. Many children may speak a variant of the language of instruction at home and in their community, and may use pronunciation, vocabulary, and syntax that are quite different from the standard language used and taught in school. Many children may have few or no role models at home or in their neighborhood to demonstrate the purpose and functions of reading and writing, or to help with home tutoring. They may have little exposure to print materials such as newspapers, magazines, and books. For younger children, this absence of home support will slow down emergent concepts about print; for older children the absence of opportunity outside the classroom could impose a ceiling on maturing skills. These diverse learning contexts across the region demand that language instruction that is meaning-focused is embedded within literacy programs.

Moreover, if the aim of language and literacy instruction is not just to nurture economically productive members of society but also a human being capable of dealing with the increasingly complex challenges of contemporary life, then early instruction needs to be as much *relational* as it is pedagogical and linguistic. The influential educational philosopher J. Krishnamurti, even as he highlights the interpersonal aspects of the learning process, asks for a thoughtful approach to instruction because "the child's natural curiosity, the urge to learn, exists from the very beginning" (Krishnamurti, 1985, p 12). If nurtured with sensitivity, language and literacy instruction can allow the child to remain curious and also connected to peers, teachers, family and community. Arguably, when teachers put up invisible walls firmly separating the child's school experiences from the home language and home experiences they impact this natural curiosity; the beginnings of literacy learning could then very well become the beginnings of a sense of alienation from community and home culture. Several scholars have described the long term consequences of such rupture, including difficulty in meeting a rapidly changing contemporary world of work (e.g., Arulmani, 2016).

We propose that a language-focused instructional approach supported by authentic interpersonal interactions between the teacher and the child, while useful for teaching *all* children to read and write in the *akshara*-based languages, is particularly useful for children from demographically diverse backgrounds. Such an approach is inherently meaning-focused within a broad curriculum that targets all key aspects of language and literacy learning, including phonological, orthographic and morpho-syntactic awareness, but with sensitivity to the socio-cultural and affective-motivational aspects of literacy development as well.

Literacy Instruction

Typical Literacy Programs

A typical literacy program in classrooms across the region may be as follows: children are taught a set of 'action rhymes' and songs which they recite in a sing-song manner. Simultaneously they are taught to decode, sounding out the symbols they see printed in their textbooks or written on the blackboard; they are then required to copy these out in their notebooks or handheld slates. Patel (2004) in his description of Gujarati classrooms from the 1970s describes 'a linguistic-cultural ecology' for language and literacy instruction imbued with various shades of recitation and sing-song lessons and learning by writing, and an assumption that fluency is sufficient for comprehension. Four decades later, Nag and Sircar (2008) describe similar practices in Bengali classrooms, as do Jayaram (2008) and Saigal (2012) in Hindi classrooms, and Geetha (2012) and Nag et al. (2014a) in Tamil and Kannada classrooms, respectively.

The rhymes and songs in these literacy programs, however, rarely form the basis for acquisition of literacy. The requirement to infer, the acts of thinking, creating, and deriving personal meaning, of making connections and expressing oneself, are most often left to the personal attributes of the learner. These are neither taught nor evaluated. The recitation of rhymes, sounding-out, and decoding-encoding of texts are audible and visible. Children's participation in these activities can be heard and seen and hence readily demonstrated and tested. Perhaps factors such as these have led to an over-emphasis on sounding-out and decoding-encoding.

Relationship of Literacy Practices to Cultural Traditions

Literacy instruction in the *akshara* languages sits uneasily with the rich folk traditions of singing, recitation and the classical traditions of chanting that are integral to the cultural sensibilities in the region. Even though there is remarkable openness to the idea of recitation and memorization among primary school teachers across the region, most either do not know the details of these traditional practices or have little appreciation for their nuances. Within these traditions, attention is paid to how words must be recited—the length of vowels, the rise and fall of syllables within words, the rhythm of verses, and the conditional rules of prosody. There is a focus on memorization of long tracts of texts in these traditions, and this is often made possible by the use of mnemonics and rhythmic chunking of sounds. In Vedic chanting, for instance, the power resides in the very act of vocalizing the sounds; hence, there is a strong emphasis on accurate pronunciation and recitation techniques rather than on the meaning for itself. This is perhaps because the meaning of the *sloka*—the most popular form of verse in Sanskrit poetry which presents itself as a couplet—is usually philosophical, with complexities that the novice young chanter is only likely to imbibe and refine over long years. What needs to be recognized is that the aim of teaching a *sloka* was for introspection about the philosophical messages embedded in the verse and not to teach the student how to read and write.

And yet, a sophisticated vocabulary available from the classical traditions captures the underpinnings of recitation, and this could arguably be deployed in an effective literacy programme. There are terms for the long pauses (*danda*), for shorter pauses the duration of a short vowel (*ekamatra*) and for pauses of even shorter duration that are equivalent to half a short vowel (*ardhamatra*). Every change of meter is reflected in the tune and rhythm that the text is sung in, and this sensitization arguably increases awareness of the meter in poetry, its verses, and its couplets. Syllable weight also plays an important role. For example, in Hindi texts (including texts in Hindi dialects such as Brajbhasha and Avadhi), syllables consisting of consonant and vowels such as a-e-u (अ, इ, उ,) carry a value of one, while a syllable consisting of a consonant and some other vowels such as */aa/–/ee/–/uu/* (आ, ई, ऊ,) is considered double in weight and carries a value of two. In other words, the folk and classical traditions appear to offer specific ways to engage with the phonology and the morphology of *akshara*-based languages.

Most primary school teachers today are, however, unable to deploy such culturally available oral practices effectively for an engaging teaching program. Recitation is usually a sing-song affair with little attention paid to either meaning or the rhythm of the meter (e.g., contemporary: P. Patel, 2004; twentieth century: G. Badheka, 1932/1986). It is unclear when and how such cultural knowledge was lost in mainstream literacy instruction; but one clear historical point of loss was when formal education systems introduced by European colonizers of the region chose to ignore such oral practices, perhaps even devaluing these for their blatant cultural mooring, emphasizing instead evaluation and testing of reading and writing (e.g., Khubchandani, 1981). Since word-level processes (decoding, legibility in writing) are easily tested, these may have gained centrality in classroom practice.

Popular Images of Literacy Instruction

Typical classroom practices of the type described earlier have been rendered eternal in popular media. For example, there is the classroom scene built around *akshara* riddles of a popular song, *Eechak Daanaa, Beechak Daanaa,* in the 1955 Hindi film, *Shri 420*. The teacher reflects shades of a deeply embedded tradition of oral culture across South and Southeast Asia, but her pedagogy of literacy instruction is questionable. The language of the riddles is of a high order, involving connected but complex metaphors. Any first grader, even if a native speaker of Hindi, would struggle with these advanced linguistic devices. Yet, in the film sequence students are shown singing lustily after the teacher and shouting the answer in unison, even as the adult onlooker is shown struggling to make sense of the metaphors in the riddle and find the answer. So presumably, the teacher has provided the children with the answers beforehand and the inferential aspect of the riddles has been put aside. In parallel, the teacher draws a picture of the answer to each riddle on the blackboard. She writes the word under the picture. Then she proceeds to write the first *akshara* of the word above the picture. It turns out that introducing the *akshara* through high frequency, early acquired words, is the objective of her lesson. A pictorial chart of the symbol set of the Devanagari script hangs nearby from a tree, attesting to this.

The pedagogical movement in this illustrative film sequence goes from the singing of the riddle, to providing the answer in multiple modalities (recited, drawn as a picture, and written as a word), to finally highlighting the writing routine for the first *akshara* in the word. Just as the descriptions of the typical language classroom given above indicate, the problem with this—the *Eechak Daanaa Beechak Daanaa* form of pedagogy—is in its assumptions. The teaching practice assumes, for example, that songs for children need not be sensitive to the linguistic and cognitive levels of the children, that the structure inherent in a song or rhyme is sufficient and coherence in the narrative is not necessary, that sing-along is sufficient for insight, and that demonstrating fluency implies understanding. Beyond popular media, this form of pedagogy at best assumes comprehension and at worst pays scant attention to meaning-making.

Pedagogy that prioritizes decoding text over reading with comprehension and the mechanics of writing over the skill to express ideas orally, or to negotiate and construct meaning, amounts to an emaciated form of literacy instruction. We suggest that the absence of this awareness in the teacher has serious and widespread consequences (e.g., Annual Survey of Education Report, 2017; Education Initiatives, 2010; National Achievement Survey, 2014). It directly blocks attainments in reading comprehension and written expression. Less visible, but equally devastating, is its impact on children's basic impulse to actively infer and engage with ideas. For the first language learner, this fragmentation can pose serious challenges in successful literacy development. When this approach is adopted for second language learners, it amounts to dysfunctional practice.

Encouraging Conversation, Discussion, and Negotiation of Meaning

The natural fallout of an emphasis on sing-song repetition and the mechanics of reading and copywriting, described in the previous sections, is that guided opportunities for active, child-led conversations are rarely promoted (Bhattacharjea, Wadhwa, & Banerji, 2011; Nag, Snowling, & Asfaha, 2016). There is good evidence to show that enabling children to 'talk' draws upon and develops their oral language resources, and this is associated with better attainments in literacy and the development of higher order skills (Bangladesh: Aboud, 2006; India: Nag, 2013; Sri Lanka: Aturupane, Glewwe & Wisniewski, 2013). It therefore follows that opportunities for practicing oral language need to find prominence in the curriculum. And yet, it is a rare teacher who uses class time to *teach* children to talk, to listen, and to respond. This is qualitatively different from the oft-seen tendency to lapse into a monologue; or to simply value children's verbalizations and listen patiently to whatever they have to say.

In the early years of pre- and primary school, a language-focused program needs to be alert to supporting children's ability to converse, discuss, and negotiate common understandings. The teacher may use a variety of contexts as a springboard for talk. There could be, for instance, discussions to find a solution to a problem that the class is facing, such as monkeys coming into the classroom and sampling the children's lunch while the children are at assembly. There could be interactive activities such as making a model of an animal using available materials. During such sessions, the teacher can explicitly instruct use of scripted expressions, for example, sentences and phrases for suggesting something ('*how about doing it this way…', 'let's use x instead of y'*), or for agreeing or disagreeing ('*I like your idea*', '*Yes, that's a good plan.*' or '*No, that may not work.*', '*That doesn't sound like a good idea because….*'). An important assumption in a language-focused classroom is that there will be a lot of open-ended questions (e.g., '*We are going for a picnic tomorrow. What do you think we should each carry with us? How should we behave once we reach the picnic spot?*'). Unlike a closed question, which elicits only yes/no or allows for only one 'correct' answer, the open-ended question encourages thoughtful, individual opinion. All children gain from opportunities for such conversations that arguably support active meaning-making and meaning-negotiation. Importantly, children who are the least exposed to the language of literacy instruction outside of the classroom stand to gain the most (e.g., Amritavalli, 2007; Kumar, 1994; Nag, 2013).

In the rest of this chapter we elaborate on the language- and meaning-focused approach to literacy instruction. We illustrate this approach using Hindi as the language of literacy learning after first describing the language and its orthography. Hindi is an exemplar language for illustrating a language-focused approach to literacy instruction because several languages share its orthography (e.g., Nepali and Marathi), its grammar (e.g., Urdu) and through the common historical links with Sanskrit, its vocabulary (e.g., Malayalam and Bengali). Hindi also has one of the largest numbers of first and second language learners.

Hindi and the *Devanagari* Orthography

Hindi is an official language of India, and the common language across much of the north and west of the country. Counted among the top five languages in the world for size of native and second language speakers, Hindi is also spoken in Nepal and Singapore, and in Fiji, Guyana, and South Africa. Hindi has several dialects including the *Khari Boli* and *Dakhini* varieties and has a vocabulary that draws from both Sanskrit and Persian. In more recent times Hindi has also been drawing heavily from English.

Morphology and Morpho-Syntax

All nouns in Hindi, animate and inanimate, have an intrinsic gender, which is assigned in a more-or-less arbitrary manner. For example, the word चट्टान (*chattan*; rock) is feminine but पत्थर (*patthar*; stone) is masculine. Hindi uses distributed grammatical information: the gender of the noun is reflected in different parts of the sentence, a phenomenon also seen in languages like Kannada and Bengali. Word order in Hindi typically follows the subject-object-verb structure: मैं आम खाती हूँ। (*Main aam khaatee hoon;*I-mangoes-eat.). The verb reflects the number and gender of the subject (which in this case is first person-singular-feminine). This pattern is visible in all tenses except the simple past tense. In the simple past tense, if a direct object is present in the sentence, the subject is followed by a subject-marker preposition (ने– *ne*) and the verb reflects the number and gender of the direct object: मैंने आम खाए। (*Maine aam khaaye;* I-mangoes-ate, with the verb reflecting the first person-plural-masculine of the direct object, 'mangoes'). A point to note is that in Hindi, certain verbs do not form the sentence with the subject form of the pronoun or the noun; instead, they require the indirect object form: मुझे आम अच्छे लगते हैं। (*Mujhe aam achhe lagte hain.* To-me-mangoes-like or 'I like mangoes'). Here, the verb अच्छा लगना (*achhaa lagnaa*; to like) does not accept the subject form of the pronoun (मैं– *main*). Instead, this verb requires the indirect form of the object to be used (मुझे– *mujhe*). In such cases, the verb reflects the number and gender of the direct object (आम – *aam*). Hindi has several other morpho-syntactic features; but these basic characteristics of gender markers, distributed morpho-syntactic information across the sentence, word order, and non-standard syntactic structures are particularly important to learn in the early phases of language acquisition, especially for a learner who is new to the language. These need to be made explicit to the learner, along with opportunities to practice their usage.

Orthography

A symbol unit is called *akshara* in Hindi and the orthography is called *Devanagari*. The symbol set comprises consonants (*vyanjans*), vowels (*swara*), consonant + vowel pairs, and consonant clusters. The first set is composed of consonants and vowels. There are 33 consonants or *vyanjan*: क्-ख्-ग्-घ्-ङ, च्-छ्-ज्-झ्-ञ, ट्-ठ्-ड्-ढ्-ण्, त्-थ्-द्-ध्-न्, प्-फ्-ब्-भ्-म्, य्-र्-ल्-व्-श्, ष्-स्-ह् of which two – ङ्, ञ् – are no longer used in modern Hindi. Besides these, there are 2 allophones which are commonly used consonants: ड़, ढ़. There are 13 vowels: अ, आ, इ, ई, उ, ऊ, ए, ऐ, ओ, औ, अं, अः, ऋ. The consonants and vowels together form the *aksharmaalaa* or the garland of *akshara*.

The vowel अ /a/ is inherent in all written consonants (e.g., क /ka/ *akshara*). The rest of the twelve vowels have a *matra* each or a vowel marker that is to be attached (ligatured) to a consonant to form a consonant-vowel or CV *akshara*. The second set which consists of the full complement of consonants with vowels is called the *baaraakhadi,* for example the consonant क् /k/ written with the inherent vowel अ /a/ and the full set of vowel markers: क, का, कि, की, कु, कू, के, कै, को, कौ, कं, कः, कृ. Modern Hindi uses two newer vowels to transcribe into Devanagari sounds that are foreign to Hindi: ऑ, ऍ. These are used in transcribing borrowed phonology shown in bold in words such as ऑटो (**au**to), चॉकलेट (ch**o**colate), लैपटॉप (lapt**o**p) and ऍक्टर (**a**ctor). Though well-accepted in typography—there are keys on the Hindi keyboard assigned to them—and appearing with increasing frequency to transcribe loan words, these vowel *akshara* have typically not yet entered most early grade readers or teaching programs (for exceptions, see Mathur, 2013d).

A third set is the consonant clusters (the CCa and CCV *akshara*). There are 4 consonant clusters for which a ready-made, designated *akshara* is available: क्ष-त्र-ज्ञ-श्र. But most often consonant clusters need to be transcribed according to the role of the first and second sound within the cluster. In Hindi, these consonants may be half-symbols or may be formed by the dropping of the inherent vowel. In addition, reflecting the phonology of Hindi, some consonant clusters shorten the sound of a consonant and merge with another different consonant e.g., वक्त (/kt/ in *vakt*; time) or double the sound to form geminates e.g., इक्कीस (/kk/ in *ikkees*; twenty-one).

A fourth set of *akshara* are the off-the-line 'dot' markers found above or below a base *akshara*. When used for nasal sounds they are called *bindu* as in हैं (*hain*; are), and *chandrabindu* as in आँख (*aankh*; eye). The *bindu* is increasingly being used to represent the sound of half-न (*/n/)* in contemporary Hindi, especially before the set of consonants त्-थ्-द्-ध्. For example, बन्दर (*bandar;* monkey)**,** चुकन्दर (*chukandar;* beet-root)**,** मन्दिर (*mandir;* temple)**,** अन्त (*ant;* end)**,** can most often be seen printed as बंदर, चुकंदर, मंदिर, अंत. The other off-the-line 'dot' marker is the *nukta*. It is used to bring variation to the pronunciation of certain consonants: जहाज़ (*jahaaz,* ship), where /j/ changes to /z/. For various reasons, the Hindi publishing industry has been systematically omitting the use of *nukta* from printed texts. However, we suggest that *nukta* needs to be re-introduced into the early literacy curriculum to teach accurate grapho-phonic correspondence, even if the symbol is later dispensed with for mature readers. The *bindu, chandrabindu,* and *nukta* illustrate the rich use of visuo-spatial space in this writing system, and there is growing research evidence that young learners differ in how much attention they pay to this aspect of the symbol set (for further details on the architecture of Hindi and other *akshara* orthographies see Nag, 2017; for patterns of Hindi acquisition see Gupta, 2004; Vaid & Gupta, 2002).

The Foundation Program

Literacy instruction in schools across South Asia often occurs in more than one language. Some learners receive literacy instruction for the first time in an *akshara* language while for others it may be the second or third language that they are learning to read. Given the nature of the linguistic landscape, children of the region may come to the task of literacy learning speaking a different language or a dialect at home, and because of this, literacy instruction cannot assume firm oral language foundations in the target language even among those learning to read for the first time. Some key features of a robust foundation program for *first-time literacy instruction* are that it draws upon and/or builds strong oral skills of the child, it is accompanied by the creation of a print-rich environment which leads to a curiosity about the written word, it begins with activities involving sorting, pattern spotting, and drawing rather than activities directly linked with print instruction, and it validates and creates links across home and school experiences. Important evidence-bases that inform the principles of such a foundational literacy program include the research on the role of the home environment on children's acquisition of language and literacy, the types of programs that work well for consolidating language learning in early childhood settings, and the linkages between early emergent literacy skills and later literacy attainments, (e.g., Aboud, 2006; Vagh, 2009).

Second-time literacy instruction, in contrast, can focus on a different set of principles but only if the first-time literacy instruction has been effective. We suggest that many aspects of the foundation program are transferable from one language to another. Examples include concepts about print including the knowledge that the printed text moves in a given direction, that blank spaces separate out different words, that certain symbols indicate the ending of an idea, and that different symbols can be joined in different combinations to produce new sounds. The notion that well-established foundation skills in one language transfer with ease to learning to read in a second or a third language has a growing evidence base in the alphabetic languages (e.g., Reese, Garnier, Gallimore, & Soldenberg, 2000; Verhoeven, 1994; for gaps in the literature see Hammer et al., 2014). Such cross-linguistic studies capturing transfer from one *akshara*-based language to another are not yet available, although some insight may be drawn from the available research in English language pre-schools in the region (e.g., Kalia & Reese, 2009; Sen & Blatchford, 2001).

Talking About Daily Experiences

Much before the child is introduced to written language, several knowledge bases are already developing from everyday experiences. At the simplest level, the inferences may be about simple sequences in space and in time. The following are examples of topics drawn from daily life that can consolidate these knowledge bases in a school context (also see Kurrien, 2014).

1. इस लाइन में सबसे आगे कौन बैठा है? उसके पीछे कौन बैठा है? लाइन के बीचोंबीच कौन बैठा है? सबसे पीछे कौन बैठा है?
   Is laain mein sabse aage kaun baithaa hai? Uske peechhe kaun baithaa hai? Laain ke beechonbeech kaun baithaa hai? Sabse peechhe kaun baithaa hai?
   In this line, who is sitting right in front? Who is sitting behind him or her? Who is sitting right in the middle of this line? Who is sitting right at the back?
2. पहले हम हाथ धोते हैं। फिर हम खाना खाते हैं। फिर हम अपनी प्लेट धोते हैं।
   Pehle hum haath dhote hain. Phir hum khaana khaate hain. Phir hum apnee plate dhote hain.
   First, we wash our hands. Then we eat food. Then we wash our plate.

Taking daily experiences as topics for consolidating children's knowledge bases (such as examples 1 and 2) rest on the assumption that teachers will ask and learn about the lives of their students outside of school hours.

Making Home-School Connections

An important principle to support emergent literacy is to use talk that can link stories in the classroom to experiences that the children may have had in other situations.

3. इस कहानी को सुनकर तुम्हें कुछ ध्यान में आता है, जो तुम्हारे साथ हुआ था, या जो तुमने देखा है?
   Is kahaani ko sunkar tumhe kuch dhyaan mein aataa hai, jo tumhare saath huaa tha, ya jo tumne dekhaa hai?
   Does this story remind you of something that has happened to you, or something that you have seen happening to someone else?

Such questions prompt children to connect in a personal manner to a story narrated by the teacher. We argue that a regular practice of such talk gives children confidence to bring experiences from outside the classroom into the class. In such an approach, where the classroom is made a safe space for talking and sharing experiences, the school experience can begin to be integrated with the child's home life. We propose that supporting this home-school connection is a true motivator for all learning, and most certainly so for literacy development.

Promoting Concepts About Print and Purposes of Literacy

A key principle in emergent literacy instruction is that teachers and, where possible, family members provide opportunities for the child to see the purposes and functions of literacy (e.g., Adams, 1994; Amritavalli, 2007; Teale & Sulzby, 1986). Reading can be a life-long source of enjoyment and learning. It is important that young children perceive these functions early. When children see their parents and siblings, or else their teachers and school mates, visibly enjoying the act of reading, they too wish to read. In homes where reading is a daily activity, very young children have often been seen holding up a book and pretend-reading, even if the book is held upside down! In school, a child who doesn't yet know how to read, but who has enjoyed a story being read aloud and discussed, will want to hold the book, scan the book with her eyes, and turn pages, as she has seen the teacher do. Developing concepts about print from such experiences is a well-documented key step in gaining literacy. Further, when parents or teachers model the act of reaching out to books such as dictionaries or informative books to look for something they do not know, it is a sure sign that books are a source of new knowledge. Children too then tend to display such adult literacy practices and, arguably, this in turn strengthens literacy acquisition (for a meta-analysis see Nag, Vagh, Dulay & Snowling, 2018). The creation of this 'interest' in reading is thought to be a key to subsequent engagement with literacy learning activities.

Dialogic Reading

A story provides opportunity for dialogue and talk (Vygotsky, 1986). Take the example of '*Mein Bhee'* (मैं भी, Me too, Suteyev, 2002). This is a story of a hen's chick that is born at the same time as a duckling. The hen-chick decides to copy the duckling in everything, until he finds out that this is simply not a good idea. Examples of simple questions to make children alert to finding the sequence of events in the story are:

4. सबसे पहले बतख के बच्चे ने क्या किया? और चूज़े ने क्या किया?
   उसके बाद बतख के बच्चे ने क्या किया? और चूज़े ने क्या किया?
   फिर बतख के बच्चे ने क्या किया? और चूज़े ने क्या किया?
   Sabse pehle batakh ke bacche ne kyaa kiyaa? *Aur chooze ne kyaa kiyaa?*
   Uske baad batakh ke bacche ne kyaa kiyaa? Aur chooze ne kyaa kiyaa?
   Phir batakh ke bacche ne kyaa kiyaa? Aur chooze ne kyaa kiyaa?
   First of all, what did the duckling do? And what did the hen-chick do?
   After that what did the duckling do? And what did the hen-chick do?
   Then what did the duckling do? And what did the hen-chick do?

The use of the terms '*first*', '*after that'* and '*then*' allow children to become aware of the unfolding sequence of ideas in the story (see also talk in examples 1 and 2 above). A parallel dialogue draws attention to details of the pictures that accompany the text. Teacher-child talk in the above story could further be about the physical similarities and differences between the duckling and the hen-chick:

5. बतख का बच्चा कहाँ से निकला? मुर्गी का चूज़ा कहाँ से निकला?
   बतख के बच्चे के पाँव कैसे हैं? मुर्गी के चूज़े के पाँव कैसे हैं?
   बतख के बच्चे की चोंच कैसी है? मुर्गी के चूज़े की चोंच कैसी है?
   Batakh ka bacchaa kahaan se niklaa? Murgee ka chooza kahaan se niklaa?
   Batakh ke bacche ke paanv kaise hain? Murgee ke chooze ke paanv kaise hain?
   Batakh ke bacche kee chonch kaisee hai? Murgee ke chooze kee chonch kaisee hai?
   Where did the duckling come out from? Where did the hen-chick come out from?
   How do the duckling's feet look? How do the hen-chick's feet look?
   How does the duckling's beak look? How does the hen-chick's beak look?

Prompting attention to similar and dissimilar features is useful for tasks the child will encounter during subsequent literacy instruction. Somewhat similar, though more sophisticated, perceptual processes will be needed when the child has to distinguish *akshara* that are visually close to each other (e.g., अ */a/* from आ */aa/*, भ */bha/*

from म */ma/*, य */ya/* from थ */tha/*). There is some research evidence to suggest that visually confusable aspects of symbols influence *akshara* recognition (e.g., Nag, Snowling, Quinlan & Hulme, 2014b).

Visuo-Motor Activities

The *akshara* of Hindi comprise a set of lines and curves packed into a small space. Learning these symbols through copying helps to build an awareness of these various visual features (e.g., Bhide, 2018). However, beginning to write this symbol set too early, as seen in many kindergartens, serves little purpose. Drawing in sand with a stick, drawing with a thick chalk or crayon on a large paper are sensible precursors to putting a pencil in children's hands. These foundation tasks may appear to have little to do with 'writing' but are firmly linked. Surrounding the child with a print-rich environment also plays an important role in learning to write. This helps build awareness of the physical traits or form of written words: is a word long or short? (e.g. octopus – owl in English or मगरमच्छ – मच्छर in Hindi); does a word occupy space only along the line or does it occupy space above or below the line? (e.g. मच्छर – **मछली,** कुत्ता – कमल). These early skills are valuable when the child comes later to the task of transcribing *akshara* and spelling words.

Summary

Talk is at the heart of the foundation program for literacy learning. What could be prioritized in a language classroom is talk that:

- supports the ability to recall and reconstruct sequences;
- brings attention to elements that are similar and those that are dissimilar;
- connects new ideas learned in school with the prior knowledge and experiences that the child brings from home;
- promotes concepts about print.

These opportunities for conversation, discussion, and exposure to print increase personalized connections with the foundation literacy program and can help demonstrate the purposes of literacy.

The Early Reading Program

Within a language-focused framework for literacy instruction, an early reading program explicitly addresses orthographic, morpho-semantic, syntactic and metalinguistic levels of the language. Evidence for the importance of these domains of learning comes from research that shows that children who struggle to read have poor *akshara* knowledge, lower levels of phonological skills, limited vocabulary knowledge and a shaky grasp of the morpho-syntax of the language (e.g., Nag & Snowling, 2011). Apart from this, and in line with emerging trends in multilingual contexts, we suggest that an early reading program ought to engage with the psycho-social aspects of learners, including their home culture, their perceptions about the language to be learned, and the opportunities they have to listen to and practice this language outside the classroom. As highlighted in our introduction, while such a focus is useful for all learners, it becomes a particularly powerful pedagogical approach for learners who are new to the language or have little support at home for tutoring. One assumption we make is that children who learn to read well will reach out to books on their own to read for pleasure and to learn something new, and in doing so set forth on a self-teaching path.

The following sections will take examples from a Hindi teaching program (*Hindi ki Duniya 1–5*, Mathur, 2013a, 2013b, 2013c, 2013d), and be supplemented with examples from other languages, to highlight key details of literacy instruction with a focus on language and meaning.

Alliterative Rhymes

Alliterative rhymes form the backbone of the *Hindi ki Duniya* program. Bringing attention to the sounds embedded within words is an important target for literacy instruction. *Anupras alankar*, a literary device akin to alliteration in English, is well suited for this as a teaching device. In alliteration, a specific sound is repeated in close contiguity in a string of words. The sound repeats itself mostly in the first position of the word, but it could also occur in the middle and word-final position (e.g., "While I **n**odded, **n**early **n**apping, sudde**n**ly there came a tapping" in The Raven by Edgar Allen Poe, Poe, 1962). Consonants lend themselves readily to the construction of such rhymes which are used, in this program, to introduce the child to the sounds and symbols of Hindi. Alliterative rhymes may depict an oft-seen reality in children's lives such as situations shared within the family. The rhymes can also echo the tradition of the fable (e.g., India's *Jataka* and *Panchatantra* tales) where animals are routinely anthropomorphized and endowed with human desires and disappointments, or a fantasy world that could appeal to a young child (examples 6 and 7 below, respectively).

6. **म** से **मि**याँ **म**गर**म**च्छ को, **मो**टी **म**छली का सपना आया,
   पर जो आँख खुली तो केवल, **म**कड़ी और **म**च्छर ही पाया।
   Ma *se* ***mi****yaan* ***ma****gar****ma****chh ko,* ***mo****tee****ma****chhlee ka sapnaa aayaa,*
   parjo aankh khulee to keval, ***ma****kadee aur* ***ma****chhar hee paayaa.*
   Ma is for… Mr. Crocodile, who dreamed of a fat fish,
   But when he opened his eyes, all he found was a spider and a mosquito. (for illustration see Fig. 1)

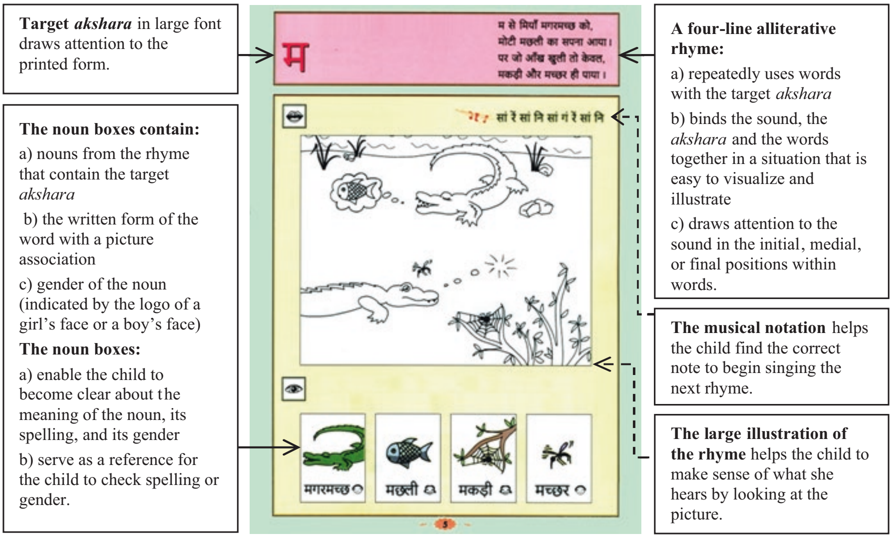


Fig. 1 Connecting orthography, phonology, semantics and syntax with culturally embedded materials: the case of the alliterative rhyme on *akshara* म /ma/
Note: The teacher supports comprehension further by pointing to relevant parts of the picture as the rhyme is sung or recited or a recording is played. Children also colour the illustration, and as they do so are likely to look at details and ask for supplementary vocabulary. To read the alliterative rhyme see example 6 above. (Source: Mathur, 2013a, *Hindi ki Duniya*, Textbook 1, p 5)

7. **त** से **ती**र-कमान पकड़कर, **तोता** दे**ता** पहरा किसपर?
तितली रानी **त**किया लेकर, सो**ती** देखो **तान**कर।
Ta *se* ***tee****r-kamaan pakad kar,* ***totaa*** *de**taa** pehraa kispar?*
Tit*lee raanee* ***ta****kiyaa lekar, so**tee** dekho* ***taan****kar.*
Ta is for…whom does the parrot guard, holding a bow and an arrow?
Look, the queen butterfly sleeps soundly, with a pillow.

Alliterative rhymes set to rhythmic tunes are thus used for introducing children—both native speakers and non-Hindi speaking children—to the distinctive sounds and shapes of Hindi *akshara*. The complexity of the language in these rhymes is within the grasp of a native Hindi-speaking six-to-seven-year-old. Moreover, no attempt is made to simplify the language of the rhyme even when introducing it to the non-Hindi speaking child. Instead, the attempt is to ensure that there is a cultural embedding of the content. Situations depicted allow for already available knowledge bases of South Asian children to scaffold meaning-making, even when vocabulary in a new language is still limited.

Besides the use of alliteration, a wide range of sound games exists in *akshara* languages, and these can be profitably used to support phonological processing. Patel (2004) describes the Gujarati syllable insertion game where high frequency clusters are systematically inserted within multi-syllabic words (e.g., inserting /sm/ in *mari-nishal* to become *ma**sm**ari*, *nish**smi**shal;* note the vowel changes on /sm/ to maintain the phonological rules of Gujarati). Another phonological manipulation is seen in the Malayalam /pa/ game where words like *susmita* become ***pa**su**pa**smi**pa**ta* (in Mohanan, 1986). Some of the phonological processes covered in these sound games relate to the construct of *sandhi*. Within the Indic languages, *sandhi* describes the changes that occur—due to the specific rules of pronunciation in a given language—in the way an *akshara* gets pronounced after it is placed in close proximity to another specific *akshara*. In French, similar rules about pronunciation are called 'liaison' (for example, whereas usually the last consonant is not pronounced in French, it does get pronounced if it is followed by a vowel; hence one would need to compulsorily pronounce the 'n' sound while saying the phrase *mon ami*).

A further level of cultural embedding of the material in Mathur (2013a, 2013c) is that all rhymes are sung in the local classical music tradition (the *Hindustani* music tradition). This introduces children, especially non-native speakers, to the Hindi language in all its literary and musical richness. Taken together, the use of alliterative rhymes, a musical idiom and rich accompanying illustrations result in a lively, multisensory program that supports not just phonological awareness but also awareness of the cultural mores of the language and its rich vocabulary. Figure 1 gives the further domains of learning that are tagged to these alliterative rhymes.

Akshara Single and Within Words

We suggest that an *akshara* needs to be first encountered within words and within meaningful contexts, and then shown as singletons. This approach differs quite radically from the artificial sequence common to literacy teaching in all *akshara*-based languages where *akshara* charts are taught first and encounters with simple texts delayed. Introducing the *akshara*-in-context early in the literacy program allows the abstract form (the symbol) to hook on to individual words situated within sentences and longer text. Again, a focus on broader language, with meanings that the child can relate to, is at the heart of this pedagogical approach. Figure 1 shows one example of how this could be accomplished. Through each rhyme, connections are established between singleton *akshara* and the corresponding sound. While this sound may often be a syllable, it can also be a phonological unit smaller than a syllable (e.g., an onset or coda). Two points deserve to be highlighted to contextualize this approach to teaching decoding. First, even though there is substantial polarization around use of the whole language approach *versus* phonics, there is a need to go beyond such binaries. For skill attainment to progress smoothly, neither approach is a sufficient intervention by itself. The mix of *akshara* within words and as single units is therefore an example of a departure from the whole language versus phonics debate. Second, our approach may appear as falling within the traditional approach with *akshara* charts. However, a distinct difference we propose is that *in parallel* with *akshara* drills and memorization, literacy instruction must give substantial time and value to reading and talking about words in meaningful texts and contexts. This is what makes the language come alive for the child. Thus, another equally plausible formulation of our approach is to start with linking the sound of the *akshara* to a meaningful context (as in Fig. 1), then linking it to specific words in which it occurs, and finally bringing focus to a single *akshara* as a written unit.

A common question in teaching *akshara*-based languages is the sequence in which the extensive symbol set should be introduced. The received wisdom is to follow the phonics-based *akshara* charts (the *aksharmaalaa* and *baarakhadee* described earlier). In contrast, within the language-focused approach to literacy instruction, the sequence of introducing the *akshara* is around the early vocabularies of children. *Akshara* that occur at the beginning of more frequently occurring words are taught first, the rationale for which we will see in the next section.

For second language learners a further variation may need to be introduced in the sequence of instruction since particular sounds in the target language may not exist in the child's native language. An example of unfamiliar sounds for second language learners is one of the /l/ sounds in Tamil. Many children coming from non-Hindi speaking backgrounds too experience difficulty in recognizing and pronouncing certain sounds in Hindi, especially the aspirated sounds. The program can therefore begin with those *akshara* which most non-Hindi speaking children can hear and pronounce with relative ease: the unaspirated unvoiced consonants (क, च, प), unaspirated voiced consonants (ग, ब, ड), nasal consonants (न, म), semi vowels (य, र, ल), the sibilant consonant (स) and the fricative consonant (ह). The *akshara* of the more difficult aspirated consonants, both unvoiced (ख, छ, ठ) and voiced (घ, झ, भ), are introduced later, for example in the second year of learning the language (Mathur, 2013c). Limiting instruction to a small set of *akshara*, in the initial stages, is only for purposes of initiating reading and writing. *All* sounds of the language—even if they are novel for new learners—are naturally part of the rhymes, dialogic reading and talk, right from the beginning. Such early encounters with oral language allow for a sufficiently long period of exposure and sensitization to the sounds of the language in meaningful contexts. The expectation is that, similar to natural language learning contexts, the immersion will support receptive and expressive phonology in the new language.

A further aspect in *akshara* instruction is the sequence to be adopted in the teaching of the two forms of the vowel, namely, the primary form and the secondary form (the vowel marker). Although the common practice is to introduce these in separate sets (teaching the primary set first, e.g., Sinhala: Wijayathilake & Parrila, 2014), a better approach is to introduce both forms of each vowel simultaneously within a language-rich context. Two-line alliterative rhymes is one such method to introduce both the primary vowels and their phonemic markers (see examples 8 and 9 below, where the primary forms are given in bold and the *akshara* with the vowel markers are underlined).

8. **आ** से **आ**<u>धा</u> लो मे<u>रा</u> **आ**म, <u>खा</u>कर इसे करेंगे **आ**<u>रा</u>म।
   Aa *se* ***aa***<u>*dhaa*</u> *lo mera* ***aa****m,* <u>*khaa*</u>*kar ise karenge* ***aa***<u>*raa*</u>*m.*
   Aa is for… take half of my mango, we will eat it and then rest.
9. **ई** से **ई**टों का घर है एक, सामने <u>मीठी</u> **ई**ख उ<u>गी</u> देख।
   Ee *se* ***ee****nton ka ghar hai ek, saamne* <u>*meethee*</u> ***ee****kh u*<u>*gee*</u> *dekh.*
   Ee is for… a house made of bricks, look, the sweet sugar-cane has grown there in front.

Such rhymes draw the child's attention to each orthographic form of the vowel. More importantly, these rhymes allow for targeted discussion on valid sound- and symbol-level sequences in the language (the phono-tactics and ortho-tactics of the language). An important nuance that needs to be pointed out to children is the distinction between 'independent sounds' (the primary vowel) and 'sounds that modify consonants' (the vowel marker). Thus, when the sound of आ is an independent sound, it is written as a primary vowel as in the initial आ /*aa*/ in आम (*aam*; mango) and आधा (*aadhaa;* half). When the sound of आ modifies the sound of a preceding consonant like in मेरा (*meraa;* my) and खाकर (*khaakar;* after eating), then a vowel marker or *matra* is used. Such an explanation is necessary because the common misconception among children, and indeed among many teachers and researchers, is that the primary form is used only when the vowel is in the word-initial position. One prominent counter-example is where the primary vowel form is *always* used when a vowel follows the sound of another vowel, both when the first vowel is marked—as in आईना (*aaeenaa;* mirror), आओ (*aao;* come), दाएँ-बाएँ (*daaen-baaen;* right-left)—as well as unmarked—as in कई (*kaee;* many), मई (*maee;* May). In both instances, the second vowel has an independent syllable weight. Similar word-level prosody can be assumed to govern *akshara* choice in several other languages (e.g., for Bengali examples, see Nag, 2017; Sircar & Nag, see book chapter, this volume).

A final point is related to *consolidation* of *akshara* recognition. Clearly, a lot of practice is needed to establish sound-symbol correspondences. An easily accessible aid is the black board and chalk and talk to introduce each singleton *akshara*. Once children are proficient at these single correspondences, they could play simple board games: for example, the teacher calls out an *akshara* and the child marks it on a card with two-four-six-eight-or-more distracter *akshara*. The learning objective can drive the mix of *akshara* that are chosen as distracters in these practice cards. Some sets may comprise visually similar *akshara*; others could be with phonologically confusable *akshara*, and yet others can target vowel learning by mixing the primary form with *akshara* that carry the vowel markers. Further reinforcement of *akshara* learning can be through filling in blanks using continuous and dotted lines as place holders to prompt use of primary or secondary forms, respectively (see left page in Fig. 3).

Reading Dialogues, Stories, and Expository Texts

The approach to *akshara* instruction described above allows for drill and practice but is also language-rich in order to simultaneously support broader language development. This approach departs from the traditional sequence where all the vowels in primary form and consonants with inherent vowels are taught in the first grade and the consonant-vowel pairs are introduced in the second grade. There is a very good reason to drop such a segmented practice: most words that children acquire early contain vowels which in the written form are represented with a vowel marker.

If one expects the child to read with focus on meaning right from the beginning, then it is essential to introduce the vowel markers along with the primary form of the vowel. One suggested entry point into reading is to introduce short texts for reading, after introducing just four consonant and two vowel *akshara* along with their markers. These texts can be imbued with all the characteristics of a typical dialogue: there is a person who is speaking, a person to whom s/he is speaking, and something about which s/he is speaking. Embedding these dialogues within simple yet culturally resonant illustrations supports the child in making meaning of the text even as s/he learns to read it (see Fig. 2).

One last concept, rarely taught but very important for correct pronunciation, is the phenomenon of *schwa* deletion. In Hindi words, as with Indic languages such as Bengali, the orthographic unit might imply presence of the inherent vowel (the *schwa*) but should be ignored (*schwa* deletion). It is essential to give broad rules of schwa deletion early to the learner for these help ensure a correct unaccented

Fig. 2 Connecting orthography, phonology, semantics and syntax through simple texts in speech balloons (Sp-B[a])
Note: This text uses high frequency *akshara*: four consonants with the inherent vowel (न, क, ल, म), two other vowels (आ, ई) and their markers (ा -ी)
Sp-B[1]:/*Naanaa kaa kaan*. /Grandpa's ear. Sp-B[2]: /*Naanaa kee naak.*/ Grandpa's nose. Sp-B[3]: /*Leelaa, Naanee aayee. Leelaa kee naanee aayee.*/ Leela, here comes Grandma. Here comes Leela's Grandma. Sp-B[4]: /*Naanee kaa kaan. Naanee kee naak. Leelaa kee naanee kee naak.*/ Grandma's ear. Grandma's nose. Leela's Grandma's nose. Sp-B[5]: /*Aa Leelaa aa. Naanee aam laayee.*/ Come Leela come. Grandma has brought some mangoes. Sp-B[6]: /*Aa Leelaa. Naanee laal-laal aam layee.*/ Come Leela. Grandma has brought some red mangoes. Sp-B[7]: /*Laal-laal aam! Aayee naanee, aayee.*/ Red-red mangoes! I'm coming Grandma, I'm coming. (Source: Mathur, 2013a, *Hindi ki Duniya*, Textbook 1, p 14–15,)

pronunciation of Hindi. In the absence of explicit instruction, this phenomenon can pose serious challenges to non-native learners of Hindi (Gupta, 2004; also see Bengali: Sircar & Nag, this volume). Broadly speaking, learners can be made aware that:

- If a word ends with a consonant the /a/ inherent in the consonant is dropped. For example आम (*aama;* mango) is pronounced आम् (*aam*) in Hindi because of the deletion of the sound of अ from the consonant in the last position of the word.
- When a word is made up of suffixes such as ना, ता-ती-ते, in words such as खेलना (*khelnaa;* to play), खेलता-खेलती-खेलते (*kheltaa-kheltee-khelte;* he plays-she plays-they play) or contains a preposition such as ने-को-से-के लिए-का-के-की-मे-पर in words such as मुझसे (*mujhse;* by me), उसका (*uskaa;* his/her), उसकी (*uskee;* his/her), and उसके (*uske;* his/her), then the inherent vowel of the consonant preceding the suffix or the preposition is not pronounced (schwa deletion).

Word Gender, Prepositions and Other Grammar Insights

At the morpho-semantic, syntactic, and meta-linguistic levels, an important realization for all learners is that nouns in Hindi are either masculine or feminine and that their gender influences other words in a sentence such as the adjectives, prepositions, and verbs, which could occur either before or after the occurrence of the noun. Similar meta-linguistic insights are required in all *akshara* languages. One simple way to bring children's attention to gender markers in Hindi, for instance, is through matching nouns with gender icons. See for example the bottom panel in Fig. 1 where nouns occurring in the alliterative rhymes are isolated and presented alongside a picture and a gender icon of a boy-face or a girl-face. Each noun can therefore be seen in its written form, linked to its meaning and its gender. Subsequent language work can focus on making sentences by applying the gender-identity of key nouns to other words in the sentence such as prepositions का or की (*kaa, kee*; of) and color adjectives काला-काली, नीला-नीली (*kaalaa-kaalee, neelaa-neelee*; black, blue). Awareness of and insight into these and other similar forms of grammar rules can be supported using display charts, targeted classroom talk, hands-on work with flash cards, and written exercises such as the ones shown in Fig. 3.

Talk, Dialogue, and Thinking Beyond the Written Text

When children begin to first get exposed to written texts, teachers could encourage them to make inferences about the thoughts, actions, motivations, and emotions of the characters in the text. As in the foundation program, it is possible to encourage children to make connections between situations in texts and their personal experiences and memories. Illustrated rhymes may also be used to trigger a conversation.

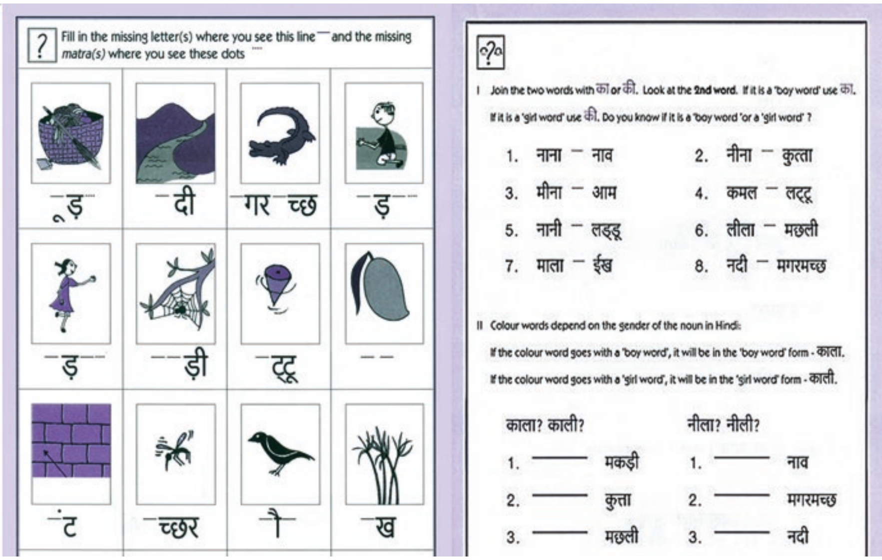


Fig. 3 Connecting semantics to orthography and morpho-syntax through fill-in-the-blank activities
Note: The use of continuous and dotted lines is to support meaning-focused transcription practice (left page). The morpho-syntax practice is for connecting gender identity of the noun to the choice of the preposition and adjective (right page). (Source: Mathur, 2013b, *Hindi ki Duniya*, Workbook 1, p 16–17)

Below, we give possible conversation-starters based on the alliterative म /ma/ rhyme from Fig. 1 and example 6.

10. मगरमच्छ के बारे में – क्या तुमने असली का मगरमच्छ देखा है? कब? कहाँ? वो किस रंग का था? क्या कर रहा था? मगरमच्छ कहाँ रहते हैं? वे क्या खाते हैं?
 Magarmachh ke baare mein – kyaa tumne aslee kaa magarmachh dekhaa hai? Kab? Kahaan? Vo kis rang kaa thaa? Vo kyaa kar rahaa thaa? Magarmachh kahaan rehte hain? Ve kyaa khaate hain?
 About the crocodile – Have you seen a real crocodile? When? Where? What color was it? What was it doing? Where do crocodiles live? What do they eat?
11. मछली के बारे में – तुम्हें क्या लगता है – क्या सपने में मगरमच्छ मोटी मछली के साथ खेलना चाहता था? या उसे खाना चाहता था?
 Machhlee ke baare mein – Tumhen kyaa lagtaa hai – kyaa sapne mein magarmachh motee machhlee ke saath khelnaa chaahtaa thaa? Yaa use khaanaa chaahtaa thaa?
 About the fish – What do you think, in his dream, did the crocodile want to play with the fat fish? Or, did he want to eat it?
12. सपनों के बारे में – क्या तुम्हें सपने आते हैं? तुम्हें अपना कोई सपना याद है?
 Sapnon ke baare mein – kyaa tumhen sapne aate hain? Tumhen apna koyee sapnaa yaad hai?
 About dreams – Do you get dreams? Do you remember any of your dreams?

When activities such as the above are done with second language learners, focus needs to be brought on the recall and use of topic-specific vocabulary. For example, the rhyme in Fig. 1 and example 6 lends itself to eliciting specific content and function words, even if the response comes in sentence fragments. The teacher could ask, for instance, while pointing to the picture of the fish: मगरमच्छ को किसका सपना आया? (*Magarmachh ko kiskaa sapnaa aayaa?;* Whom did the crocodile dream of?). Children could be cued to tag the appropriate preposition to the adjective–noun cluster: मोटी मछली का (*motee machhlee kaa*; of the fat fish); this promotes practice of the morpho-syntax. Broadly speaking, at the early stages, questions may remain limited to 'who', 'what', 'whose', 'which color' and 'where' questions (कौन, क्या, किसका/किसकी, किस रंग का, and कहाँ). The 'how' and 'why' questions (कैसे and क्यों) are generally more difficult to answer. Hence, when these are posed, teachers need to be alerted that children will require help with vocabulary and syntactic structures. Yet, creating the space for dialogue for the first language learner as well as the second language learner is a must at all stages of a language-focused pedagogy.

Summary

The meaning- and language-focused approach undergirded by a supportive interpersonal pedagogy is deeply influenced by early educational experiments including by Shivarama Karanth (e.g., *odova ata* 'Reading Games' Karanth, 1995.), Gijubhai Badheka (Badheka, 1932/1986), Ravindranath Tagore (e.g., Tagore, 1918) and Sukumar Ray (e.g., *abol tabol*, 'The Weird and Absurd', 1923/2004) as well as in schools established by the philosopher J. Krishnamurti. An overarching objective for an early reading program that derives from these thinkers and practitioners is that of deepening comprehension and making personal meaning. Culturally rich stories and texts provide the context for deepening comprehension and children learn to self-monitor the meaning of what is heard and read; questions and talk help clarify and build on this understanding. The program has several further layers:

- Target consonants and vowels are introduced in words and singly, within a clear context that is age- and culture-appropriate and easy to visualize for the learner;
- Specifically, for vowels, the primary form and corresponding marker are taught in close proximity to support insights about when each orthographic form is used;
- Consolidation is through *akshara* games and meaningful texts. In parallel, there is practice with picture- and word cards (meaning-word correspondences) and coherent texts for reading made up exclusively of already taught *akshara* and whole words;
- Linked to all of this is support for children to notice and apply meta-linguistic information such as the gender of nouns and how this modifies prepositions and adjectives.

The Early Writing Program

Successful early writing programs rest on three assumptions: children's skills for fine motor control are appropriately developed to deal with the demands of writing print, children speak and understand the language they are being asked to write in, and children see the need to write. Many aspects of learning to write are perhaps transferable from one language to the other. Children who have already begun writing in one language may find the mechanical skills of writing in a second or third language that much easier to master. Nevertheless, there are language-specific challenges in the *akshara* languages that need to be supported. We detail some of these areas of support in learning to write the Hindi orthography.

Sound-Symbol Correspondence and Writing Routines

Writing has a communicative purpose, and multiple learning steps are required for the child to become a willing and eager writer. Children need to learn the linkages between sounds and symbols – this includes the linkages of name-sounds with the written form, and one way to approach teaching this is through the alliterative rhyme page (see Fig. 1); they need to recognize individual *akshara* and practice the writing routines for each symbol. Writing with the correct series of strokes is another teaching target, though this area of teaching is often left unattended in many *akshara*-language classrooms. Studies on stroke sequences in *akshara* writing show confusion. Not only do different children adopt different sequences within the same classroom but there is also evidence that when there is no explicit instruction, an individual child may use a different sequence to write the same *akshara* at different times, such as /ka/ in Marathi, /ma/ in Kannada, /so/ in Tamil (Marathi: Kshirsagar & Poovaiah, 2015; Kannada: Nag et al., 2014a; Tamil: Nag & Narayanan, book chapter, this volume).

A specific teaching focus of writing routines must be on non-linear arrangements such as vowels that are written before, below, above, or around the consonant. Children need to be explicitly shown that vowel markers (the *matra*) are always written *after* the consonant has been written, irrespective of where they are located in the visuo-spatial arrangement. Thus in Hindi, for the non-linearly placed इ /i/ marker (the *matra* ि◌) the pencil needs to go backwards to print the vowel marker after writing the base consonant. This appears like a simple rule but is typically left unexplained in many language programs, bringing an unnecessary layer of mismatch between phonology and transcription (Hindi: Vaid & Gupta, 2002; also see Thai: Winskel & Iemwanthong, 2010; Tamil: Nag & Narayanan, see chapter this volume). These early writing instructions for vowel markers—the valid sequence of writing down the vowel markers and the valid form for particular sounds—are both examples of teaching children the ortho-tactics of a language.

Beyond the simple *akshara* of vowels, consonants and consonant-vowel pairs are the complex *akshara* for consonant clusters (often called 'half-letters' by teachers). In the box below is given a comprehensive list of rules for construction of these clusters in Hindi. Similar rules are possible to extract in all *akshara* orthographies. In a language-focused early writing program, specific practice with consonant clusters is done because some clusters appear frequently in early vocabularies. However, the majority of the consonant clusters will occur naturally in the continuing program in later grades and therefore need not be taught at the early stage. This reduction of instruction load is wise and seen in all current language and literacy programs (e. g., see Kannada textbook analysis in Nag, 2014).

Simple rules for constructing consonant clusters in Hindi

- Visual aspects of the *akshara*—whether there are elements on both sides of a long vertical line or on just one side, or whether elements hang from a short vertical line—play a role in how a given *akshara* forms the consonant cluster.
- Consonant clusters in Hindi are constructed in four basic ways.

1. By attaching horizontally, a part of the first *akshara* to the second *akshara*. The following types of *akshara* always form consonant clusters like this:

   (a) All *akshara* written left of the long vertical line (called *khadi paai*, खड़ी पाई, literally ‘standing leg’)—ख, ग, घ, च, ज, झ, त, थ, ध, न, प, ब, भ, म, य, ल, व, श, ष, स, क्ष, त्र, ज्ञ—drop the long vertical line and attach to the nearest part of the next consonant.

   1. /k^h^y/ ख्याल (*khyaal;* thought)
   2. /gy/ ग्यारह (*gyaarah;* eleven)
   3. /g^h^n/ विघ्न (*vighn;* obstacle)
   4. /chch/ कच्चा (*kacchaa;* raw)

   (b) All *akshara* written on both sides of the long vertical line—क and फ—drop the last downward curve and attach to the nearest part of the next consonant.

   1. /kk^h^/ मक्खी (*makkhee;* fly)
   2. /p^h^th/ हफ़्ता (*haftaa;* week)

2. By writing both *akshara* of the consonant cluster one after the other and using the sub-script slash marker—्‌ (called *halant*)—to suppress the inherent vowel of the first consonant. The rounded *akshara* that hang from the short vertical line—ट, ठ, ड, ढ, द—sometimes form consonant clusters like this.

   1. /tt/ लट्टू (*lattoo;* spinning top)
   2. /dd/ लड्डू (*laddoo;* round sweetmeat)
   3. /dy/ विद्यालय (*vidyaalaya;* school)

(continued)

3. By attaching vertically, the first *akshara* to the second *akshara* in a consonant cluster. The rounded *akshara* that hang from the short vertical line—ट, ठ, ड, ढ, द—sometimes appear in this form in consonant clusters.

    1. /tt/ लट्टू (*lattoo*; spinning top)
    2. /dd/ लड्डू (*laddoo*; round sweetmeat)
    3. /tt^h/ इकट्ठा (*ikatthaa*; together)
    4. /dd/ रद्दी (*raddee*; waste)

4. By joining in other ways.

    (a) Sometimes /d/ द remains full in a consonant cluster with a miniature of the following consonant attaching itself to either the lower left or right part of the *akshara*:

    1. /db^h/ अद्भुत (*adbhut*; strange)
    2. /db^h/ युद्ध (*yuddh*; war)
    3. /dy/ विद्यालय (*vidyaalaya*; school)

    (b) Representation of /r/ र changes depending both on its position in a cluster and the characteristics of the next *akshara*:

    (i) When /r/ र is the first sound in a consonant cluster, a small curl is attached above the following consonant.

    1. /rp^h/ बर्फ़ (*barf*; ice/snow)
    2. /rv/ पर्वत (*parvat;* mountain)

    (ii) When /r/ र is the first sound in a consonant cluster and the second sound includes a vowel marker (*matra*) the curl is placed *above or beside* the *matra* producing a mis-sequence in the sound-symbol mapping.

    1. /rt^h/ कुर्ता (*kurtaa;* a type of shirt)
    2. /rt^h/ मूर्ति (*moorti;* statue)

    (iii) When /r/ र is the second sound in a cluster, the first consonant is written in full and a slanting line is attached in-line but in the lower corner of most consonants.

    1. /kr/ क्रान्ति (*kraanti;* revolution)
    2. /gr/ अंग्रेज़ी (*angrezee;* English)
    3. /dr/ पंद्रह (*pandrah;* fifteen)
    4. /pr/ प्रकाश (*prakaash;* light)

(continued)

(iv) Two exceptions to point iii occur when /r/ र follows the rounded consonants /t/ ट and /d/ड in a consonant cluster. Then, /r/ र takes a third form of an inverted ‘v’ attached off-the-line, below the first consonant.

1. /tr/ ट्रक (*truck;* truck)
2. /dr/ ड्रम (*drum;* drum)

Drawing explicit attention to patterns in the orthography (e.g., in the manner shown in the box) reduces apparent ambiguities in the symbol set. It is best to begin drawing children’s attention to consonant clusters several months before they are expected to transcribe them. Once children are familiar with all the consonants, vowels and *matra*, they can be introduced systematically to each set. A suggested sequence, typically introduced in the last 6–8 weeks of the second year of instruction, moves from visual recognition of such *akshara* in already familiar words to sounding out each consonant cluster as a singleton *akshara*, and from understanding how the *akshara* are joined or ligatured to listening to and transcribing several words that follow the same pattern of *akshara* construction. Given the multiple units to be learned, the rules of construction need to be re-visited periodically over the next 2 years.

Meaning-Focused Transcription

From the above list it is evident that learning Hindi transcription deserves due pedagogical attention with careful sequencing, sufficient practice, and time for assimilation. Yet, focusing on transcription in isolation, without linking frequently to broader language experiences can work against successful literacy acquisition rather than for it. Dictation and copywriting, which are routine tasks across the region, gain value only when they are balanced with the use of the same words within meaningful contexts (e.g., Amritavalli, 2007; Kumar, 1994; Nag, 2013; Sahi, 2015). Examples of authentic meaning-focused transcription activities include the following: writing one’s name and surname on notebooks and other personal items, making labels for objects in the class and at home, making birthday cards for classmates (each child’s intended message could be expressed to the teacher who writes out a sample for the child), and copying simply worded, brief functional messages for communication within the school and at home (e.g., invitations for celebrations, list of items required for different tasks such as a festival, a party, a group project). These activities reinforce the communicative nature of writing while allowing for practice of writing the *akshara* symbol set which is packed with visual elements.

Summary

The early writing program described above consists of the following elements for teaching and practice:

- Writing target consonants with inherent vowels accompanied by:
  - Explicit instruction of sound-symbol correspondence;
  - Explicit instruction in strokes, correct formation and relative size of the *akshara*;
- Writing target vowels and vowel markers accompanied by:
  - Explicit instruction regarding strokes, including mis-sequence of strokes for certain vowel markers;
  - Explicit instruction about when a vowel marker is required and when the primary form of the vowel is to be written.
- Writing target consonant clusters with explicit instruction about form (*when* is a consonant written as a 'half-letter') and its writing routine (*how* is the akshara constructed).
- Writing meaningful words, labels and messages which fulfill a genuine communicative function.

Pedagogical Challenges

In this chapter, we have explored some of the key features of an engaging language- and meaning-focused approach to literacy instruction for pre-primary and the first 2 years in primary school (kindergarten and grades 1 and 2). A basic pre-requisite for this is a safe and secure classroom environment, which encourages children to "relate with each other, with the teacher and with the cultural world of the language" (Mathur, 2013e, p. 43). We now discuss some pedagogical challenges that will need to be addressed for such classrooms to become a reality across the region.

Curricular Materials and Enrichment Materials

The relational and language-focused approach to teaching *akshara* languages needs to be undergirded by curricular materials that systematically address the needs of heterogeneous learners in a culturally appropriate manner. A key assumption is the availability of child-friendly poems, songs and other texts. These could be in folk or classical idiom, oral or written, inherited or borrowed, full of history or very contemporary. A second assumption is that these textual sources are available in a style that is coherent and accessible to teachers and children. Finding or creating such textbooks, workbooks and similar child-friendly resources is a challenge in most *akshara*-based languages, and clearly an area that needs significant attention. Two initial criteria for the development and evaluation of appropriate curricular materials are: (1) does the material support *active engagement* of learners; and (2) do the activities proposed in the material have teaching-learning objectives that are *precise, and explicit*?

A further point to be considered is whether the material introduces and develops a sequence of concepts in a manner that avoids 'big jumps' in expectations from learners, for these are liable to make them lose motivation. Related to smooth transitions are the criteria for *balance* between meaning-focussed, culturally responsive activities and practice activities that may be potentially de-contextualised and formulaic. Another, perhaps obvious but surprisingly difficult to execute, criterion is whether the curricular materials are well laid out, with text and illustrations supporting each other and representing a range of cultures that young learners can relate to.

Teacher Preparation and Teacher Support

However, good material is simply not enough to actively support higher attainments. Even when materials are available, for example, to encourage child talk in the classroom or celebrate diversity, they can be rendered dull and lifeless (e.g., Nag et al., 2014a). Leadership by a skilled teacher is therefore a critical second assumption of the language-focused approach to literacy instruction.

Professional development for teachers in the language-focused approach covers several areas. A primary skill is to maintain a balance between listening, speaking, reading and writing instruction. Insights are required for teaching not only the sound, symbol, and semantic domains, but also the implicit details of morphosyntax (as outlined, for example, in the sections above). Another crucial skill is the ability to provide opportunities to all learners, but particularly to support the struggling learner, to experience success because, in our understanding, success experiences help learners to remain engaged with a complex learning task.

The idea of having precise language learning objectives for individual classroom activities is not well-established among many language teachers (and even teacher educators!), and needs to be inculcated systematically. For instance, learning to recite a poem is common in most pre-primary and primary classrooms. Yet few teachers stop to question why children are learning a particular poem, on what basis a poem should be chosen, how it should best be recited to bring out its meaning, and what further language work can be linked to a poem. Once such awareness develops, teachers will be able to make connections across orthography, phonology, semantics, and syntax, and with culturally embedded materials such as rhymes, songs and stories.

An important pre-requisite for enabling such an approach to literacy instruction is that teachers are open to becoming learners themselves and are willing to engage in reflexive practice, deriving personal meaning from the acts of learning they are able to facilitate in their classrooms. While this can be supported by reading of professional development materials—like teachers' manuals, professional journals, and research reports—nurturing a culture of discussion and dialogue among teachers is perhaps the most promising available strategy for encouraging introspection and reflection. Here, teachers are encouraged to bring up real issues from their classrooms and seek wholesome solutions. Shifting to such a culture of learning within the language faculty in schools is in itself a very real challenge.

Child Interest and Motivation

Children's attitude to learning a language significantly impacts their engagement with the teaching-learning process and, unsurprisingly, the pace of their learning (see Vagh, 2009). Hence, it becomes important that the teaching-learning materials as well as teacher-mediated processes that children experience in the classroom support and sustain a positive engagement with the language.

In our experience, children enjoy learning something when they can relate to it at the level of content, format, and aesthetic experience. Children have an innate appreciation of order and beauty. Classrooms as well as teaching-learning materials need to be responsive to this sensibility. The themes and topics presented need to reflect, in an age-appropriate manner, the areas of familiarity, interest and curiosity of learners. When learning experiences bring into play activities that, by and large, children enjoy doing, such as singing, playing games, or play-acting, then they bring their entire persona to the classroom and learn in a variety of non-linear ways, alongside the expected learning outcomes.

The other factor which seems to help in keeping children motivated, as mentioned above, is linked to the experience of success in meeting the explicit and implicit expectations of the teacher. Accomplishing set tasks which they can do independently encourages children to be more active; conversely, having to do tasks that they are not yet ready to do, or engage with texts that make little sense to them, or not understanding why answers are marked wrong, are clear reasons for demotivating students.

Last, but not the least, the language and literacy program needs to be in a space that is free of fear, of teacher domination, and of peer ridicule. In such an atmosphere of security children can process events and occasions from their own lives through meaningful reading and writing tasks. It is through such processing that children learn to own the language, to negotiate the complexities of both its spoken and written forms, as well as to communicate effectively with others.

Conclusion

We have argued that literacy acquisition in *akshara*-based languages needs a relational, meaning-based, language-focused approach which addresses first and second language learners, as well as children with diverse levels of opportunities outside of the classroom. This demands a clearer understanding of the specific needs of first-time literacy instruction and also of how these requirements shift when it comes to second- and third-time literacy instruction in second and third languages. There is a need to understand what happens in classrooms with an *akshara* heritage and why some practices are harder to shift than others. We have further attempted to highlight how some of the available knowledge-bases related to morpho-phonology could be deployed for a more effective literacy instruction anchored in cultural-bases specific to *akshara* languages. The assumptions of this approach require that teachers are supported in rethinking the ways in which they structure their classes, in updating their pedagogical content knowledge, and in appropriately modulating the tone and variety of their classroom discourse. More broadly, there is a need to build on the available descriptive and anecdotal reports of good practice with a more robust evidence-base of what works.

References

Aboud, F. E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly, 21*(1), 46–60.

Adams, M. (1994). *Beginning to read: Thinking and learning about print*. Cambridge, MA: The MIT Press.

Amritavalli, R. (2007). *English in deprived circumstances: Maximising learner autonomy*. Hyderabad, India: Foundation Books. https://doi.org/10.1017/UPO9788175968684.

Annual Survey of Education Report. (2017, Jan). *Annual survey of education report, 2016*. ASER Center: New Delhi, India.

Arulmani, G. (2016). Contexts and circumstances: The cultural preparation process approach to career development. In M. McMahon (Ed.), *Career counselling: Constructivist approaches* (2nd ed., pp. 79–90). Abingdon, UK: Routledge.

ASER 2016, (2017). *Annual Status of Education Report (Rural) 2016: Provisional Report*. http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%202016/aser_2016.pdf. retrieved on 20th April, 2018.

Aturupane, H., Glewwe, P., & Wisniewski, S. (2013). The impact of school quality, socioeconomic factors, and child health on students' academic performance: Evidence from Sri Lankan primary schools. *Education Economics, 21*(1), 2–37. https://doi.org/10.1080/09645292.2010.511852.

Badheka, G. (1932/1986). *Divaswapna*. http://www.mavjibhai.com/EBOOKS/Divaswapna_Eng.pdf

Bhattacharjea, S., Wadhwa, W., & Banerji, R. (2011). *Inside primary schools. A study of teaching and learning in rural India*. Mumbai, India: Pratham.

Bhide, A. (2018). Copying helps novice learners build orthographic knowledge: Methods for teaching Devanagari akshara. *Reading and Writing, 31*, 1–33.

Educational Initiatives. (2010). *Student learning study*. Ahmedabad, India: Educational Initiatives.

Geetha, V. (2012). Literacy and reading: A Tamil experiment. *Contemporary Education Dialogue, 9*, 63–84.

Gupta, A. (2004). Reading difficulties of Hindi speaking children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 17*, 79–99.

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly, 29*(4), 715–733.

Jayaram, K. (2008). Early literacy project: Explorations and reflections. Part 1: Theoretical perspectives. *Contemporary Education Dialogue, 5*(2), 133–174.

Kalia, V., & Reese, E. (2009). Relations between Indian children's home literacy environment and their English oral language and literacy skills. *Scientific Studies of Reading, 13*(2), 122–145.

Karanth, S. (1995). *oduva ata* 'Reading Games'. Bangalore, India: SBs Publishers.

Kshirsagar, S., & Poovaiah, R. (2015). *Investigating frequently used stroke sequence, of handwritten Devanagari letters, by observing second script learners.* Paper presented in the International Conference on Research into Design, Indian Institute of Science, Bangalore.

Krishnamurti, J. (1985). *Life ahead*. London, UK: Victor Gollancz Ltd.

Khubchandani, L. M. (1981). *Language, education and social justice*. In search of tomorrow, Issue 2. Pune, India: Centre for Communication Studies.

Kumar, K. (1994). *Child's language and the teacher*. New Delhi, India: National Book Trust.

Kurrien, Z. (2014). *Helping young children learn*. Hyderabad, India: Orient Blackswan.

Mathur, C. (2013a). *Hindi ki Duniya 1, Textbook*. Hyderabad, India: Orient Blackswan.

Mathur, C. (2013b). *Hindi ki Duniya 1, Workbook*. Hyderabad, India: Orient Blackswan.

Mathur, C. (2013c). *Hindi ki Duniya 2, Textbook*. Hyderabad, India: Orient Blackswan.

Mathur, C. (2013d). *Hindi ki Duniya 3, Textbook*. Hyderabad, India: Orient Blackswan.

Mathur, C. (2013e). *Hindi Ki Duniya, A Guide of Teachers 1, For Course Book 1 & 2*. Hyderabad, India: Orient Blackswan.

Menon, S., & Thirumalai, B. (2016). Curricular materials in early language and literacy classrooms in Karnataka and Maharashtra. In D. Nawani (Ed.), *Teaching-learning resources for school education* (pp. 394–404). New Delhi, India: Sage Publications.

Mohanan, K. P. (1986). *The theory of lexical phonology*. Dordrecht: Reidel.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22.

Nag, S. (2013). Low literacy attainments in school and approaches to diagnosis: An exploratory study. *Contemporary Education Dialogue, 10*(2), 197–221.

Nag, S. (2014). Akshara-phonology mappings: The common yet uncommon case of the consonant cluster. *Writing Systems Research, 6*, 105–119.

Nag, S. (2017). Learning to read Kannada and other languages of South Asia. In L. Verhoeven & C. Perfetti (Eds.), *Reading acquisition across languages and writing systems*. Cambridge, UK: Cambridge University Press.

Nag, S., & Narayanan, B. (this volume). Orthographic knowledge, spelling and reading development in Tamil: The first three years. In M. Joshi & C. McBride (Eds.), *Handbook of akshara languages*. Cham, Switzerland: Springer.

Nag, S., & Sircar, S. (2008). *Learning to read in Bengali: Report of a survey in five Kolkata primary schools*. Bangalore, India: The Promise Foundation.

Nag, S., & Snowling, M. J. (June, 2011). Reading comprehension, decoding and oral language. *The EFLU Journal, English and Foreign Languages University, 2*, 75–93.

Nag, S., Snowling, M. J., & Asfaha, Y. (2016). Classroom literacy practices in low- and middle-income countries: An interpretative synthesis of ethnographic studies. *Oxford Education Review, 42*(1), 36–54. https://doi.org/10.1080/03054985.2015.1135115.

Nag, S., Ramkumar, S., Miranda, R., Sutar, L., Krishna, M., Aravind, S., Kala, B., & Arulmani, G. (2014a, October). *Home and school learning environment: Field notes from eight urban, public-funded nursery and primary schools* (Working Paper). Bangalore, India: The Promise Foundation.

Nag, S., Snowling, M. J., Quinlan, P., & Hulme, C. (2014b). Child and symbol factors in learning to read a visually complex writing system. *Scientific Studies of Reading, 18*(5), 309–324.

Nag, S., Vagh, S. B., Dulay, K. M., & Snowling, M. J. (2018). Home language, school language and children's literacy attainments: A systematic review of evidence from low- and middle-income countries. *Review of Education: An International Journal of Major Studies in Education*. https://doi.org/10.1002/rev3.3130.

National Achievement Survey. (2014). *Class V: Subject wise report*. New Delhi, India: NCERT.

Patel, P. G. (2004). *Reading acquisition in India: Models of learning and dyslexia*. New Delhi, India: Sage Publishers.

Poe, E. A. (1962). The raven. In *Selected stories and poems* (pp. 223–227). New York, NY: Airmont Books.

Ray, S. (2004). *Abol Tabol: The nonsense world of Sukumar Ray* (Translated from Bengali into English by Sampura Chattarji). New Delhi, India: Penguin.

Reese, L., Garnier, H., Gallimore, R., & Gldenberg, C. (2000). Longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of Spanish-speaking students. *American Educational Research Journal, 37*(3), 633–662. https://doi.org/10.3102/00028312037003633.

Sahi, J. (2015). *'In our own words: A resource book for primary school teachers'*, Eklavya, 2015, Bhopal, India.

Saigal, A. (2012). Demonstrating a situated learning approach for in-service teacher education in rural India: The Quality Education Programme in Rajasthan. *Teaching and Teacher Education, 28*, 1009–1017.

Sen, R., & Blatchford, P. (2001). Reading in a second language: Factors associated with progress in young children. *Educational Psychology, 21*(2), 189–202.

Sircar, S., & Nag, S. (this volume). Spelling and reading words in Bengali: The role of distributed phonology. In M. Joshi & C. McBride (Eds.), *Handbook of akshara languages*. Cham, Switzerland: Springer.

Suteyev, V. (2002). *Me too: A picture story*. Translated into Hindi from Russian. Bhopal, India: Eklavya Publishers.

Tagore, R. (1918). *tootakaahini* 'The Parrot's Tale'. Sobuj Patra, Kolkata: India http://dsal.uchicago.edu/digbooks/digpager.html?BOOKID=PK1664.D6_1988&object=25. Retrieved on 20th April, 2018.

Teale, W. H., & Sulzby, E. (1986). Introduction: Emergent literacy as a perspective for examining how children become writers and readers. In W. H. Teale & E. Sulzby (Eds.), *Emergent literacy: Writing and reading*. Norwood, NJ: Ablex Publishing Corporation.

Vagh, S. B. (2009). *Learning at home and at school: A longitudinal study of Hindi language and emergent literacy skills of young children from low-income families in India* (Unpublished doctoral dissertation). Graduate School of Education, Harvard University.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690.

Verhoeven, L. T. (1994). Transfer in bilingual development: The linguistic interdependence hypothesis revisited. *Language Learning, 44*(3), 381–415.

Vygotsky, L. S. (1986). In A. Kozulin (Ed.), *Thought and language*. Cambridge, MA: MIT Press.

Wijayathilake, D., & Parrila, R. (2014). Predictors of word reading in good and struggling readers in Sinhalese. *Writing Systems Research, 6*, 120–131.

World Declaration on Education For All: Meeting Basic Learning Needs. (1990). World Conference on Education for All, Jomtein, Thailand, March 1990. http://www.un-documents.net/jomtien.htm. Retrieved on 8th April, 2018.

Winskel, H., & Iemwanthong, K. (2010). Reading and spelling acquisition in Thai children. *Reading and Writing: An Interdisciplinary Journal, 23*, 1021–1053.

Challenges in Learning Akshara orthographies for Second language Learners

Adeetee Bhide and Charles Perfetti

Abstract Marathi is a language derived from Sanskrit and spoken in the state of Maharashtra, India. In this chapter, we describe Marathi's orthographic and phonological systems. The script used to write Marathi is Devanagari. Because Devanagari is described in the Hindi chapter (Singh & Sumathi), we focus our description on how it is implemented in Marathi and how this differs from how Devanagari is implemented in Hindi. In addition to describing the orthographic and phonological systems separately, we explain how Marathi's orthography codes phonology. Written Marathi is generally transparent. However, there are some exceptions including graphs that code multiple phonemes and an inconsistent expression of the schwa vowel. After describing the language and its writing system, we summarize experiments with both native Marathi speakers and second language learners of Marathi. For native speakers, we explain the consequences of inconsistent schwa expression for speakers' awareness of the vowel. For second language learners, we review experiments comparing different methods for teaching Marathi's orthographic and phonological systems.

Keywords Devanagari · Marathi · Orthography · Phonology · Second language acquisition

India has two official languages, Hindi and English. Additionally, people in many of India's 29 states speak their own languages. The result is 417 living, indigenous languages (Lewis, Simons, & Fennig, 2016), but only 22 scheduled[1] languages

[1] The eighth schedule to the Constitution of India lists the 22 official languages of India. The Indian government is obligated to support these languages so they become an effective means for modern communication. Furthermore, candidates may use any of these languages when taking public service examinations ("Official Languages: Constitutional Provisions," 2016).

A. Bhide (✉) · C. Perfetti
Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

Learning Research and Development Center, Pittsburgh, PA, USA
e-mail: arb135@pitt.edu; arbhide89@gmail.com; perfetti@pitt.edu

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_16

(Bondale et al., 2013). The northern states speak languages mostly in the Indo-Aryan language family and the southern states speak languages generally in the Dravidian language family (Vaid & Gupta, 2002).

Marathi is an Indo-Aryan language and, like Hindi, is descended from Sanskrit. It is the state language of Maharashtra and is one of the 22 scheduled languages. In addition to Maharashtra, Marathi is widely spoken in several other Indian states including Andhra Pradesh, Chhattisgarh, Goa, and Karnataka. There are 42 dialects of Marathi including Gawdi of Foa, Kasargod, Kosti, Kudali, and Nagpuri Marathi. In India, there are 71,700,000 and 3,000,000 L1 and L2 speakers of Marathi, respectively. Marathi is an immigrant language in Canada, Israel, Mauritius, the United Kingdom, and the United States; there are 75,760 L1 speakers of Marathi outside of India (Lewis et al., 2016). Maharashtra is centrally located, and is bordered to the North and East by states that speak Hindi and Gujarati (Indo-Aryan languages) and to the south by states that speak Kannada and Telugu (Dravidian languages).[2] Due to Maharashtra's central location near Kannada and Telugu speaking areas, its Marathi language is strongly influenced by Dravidian languages (Bondale et al., 2013). In fact, there are some dialects that are very strongly influenced by the surrounding languages; for example, a dialect spoken in Tamil Nadu is strongly influenced by both Tamil and Kannada (Lewis et al., 2016).

Marathi's Orthographic and Phonological Systems

Marathi Orthography

Marathi is written in the Devanagari script, which is also used to write Hindi, Sanskrit, and Nepali languages (Sinha & Mahabala, 1979). Devanagari is an abugida, written with graphs called aksharas. Aksharas represent open syllables; however, some subcomponents of the aksharas represent phonemes. There are three types of aksharas: simple, consonant-vowel (CV), and complex. There are two types of simple aksharas—consonant and vowel. Consonant aksharas (e.g., स /sə/) represent consonants and inherent schwa vowels. Vowel aksharas (e.g., उ /u/) represent vowels. CV aksharas represent consonants and vowels (e.g., तु /t̪u/) and consist of consonants with attached vowel diacritics. Note that the vowel replaces the inherent schwa. Complex aksharas have two or more consonants and may also have a vowel diacritic (e.g. स्तु /st̪u/).

[2] It is also bordered to the South by Goa, whose official language is Konkani (an Indo-Aryan language). However, because that border is very short, we omitted it for simplicity's sake.

Vowels are represented with simple vowel aksharas only when the vowels do not follow a consonant. Thus, vowels are most commonly found in their simple vowel akshara form at the beginnings of words (e.g., ईडलिंबू /i:ɖəlimbu:/ (lime)[3] where /i:/ is represented by the simple vowel akshara ई). However, simple vowel aksharas can be found in other word positions when there are two consecutive vowels (e.g., सुई /sui:/ (needle) where /i:/ is represented by the simple vowel akshara ई). When a vowel follows a consonant, it is expressed with a diacritic. Because vowels typically follow consonants, the diacritic vowel form is more common than the simple vowel akshara form.

Whereas the simple vowel akshara is the full form, diacritics can be considered the "half" form of the vowel. In the same way, consonants also have half and full forms; the full forms are used in simple consonant and CV aksharas and the half forms are used in complex aksharas. The most common way to form a complex akshara is to remove the right-most portion of the first consonant and physically concatenate it to the second consonant (e.g., स + त = स्त). For some of the rounded consonant graphs, the consonants are vertically joined (e.g., ट + ट = ट्ट ; ड + ड = ड्ड). When द is the first consonant, the second consonant is often attached to its round part (e.g., द + व = द्व ; द + ध = द्ध). Complex aksharas including र (/r/) are unique in that the र is not easily visible. When the र is the first consonant, it is always written as a curved line above the second consonant (e.g., र + द = र्द ; र + त = र्त). When the र is the second consonant, it is written as one diagonal line when attaching to a vertical line and two diagonal lines when attaching to a curve (e.g., प + र = प्र ; क + र = क्र ; ट + र = ट्र ; ड + ड = ड्र). For some complex aksharas, there are two correct ways of writing them. For example, both ल्ल and ल्ल are acceptable ways of joining ल and ल. Finally, some complex aksharas do not look like either of their components and must be memorized (e.g., क + ष = क्ष ; श + र = श्र) (Bhide, 2017a). The result of this coding complexity is a high degree of visual complexity. This complexity and the size of the grapheme inventory can slow orthographic learning (Nag, 2014).

Although complex akshara formation is largely rule-based, these rules can be quite complex and result in non-linear graphs. For example, in the complex akshara स्का /ska/, the three phonemic components are ordered left to right; in the complex akshara द्वि /d̪wi/, the three phonemic components are ordered right to left; and in the complex akshara ट्टु /ʈʈu:/, the three phonemic components are ordered top to bottom.

Although aksharas are supposed to represent syllables, the coding of syllabic structure by the orthography is more complex. For example, the aksharas do not represent all syllables; they represent only open syllables. Therefore, any CVC syllable (e.g., दम /d̪əm/) must be represented by two aksharas. The inherent schwa

[3] All Marathi pronunciations and definitions were checked using the following Marathi-English dictionary: सुबोध मराठी–इंग्रजी शब्दकोश written by कृष्णाजी भास्कर वीरकर and published by अनमोल प्रकाशन in पुणे in २००४.

vowel is also quite complex. In Sanskrit, the inherent schwa vowel was always pronounced unless its absence was marked with a halanta or virama, a diagonal line attached to the bottom of the consonant that indicated omission of the schwa (e.g., म् is pronounced /m/, not /mə/).[4] In Marathi, the inherent schwa vowel is often not pronounced at syllabic breaks or at the ends of words, but its omission is not marked. For example, if all of the inherent schwas in the word पटकन (quickly) were pronounced, it would be pronounced /pə.ʈə.kə.nə/. However, the inherent schwas at syllabic breaks and at the end of the word are dropped, and the word is pronounced /pəʈ.kən/. The omission of these schwa vowels is not marked; native speakers have an intuitive understanding of when to pronounce the schwa (Salomon, 2000). Finally, complex aksharas can either represent consonantal blends (which are within the same syllable) or consecutive consonants that cross a syllabic boundary. For example, in the word स्तुति /st̪ut̪i/ (praise), स्तु /st̪u/ is written with a complex akshara that represents a consonantal syllable blend. In contrast, in the word रक्त /rək.t̪ə/, /k/ and /t̪/ are consecutive consonants separated by a syllable boundary but are written as a complex akshara.

Marathi Phonology

Marathi has six monophthongs (ə, a, i, u, e, o) and two diphthongs (əi, əu) (Bondale et al., 2013). Both *i* and *u* have two length variations (short and long) that are differentiated in the script. Dhongade and Wali (2009) claim that these variations are positionally-dependent and not phonemic. Pandharipande (1997) notes that this is not necessarily the case: there are some minimal pairs that are only distinguished on vowel length. For example, पुर /pur/ and पूर /pu:r/ mean town/bury(IMP) and flood, respectively.

Marathi has also added two vowels (æ, ɔ) that are used only in borrowed words. Additionally, it has three sounds (ru, nasalization, and breathy ending) that are represented by vowel-like diacritics, although they are not vowels per se. Marathi has 46 consonants that are represented by 34 consonantal aksharas. There are 28 consonantal aksharas that represent only one consonant and 6 consonantal aksharas that represent two consonants. Additionally, there are 6 aspirated consonants that are represented by a complex akshara rather than a simple akshara (Bondale et al., 2013). All of the graphs with their corresponding sounds are in the tables below.

[4] The halanta/virama also provides an alternative method for representing complex aksharas. Rather than physically attaching the consonants, one can also put a halanta/virama on the first consonant. For example, स्त is the same as स्‌त.

Marathi vowels

Akshara	Phoneme	Vowel diacritic written with प	Akshara	Phoneme	Vowel diacritic written with प
अ	ə	प (inherent)	ए	e	पे
आ	a	पा	ऐ	əi	पै
इ	i	पि	ओ	o	पो
ई	i:	पी	औ	əu	पौ
[3]उ	u	पु	[2]अं	nasalization	पं
[3]ऊ	u:	पू	[2]अः	breathy ending	पः
[2]ऋ	ru	पृ	[1]ॲ	æ	पॅ
			[1]ऑ	ɔ	पॉ

[1]These vowels are used only in borrowed words
[2]Although these are not vowels per se, they are represented with vowel-like diacritics
[3]These vowel diacritics look different when attached to र /r/; the CV aksharas are रु /ru/ and रू /ru:/

Marathi consonants

Akshara	Phoneme	Akshara	Phoneme	Akshara	Phoneme	Akshara	Phoneme
क	k	ट	ʈ	प	p	[2]ष	ʂ
ख	k^h	ठ	ʈ^h	[4]फ	p^h/ f	स	s
ग	g	[4]ड	ɖ/ ɽ	ब	b	ह	h
घ	g^h	ढ	ɖ^h	भ	b^h	ळ	ɭ
[1]ङ	ŋ	ण	ɳ	म	m	[3]म्ह	m^h
[4]च	t͡ʃ/ t͡s	त	t̪	य	j	[3]न्ह	n^h
छ	t͡ʃ^h	थ	t̪^{h}	र	r	[3]ण्ह	ɳ^h
[4]ज	d͡ʒ/ z	द	d̪	ल	l	[3]ल्ह	l^h
[4]झ	d͡ʒ^h/ z^h	ध	d̪^{h}	[4]व	w/v	[3]व्ह	w^h
[1]ञ	ɲ	न	n	[2]श	ʃ	[3]ऱ्ह	r^h

[1]These aksharas are never written out in Marathi words; instead their sounds are represented with a nasalization vowel diacritic
[2]Although these sounds were pronounced differently in Sanskrit, modern Marathi has lost the pronunciation distinction. However, the distinction is still maintained in the orthography (Bondale et al., 2013)
[3]Marathi has 16 aspirated phonemes. Of these, 10 are represented by a single akshara. These six are represented by complex aksharas combining the consonant and /h/ (ह) (Bondale et al., 2013)
[4]These six aksharas have two possible pronunciations

The nasalization and breathy ending diacritics are unique in that they are the only diacritics that can be used in conjunction with another. For example, one would never attach both the /u/ and /e/ diacritics to the same consonant. However, it is possible to attach a vowel diacritic and the nasalization diacritic to the same consonant; for example हिं /hin/ has both the /i/ and nasalization diacritics. Similarly, one can attach both the /i/ and breathy ending diacritics to form निः (Dhongade & Wali, 2009).

Furthermore, one typically cannot attach a vowel diacritic to a simple vowel akshara; to write /ue/ one would never write उे. However, it is possible to put the nasalization diacritic over a simple vowel akshara; for example in the word आंबा /amba/ (mango), the nasalization diacritic is over the simple vowel akshara आ.

There are a few complex aksharas that look nothing like their components; for example क्ष is a blend of क /k/ and ष /ʂ/ and श्र is a blend of श /ʃ/ and र /r/. Because these complex aksharas are special in their opaque configuration, they are often listed on the consonant chart as individual entries.

Marathi Is Not Completely Transparent

Marathi is relatively transparent, with one akshara representing every phoneme. However, there are some notable exceptions. Below, we detail the opaque aspects of Marathi separately for consonants, vowels, and nasals.

Consonants There are six aksharas (see consonants table, footnote 4) that have two possible pronunciations. Often, these dual pronunciations stem from the representation of foreign phonemes with existing graphs. For example, च is pronounced as /t͡s/ in most Marathi words, and the /t͡ʃ/ pronunciation is only used in words borrowed from Sanskrit (Dhongade & Wali, 2009). Similarly, the /f/ pronunciation of फ is reserved for English or Perso-Arabic loan words (Pandharipande, 1997). In some instances, these differences are allophonic and thus do not affect the meaning of the word. For example, both /wiʃran̪t̪i/ and /viʃran̪t̪i/ are acceptable pronunciations of विश्रांती (rest). In other instances, changes in pronunciation do affect the meaning of the word. For example, जग is a homograph that can mean either *world* when pronounced as /dʒəg/ or *to live* when pronounced as /zəg/ (Bondale et al., 2013).

There is one case in Marathi in which two aksharas represent the same phoneme. In Sanskrit, श and ष represent the sounds /ʃ/ and /ʂ/, respectively. However, the pronunciation difference between the aksharas is diminishing in Standard Marathi, but the orthography continues to distinguish between the two (see consonant table, footnote 2) (Bondale et al., 2013).

Furthermore, Marathi has 16 aspirated consonants. Although 10 of them are represented by their own aksharas, 6 of them are represented with complex aksharas combining the consonant and /h/ (see consonant table footnote 3) (Bondale et al., 2013).

Finally, although the orthography represents many geminates (two identical adjacent consonants) it does not represent them all. For example in the word किल्ली /kil.li:/ (key), there are two adjacent /l/ sounds in adjacent syllables. The orthography represents this geminate by attaching ल /l/ to itself in the consonant cluster ल्ल. However, other geminates are not represented in the orthography. For example, the consonant cluster व्य can either be pronounced as /vj/ (no geminate) or /v.vj/ (geminate). These alternative pronunciations can result in homographs. For example, दिव्याजवळ can be pronounced as either /div.vja.zə.wəɭ/ (with the geminate) meaning *near the girl named Divya* or as /di.vja.zə.wəɭ/ (without the geminate) meaning *near the lamp* (Bondale et al., 2013).

Vowels A small source of vowel opacity is that some words borrowed or derived from Sanskrit are written with an /u/ or /i/ in the word final position, but they are pronounced as /u:/ and /i:/ (e.g., गुरु /guru/ (religious leader); मति /mət̪i/ (intellect)) (Pandharipande, 1997).

The largest source of vowel opacity is the inherent schwa vowel. In Marathi, every consonantal akshara includes an inherent schwa vowel. If there is no vowel diacritic, the schwa is assumed to be present. In Sanskrit, the schwa was always pronounced unless the akshara was marked with a halanta or virama. In Modern Marathi, schwa vowels often are not pronounced at syllable boundaries or at the ends of words. However, the schwa syncope is not marked, so native speakers use semantic information to derive the correct pronunciation for words. Because of schwa syncope, the same sequence of aksharas can be pronounced differently. For example, in the word हरक /hərək/ (a kind of cotton cloth), the sequence रक is pronounced as /rək/. In contrast, in the word हरकत /hər.kət̪/ (an objection, an opposition, a hindrance), the sequence रक is pronounced as /r.k/. Furthermore, syllabic breaks between consonants can be represented either as two consonantal aksharas (and the schwa between them is understood to be deleted) or as a consonantal conjunct (the syllabic break between the two consonants is understood). These two modes of representation can lead to a case in which the same phonological sequence is represented by two orthographic representations. For example, the phonological sequence /r.t̪a/ can either be written as रता or as र्ता in the words हरताळ /hər.t̪aɭ/ (closing of shops or business, a strike) and हर्ता /hər.t̪a/ (remover), respectively.

Finally, in historical Marathi, some words ended with a nasalized /e/ (/ẽ/). In most dialects of Marathi, this vowel has been shortened to a schwa vowel. The anuswara that represents that nasalization is still retained in the orthography, but in contemporary Marathi it represents the schwa vowel (Pandharipande, 1997).

Nasals The nasalization vowel diacritic (anuswara) can represent either a vowel or a consonant. Thus, it does not fit neatly into any of the above categories and needs to be addressed separately.

The anuswara can represent both a nasalized vowel and the nasal in a consonantal cluster that has both a nasal and non-nasal consonant. When representing a nasal consonant, the anuswara can represent any of the five nasal sounds (m, n, ɳ, ɲ, or ŋ). Which nasal sound mostly depends on which consonant follows the nasal and, to a small extent, the speaker (Bondale et al., 2013). One example of the following consonant affecting the pronunciation of a nasal consonant is in the words आंबा /amba/ (mango) and हिंदी /hind̪i:/ (Hindi). In /amba/, the anuswara is realized as the bilabial nasal /m/ because it is followed by the bilabial stop /b/. In /hind̪i/, it is realized as the alveolar nasal /n/ because it is followed by the dental stop /d̪/. An example of the speaker affecting the pronunciation of the nasal consonant is in the word किंचित (very little); both /kint͡ʃit̪/ and /kiɲt͡ʃit̪/ are correct pronunciations (Bondale et al., 2013). An example of the speaker affecting the pronunciation of the nasalized vowel is in the word सिंह (lion); both /sĩhə/ or /siũhə/ are correct pronunciations (Bondale et al., 2013).

When nasal consonants are not part of a consonant cluster, they are always written with an akshara, never an anuswara. Although m, n, and ɳ can occur outside of consonant clusters, ɲ and ŋ are only found within consonant clusters (Pandharipande, 1997). When ɲ and ŋ occur in consonant clusters, they are always represented with an anuswara. Therefore, although there are aksharas that represent these sounds, the aksharas are never used when writing Marathi (see consonant table, footnote 1). When m, n, and ɳ occur in consonant clusters, they can often be represented either with an anuswara or a conjunct consonant. Therefore, हिन्दी and हिंदी are both correct spellings of *Hindi*.

Differences Between Hindi and Marathi

Although Hindi and Marathi are both written in the Devanagari script, there are some orthographic differences between the two languages. Some of these orthographic differences stem from phonological differences, whereas others arise because Hindi, but not Marathi, disambiguates consonantal aksharas that have two pronunciations. Marathi has one phoneme(ɭ), that is used in Dravidian and Old Sanskrit (Pandharipande, 1997), but not in Hindi. Thus, the graph that represents the phoneme (ळ) is present in Marathi but not in Hindi (Rathod, Dhore, & Dhore, 2013). Furthermore, several aksharas are pronounced differently in Hindi and Marathi. In Marathi, ऐ is pronounced as /əi/ and औ is pronounced as /ou/. In contrast, in Hindi, ए is typically pronounced /ɛ/, although it has the same diphthong /əi/ pronunciation as Marathi when preceding /j/. Similarly, in Hindi, औ is typically pronounced as /ɔ/, but has the same diphthong /əu/ pronunciation as Marathi when preceding /v/ (Kachru, 2006). The pronunciation of ऋ is also different in the two languages. In Sanskrit, that akshara was pronounced as a retroflex vowel. In modern languages, it is no longer pronounced as a vowel (Kachru, 2006). In Marathi ऋ is pronounced as /ru/ whereas in Hindi it is pronounced as /ri/ (Kachru, 2006). The spelling of *Sanskrit* has ऋ in it. In Hindi, the word is pronounced as /Səũskrit̪/ whereas in Marathi it is pronounced as /Səũskrut̪/. The pronunciation of ज्ञ is also different in Hindi and Marathi. In Devanagari, ज्ञ is an irregular ligature of ज /d͡ʒ/ and ञ /ɲ/. In Hindi, it is pronounced as /gjə̃/ (Kachru, 2006), whereas in Marathi it is pronounced as /d̪ɲjə/ (Pandharipande, 1997).

Many of the consonants that are ambiguous in Marathi are not ambiguous in Hindi. For example, in Marathi, च can be pronounced as either /t͡ʃ/or /t͡s/. In Hindi, it can be pronounced only as /t͡ʃ/; the /t͡s/ sound does not occur in Hindi. Further, in Marathi, फ, ज, and झ are ambiguous; all three graphs have two possible pronunciations. In contrast, in Hindi, dots known as *nuktas* are added to disambiguate the aksharas. For example, in Marathi, फ can be pronounced as either /pʰ/ or /f/. In contrast, in Hindi, फ is pronounced as /pʰ/ and .फ is pronounced as /f/. Similarly, in Hindi, ज and .ज are pronounced as / d͡ʒ/ and /z/ respectively and झ and .झ are pronounced as / d͡ʒʰ/ and /zʰ/ respectively.

Studies with Native Speakers and Second Language Learners of Marathi

In this section, we describe some of the consequences of Marathi's phonology and orthography, drawing on our empirical studies of native speakers and L2 learners. First, we describe how Marathi's orthographic and phonological systems affect the perception of spoken Marathi by native speakers. We then summarize two experimental studies of learning a set of aksharas and one study of learning Marathi phonemes by English background speakers.

Phonological Awareness of Native Marathi Speakers

As we described above, the Marathi orthography inconsistently expresses the schwa vowel. The schwa vowel is written as a full akshara only when it does not follow a consonant, typically at the beginning of words. Almost all medial vowels are unexpressed; they are inherent in the consonantal aksharas. Many words with a final schwa express the schwa with an anuswara to reflect the fact that they historically ended with the /ẽ/ sound.

Bhide, Gadgil, Zelinsky, and Perfetti (2014) examined how these various orthographic representations affect native speakers' awareness of the schwa vowel. In the study, participants had to divide words into their constituent phonemes. For example, if a participant heard the word *cat*, s/he had to respond with */k/, /æ/, /t/*. All of the participants could speak both English and Marathi. The stimuli were both Marathi and English words and were chosen such that there were schwa vowels and other vowels in word initial, medial, and final positions. For the Marathi words, when the word began with a schwa or other vowel, they were represented with a simple vowel akshara. Medial schwas were unexpressed but other medial vowels were expressed with diacritics. Final schwas were expressed with anuswara and other final vowels were expressed with diacritics. For the English words, all vowels were expressed with letters.

The data on Marathi words show that orthographic properties significantly affected participants' performance on this phonological awareness task. Specifically, participants were highly accurate for non-schwa vowels in all three positions. For the schwa vowel, participants' responses fell into four categories. Fifteen participants were highly accurate on initial schwas but omitted almost all of the medial and final schwas. For these participants, the orthographic support provided by the simple vowel aksharas helped them recognize the initial schwas. However, they failed to produce the medial schwas (unexpressed) and the final schwas (expressed by an anuswara). We know that this accuracy difference was primarily driven by orthography rather than phonological saliency because a difference of the same magnitude was not seen for English words. Four participants omitted almost all medial schwas, but were highly accurate on the initial and final schwas. The orthographic explanation for this is that the anuswara at the end signified a vowel, leading these participants to correctly recognize the final schwas. However, they too failed to produce the unexpressed medial schwas. Three participants did not omit schwas, but they inserted schwas at syllabic boundaries and at the ends of words where they are not pronounced. These participants seemed to know that consonantal aksharas have inherent schwas. However, their overproduction of schwas indicates either a failure to recognize that schwas are often not pronounced at syllabic boundaries and at the ends of words or that they used they simply "read" the vowel that was part of the akshara rather than attending to the whole word. Finally, one participant was highly accurate; she did not omit or insert schwas. This participant reported that she had learned Sanskrit and the fact that Sanskrit marks schwa omission with an orthographic indicator (halanta/virama) increased her awareness for the vowel.

Although there were four different response patterns, orthography drove performance for all three of the patterns in which errors frequently occurred. These performance patterns reflect properties specific to Marathi orthography; they were not seen with the English words.

Our study of Marathi adds to other demonstrations that phonological awareness at the level of the phoneme is influenced by orthographic representations. In Chinese, Read, Zhang, Nie, and Ding (1986) found that Chinese adults who learned to read with alphabetic pinyin were successful on phonemic awareness tasks, but those who learned only characters without alphabetic learning were not successful. Further, Morais, Cary, Alegria, and Bertelson (1979) found that adults from illiterate communities in Portugal with no alphabetic learning (but with otherwise sufficient cognitive skills) failed to show phonemic awareness. In our case, we extend this general conclusion to the case of a single bilingual individual whose awareness of a phoneme (the minimal vowel, to be sure) is affected by which language is required.

Instructional Methods for Teaching Second Language Learners the Orthography of Marathi

Whole and Part Learning of Aksharas Is Equally Effective It is possible to teach Marathi's CV aksharas in two ways—a focus on the akshara as a whole or a focus on its consonant and vowel subcomponents. Schools in India differ in terms of the approach they use (Nag, 2007; Nag & Sircar, 2008) and it would be useful to have an *in-vivo* comparison of these two approaches. However, when such studies are lacking, then an experimental study with novice learners can shed light on the relative advantages of these two methods.

A first attempt to address this question was not informative. We[5] taught adult native English speakers an artificial alphasyllabic orthography that consisted of Cə and CV graphs. Half the participants were taught the graphs as whole units, whereas other participants were taught to recognize the graphs' subcomponents. Measures of graph learning and word reading failed to show any effect of instruction. Importantly, participants in both groups performed equally well on transfer tasks; they could read novel graphs that consisted of new combinations of learned subcomponents. Thus, at least for adult second language learners, it seems that both learning methods are equally effective because adults are able to identify the subcomponents without explicit instruction. This finding mirrors prior research suggesting that children taught whole aksharas implicitly learn the ligaturing rules (Nag, Treiman, & Snowling, 2010).

Note that this study used highly transparent two component aksharas in which it was easy to infer subcomponent-phoneme mappings. The results may have been different if we had used more complex orthographic forms in which it is difficult to identify the subcomponents. Furthermore, our participants were adults who had experience with an alphabetic orthography. Training on alphabetic-like decoding may allow participants to transfer that strategy to a novel orthography (Bitan & Karni, 2003). Therefore, although our research suggests that people are able to infer subcomponent-phoneme mappings, more research is needed with diverse participants and complex orthographic forms.

Motor Encoding Benefits Complex Akshara Learning Learning complex aksharas is difficult across different languages written with aksharas (Telugu: Vasanta, 2004; Bengali: Nag & Sircar, 2008; Malayalam: Tiwari, Nair, & Krishnan, 2011; Kannada: Nag, 2007, Nag et al., 2010, and Joshi, 2013). This is because there are many visually complex graphs and each graph occurs relatively infrequently in text (Nag, 2014). Bhide (2017a) compared learning methods for teaching multi-component aksharas, with a focus on four different learning methods that draw attention to components in different ways. Specifically, this study examined the effects of testing, motor encoding, and part/whole focus on learning complex aksharas. The four learning methods were: (1) Participants were shown two simple aksharas and had to choose the complex akshara they comprise from among several choices; (2) Participants were shown a complex akshara and had to choose the simple aksharas that it contains from among several choices; (3) Participants were shown both the simple and complex aksharas and had to copy the complex akshara onto a sheet of paper; (4) Participants were shown two simple aksharas and had to write the complex akshara they comprise from memory onto a sheet of paper. A comparison of methods 1 and 2 examined composition/decomposition effects, a comparison of methods 3 and 4 examined testing effects, and a comparison of methods 1/2 and 3/4 examined motor effects.

[5] This study was a 2015 collaboration of the first author, Li-Yun (Wendy) Chang, and Elizabeth Hirshorn.

The results were that, although there were minimal effects of testing and the composition/decomposition contrast, there were large effects of motor encoding. The two learning methods that incorporate motor encoding (methods 3 and 4) outperformed the two multiple choice methods (methods 1 and 2) on most post-test measures. Specifically, motor encoding helped on post-tests that required pure orthographic knowledge (e.g., judging orthographic legality) or required participants to produce an orthographic form given a phonological form (e.g., writing to dictation). However, motor encoding did not help on tasks that require participants to produce a phonological form given an orthographic form (e.g., reading aloud).[6]

Writing incorporates both motor encoding and testing, whereas copying incorporates only motor encoding. Although the copying (method 3) and writing (method 4) learning methods generally produced equivalent levels of learning, copying was much faster than writing. This suggests that testing people on material (i.e., the testing effect, Roediger & Karpicke, 2006) is not particularly powerful during early learning. The lack of a testing effect could have been caused by the fact that participants got many answers incorrect during the learning phase, thus strengthening the incorrect response (the correct answer was immediately provided, but this did not seem to help). Therefore, the results of our study suggest that copying is an efficient and effective method for teaching novice adult learners complex aksharas. Our results provide further evidence of the benefits of motor encoding, a benefit that has been demonstrated in many orthographies (English: Cunningham & Stanovich, 1990, Longcamp, Zerbato-Poudou, & Velay, 2005, and Ouellette, 2010; Chinese: Guan, Liu, Chan, Ye, & Perfetti, 2011, Naka, 1998, and Xu, Chang, Zhang, & Perfetti, 2013; Bengali and Gujarati: Longcamp et al., 2008; Arabic: Naka, 1998, and Naka & Naoi, 1995).

Instructional Methods for Teaching Second Language Learners the Phonology of Marathi

Marathi's phonological system contains many consonants that vary by place of articulation (e.g., t̪/ʈ) and aspiration (e.g., k/kʰ). These differences are phonemic (e.g., ताक /t̪ak/ and टाक /ʈak/ mean *buttermilk* and *throw*(IMP) respectively; कांदा /kand̪a/ and खांदा /kʰand̪a/ mean *onion* and *shoulder* respectively). It is very difficult for many second language learners to differentiate these phonemic contrasts (Polka, 1991; Tees & Werker, 1984; Werker & Tees, 1984), impeding second language acquisition.

Bhide, Ortega-Llebaria, Fraundorf, and Perfetti (2017a) (see also Bhide, 2017b; Bhide, Ortega-Llebaria, Fraundorf, & Perfetti, 2017b) examined whether computer manipulation and orthographic supports could help learn these difficult contrasts. Participants were able to learn the k/kʰ contrast with natural tokens, but not the d̪/ɖ

[6] In fact, testing appeared to benefit reading aloud.

and t̪/ʈ contrasts. To help people with the place of articulation contrasts, we manipulated the sounds by increasing the voice onset time of the dental consonants. This change helped people discriminate the phonemes, suggesting that computer-manipulated phonemes may help second language learners learn these contrasts. A similar method has been shown to help speakers learn other contrasts (Jamieson & Morosan, 1986).

Furthermore, we found that participants were better able to discriminate the phonemes when they had Marathi or no orthographic support than when they had English orthographic support (i.e., English transliterations).[7] This finding suggests that transliterations can impair contrast learning because they induce L1 interference. In contrast, both types of orthographic support helped people remember which phonemes were in a given vocabulary word, in line with previous research on orthographic facilitation (e.g., Rosenthal & Ehri, 2008). This suggests that Marathi orthographic support is most beneficial to learners because it aids both discrimination and memory.

Summary

Learning both the language and its writing system present challenges for the second language learner. The native speaker faces similar challenges in learning the writing system, although with the advantage of knowing the spoken language. In the case of Marathi, as with any of the languages written in Devanagari, the complexity of akshara-based writing produces unfamiliar forms with complex mapping variations. For second language akshara literacy to develop, effective instructional methods are needed. Research on effective methods is relatively scarce, especially those of the *in vivo* type—well-controlled studies in actual classrooms.

To gain some information on how instructional principles might be implemented in teaching aksharas, artificial orthography experiments have some value. The Bhide (2017a) study described here tested whole vs. part learning, the principle of testing, and the principle of motor encoding with adult learners who confronted a Marathi-like (also Hindi-like) Devanagari script. The results suggest that motor encoding through copying helps support orthographic learning of aksharas. As a practical matter, writing from memory may not be useful relative to the instruction time it takes, although this conclusion may not hold when learners have more knowledge of the orthography.

For second language learners, Marathi phonology also creates a challenge that presumably varies with learners' first language. For example, Marathi's consonants are much more numerous than those of English and most other European languages. This greater variety of consonants requires the use of places of articulation that are not phonemic in English but are in Marathi, where, for example, the contrast between /t̪/ and and retroflex /ʈ/ (and their voiced counterparts /d̪/ and /ɖ/) affects

[7] Note that these findings were marginal.

meaning. The Marathi use of aspiration as a phonemic contrast (e.g., k/k^h) is unknown in English, where such variation would be incidental. Bhide et al. (2017a)'s study suggests that mere exposure to these phonemes is not effective where place of articulation is involved, but that manipulations of articulations that make subtle differences more prominent can be effective. Furthermore, the study suggests that avoiding transliterations, which can induce L1 interference, can aid discrimination learning.

Compared with work on other writing systems—alphabets, Abjads (Arabic, Hebrew), syllabaries (Kana), and morphosyllabaries (Chinese)—the study of reading in alphasyllabaries or Abugida (Daniels, 1996) is less developed. Much more will be learned in the coming years about effective instruction for both children learning to read in their native Marathi (or other Devanagari languages) and for second language learners.

References

Bhide, A. (2017a). Copying helps novice learners build orthographic knowledge: Methods for teaching Devanagari akshara. *Reading and Writing: An Interdisciplinary Journal*. https://doi.org/10.1007/s11145-017-9767-8.

Bhide, A. (2017b). *Instructional methods for promoting the development of orthographic and phonological knowledge in second language learners of Indic languages* (Doctoral dissertation). University of Pittsburgh, Pittsburgh, PA, USA.

Bhide, A., Gadgil, S., Zelinsky, C. M., & Perfetti, C. A. (2014). Does reading in an alphasyllabary affect phonemic awareness? Inherent schwa effects in Marathi-English bilinguals. *Writing Systems Research, 6*(1), 73–93. https://doi.org/10.1080/17586801.2013.855619.

Bhide, A., Ortega-Llebaria, M., Fraundorf, S. H., & Perfetti, C. A. (2017a). The contribution of orthographic support, phonological skills, rise time discrimination, and acoustic enhancement to the learning of non-native phonemic contrasts. *Manuscript Submitted for Publication*.

Bhide, A., Ortega-Llebaria, M., Fraundorf, S. H., & Perfetti, C. A. (2017b). *The effect of orthographic support, rise time discrimination, and phonological skills on second language learners' abilities to learn Marathi pronunciations.* Paper presented at the 24th annual conference for the Society for the Scientific Study of Reading, Halifax, Canada.

Bitan, T., & Karni, A. (2003). Alphabetical knowledge from whole words training: Effects of explicit instruction and implicit experience on learning script segmentation. *Cognitive Brain Research, 16*(3), 323–337. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12706213

Bondale, N., Surve, V., Nadkarni, M., Parkhi, O., Joshi, P., & Pandey, A. (2013). Issues in developing pronunciation lexicon for Marathi. In *Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International conference* (pp. 1–4). IEEE.

Cunningham, A. E., & Stanovich, K. E. (1990). Early spelling acquisition: Writing beats the computer. *Journal of Educational Psychology, 82*(1), 159–162. https://doi.org/10.1037/0022-0663.82.1.159.

Daniels, P. T. (1996). The study of writing systems. In P. T. Daniels & W. Bright (Eds.), *The World's writing systems* (pp. 3–17). New York, NY: Oxford University Press.

Dhongade, R., & Wali, K. (2009). *Marathi*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Guan, C. Q., Liu, Y., Chan, D. H. L., Ye, F., & Perfetti, C. A. (2011). Writing strengthens orthography and alphabetic-coding strengthens phonology in learning to read Chinese. *Journal of Educational Psychology, 103*(3), 509–522. https://doi.org/10.1037/a0023730.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics, 40*(4), 205–215. https://doi.org/10.3758/BF03211500.

Joshi, R. M. (2013). Literacy in Kannada, an alphasyllabic orthography. In H. Winskel & P. Padakannaya (Eds.), *South and Southeast Asian psycholinguistics* (pp. 184–191). New York, NY: Cambridge University Press.

Kachru, Y. (2006). *Hindi*. Philadelphia, PA: John Benjamins Publishing Company.

Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.). (2016). *Ethnologue: Languages of Asia* (19th ed.). Dallas, TX: SIL International.

Longcamp, M., Boucard, C., Gilhodes, J.-C., Anton, J.-L., Roth, M., Nazarian, B., & Velay, J.-L. (2008). Learning through hand- or typewriting influences visual recognition of new graphic shapes: Behavioral and functional imaging evidence. *Journal of Cognitive Neuroscience, 20*(5), 802–815. https://doi.org/10.1162/jocn.2008.20504.

Longcamp, M., Zerbato-Poudou, M.-T., & Velay, J.-L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica, 119*(1), 67–79. https://doi.org/10.1016/j.actpsy.2004.10.019.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323–331.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22. https://doi.org/10.1111/j.1467-9817.2006.00329.x.

Nag, S. (2014). Akshara-phonology mappings: The common yet uncommon case of the consonant cluster. *Writing Systems Research, 6*(1), 105–119. https://doi.org/10.1080/17586801.2013.855621.

Nag, S., & Sircar, S. (2008). *Learning to read in Bengali: Report of a survey in five Kolkata primary schools*. Bangalore, India: The Promise Foundation.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*(1), 41–52. https://doi.org/10.1093/wsr/wsq001.

Naka, M. (1998). Repeated writing facilitates children's memory for pseudocharacters and foreign letters. *Memory & Cognition, 26*(4), 804–809. https://doi.org/10.3758/BF03211399.

Naka, M., & Naoi, H. (1995). The effect of repeated writing on memory. *Memory & Cognition, 23*(1), 201–212.

Official Languages: Constitutional Provisions. (2016). Retrieved from https://iaspoint.gktoday.in/article/official-languages-constitutional-provisions/.

Ouellette, G. (2010). Orthographic learning in learning to spell: The roles of semantics and type of practice. *Journal of Experimental Child Psychology, 107*(1), 50–58. https://doi.org/10.1016/j.jecp.2010.04.009.

Pandharipande, R. (1997). *Marathi*. New York, NY: Routledge.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *The Journal of the Acoustical Society of America, 89*(6), 2961–2977. https://doi.org/10.1121/1.400734.

Rathod, P. H., Dhore, M. L., & Dhore, R. M. (2013). Hindi and Marathi to English machine transliteration using SVM. *International Journal on Natural Language Computing, 2*(4), 55–71.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24*, 31–44.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181–210.

Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology, 100*(1), 175–191. https://doi.org/10.1037/0022-0663.100.1.175.

Salomon, R. (2000). Typological observations on the Indic script group and its relationship to other alphasyllabaries. *Studies in Linguistic Sciences, 30*(1), 87–103.

Sinha, R. M. K., & Mahabala, H. N. (1979). Machine recognition of Devanagari script. *IEEE Transactions on Systems, Man, and Cybernetics, 9*(8), 435–441.

Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology, 38*(4), 579–590. https://doi.org/10.1037/h0080868.

Tiwari, S., Nair, R., & Krishnan, G. (2011). A preliminary investigation of akshara knowledge in the Malayalam alphasyllabary: Extension of Nag's (2007) study. *Writing Systems Research, 3*(2), 145–151. https://doi.org/10.1093/wsr/wsr013.

Vaid, J., & Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690. https://doi.org/10.1006/brln.2001.2556.

Vasanta, D. (2004). Processing phonological information in a semi-syllabic script: Developmental data from Telugu. *Reading and Writing: An Interdisciplinary Journal, 17*(1), 59–78. https://doi.org/10.1023/B:READ.0000013830.55257.3a.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866–1878. https://doi.org/10.1121/1.390988.

Xu, Y., Chang, L.-Y., Zhang, J., & Perfetti, C. A. (2013). Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Language Annals, 46*(3), 423–444. https://doi.org/10.1111/flan.12040.

Part V
Research Implications

Learning to Read in Their Heritage Language: Hindi-English Speaking Children Reading Two Different Orthographies

Amna Mirza and Alexandra Gottardo

Abstract In the past few decades, research in reading development has begun to focus on reading in different languages to establish language universals. However, the focus has remained largely on alphabet-based orthographies as compared to other levels of sound-symbol mappings (Share DL, Front Psychol 5:1–3, 2014). The assumption is that the theoretical models of reading acquisition in English are generalizable to other languages, including Spanish, German and Chinese (Leong CK, Tamaoka K, Read Writ Interdiscip J 10:155–164, 1998; Wimmer H, Goswami U, Cognition, 51(1):91–103, 1994; Vaid J, Padakannaya P, Read Writ Interdiscip J 17(1):1–6, 2004). However, Indic writing systems, such as Hindi, which are written, using symbol units called akshara, are largely ignored. This study examined variables related to word reading in Hindi and English for Hindi-English speaking children. The children resided in Canada, were fluent English speakers, and attended heritage language classes in Hindi, their first language. Word reading and vocabulary knowledge were measured in both languages, while phonological awareness was measured in English. Within- and cross-language relations revealed that similar processes were related to word reading in each language for this group of participants.

Keywords Abugida orthography · Bilingualism · Biliteracy · Hindi-English speakers

A. Mirza
Child and Youth Studies, Brock University, St. Catharines, ON, Canada
e-mail: mirz5320@gmail.com

A. Gottardo (✉)
Psychology Department, Wilfrid Laurier University, Waterloo, ON, Canada
e-mail: agottardo@wlu.ca

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*, Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_17

Introduction

Exploring reading skills across languages has allowed researchers to determine language-specific and language-universal processes and skills. The search for language universals in reading involves determining similarities and differences among variables related to reading in languages with different orthographic units. Research examining skills associated with reading in less commonly studied orthographies can guide the development of universal models of reading and enhance reading theory beyond models that incorporate only data based on reading acquisition in English (Nag & Snowling, 2012; Share, 2008). In many cases, researchers have compared skills of bilingual readers of alphabetic languages with similar orthographies (e.g., Spanish & English) (Durgunoğlu, Nagy, & Hancin-Bhatt, 1993; Gottardo, 2002; Lindsey, Manis, & Bailey, 2003) or alphabetic languages with different writing systems (e.g., Arabic & Hebrew; Russian & Hebrew) (Abu-Rabia, 2007; Schwartz, 2008). These studies allow researchers to determine processes associated with reading alphabetic orthographies when specific grapheme-phoneme (symbol-sound) relationships are similar across languages versus reading alphabetic orthographies where specific grapheme-phoneme correspondences cannot be "transferred" (Geva & Siegel, 2000). In some cases, a reader's understanding of the specific orthographic constraints of a given language is not related to learning to read the other language. For example, in the case of English and Chinese, understanding the alphabetic principle, which is necessary to become a fluent reader of English, is less useful for understanding the morpho-syllabic mappings between Chinese characters and their pronunciations and meanings (Perfetti, 2003). Languages that code symbol-sound relations at the level of the syllable as well as marking phonemes afford an interesting and less commonly studied contrast to alphabetic languages. Thus, the current study examines relations across and within languages for bilingual children learning to read an alphabetic orthography, English, and an abugida orthography, Hindi.

The *psycholinguistic grain size theory* has been posited to test the role of different orthographic constraints on the development of reading (Ziegler & Goswami, 2005). The theory states that the *grain size*, or size of the phonological and orthographic unit that must be learned and used in order to read, has an impact on reading development. According to the *psycholinguistic grain size theory,* reading acquisition requires that children learn to optimally map phonological units of a given language, usually their native language, to the symbol system of that language (Ziegler & Goswami, 2005). The *psycholinguistic grain size theory* identifies the *availability* or accessibility of sound units represented in the orthography, the *consistency* of sound-symbol units, and the *granularity* or size of the consistent units as key in determining processes used to learn to read a given orthography. This optimal mapping likely influences processes related to reading acquisition.

Previous characterizations of the link between sounds in speech and written symbols were described using the *orthographic depth hypothesis* (Frost & Katz, 1992). Orthographic depth refers to the regularity and consistency of sound-symbol mappings with languages with high consistency and regularity being classified as having a shallow orthography and languages with low consistency and regularity being classified as having a deep orthography. This continuum usually is used to refer to alphabetic orthographies with English being described as a deep alphabetic orthography (Venezky, 1970). However, some researchers argue that the *orthographic depth hypothesis* can be used to explain processes used to read Hindi, an abugida orthography (Rao, Vaid, Srinivasan, & Chen, 2011).

The symbols, or aksharas, used to write Hindi are considered to have a one-to-one correspondence to the oral syllable, suggesting that Hindi is written using a shallow orthography (Nag & Perfetti, 2014). However according to the *psycholinguistic grain size theory,* differences are expected between Hindi and English based on the size of the orthographic units used to represent print (see below). Because of the distinct differences in orthographic representations across Hindi and English, findings of relations among English and Hindi reading skills would point to universal processes being used to read these two languages by learners of these languages.

Hindi Language and Literacy

Oral Language Hindi is one of the official national languages of India and the most widely spoken language in India with 40% of the population speaking Hindi (2001 Census). Survey data has revealed that based on the total number of native Hindi speakers, it is among the top five most widely spoken languages of the world (Pandey, 2014; Statista, 2017). However, estimating the number of Hindi speakers is difficult because many people speak Hindi as their second or additional language. This is because India has a highly diverse population with citizens who are bilingual or multilingual. Within India, Hindi is a widely spoken first language in northern and central regions of the country. Hindi also is spoken in some regions of Nepal and Bangladesh. Other communities of Hindi speakers have immigrated to and live in the United Kingdom, Australia, New Zealand, Canada, South Africa and the United States of America.

Hindi is closely related to Urdu, with the two spoken languages being mutually intelligible. Despite differences in orthography and associated demographics (i.e., geographic location, religion of speakers), some people refer to the languages as the “Hindustaani language”, and consider Hindi and Urdu to be two primary dialects of this language. Hindi vocabulary and grammar are based on its Sanskrit roots. Due to its colonial past, Hindi is influenced by English, adopting English vocabulary as loan words. Literary Hindi differs somewhat from oral Hindi in terms of formal style and draws more vocabulary items from Sanskrit. Despite some overlapping vocabulary, Hindi and English have very different linguistic roots and typologies.

Written Language Orthographies differ not only in the appearance of the script, but also in a way in which the symbols map onto sounds in the speech stream. Hindi is written with the Devanāgarī script. Hindi orthography has elements of an alphabetic script and a syllabary, resulting in it being characterized by some researchers as an alpha-syllabic script or alphasyllabary (Nag, 2011; but also see Share & Daniels, 2016). Similar to English, Hindi is written from left to right. Abugida orthographies, such as Hindi, represent speech at two levels, the syllabic level and the phonemic level (Salomon, 2000). Each orthographic symbol is referred to as an akshara, which contains elements of the consonant and the vowel. In Hindi texts, all basic aksharas represent a syllable.

The surface organization of each akshara is typically based on a symbol block with one or more phonemic markers added to the base. Specifically, the most basic unit represented by an akshara is a consonant with the inherent or default vowel /a/ (i.e., Ca) (Nag, 2011). Modifications of these basic units include the addition of consonants or the diacritic vowel markers other than the default vowel in the syllable. The diacritic vowel markers must be incorporated into the basic syllable blocks and cannot stand alone. These modifications encode phonemic markers and preserve the basic orthographic unit while adding smaller visual components that code the vowels or consonants to create more complex syllables. All written syllables have an open syllable structure and include CCa, CCV or CCCV. When an akshara appears as a single unit, then it is typically an orthographic syllable, but when it appears in a string, then language-specific rules are applied to the akshara. For example, a rule of re-syllabification determines the mapping of word level phonology for each specific akshara. Consequently, aksharas map to multiple levels of phonology (Patel, 1996, 2004; Patel & Soper, 1987). Because modifications to basic aksharas that create new syllables are often visually subtle, each syllabic unit must be learned separately, resulting in 200–500 symbols to be learned. Because individual consonants or vowels do not tend exist in isolation in the written form of Hindi, some researchers argue against characterizing components of Hindi as alphabetic (Share & Daniels, 2016). Alternately, researchers argue that although the visual form of akshara is syllabic, the representation of phonemes in the Hindi script suggests that it is functionally predominantly alphabetic (Rimzhim, Katz & Fowler, 2014).

Other factors, such as directionality of script, influence strategies used by readers, both for reading and for other cognitive tasks (Vaid, 1995). In the case of Hindi, directionality of syllables and consonants occurs in a left to right fashion, while vowel markers can occur above, below or beside consonants (Das et al., 2014). In contrast to some linguistic classifications of Hindi orthography, neuro-imagining research shows that cortical areas activated in reading Hindi include areas associated with reading an alphabetic language as well as areas associated with reading a syllabic language (Das et al., 2014). It has also been argued that representing Hindi's rich morphology, that is instantiated by phonemic contrasts, requires coding of phonemes in the written language (Rimzhim et al., 2014). The lack of consensus as to the nature of Hindi orthography, suggests that it is a good candidate for further research. Therefore examining whether skills related to reading in Hindi are similar to skills related to reading in English in bilingual children can inform the debate as to the nature of Hindi reading.

The Role of Orthography in Reading

The connectionist model originally posited by Seidenberg and McClelland (1989) suggests that the process of learning to read words depends on establishing mappings among phonology, orthography and semantics. However, learning to read an orthography is also dependent on whether it is an alphabetic or non-alphabetic writing system and the consistency of sound-symbol mappings (Katz & Frost, 1992; Perfetti & Harris, 2013; Share, 2014). Languages represent units of speech of different sizes from syllables to smaller units, specifically phonemes. In order to become skilled readers of an alphabetic orthography, readers must learn how to map phonemes onto graphemes (Share, 1995). Other units can also be represented by orthographies and are perceived as psychologically real by speakers of those languages; for example, native speakers of English perceive onsets and rimes as psychologically real (Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000). If the same processes are used to read words in different languages, this would suggest a general reading mechanism.

When examining different writing systems, script-specific differences in relation to typological features will affect reading development. Research on learning to read an abugida language is in the initial stages, with most recent research on learning to read being conducted in India (Nag & Perfetti, 2014). This research highlights the importance of orthography-specific investigations in the reading science. Because additional consonants and vowel diacritics are represented as modifications to the base form of akshara (see above), a larger number of aksharas must be learned to read this writing system. The total number of symbols to be learned in Indic writing systems is usually much larger, when compared to the number of symbols to be learned in an alphabetic system. Research studies in reading acquisition suggest that the pace of learning depends on the size of the symbol set such that Latin scripts with 20–40 graphemes are expected to be learned by the end of the first year in school with some variability based on the consistency of the grapheme-phoneme correspondences (Seymour, Aro, & Erskine, 2003). However languages with akshara symbols have between 200 and 500 symbols that vary in frequency of appearance in script. The large number of symbols and the relatively low frequency of some sound-symbol correspondences may result in children mastering aksharas by fourth grade or even later (Nag, 2007).

Learning to Read an Abugida Orthography At present, all known studies of reading acquisition in Indic writing systems were conducted in India, most with monolinguals. Studies with bilinguals have usually contrasted languages with similar linguistic roots. These studies examined challenges encountered by children learning to read Kannada as well as one other Indian language, specifically Bengali, Hindi, and Tamil. The results of the studies have concluded that the causes of reading difficulties in these languages could be explained by multiple factors (Nag, Treiman, & Snowling, 2010). These studies usually deal with children with reading disabilities, as the identification and remediation of reading disabilities is the most pressing need for schools. For instance, Prakash and Joshi (1995) reported that children with dyslexia had poor knowledge of the aksharas and experienced additional difficulties in auditory sequential memory, syllable processing, visual-verbal processing, and visual processing. The results were replicated in other studies that explained children's challenges with learning Kannada aksharas (Prema & Karanth, 2003). Gupta (2004) found similar results for Hindi speakers, specifically that children with dyslexia not only struggle with reading accuracy and speed as compared to skilled readers but that they also face difficulties with orthographic learning of the phonemic markers in the language. Considering the large number of symbols (between 200 and 500) in an Indic writing system, it is clear that the orthography plays an important role in predicting performance of good and poor readers. Additional factors related to weaknesses found in poor comprehenders can be accounted for by the difficulties with visual learning (Nag et al., 2010). Although these findings were replicated in many studies, the research did not examine factors related to word reading across languages in these bilingual and multilingual learners.

Word Reading Development in Bilinguals

Studies that have examined reading development suggest that language and literacy skills are related to each other and that first language (L1) and second language (L2) skills can influence each other (Chang, 2013; Jiang, 2004; Koda, 1996). For example, Spanish-English speakers who have good language and literacy skills in Spanish tend to have strong skills in English, their L2 (Durgunoğlu et al., 1993; Gottardo, 2002; Lindsey et al., 2003). The *linguistic interdependence hypothesis* states that strong L1 skills are related to strong L2 skills (Cummins, 1979). The *script dependent hypothesis* highlights the role of differences in script with cross-linguistic relations being greater for languages with similar orthographies than for languages with different orthographies (Geva & Siegel, 2000). Strong versions of the *linguistic interdependence hypothesis* suggest cross-linguistic relationships for all constructs (e.g., morphological skills, phonological awareness), while other researchers have suggested that some skills are more likely to be related across languages than other skills (Durgunoğlu, 2002; Geva & Wang, 2001 for reviews). For example, lower level phonological skills and higher level comprehension skills are more likely to be related across languages for each construct (Durgunoğlu, 2002; Gottardo, Yan, Siegel, & Wade-Woolley 2001; Lindsey et al., 2003; Verhoeven, 2007). In contrast, skills that deal with linguistic structures such as syntax and morphology as well as vocabulary show differential levels of transfer based on similarities between languages (Gottardo, 2002; Pasquarella, Chen, Lam, Luo, & Ramirez, 2011; Ramirez, Chen, Geva, & Kiefer, 2010; Verhoeven, 2007). Languages with different orthographies are also less likely to be related in terms of knowledge of grapheme-phoneme correspondences and in terms of spelling skills (Geva & Sigel, 2000; Pasquarella, Chen, Gottardo, & Geva, 2015). Although extensive research has been conducted on the role of L1 skills on L2 skills, examination of the role of L2 skills on L1 skills is less common (Bialystok & Herman, 1999; Cook, 2003; Gottardo, Javier, Farnia, Mak, & Geva, 2014).

Bidirectional cross-linguistic relationships between languages with different linguistic typologies and orthographic systems provide the opportunity to examine language-specific and language-general mechanisms. Although cross-linguistic relationships have been found for languages with similar orthographies or linguistic typologies (e.g., the Roman script), do language-general mechanisms influence the relationships across typologically different orthographies (e.g., reading an alphabetic script vs. an abugida)?

When young children begin the process of learning to read, they learn the code used by their language to represent speech and how the symbols map onto speech. The key precursor to word reading in an alphabetic language is phonological awareness (Hulme et al., 2002). Many researchers accept the notion that phonological awareness includes a range of linguistic subcomponents from syllables, to onsets and rimes to phonemes (Anthony & Lonigan, 2004; Stanovich, 1990). The size of the phonological unit that is most highly related to reading a specific orthography might be related to the specific language being read or might be related to the learner's first language (L1) (Gottardo, Pasquarella, Chen, & Ramirez, 2016; Ziegler & Goswami, 2005). For example, phonemic awareness is related to reading a shallow alphabetic orthography such as Spanish (Jimemez, 1997).

Even in an irregular language such as English, this relationship between phonemes and graphemes is usually systematic (e.g., the symbol L is usually pronounced /l/). Phonological recoding is considered to be a crucial element of reading, because this process allows children to decode words that are heard but have not been seen before (Ehri, 2011; Share, 1995). Successful decoding of an alphabetic language requires mapping of graphemes to phonemes and determining the rules of the cipher to read accurately and fluently (see discussion of the psycholinguistic grain size theory). However, the size of the phonological unit that maps onto the symbol in Hindi, and therefore the processes used to read Hindi, are less clear.

Other variables are also related to word reading skills beyond the development of initial decoding skills. For example, the ability to read words fluently is a necessary next step after word reading accuracy is mastered (Bowers & Wolf, 1993; Pasquarella et al., 2015). This skill is in part related to rapid access to phonological information that includes rapid naming of lexical items (Kirby, Parrila, & Pfeiffer, 2003; Wagner & Torgesen, 1987). Rapid naming is related to reading in a shallow orthography (Wimmer & Goswami, 1994).

Linguistic theory has examined relations between oral proficiency in the L1 and L2 in an attempt to build theoretical models of bilingualism (Cook, 2003). Because both languages are in one 'mind', they must interact in bilinguals. However, the degree and direction of overlap has been the subject of debate in theories of second language acquisition. For example, Cook (2003) has suggested that L1 and L2 relations are bidirectional and has provided evidence of L2 influences on the L1 in highly skilled users of each language (also see Chow, McBride-Chang, & Burgess, 2005). However, less is known about bidirectional relations in young learners with Hindi as an L1, who are learning to read two or more writing systems.

Bilingual Hindi learners Although census information points to a large number of bilingual Hindi speakers in India, very little research has been conducted on reading in bilinguals or multilinguals who speak Hindi as one of their languages. This research is particularly interesting because most Hindi speakers learn to speak and read additional languages that are represented by different orthographies. For example, many Hindi-English speakers learn to read their native language written in an abugida or Indic writing system as well as reading English, which is written in the Roman script. Hindi-Urdu speakers who read Urdu must learn the Arabic alphabet, which is represented in a shallow and deep form, specifically with and without vowel markers.

A series of studies has examined the role of orthographic depth in shaping visual word recognition in bilinguals who spoke Hindi and Urdu (Rao et al., 2011). Although these two languages share a common spoken form, the written forms differ in terms of orthographic, structural and visual differences as well as directionality (Kelkar, 1968). Two experiments were conducted with Hindi/Urdu biliterate university students (Rao et al., 2011). The first experiment examined the effects of providing the same form/orthography on priming (i.e., Hindi prime – Hindi target; Urdu prime – Urdu target). In all cases, the phonology overlapped between the prime and the target. Results of the study showed that form-related primes increased speed and accuracy for words written in Hindi orthography to a greater extent than for Urdu words (Rao et al., 2011). These effects supported the hypothesis that Hindi is represented by a more consistent mapping between symbols and sounds than Urdu. The purpose of the second experiment was to isolate the effects of phonological overlap and visual script overlap in priming. Therefore, primes were presented in Roman script while the targets were presented in Hindi or Urdu. This manipulation was designed to separate the visual form from the phonological form of an item. Consistent with researcher expectations, the results of the study showed greater naming speed and accuracy for the Hindi items than the Urdu items (Rao et al., 2011). These results suggest the benefits of reading a shallower orthography such as Hindi as compared to Urdu. Although research has been conducted examining cross-linguistic effects of reading Hindi and Urdu, the effects of reading English and Hindi, languages commonly spoken by Hindi bilinguals, have not been examined. The current study examined whether skills related to reading a deep alphabetic orthography, English, are related to reading a shallow orthography, Hindi, in bilingual children.

Bilingualism in the Canadian Context

Although Canada is officially a multicultural country with a policy of multilingualism (Statistics Canada, 2016), in reality, many children born to immigrant parents show a pattern of language loss. These children begin school fluent in their L1. However, due to the large number of first languages known by children in many urban classrooms, and the lack of a single, common minority group, the language of the classroom is English[1]. Not only is English immersion instruction conducted in the classroom, but English is also the common language of the playground. Therefore, immigrant children often show a pattern of L1 loss and dominance in their L2 after having attended school in Canada for several years (Statistics Canada, 2016). Immigrant parents in Canada attempt to preserve their L1 at home and encourage L1 literacy through heritage language classes, which are held after school or on weekends. Therefore, although these immigrant children are able to communicate orally in their L1 at various levels of proficiency, they often have strong oral language skills in English and acquire English literacy prior to literacy in their L1.

Goals of the Present Study

The present study extends the research on skills related to reading in non-European languages to include reading in Hindi, an Indic writing system. Specifically, this research project extends existing literature by determining whether learning to read an abugida orthography is related to reading an alphabetic orthography in children who are bilingual speakers of English and Hindi. Preliminary data are reported, examining within-language and cross-language relations for children who are bilingual speakers of Hindi and English.

Methods

Participants

Twenty Hindi-speaking children (8–10 year olds) living in Southern Ontario, Canada, participated in this study. The average age of these children was 9.50 years, *SD* = .59. The sample had eleven male and nine female participants. Participants had learned English in their local public schools and were enrolled in grades three to five. The participants attended weekend language school for about three hours on Saturdays to learn Hindi. Demographic information was collected through a questionnaire completed by the parents of each participant. This questionnaire provided information about the home language and literacy environment.

Parents were asked about the languages spoken at home. The responses revealed that most of these children were exposed to multiple languages in the home. Ten families mentioned that they speak Hindi and Tamil at home. Among these ten families, four of them mentioned that the parents had different first languages. Four families spoke Hindi at home, and six families spoke Hindi and Punjabi at home. All

[1] Although French and English are both official languages in Canada, in the context of this study the majority and societal language is English.

families spoke some English at home. All children but one attended Hindi language classes on weekends. On average children were enrolled in the weekend Hindi language program for more than three years. Parents were also asked to judge their child's best language and all of the parents mentioned that English was their child's best language. Parents were asked about their child's use of Hindi with other family members at home (parents, other siblings or grandparents if they lived within the same household). The same question was repeated for the child's frequency of English use with family members at home and with friends outside the home. Seventy-one percent of the parents reported that their children speak English (L2) more frequently with their peers and siblings and speak in their first language with their family members (i.e., parents, grandparents, uncles and aunts). The remaining parents (29%) mentioned the use of both first and second languages equally with family and peers.

Information was collected regarding child's record of attending school within or outside of Canada. Seventeen out of twenty families reported that their child had attended school outside of Canada. Fourteen of them mentioned that their child previously had attended school in India and three had attended kindergarten in the United States of America.

Also, parents reported on their child's frequency of watching television and reading books in the first and second languages. Fifty-four percent of the families reported that their children watch television in their second language more often than watching television in first language. The rest of the families (46%) mentioned that their children watch television in both languages equally. For book reading, 78% of the families reported that they spend 2–5 h per week with their children on reading books in both languages. Twenty-two percent of parents reported that they do not have books in their first language, but that they spend approximately one hour per day with their children on reading in English.

Finally, parents responded when asked about the amount of exposure their child had received to his/her first language. Fifty-seven percent of the families reported that they speak in their first language to their children. Forty-three percent of parents mentioned speaking in either language with their children. Questions also included how many hours a day the child received instruction in Hindi and the reasons parents had for deciding to send their child to international language school for Hindi classes. Almost all of the parents reported that they chose to send their children to the weekend language schools because they wanted their children to be fluent in their first language (both oral and written) and to be able to communicate with their other family members (e.g., grandparents, aunts and uncles).

Measures

A battery of English and Hindi measures was administered to each participant. Participants were assessed on word reading, phonological awareness and vocabulary knowledge in English. The English measures were standardized tests that exhibited high reliability and validity. Standardized measures were not available in Hindi; therefore, measures were created for the constructs of word reading, vocabulary knowledge, and rapid naming.

Demographic/Family Language Questionnaire

This questionnaire was given to the parents along with the consent forms in order to determine what language(s) the parents and children speak at home. The questionnaire focused on information about the factors that influence a child's ability to learn a second language and their verbal ability (see demographic information presented above). The questionnaire was written in English but parents were offered help and items were orally translated as necessary.

English measures

English Word Level Reading Two subtests from the Woodcock Reading Mastery Test-Normative Update (Woodcock, 1998) were used to measure English word and pseudoword reading ability. Word reading was measured using the Woodcock Word Identification subtest (Woodcock, 1998). The words in the list increased in difficulty from high frequency, monosyllabic words to lower frequency multisyllabic words (e.g., *petroleum*). Participants were asked to read the words out loud. Testing was discontinued after six consecutive errors in a set. Based on the manual, the Cronbach's α was .92 (Woodcock, 1998).

The Woodcock Word Attack was used to measure pseudoword reading (Woodcock, 1998). This test contained 45 pseudowords of increasing difficulty (e.g., *dee* to *pnomocher*). The same procedures were used as above. The Cronbach's α from the manual was .95.

English Phonological Awareness The elision subtest of the Comprehensive Test of Phonological Processing was administered using standardized procedures to measure phonological awareness skills in English (CTOPP; Wagner, Torgesen, & Rashotte, 1999). Children were asked to repeat a word and then to delete part of the word to produce a meaningful word in English. Simpler items required the children to delete syllable segments from compound words, such as *tooth* from the word *toothbrush*. Children were also required to delete the syllable onset to produce a real word (e.g., Say *bat* without the /b/). For the more difficult items, children deleted a phoneme within the word (e.g., Say *sling* without the /l/, response *sing*). The Cronbach's α from the manual was .89.

English Vocabulary Knowledge The Expressive One Word Picture Vocabulary Test (EOWPVT-SBE, Brownell, 2001) was used to assess children's vocabulary.

This test measured the ability to name objects, actions and concepts represented by pictures. Based on the test manual, the reliability of this measure was Cronbach's $\alpha = .95$.

Hindi Measures

Hindi Word Reading Due to the lack of availability of Hindi standardized measures, the primary investigator of the study created a word list by using words from children's Hindi textbooks. These textbooks are graded and used to teach the curriculum in India. Words were selected in consultation with a registered Hindi teacher and a translator. The word list was comprised of words from the Hindi curriculum books for grades one and two in India and was consistent with the syllabus used in weekend language school. This word list consisted of 50 items. The words gradually increased in level of difficulty. All words were presented with diacritics/maatras where relevant. Five out of fifty words did not have maatras because they consisted of consonants and the default vowel, the remaining forty-five words were presented with maatras. Participants were asked to continue reading the words until the end of the list. The reliability of this measure was Cronbach's $\alpha = .70$.

Hindi Vocabulary Knowledge The Expressive One Word Picture Vocabulary Test was translated into Hindi. This test was translated from English and parallel items were used in the Hindi vocabulary test as were used in English version to assess expressive vocabulary in Hindi (EOWPVT-SBE, Brownell, 2001). A total of 170 pictures of different objects and actions were shown to the participants one picture at a time. Children were asked to name the pictures in Hindi. The pictures were arranged in order of increasing difficulty. Reliability was calculated for this measure using Cronbach's $\alpha = 78$.

Hindi Rapid Naming The rapid digit naming subtest of the Comprehensive Test of Phonological Processing was translated to measure phonological processing skills in Hindi (CTOPP; Wagner et al., 1999). This task measured the speed with which an individual named six randomly selected numbers presented in four rows and nine columns. Participants were asked to name the numbers starting at the top row from left to right in the Hindi language. There were 36 items in total. The score for this task is the number of seconds it took the participant to name all the numbers.

Procedure

Participants were tested in English and Hindi in one testing session that lasted for two hours. Participants were first tested on the English battery and then on the Hindi measures. Children were given a break between the two languages. Most of the participants were tested at their weekend language school, although some of them were tested in their homes.

Results

This study examined the reading abilities, vocabulary knowledge and phonological skills of Hindi-English bilinguals. Descriptive statistics are provided for all the main variables (See Table 1). No floor or ceiling effects were found for any of the measures. In addition to reporting the total raw scores for the elision measure, scores were analyzed based on items that measured the deletion of syllables, onsets and phonemes. Correlational analyses were conducted to examine the relationships among variables. Given the small number of participants only a small number of variable were examined in each analysis.

Within Language Relations

Hindi Within-Language Relationships Correlational analyses were conducted with all the variables administered in Hindi. Given the limited sample size, all significant correlations were large (r < .50). Hindi word reading, vocabulary knowledge and rapid naming were the constructs used in this analysis. Results showed Hindi word reading was significantly related to Hindi vocabulary $r_{(18)} = .69$, $p \leq .001$. The other within-language constructs were not significantly correlated (See Table 2).

English Within-Language Relationships Correlational analyses were also conducted to examine the relationships among English constructs that included word and pseudoword reading, vocabulary knowledge and phonological awareness. The

Table 1 Descriptive statistics for all Hindi and English variables

Construct	Mean	SD
English word reading raw score	70.70	9.21
English word reading percent correct	66.69	8.69
English non-word reading	24.40	5.85
English vocabulary	77.40	12.36
English phonological awareness	11.00	2.79
PA: Syllable (max = 3)	2.55	.60
PA: Onset (max = 3)	2.70	.47
PA: Phoneme (max = 14)	5.75	2.22
Hindi word reading raw score	24.45	5.13
Hindi word reading percent correct	48.90	10.26
Hindi vocabulary	30.80	8.07
Hindi RAN	21.02	2.41

Table 2 Correlation Matrix showing within and cross linguistic relationships

	1	2	3	4	5	6	7
1. English word reading	–						
2. English non-word reading	.90***	–					
3. English vocabulary	.86***	.81***	–				
4. English phonological awareness	.58**	.52*	.66**	–			
5. Hindi word reading	.58**	.54*	.60**	.48*	–		
6. Hindi vocabulary	.76***	.77***	.78***	.64**	.69***	–	
7. Hindi RAN	−.44*	−.49*	−.44	−.24	−.22	−.39	–

*p < .05, **p < .01, ***p < .001

results revealed large significant relationships among most of the English measures. Specifically, correlations were found between English word reading and several other constructs measured in English, including pseudoword reading, $r_{(18)} = .90$, $p < .001$, vocabulary, $r_{(18)} = .86$, $p < .001$ and phonological awareness, $r_{(18)} = .58$, $p < .01$. Word reading was more highly related to phonemic awareness than syllable or onset awareness (See Table 3). Pseudoword reading was also significantly correlated with vocabulary, $r_{(18)} = .81$, $p < .001$ and phonological awareness $r_{(18)} = .52$, $p < .05$. English vocabulary knowledge was related to phonological awareness, $r_{(18)} = .66$, $p < .01$ (See Table 2).

Cross-Language Relationships Cross-linguistic correlations were also examined (see Table 2). Hindi vocabulary was related to all English language and literacy measures, showing large significant correlations, rs $\leq .64$. Hindi word reading and rapid naming were related to English word and pseudoword reading and to English vocabulary. Hindi word reading was also significantly related to English phonological awareness $r_{(18)} = .48$, $p < .05$ (See Table 2). Hindi word reading was more highly related to English phonemic awareness than to English syllable or onset awareness (See Table 3).

Two pairwise t-tests were conducted to examine cross-linguistic skills of Hindi-English bilinguals for the constructs of word reading and vocabulary. The analyses for word reading were conducted using percent correct scores rather than raw scores to attempt to better equate and compare the measures. Percent correct scores were used because the word reading measures were developed independently with different numbers of items. Additionally, the English word reading measure was normed and standardized while the Hindi measure was an experimental measure. Analyses revealed significant differences between the two pairs of constructs (See Table 4). Scores achieved on English word reading, $M = 66.69\%$ correct, $SD = 8.7$, were significantly higher than scores on Hindi word reading, $M = 48.90\%$ correct, $SD = 10.3$; $t_{(19)} = 9.11$, $p < .001$. A comparison of participants' performance on the vocabulary test using the same items revealed a difference with English scores, $M = 77.40$,

Table 3 Correlations between types of phonological awareness and reading in English and Hindi

	English word reading	Hindi word reading
Syllable awareness	.277	.103
Onset awareness	.185	.233
Phoneme awareness	.626**	.532*

***p < .001

Table 4 Cross-linguistic comparisons of variables

			Mean	SD	df	t	Sig
Pair 1	Word reading	English	66.69	8.69	19	9.11	.000
	Percent correct	Hindi	48.90	10.26			
Pair 2	Vocabulary	English	77.40	12.36	19	26.52	.000
		Hindi	30.80	8.07			

SD = 12.4, being higher than Hindi vocabulary scores, $M = 30.80$, $SD = 8.1$; $t_{(19)} = 26.52$, $p < .001$. The higher scores in English are consistent with the demographic data, which suggests that the children showed English dominance (See Table 4).

Variables Uniquely Related to English and Hindi Word Reading Two sets of hierarchical regression analyses were conducted to determine significant predictors of word reading for each of the languages. The first analysis included the measures of English vocabulary and phonological awareness entered as independent variables and word reading as a dependent measure. Results revealed a significant model, $R^2 = .750$, $F_{(2, 17)} = 25.72$, $p < .001$. Phonological awareness was entered as the first step and then vocabulary was entered as the second step. Phonological awareness was significantly related to English word reading as the first step. English vocabulary knowledge was uniquely significantly related to participants' performance on English word reading, $\beta = .845$, $t_{(17)} = 5.23$, $p < .001$ (See Table 5). Subsequently, the variables were entered in the opposite order. Phonological awareness was no longer significantly related to English word reading as the second step.

The second set of hierarchical regression analyses was conducted to examine variables related to Hindi word reading. English phonological awareness and Hindi vocabulary were entered as independent variables and Hindi word reading as a dependent measure. Findings of the analysis showed a significant model, $R^2 = .479$, $F_{(2, 17)} = 7.83$, p = .004. Phonological awareness was significantly related to Hindi word reading as the first step. Hindi vocabulary knowledge was uniquely significantly related to Hindi word reading level, $\beta = .645$, $t_{(17)} = 2.82$, $p < .012$ (See Table 6). Subsequently, the variables were entered in the opposite order. Phonological awareness was no longer significantly related to Hindi word reading.

Table 5 English measures predicting English word reading

Total R^2 = .750				
Step	ΔR^2	β	Final β	t-value/sig
1. English Phonological awareness	.347	.589	.031	3.09**
2. English Vocabulary	.403	.845	.845	5.23***
1. English Vocabulary	.749	.866	.845	7.34***
2. English phonological awareness	.001	.031	.031	.193

***$p < .001$

Table 6 Hindi and English measures predicting Hindi word reading

Total R^2 = .479				
Step	ΔR^2	β step	Final β	t-value/sig
1. English Phonological awareness	.235	.485	.071	2.35*
2. Hindi Vocabulary	.244	.645	.645	2.82*
1. Hindi Vocabulary	.476	.690	.645	4.05**
2. English Phonological awareness	.003	.071	.071	0.31

*$p < .05$, **$p < .01$

Discussion

This study explored language-specific and language-universal processes and skills among Hindi-English speaking children. More specifically, this study examined the variables related to word reading in an alphabetic orthography (English) and abugida orthography (Hindi) in bilingual and multilingual learners living in North America. These participants showed a typical pattern of language loss with higher scores in English than in Hindi on measures of word reading and vocabulary. These between-language comparisons in terms of percent correct on tasks that measure similar constructs should be interpreted with caution. Despite the use of percent correct scores for the measures, it is not clear that the word reading lists can be equated across languages. Additionally, the extent to which the children had lower word reading scores in Hindi as compared to English are not fully captured by the analyses. Specifically, the Hindi word list was created using words from elementary school texts, while the English measure included items that were normed across a wide age range, including adults. Additionally, even though the same items were used for the vocabulary measure, word frequency and familiarity can vary across languages and cultures. However, the comparisons across these measures are illustrative of the level of proficiency across languages and show a clear pattern of L2 dominance in this sample. Despite low scores, the children did not show floor effects on any of the Hindi tasks. Additionally, the tasks showed good reliability. These indicators suggest that the tasks were valid measures of word reading and vocabulary knowledge in Hindi for the children in this study.

Parental responses to language use by the child in and outside the home suggest that both English and Hindi as well as other languages are spoken in the home. These results suggest that despite exposure to Hindi in the home and through the media as well as use of Hindi with members of the community, the children showed a pattern of language dominance in English. Therefore, the effects of English skills on Hindi performance are likely to be greater than the effects of Hindi skills on English performance. In light of the children's dominant skills in English, these participants are referred to as bilingual in the broad sense of the term, but the effects of English language and literacy skills will be discussed to a greater degree than the effects of the L1, Hindi, on the L2, English.

The Role of Phonological Awareness

In terms of predictors of English word reading, theory and research support the importance to phonemic awareness for reading an alphabetic language such as English. For these fluent bilingual readers, English phoneme awareness was the phonological measure that was most highly related to English reading. *Psycholinguistic grain size* theory would suggest that reading Hindi, an abugida, would be related to awareness of syllables and phonemes (Rimzhim et al., 2014; Share & Daniels, 2016; Ziegler & Goswami, 2005). However, previous research findings support the link between phonemic awareness and reading in Hindi (Das et al., 2014; Rao et al., 2011). The findings of the present study align with the link between phoneme awareness and reading in Hindi. However, these findings are influenced by the greater variability for the items measuring phoneme awareness than the items measuring syllable awareness. Additionally, these findings might not be definitive in terms of describing reading in all Hindi speakers, given that these children showed dominance in English, their L2.

Previous research has shown that phonological awareness skills in one language are related to reading in the other language, if the L2 is an alphabetic language. L1 phonological awareness skills are related to L2 reading even if the L1 is not an alphabetic language, such as Chinese (Gottardo et al., 2001). Previous studies conducted with Spanish-English and Portuguese-English speakers also have shown the benefits of learning to read a shallow orthography on reading a deeper orthography (Da Fontoura & Siegel, 1995; Durgunoğlu et al., 1993). Although in the previous research each language was written using an alphabet, similar principles might apply to reading an alphabet and an abugida system, where alphabetic skills can assist in reading both systems.

Other Variables Related to Reading

Interestingly, in both cases, vocabulary knowledge in the respective language was highly related to word reading. Given that models of word reading include semantic information in the process of word reading (Seidenberg & McClelland, 1989), the role of vocabulary in word reading should not be surprising. Additionally, language-specific vocabulary knowledge can help readers who are not skilled decoders to decipher a word, which they are able to partially decode, if they already are familiar with the meaning of that word (Gottardo, 2002; Share, 1995). The children in the present study had stronger oral than written skills in Hindi and had good English oral skills. Therefore, their vocabulary knowledge could assist them in reading words in each language. Additionally, it is possible that their strong reading skills enhanced the children's vocabulary skills or that the relationship is bidirectional (Cunningham & Stanovich, 1991). However, the cross-sectional nature of study does not allow us to address the direction of the relationship between word reading and vocabulary in this sample.

These findings that vocabulary knowledge was related to word reading suggest that language-general mechanisms were also related to word reading skills in these bilingual learners. Previous research has shown language-general skills and mechanisms influence reading skills in bilinguals. For example, for adolescent English L2 learners with varying amounts of experience with their L2, all decoding and vocabulary factors loaded on to one factor (Pasquarella, Gottardo, & Grant, 2012). In addition, a review by Geva (2006) found a link between vocabulary and word reading ability in English L2 learners with the level of proficiency in English explaining why vocabulary was a predictor of word reading in some studies but not in others. Geva (2006) suggests that vocabulary may be a proxy for general language proficiency in the L2 and have a greater impact on word reading in more proficient L2 learners. Similar relationships were found between word reading and vocabulary for native speakers of English (Kirby, Desrochers, Roth, & Lai, 2008). Finally, work by Frost (2012) has raised the possibility that models of word reading should consider many properties of the language at the same time, including semantic units. Even more recently, Frost and colleagues have suggested that the mechanisms used to learn to read words in a second language, involve principles of universal statistical learning (Frost, Siegelman, Narkiss, & Afek, 2013).

Limitations

Unfortunately, the lack of standardized measures in Hindi, limited the scope of measures that could be collected. Further development of measures in Hindi is required for research purposes and can eventually be used for educational purposes. Additionally, the comparisons between standardized and experimental measures are not ideal. Finally, the small sample size for these preliminary data analyses, limited the number of variables that could be entered in each regression equation.

Conclusion

Although the data from the present study cannot fully address broader universal principles, the findings suggest that language-general mechanisms are being used both in terms of levels of phonological awareness and in terms of general linguistic abilities. These findings might be applicable to samples of bilingual or multilingual immigrant children, who show a pattern of L1 loss and L2 dominance, due to their L2 being the language of instruction and the societal language. These relations across languages are interesting in light of their applicability to languages having different linguistic roots and writing systems.

References

Abu-Rabia, S. (2007). The role of morphology and short vowelization in reading Arabic among normal and dyslexic readers in grades 3, 6, 9, and 12. *Journal of Psycholinguistic Research, 36*(2), 89–106. https://doi.org/10.1007/s10936-006-9035-6.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96*(1), 43–55. https://doi.org/10.1037/0022-0663.96.1.43.

Bialystok, E., & Herman, J. (1999). Does bilingualism matter for early literacy? *Bilingualism: Language and Cognition, 2*(01), 35–44.

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing, 5*(1), 69–85.

Brownell, R. (2001). *Expressive one-word picture vocabulary test – Spanish-English* (Bilingual ed.). Novato, CA: Academic Therapy Publications.

Chang, C. B. (2013). The novelty effect in phonetic drift of the native language. *Journal of Phonetics, 41*, 520–533. https://doi.org/10.1016/j.wocn.2013.09.006.

Chow, B. W. Y., McBride-Chang, C., & Burgess, S. (2005). Phonological processing skills and early reading abilities in Hong Kong Chinese kindergarteners learning to read English as a second language. *Journal of Educational Psychology, 97*(1), 81–87. https://doi.org/10.1037/0022-0663.97.1.81.

Cook, V. (2003). The changing L1 in the L2 user's mind. In V. J. Cook (Ed.), *Effects of the second language on the first* (pp. 1–18). Clevedon, UK: Multilingual Matters.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222–251. https://doi.org/10.3102/00346543049002222.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*(2), 264–274. https://doi.org/10.1037/0022-0663.83.2.264.

Da Fontoura, H. A., & Siegel, L. S. (1995). Reading, syntactic, and working memory skills of bilingual Portuguese-English Canadian children. *Reading and Writing: An Interdisciplinary Journal, 7*(1), 139–153.

Das, N., Acharya, K., Sarkar, R., Basu, S., Kundu, M., & Nasipuri, M. (2014). A benchmark image database of isolated Bangla handwritten compound characters. *International Journal on Document Analysis and Recognition, 17*(4), 413–431. https://doi.org/10.1007/s10032-014-0222-y.

Durgunoğlu, A. Y. (2002). Cross-linguistic transfer in literacy development and implications for language learners. *Annals of Dyslexia, 52*(1), 189–204.

Durgunoğlu, A. Y., Nagy, W. E., & Hancin-Bhatt, B. J. (1993). Cross-language transfer of phonological awareness. *Journal of Educational Psychology, 85*(3), 453. https://doi.org/10.1037/0022-0663.85.3.453.

Ehri, L. C. (2011). Grapheme—phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3–40). New York, NY: Routledge.

Frost, R. (2012). A universal approach to modeling visual word recognition and reading: Not only possible, but also inevitable. *Behavioral and Brain Sciences, 35*(05), 310–329. https://doi.org/10.1017/S0140525X12000635.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*(7), 1243–1252. https://doi.org/10.1177/0956797612472207.

Geva, E. (2006). Second-language oral proficiency and second-language literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 123–139). Mahwah, NJ: Lawrence Erlbaum.

Geva, E., & Siegel, L. S. (2000). Orthographic and cognitive factors in the concurrent development of basic reading skills in two languages. *Reading and Writing: An Interdisciplinary Journal, 12*(1), 1–30.

Geva, E., & Wang, M. (2001). The development of basic reading skills in children: A cross-language perspective. *Annual Review of Applied Linguistics, 21*, 182–204. https://doi.org/10.1017/S0267190501000113.

Gottardo, A. (2002). The relationship between language and reading skills in bilingual Spanish-English speakers. *Topics in Language Disorders, 22*(5), 46–70.

Gottardo, A., Javier, C., Farnia, F., Mak, L., & Geva, E. (2014). Bidirectional cross-linguistic relations of first and second language skills in reading comprehension of Spanish-speaking English learners. *Written Language & Literacy, 17*(1), 62–88. https://doi.org/10.1075/wll.17.1.04got.

Gottardo, A., Pasquarella, A., Chen, X., & Ramirez, G. (2016). The impact of language on the relationships between phonological awareness and word reading in different orthographies: A test of the psycholinguistic grain size theory in bilinguals. *Applied Psycholinguistics, 37*(05), 1083–1115. https://doi.org/10.1017/S0142716415000508.

Gottardo, A., Yan, B., Siegel, L. S., & Wade-Woolley, L. (2001). Factors related to English reading performance in children with Chinese as a first language: More evidence of cross-language transfer of phonological processing. *Journal of Educational Psychology, 93*(3), 530. https://doi.org/10.1037/0022-0663.93.3.530.

Gupta, A. (2004). Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal, 17*(1–2), 79–99. https://doi.org/10.1023/B:READ.0000013823.56357.8b.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82*, 2–28. https://doi.org/10.1006/jecp.2002.2670.

Jiang, N. (2004). Semantic transfer and its implications for vocabulary teaching in a second language. *The Modern Language Journal, 88*(3), 416–432. https://doi.org/10.1111/j.0026-7902.2004.00238.x.

Jiménez González, J. E. (1997). A reading-level match study of phonemic processes underlying reading disabilities in a transparent orthography. *Reading and Writing: An Interdisciplinary Journal, 9*(1), 23–40. https://doi.org/10.1023/A:1007925424563.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology, 94*, 67–84. https://doi.org/10.1016/S0166-4115(08)62789-2.

Kelkar, A. R. (1968). *Studies in Hindi-Urdu, Introduction & Word Phonology*. Deccan College Postgraduate and Research Institute.

Kirby, J. R., Desrochers, A., Roth, L., & Lai, S. S. (2008). Longitudinal predictors of word reading development. *Canadian Psychology/Psychologie Canadienne, 49*(2), 103–110.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95*(3), 453–464. https://doi.org/10.1037/0022-0663.95.3.453.

Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal, 80*(4), 450–460. https://doi.org/10.1111/j.1540-4781.1996.tb05465.x.

Leong, C. K., & Tamaoka, K. (Eds.). (1998). Cognitive processing of the Chinese characters, words, sentences and the Japanese kanji and kana: An introduction. *Reading and Writing: An Interdisciplinary Journal, 10*, 155–164. https://doi.org/10.1023/A:1008044304939.

Lindsey, K. A., Manis, F. R., & Bailey, C. E. (2003). Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology, 95*(3), 482–494. https://doi.org/10.1037/0022-0663.95.3.482.

Nag, S. (2007). Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading, 30*(1), 7–22. https://doi.org/10.1111/j.1467-9817.2006.00329.x.

Nag, S. (2011). The akshara languages: What do they tell us about children's literacy learning? In R. Mishra & N. Srinivasan (Eds.), *Language-cognition: State of the art* (pp. 291–310). Munich, Germany: Lincom Publishers.

Nag, S., & Perfetti, C. A. (2014). Reading and writing: Insights from the alphasyllabaries of South and Southeast Asia. *Writing Systems Research, 6*(1), 1–9. https://doi.org/10.1080/17586801.2014.883787.

Nag, S., & Snowling, M. J. (2012). Reading in an alphasyllabary: Implications for a language universal theory of learning to read. *Scientific Studies of Reading, 16*(5), 404–423. https://doi.org/10.1080/10888438.2011.576352.

Nag, S., Treiman, R., & Snowling, M. J. (2010). Learning to spell in an alphasyllabary: The case of Kannada. *Writing Systems Research, 2*(1), 41–52. https://doi.org/10.1093/wsr/wsq001.

Pandey, A. (2014). Using mother tongues as building blocks in childhood education. *Childhood Education, 90*(1), 61–67. https://doi.org/10.1080/00094056.2014.872517.

Pasquarella, A., Chen, X., Gottardo, A., & Geva, E. (2015). Cross-language transfer of word reading accuracy and word reading fluency in Spanish-English and Chinese-English bilinguals: Script-universal and script-specific processes. *Journal of Educational Psychology, 107*(1), 96. https://doi.org/10.1037/a0036966.

Pasquarella, A., Chen, X., Lam, K., Luo, Y. C., & Ramirez, G. (2011). Cross language transfer of morphological awareness in Chinese–English bilinguals. *Journal of Research in Reading, 34*(1), 23–42. https://doi.org/10.1111/j.1467-9817.2010.01484.x.

Pasquarella, A., Gottardo, A., & Grant, A. (2012). Comparing factors related to reading comprehension in adolescents who speak English as a first (L1) or second (L2) language. *Scientific Studies of Reading, 16*(6), 475–503. https://doi.org/10.1080/10888438.2011.593066.

Patel, P. G. (1996). Linguistic and cognitive aspects of the orality-literacy complex in ancient India. *Language & Communication, 16*(4), 315–329. https://doi.org/10.1016/S0271-5309(96)00020-1.

Patel, P. G. (2004). *Reading acquisition in India: Models of learning and dyslexia* (Vol. 6). New Delhi: Sage.

Patel, P. G., & Soper, H. V. (1987). Acquisition of reading and spelling in a syllabo-alphabetic writing system. *Language and Speech, 30*(1), 69–81. https://doi.org/10.1177/002383098703000106.

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*(1), 3–24.

Perfetti, C. A., & Harris, L. N. (2013). Universal reading processes are modulated by language and writing ystem. *Language Learning and Development, 9*(4), 296–316. https://doi.org/10.1080/15475441.2013.813828.

Prakash, P., & Joshi, R. M. (1995). Orthography and reading in Kannada: A Dravidian language. In *Scripts and literacy* (pp. 95–108). Dordrecht: Springer. https://doi.org/10.1007/978-94-011-1162-1_7.

Prema, K. S., & Karanth, P. (2003). Assessment of learning disability: Language-based tests. In P. Karanth & J. Rozario (Eds.), *Learning disabilities in India: Willing the mind to learn* (pp. 138–149). New Delhi, India: Sage Publications.

Ramirez, G., Chen, X., Geva, E., & Kiefer, H. (2010). Morphological awareness in Spanish-speaking English language learners: Within and cross-language effects on word reading. *Reading and Writing: An Interdisciplinary Journal, 23*(3–4), 337–358. https://doi.org/10.1007/s11145-009-9203-9.

Rao, C., Vaid, J., Srinivasan, N., & Chen, H. C. (2011). Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. *Reading and Writing: An Interdisciplinary Journal, 24*(6), 679–695. https://doi.org/10.1007/s11145-010-9256-9.

Rimzhim, A., Katz, L., & Fowler, C. A. (2014). Brāhmī-derived orthographies are typologically Āksharik but functionally predominantly alphabetic. *Writing Systems Research, 6*(1), 41–53. https://doi.org/10.1080/17586801.2013.855618.

Salomon, R. (2000). Typological observations on the Indic script group and its relationship to other alphasyllabaries. *Studies in the Linguistic Sciences, 30*(1), 87–103.

Schwartz, M. (2008). Exploring the relationship between family language policy and heritage language knowledge among second generation Russian–Jewish immigrants in Israel. *Journal of Multilingual and Multicultural Development, 29*(5), 400–418. https://doi.org/10.1080/01434630802147916.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*(4), 523–568.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174. https://doi.org/10.1348/000712603321661859.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: the perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584.

Share, D. L. (2014). Alphabetism in reading science. *Frontiers in Psychology, 5*, 1–3.

Share, D. L., & Daniels, P. T. (2016). Aksharas, alphasyllabaries, abugidas, alphabets and orthographic depth: Reflections on Rimzhim, Katz and Fowler (2014). *Writing Systems Research, 8*(1), 17–31. https://doi.org/10.1080/17586801.2015.1016395.

Stanovich, K. E. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10*(1), 72–100. https://doi.org/10.1016/0273-2297(90)90005-O. https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/.

Statista. (2017). The statistics portal: Statistics and studies from more than 22,500 sources. https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide.

Statistics Canada. (2016). Data tables, 2016 census admission category and applicant type (47), immigrant status and period of immigration (11B), age (7A), sex (3) and selected demographic, cultural, labour force and educational characteristics (825) for the population in private households of Canada, provinces and territories and census metropolitan areas, 2016 census – 25% sample data. https://www12.statcan.gc.ca/census-recensement/2016/

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. *Papers in laboratory phonology V: Acquisition and the lexicon,* 269–282.

Vaid, J. (1995). Script directionality affects nonlinguistic performance: Evidence from Hindi and Urdu. In *Scripts and literacy: Reading and learning to read alphabets, syllabaries, and characters* (Vol. 7, pp. 295–310). Dordrecht, The Netherlands: Kluwer.

Vaid, J., & Padakannaya, P. (2004). Reading and writing in semi-syllabic scripts: An introduction. *Reading and Writing: An Interdisciplinary Journal, 17*(1), 1–6. https://doi.org/10.1023/B:READ.0000013861.42512.b0.

Venezky, R. L. (1970). *The structure of English orthography* (Vol. 82). Walter de Gruyter/Mouton: The Hague, The Netherlands.

Verhoeven, L. (2007). Early bilingualism, language transfer and phonological awareness. *Applied PsychoLinguistics, 28*(3), 425–439. https://doi.org/10.1017/S0142716407070233.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: Pro-Ed.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition, 51*(1), 91–103. https://doi.org/10.1016/0010-0277(94)90010-8.

Woodcock, R. W. (1998). *Woodcock reading mastery tests, revised, examiner's manual*. Circle Pines, MN: American Guidance Service.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3.

Cross-Modal Processing of Orthography-Phonology Interface in Hindi-English Bilinguals

Ramesh Kumar Mishra

Abstract The chapter presents an overview of Devanagari script. Later it presents experimental data from one eye tracking experiment that examined interface between spoken words and orthographic activation. The chapter also touches upon discussion of dyslexia in the Indian setting. The chapter emphasizes the point that much psycholinguistic work needs to be done in Indian languages that link orthographic, phonological and visual processing.

Keywords Akshara · Orthography · Eye movements · Visual world · Bilingualism

Introduction

Every script is unique in many ways. Linguistic theories describe scripts in terms of their orthographic, phonological and morphological terms. Scripts are also described in terms of their design- the way they look. All scripts carry in them their historical legacy. Scripts evolve over time and undergo similar changes as verbal sounds as a matter of evolution. While many have described scripts, classifying them around orthographic or phonological structures, others have explored their consequences for literacy acquisition and writing development as indicators of overall language development (Ellis et al., 2004). Do scripts constrain or facilitate the way they are learned and used? This has been a central question in modern psycholinguistics of literacy learning and also in understanding many developmental and acquired disorders of literacy (Leong, 1986; Weekes, 2005). In recent times, various hypotheses have been proposed that relate different types of scripts and their ortho-phonological organization to processing (e.g., Abu-Rabia & Taha, 2005). Although today we have a good amount of theoretical description of different scripts and their histories, empirical attention has not been given to many scripts. For example, the studies and understanding of dyslexia in Hindi or any other Indian script has not received the

R. K. Mishra (✉)
Center for Neural and Cognitive Sciences , School of Medical Sciences,
University of Hyderabad, Hyderabad, Telangana, India
e-mail: rkmishra@uohyd.ac.in

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*,
Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_18

same attention as in English (An akshara script). The reasons are many and have to do with a range of issues, the most central being the correct characterization of the script and its psycholinguistic features. Most Indian scripts that are derived from "Brahmi" have been labeled as 'akshara' (Padakannaya, Pandey, Saligram, & Ranga Rao, 2016). Akshara is different from either the syllabic or the phonemic scripts. Therefore, many researchers are taking up akshara-based studies to understand literacy acquisition and literacy difficulties. This movement has led to the development of newer assessment tools in Indian languages and also to theoretical understanding in many of the scripts. My main aim in this chapter is to first discuss briefly Hindi (a Devanagari script and classified as akshara type) and then to review work spanning several fields. I will focus on experimental work on Hindi that has revealed certain unique properties of this script using contemporary techniques like brain imaging and eye tracking. Then I will present data from a bilingual cross-modal study where listening to spoken words led to the activation of an orthography.

Hindi: An Akshara Script and Its Processing

Akshara is an orthographic unit that is part of many scripts found in South Asia (Nag, 2011). English is an alphabetic script; each letter corresponds to a phoneme. The same is not true for Devanagari, a Brahmi derived script used widely across the Indian sub-continent in writing the Hindi orthogaphy. Hindi is also not entirely syllabic in the sense that one can denote its aksharas as syllables which contains both phonemic and syllabic properties. Figure 1 illustrates the akshara nature of some Indian orthographies.

What is so different and unique about the akshara? A central consideration in the psycholinguistic classification of scripts has been to identify its orthography to phonology conversion. The much studied English is an opaque language where its letters do not always correspond to the same sounds. Depending on their orthographic context, how one pronounces the letters differ widely. This has been taken as a possible causal factor in its being difficult for literacy acquisition. On the other hand, Indian scripts like Hindi do not pose such a problem once the decoding of aksharas are mastered? That is, there is a clear and consistent mapping between orthography and sounds associated with them. The learner need not remember any typical

Fig. 1 The akshara nature of Indian languages: Telugu, Hindi, Malayalam and Kannada

contextual variation. This distinction has much impact on the way one understands the acquisition of reading and analysis of disorders of reading in such scripts. Therefore all our current understanding derived from alphabetic scripts may not be applicable in understanding literacy in akshara. Does the transparency of akshara lead to faster acquisition and lesser developmental disorders? No. It appears that both developmental and acquired deficits of script processing manifest in all such languages as they very much manifest in other European transparent orthographies such as Finnish and Czech. However, what is important is a framework to describe and understand the nature of the deficits given that they are qualitatively very different.

Hindi aksharas are composed of consonants and vowels. For example the sound 'क'('Ka') is made of a consonant and a vowel. Therefore vowels are built into aksharas along with one or more consonants. Often consonant clusters are also combined with vowels, as in the sound of Kri. Here one has two consonant sounds and one vowel sound. Most fundamentally the akshara system can be just a consonant "C" with abuilt in vowel. Further on one can add additional vowels and also consonants. Therefore in a multisyllabic word in Hindi such as "matrutwa" (motherhood) there are several distinct aksharas that have consonant clusters. These are not mere syllables but phonological units that consist of larger clusters. This way when the child learns to read, s/he acquires the knowledge of aksharas and not individual syllables or phonemes. This unique composition is seen in all Indic scripts and has now been included in the studies of literacy acquisition (see Nag, 2011 for a fuller description).

Another very important feature about aksharas which has to be noted is the way in which consonants and vowels that make up akshara units are arranged (Nag, 2011). Most vowels and consonants that are attached to the main consonant appear as diacritics with unique symbols, for example, the Hindi akharas "क" ("Ka"), "कि"("Ki") or "कु" ("Ku"). However for Hindi the position of the vowel diacritic "I" appears positioned to the left of the consonant it modifies such as in the word "Kitab." This is a unique property of Hindi and children learn it that way. Thus, one cannot say that Hindi sounds are added incrementally in terms of visuo-spatial location. Unfortunately, not many eye tracking studies have been carried out on Hindi akshara learning in children. Does the visual analysis of akshara based orthographies differ from alphabetic orthogrpahies? Phonologically Hindi also has nasal vowels. Therefore nasality is 'phonemic' in Hindi. For example, consider the Hindi words sounding like "saas" and 'saans." These words differ in meaning as a result of the addition of the nasal vowel.

Although the akshara description of the Indic script is somewhat recent some researchers have already explored the unique neural signature of Devanagari script before. There are many ways one can look at whether scripts are processed differently by readers using different methodologies. One traditional method is to do some behavioral tasks with neuropsychological patients who have suffered a stroke. Other more contemporary methods include doing brain imaging (fMRI) with neurologically intact subjects. From the description offered in the introduction, it is clear that Hindi being an akshara based orthography has a phonology which has both syllabic and phonemic properties. Furthermore, the way Hindi is written, consonant clusters and vowels appear as diacritics attached at different spatial locations around a lead consonant. Therefore during word decoding readers have to read the akharas as units of articulation.

Vaid and Gupta (2002) in a first comprehensive examination of word decoding in Hindi studied vowel placement and reading. The authors also provide a very accurate and comprehensive description of the Hindemorthography for the reader which can be referred to easily. The issue they addressed was related to the spatial appearance of the short vowel /i/ which appears before word-initial consonants. For example, in the word 'sitar' the consonant diacritic precedes the consonant. If read phonemically then one can be misguided and may read this as 'istar'. However proficient Hindi readers (it is interesting to note that in their word reading experiments the authors examined not just adults but also children) do not make any mistakes in reading such words accurately. The authors wanted to examine if Hindi readers follow a syllabic or a phonemic route to decode such words that have vowels appearing before consonants. The authors created words wherein some of the short and long /i/ vowels were at spatial locations congruent with their pronunciations whereas, in other words, they were not. The idea was to examine if such spatial and pronunciation discrepancies will lead to a naming cost. The authors proposed that if there is a cost then these have to be interpreted as Hindi being read phonemically. The results showed that this was the case. However, the current theorization of the akshara is not in agreement with such a view. It is possible that there has not yet been much experimental work demonstrating how uniquely akshara is processed. For example, it is possible to test if readers access both phonemic and syllabic information while reading akharas. Although this was an initial study, it certainly shows that the unique properties of Hindi affect the way people read it. The authors also found such errors because of misalignment in writing to dictation.

In another study of dyslexic children and age-matched normally reading children, Gupta (2004) administered an oral reading test. The words varied in complexity and consonant clusters. Interestingly the reading patterns revealed that maximum errors were committed for vowels and consonant clusters. The author argued that this was because of the complex spatial organization of vowel and consonant diacritics in Hindi which the dyslexics find very difficult to decode. Apparently, such a result concerning higher errors on vowels had also been found earlier in Hindi reading by Nehru (2001) and in Kannada reading by Purushothama (1990). Thus, vowels and their positions are important for decoding in Devanagari script. More recent studies on Hindi orthography have examined the issue of orthographic depth. Orthographic depth relates to the clustering of consonant and vowel sounds within a syllable (Katz & Frost, 1992). The orthographically dense scripts are more densely packed with such information than the orthographically shallow ones. Writing systems that are orthographically deep will require more processing time than those that are shallow.

Rao, Vaid, Srinivasan and Chen (2011) used masked priming to examine the issue of orthographic depth in Hindi-Urdu bi- literate readers. Urdu is written from right to left, unlike Hindi. Interestingly, Hindi and Urdu share much of their phonology, and many words are common to both languages. Urdu is said to have a Persian origin and is spoken widely in many places in the sub-continent. The Hindi-Urdu bi-literates in India offer a nice test ground to examine how both these writing systems are represented in the same reader. The fundamental difference between them is that, while Hindi still has a more or less linear organization of letters, Urdu does not have this. Urdu is a modified version of the Perso-Arabic script (Rao et al., 2011). Interestingly unlike Hindi, in Urdu, vowel diacritics are often not explicitly indicated, thus making the script non-transparent. The script is also described as alphabetic. The authors found that while form based primes facilitated Hindi naming, they did not influence Urdu. This was interpreted as suggesting that Hindi word recognition is still based on some type of phonological analysis while for Urdu that is not the case. Urdu is complex with its nontransparent orthography to phonology mapping and therefore primes of either sort do not facilitate word naming in this script. The experiments thus showed that the visual, as well as the orthographic complexity of a script, does influence naming and word recognition. Importantly, the authors interpreted their findings as supporting a somewhat alphabetic decoding of Hindi although they classified the orthography as being akshara.

So is Hindi then alphabetic as a script? Even if Akshara is claimed as a combination of syllabic and phonemic identities, what about its actual processing? From the above experimental evidence presented, it appears that some authors have attempted to explore and designate Hindi as requiring an alphabetic decoding strategy, although it is transparent. The vowel and consonant clusters and the way they are depicted as diacritics, of course, play a critical role in their decoding. For dyslexics and other poor readers, they pose a special problem in reading. However, these demonstrations do not unequivocally point towards any solution regarding the script's true identity. Below I will discuss some other studies that have attempted to see the processing of Hindi mainly as Akshara (syllable) based.

Akashras can be visually complex. As stated above, this complexity is mostly because of the many diacritics that are attached differently to the letters in such scripts, of which Hindi is a good example. Although few behavioral studies exist on this issue, there have been some brain imaging studies that have suggested that neural processing is different for such scripts. While much neuroimaging work on visual world recognition of western scripts has been around the issue of phonological decoding, those on Hindi have focused on the script's visual complexity. It is well known that the area known as the visual word form area selectively shows activation to written letters or words (Cohen et al., 2000). However, the initial evidence of this area's script specific activity came from Roman scripts (alphabetic). Does neural processing for Hindi differ since it is an Akshara language?

When talking about Hindi it has to be kept it mind that irrespective of it being visually complex it is also phonologically transparent. That is, its letter to sound correspondence is simple and one to one, unlike Urdu. Since the visual word form area reacts to visual and not phonological features as such, it is possible that this area will also respond to higher orthographic complexity. Rao and Singh (2015) examined whether the visual word form area of Hindi readers shows modulation of orthographic complexity. The authors asked readers to either read simple or complex words. Complex words had vowel diacritics. The brain activation was modulated by such complexity. However, such complexity had no effect on naming latency. The authors argued that, for fluent readers, neural responses for visual complexity may be evident, but not so much with behavioral data. However, as already reviewed, other researchers who have studied word recognition in Hindi in both typically developing and dyslexics seem to suggest that orthographic complexity slows down processing.

The issue then requires further experimentation with carefully controlled stimuli. It is possible that both frequency and orthographic complexity may influence decoding speed. However, the important finding of this neuroimaging study was its novelty in being done with Hindi readers, a very rarely studied group with such methods. The visual word form area is sensitive to script specific attributes which again should be examined in many other Indian scripts. This result also shows that akshara could be the unit of neural activation when it comes to Hindi orthography. In an earlier study, Das, Kumar, Bapi, Padakannaya, and Singh (2009) had examined word recognition in Hindi in Hindi readers using fMRI. In this study, it was of interest to examine if specific neural networks or areas are active to both the script's syllabic and phonemic properties. The data showed that the left insula, fusiform gyrus, and inferior frontal gyrus were active as they are normally seen for alphabetic scripts. In addition, activation in the right superior parietal lobule was found indicative of syllabic processing.

Taken together some studies seem to suggest that the Devanagari script may have both phonemic and syllabic properties. The conceptualization of Akshara, however, does not as such point towards any alphabetic sound to letter processing. It takes whole syllables that have both consonants and vowels together as units of processing. In the absence of further empirical support, it is not yet clear how such processing manifests in psycholinguistic tasks. Although there is currently general agreement among researchers about the Akshara unit, data that suggest both phonemic and syllabic differentiation must be taken into consideration when developing a general theory for such scripts. Furthermore, although most Brahmi-derived scripts may be classified as Akshara based, their orthographic visual complexity may widely differ. More recently, Rao, Vaid and Chen (2017) have compared Hindi and Kannada and suggest that script specific complexity may not transfer to the other script of a bi-literate reader. The bilingualism question may further complicate matters when one refers to akharas of different scripts. Below I move on to a different domain of Hindi orthography processing in the cross-modal domain.

Dyslexia in Indian Languages

Dyslexia is a developmental reading disorder where one fails to correctly ideitify written letters. Dyslexia manifests differently in different scripts. This is primarily because different scripts (Alphabetic vs. non-alphabetic) types require different processing strategies. Miles (2000) suggested that even if one may think that so called transparent languages that have straight forward grapheme to phoneme rules should pose no problems to readers, one finds dyslexia even in such languages. Although, the nature of linguistic deficits in such languages may differ. Considering the fact that India has many languages with different scripts the assessment and training of dyslexics may pose special challenges for educators (Ramaa, 2000). The other most important point in this context is the prevalence of bi-literacy among readers in India where the two languages may have very different scripts i.e. English and Hindi. Karanth (1992) observed that in such bi-literate individuals, the reading failure in English manifested differently compared to other scripts such as Hindi and Kannada. Similarly, when dyslexic children in Hindi are compared to age matched controls they seem to have greater errors in Hindi readiing (Gupta, 2004).

Therefore, transparency of script may not prevent the appearance of dyslexia in a script. One may have to find other cognitive variables that are the underlying cause of developmental reading impairments even in such scripts. It is now well known that many reading deficits in developmental dyslexics result from visual deficits including working memory and attention problems (e.g. Facoetti, Paganoni, Turatto, Marzola, & Mascetti, 2000; Jeffries & Everatt, 2004). Eye tracking studies show that those with dyslexia do not often fixate on parafoveal letters which are critical in deriving information as one reads incrementally. Dyslexics also are slow in the eye movement control when they read.

While studies have investigated such phenomena in many western scripts (De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; Rayner, 1985) and also in Chinese (e.g. Huang et al., 2008), this area remains underexplored in India. It is not known how exactly a bi-literate Hindi-English dyslexic manifests the errors. It possible that given the different orthographic representations, the visual and attentional aspects of eye movement control should be different. A study by Padakannaya et al. (2016) explored reading using eye tracking in Kannada. The study was with sentences of various complexities and was not on dyslexics. However, data presented suggest that poor readers show many abnormalities in eye movement variables when it comes to reading. It is possible that dyslexics may show impairment in decoding both akshara and alphabetic scripts similarly.

Mapping Orthography Activation Using Spoken Words

Does one only know about orthographic processing by examining reading alone? This has been intuitively the only method among psycholinguists because it allows excellent manipulation of the orthographic and phonological structure of the language and is easy to interpret. The above description and review of both behavioral and neuroimaging data lead one to think that the main issue is the appropriate quantification of the scripts, including finding a processing unit for psycholinguistic interpretation. For Hindi, which has anyway receive scant empirical attention, studies so far have mostly involved oral word naming or reading. These studies at best could tell if a certain manipulation with the ordering of the letters or the script complexity leads to a behavioral effect in terms of either reaction time or accuracy.

These studies only looked at events in one modality and where reading was the main task. The outcomes certainly have many implications for understanding orthographic processing and also dyslexia. Importantly, in the Indian context, many psycholinguistic issues such as sound or script processing are invariably bilingual in nature since almost all the participants tested know two Indian languages and two respective orthographies. The educational curriculum in India follows the three language policy where most receive learning in their local language, English, and another language. Therefore, by definition, no study done in India testing language can be a monolingual study. Even if data is taken from one language only their interpretation cannot disregard the bilingual/multilingual underpinnings.

Much recent work in the psycholinguistics of bilingualism has focused on simultaneous activation of both languages in bilinguals (Mishra & Singh, 2014). In this context, researchers have used the eye tracking technique to study the links between spoken words and orthography. This approach has implications since the very acquisition of written words is through the spoken medium. So when one looks at a string of words, one also accesses the phonological form. Similarly, many studies have shown that listening to words leads to the activation of orthographic forms, although with more fluent users of languages. For the bilingual, the issue is bit complex since there are two orthographic lexicons apart from two phonological lexicons. Before going into the description of the actual experiment of Mishra and Singh (2014), I will provide some theoretical background to the issue of bilingual cross-modal processing and the visual world paradigm which has provided a lot of online data in this area recently.

The visual world eye-tracking paradigm was first used to understand how spoken words guide visual attention to objects in space (Cooper, 1974; Tanenhaus, Spivey-knowlton, Eberhard, & Sedivy, 1995). In the last decade or so many researchers have used this paradigm to study a range of psycholinguistic issues ranging from sentence processing to semantic activation. The paradigm (Mishra, Huettig, & Olivers, 2012) is simple enough to study how spoken words that refer to objects (also words) lead to eye movements that reflect cognitive processing. In most versions of it, a participant sees a display of objects (typically four line drawings or words) and the spoken word is presented. Often this word has a corresponding object or some other object that is related to it.

Immediately, with the onset of the spoken words, viewers start to direct their looks towards the object that is mentioned but most importantly towards the object which is related to it in any dimension (semantic, phonology or visual). Tracking these eye movements as time passes shows the level to which linguistic input activates cognitive processing. In contrast to the regular reaction time method which can only provide offline data (this is also true for fMRI), the visual word eye-tracking data provide very rich online data. These data show the simultaneous processing of both visual and linguistic information online (Mishra, 2009). Importantly, saccade programming as a function of linguistic and visual analysis also reveals the dynamic nature of cross-modal processing.

Researchers generally plot the proportion of fixations towards the various visual objects over time to understand how visual attentional bias develops dynamically. In his book the 'Continuity of Mind', Michel Spivey (2008) had referred to such data to suggest that they show the ongoing and unfinished nature of cognition. As cognition develops over time through competition and bias, eye movements reflect this. For example, in the case of reading for comprehension, by the time the reader has understood the sentence and given a response, he has already generated several saccades towards many letters and words in it. Admittedly, though, it is very difficult to later correlate which such fixations were particularly important for comprehension.

Take for example our issue of the phonemic vs. syllabic level of processing in the case of the Devanagari script. The RT-based studies will only provide some gross measure of the final decision with regard to the task. If the visual word paradigm is used, then it is possible to study how both these representations compete when a reader attempts to decode the Aksharas. Either way, any hypothesis can be tested which allows a model in which there is competition in the beginning and a final decision with regard to the activation. Therefore, the visual world paradigm allows us to study the continuously evolving nature of representations and the inherent competition in them as mental processing progress. This further allows an excellent method to examine dyslexic individuals as far as their linguistic and visual processing is concerned.

Among various models and hypotheses that have been offered to explain language mediated eye movements in the visual word paradigm, it appears that working memory plays a key role in binding both the visual and the linguistic information (Huettig, Olivers, & Hartsuiker, 2011). Working memory is a critical resource which allows a working pad for currently important information (Baddeley, 2000). When we see and listen to something simultaneously, both types of information need to be blended into a coherent representation of the saccade system to program. Most recently it has been shown that working memory capacity also predicts language mediated eye movements in viewers (Huettig & Janse, 2016).

Earlier, Crocker and Knoeferle (2006) had also implicated working memory in their model of the visual world. Importantly, working memory is also now theorized as selective attention which plays a key role in goal-directed actions (Engle, 2002). The eye movements that are seen in visual world eye-tracking studies, therefore, indicate to us online integration of cross-modal information and decision making. Most often in such studies, participants are not given any task (see Tanenhaus et al., 1995). This frees them from deliberately making any decision, and their eye movements are tracked in a more naturalistic scenario. More sophisticated statistical models of this paradigm that allow more fine-grained analysis and hypothesis testing have been developed lately (e.g. Barr, 2008).

The Visual World in Bilingual Language Processing

Bilinguals activate both the lexicons that they know in speaking, word recognition, and also during listening (Marian & Spivey, 2003; Spivey & Marian, 1999; Weber & Cutler, 2004). Many visual world studies have shown that bilinguals activate the non-target lexicon when they listen to words in just one language. For example, the early studies by Marian and colleagues showed that Russian-English bilinguals activate 'Marku' when they listen to words like 'marker' (Marian & Spivey, 2003). Words that are phonologically related within a language and cross-linguistically are activated. Bottom-up activation of phonological information leads to co-activation. These findings have been very robust across many languages. In the Indian context, Mishra and colleagues have studied Hindi-English bilinguals' language non-selective activation using eye tracking.

These studies have shown that Hindi-English unbalanced bilinguals activate lexicons in both the languages with input in one. At the same time, in contrast to other studies in European language pairs, even highly proficient bilinguals activate translation pairs. It must be stated that the Revised Heirarchical Model (RHM; Kroll & Stewart, 1994) model of bilingual language processing had proposed that only low proficient bilinguals may activate translation equivalents. Before getting into the issue of orthographic activation, it is important to offer some characterization of the Hindi-English bilinguals.

In the Indian context, most education happens in English, particularly in colleges and universities. Children learn to acquire both spoken and written forms of English at an early age. Continual use of English makes them very proficient users of this language. In fact, some of the qualitative data collected from Hindi-English bilinguals in Indian universities suggest that their dominant language is English (e.g. Bhatia, Prasad, Sake, & Mishra, 2017; Roychoudhuri, Prasad, & Mishra, 2016). Therefore, it is easy to find many Indian bilinguals whose second language is English where they show good proficiency. Very importantly, English and Hindi have different orthographies and phonological structures. This marks this language combinationunique to study how such very different phonology and orthography still lead to mixing of lexicons. It should be noted that in most European language pairs studied in this context, languages share the same orthography (Roman alphabet) and a large number of cognates between them. Therefore, although parallel language activation is observed in both the cases of language pairs it is not easy to explain to what extent orthographic and phonological dissimilarity or similarity plays a role. In addition, English is almost always acquired exclusively by formal training in schools. Children learn the letters and their pronunciations and also make words early on. This early acquisition through orthography creates a strong connection between letters and sounds. Therefore, it is legitimate to expect that for Indian speakers of English, spoken word processing should always lead to orthographic activation. Another matter which remains uninvestigated so far is to what extent bilingual language processing, in general, is influenced by such orthographic learning.

The issue on which I will now focus is activation of orthography through phonological input in bilinguals. Do listeners activate written words when they listen to spoken words assuming that they are also fluent readers in both the languages? It has been shown that orthographic overlap facilitates the lexical decision in spoken words (Chereau, Gaskell, & Dumay, 2007). When spoken words share orthographic units, they are better processed. Perre and Ziegler (2008) used ERPs to examine if orthographic consistency plays a role in spoken word recognition. They presented spoken words whose orthographic representations were either consistent or inconsistent in a lexical decision task. Results were that the earliest ERPs were time-locked to words that were orthographically inconsistent. Earlier, Seidenberg and Tanenhaus (1979) had observed that people judge rhyming pairs faster if they have orthographic consistency. These effects provide strong support for the notion that at least for those who know how to read, there is activation of the orthographic representation of the spoken words as they process them.

Importantly, such activation seems to be automatic. Although some psycholinguists believe that orthographic activation could be only post-lexical (coming much later in the decision-making process) others think it is a core process of speech recognition. This is very important in the context of reading acquisition and how this modulates spoken word recognition in developmental terms. While talking about the orthographic complexity hypothesis as discussed earlier, is it the case that words that have many clusters take more time to be recognized in spoken form? While many have examined this in the reading domain no one has examined this in the spoken domain at least in Indian languages.

Mishra and colleagues (Mishra & Singh, 2014; Singh & Mishra, 2015, 2016) carried out many studies using the visual world eye-tracking paradigm to understand how Hindi-English bilinguals activate their lexicons given input in either language. Mishra and Singh (2014) used written words in place of line drawings to see if listening to words in any one language also activated orthographic representations in another language. For Hindi-English bilinguals, the issue was interesting since both the languages do not share orthographies and also cognates. Earlier, Huettig and McQueen (2007) had demonstrated that just like line drawings, written words can also be used to study language mediated eye movements. They had found that Dutch-English bilinguals look at written words that share phonology given a spoken word as input.

Thus, in such a scenario, when one hears a spoken word, the linguistic representation derived is matched against semantics of the written words. Given the fast temporal nature of such eye movements and the spontaneity of the paradigm, it is understood that given little-spoken word input, most readers can activate the corresponding orthographical representations in their language. Later, Salverda and Tanenhaus (2010) also attempted to examine such language mediated eye movements to written words using the visual world paradigm. They too observed that listeners looked at written words that were rhymes of spoken words. Thus, representations of spoken words lead to activations of both phonological and semantic neighbors. This is important given the fact that just like phonological units, orthographic units have their neighbors and they too are activated automatically given any input.

Such issues are important to study in the case of translation equivalents that can be written in the different orthographies. Translation equivalents have different phonological and orthographic representations. It would be fascinating to explore in the future how English (alphabetic) and Hindi (akshara) connect to the same concept and bilinguals activate them flawlessly. Importantly, because of associations, each translation equivalent further has many phonological neighbors of its own. For example, the translation of the English word 'gun' in Hindi is 'बंदूक' ('bandook'). A phonologically related word is 'बंदर' ('bandar'). Similarly, one can find the other way around and make a list of words that are phonological neighbors of such translation equivalents.

Others researchers using the translation recognition task (Sunderman & Priya, 2012) observed that bilinguals activate translations during lexical decision making. In this task, the participant is first presented with a prime word and then a target word. Critically the target word is at times a phonological cohort of the translation of the prime word. Data from such studies show that bilinguals take more time to reject such words as real words when they happen to be related to translations. However, the translation recognition task may have some limitations since the task itself demands explicit activation of semantics. The participant has to judge if the target word is related to the prime word or not and therefore must use semantics. Such a problem is avoided in the visual world paradigm where the participant has no explicit task. Eye movements towards a translation cohort during listening can suggest such covert activation. Mishra and Singh (2014) used written words in place of pictures, probably for the first time in a bilingual context using eye tracking. Below is a sample trial of the experiment (Fig. 2).

The experiment's main stimuli consisted of translation equivalents and their phonological neighbors in both language directions. Objects that had distinct names in Hindi and English were selected for the study. A phonological neighbor was created by keeping the first syllable common. The words were written in both the scripts for later presentation on the computer screen. In the visual world eye tracking paradigm, the participant is presented with a display of pictures or words and then a spoken word. The spoken word could come at once with the display or after some delay. A little delay allowed participants to inspect the display and extract meaning. In this study, viewers saw a display of four words written either in Hindi or English and heard a spoken word. The spoken word was in a different language than the words that appeared on the display.

For example, when the spoken word was in Hindi, the display contained four words written in English and vice versa. Importantly, the display did not have the orthographic representation of the spoken word. However, it had a word which was the phonological neighbor of the translation equivalent of the spoken word among three distractors. In such studies, the distractors are presented with critical stimulus to see if eye movements are selectively biased towards the targets. Thus, participants had to merely look and listen to the words. Given the unconscious and automatic nature of eye movements, it was expected that viewers would immediately

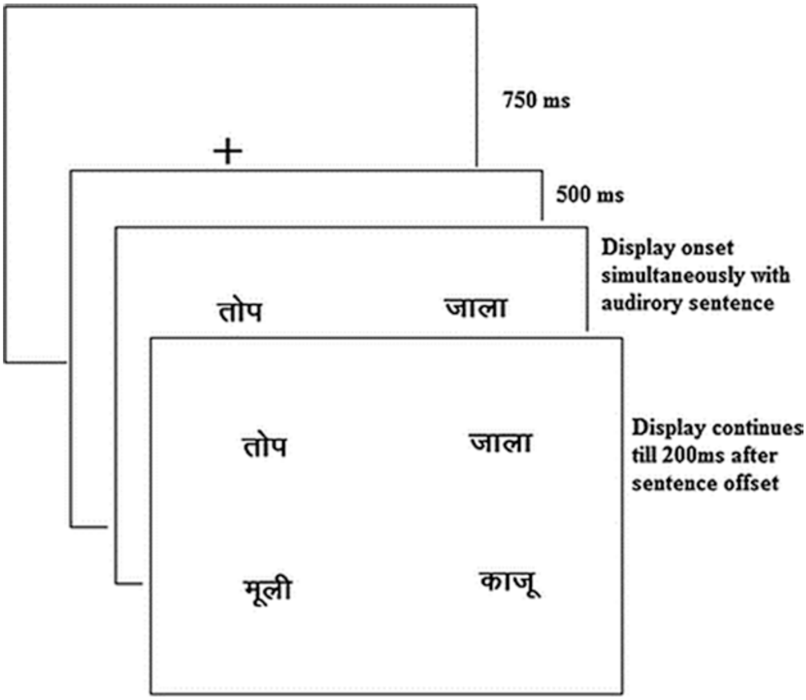


Fig. 2 Sequence of events on a sample trial (Mishra & Singh, 2014). A sentence containing the word "parrot" ("तोता" in Hindi) is presented along with a visual world display with four written words. The display consists of the word "thop" which is a phonological competitor of "तोता"

look at the written word which was a phonological cohort of the translation equivalent compared to the distractors. Such bias in attention would suggest that these bilinguals were spontaneously translating the spoken words and were further mapping their orthographic representations.

A few things must be first presented about our predictions in the experiment. Before us, Huettig and McQueen (2007), as well as Salverda and Tanenhaus (2010), had shown that written words could be used in place of pictures in a visual world study. Thus, we were certain that an ample amount of activation would be seen through eye movements. With regard to subjects who participated, they were all Hindi native speakers with English as the second language. They all were proficient in English and could be termed as unbalanced. That is because they were dominant in one of the languages because of a higher frequency of use of that language. Previous visual world studies with such bilinguals have shown that individual factors such as language proficiency and dominance and also the age of acquisition may influence the level of activation in such studies.

Theoretically, since our stimuli involved cohorts of translations, and a dominant model in the bilingual semantic organization (RHM) predicts that translation activation should be seen only in learners, it was interesting to examine if such effects would also be seen in participants who had a high degree of proficiency in L2. Previously as said before, Sunderman and Priya (2012), as well as Guo and Peng (2006), had observed translation activation in highly proficient bilinguals. However, prior to this study none had shown these effects in Hindi-English bilinguals. This made the study novel in its design as well as theoretical contribution. Most importantly it focused on a group of bilinguals who were biscriptal and whose two languages did not share any orthography or phonology as such. Therefore, it was interesting to examine the translation activation particularly when there was no bottom up similarity between the languages.

It is important also to briefly state how the visual world eye-tracking data is analyzed in general, since any inference of activation of a perceptual or cognitive representation is linked to how the data is analyzed. The main idea is to track viewers' eye movements as they look and listen to spoken words. One word is a phonological cohort of the translation equivalent (TE cohort) and others are distractors. The main interest in such studies is to see how early with the onset of the spoken word eye movements are biased towards the TE cohort compared to the distractors. Attentional bias towards any object is contingent upon the bias towards other objects. Thus, proportions of fixations are plotted against time for these objects.

Since there are three distractors, for statistical reasons, proportions of fixations to the three distractors are averaged. Further, to know at what time point the eye movements were significantly biased towards the TE cohort, fixations over certain windows are compared against one another. Such an analysis gives a clear picture of attentional bias over time as it develops. Many researchers also seek global measures like a total number of fixations on objects or total gaze durations. However, the proportion of fixations over time remains the gold standard in visual world eye tracking data analysis. Some researchers also use mixed effect models on such data since these data are continuous and not discrete. In our case, we have resorted to simple ANOVA analysis of the proportion of fixations.

The following proportion of fixation plot gives a visual depiction of eye movement events as they unfolded over time with the spoken word onset (Fig. 3). As can be seen, we considered a time line starting with the onset of the word till 1200 ms. This is a reasonable timeline in which to observe attentional bias as it develops and then continues and finally subsides. At some point in time attention towards the target object reaches a peak compared to the distractors and this gives an idea as to how effective the experimental manipulation was as well as the magnitude of activation. We have plotted proportion of fixations lines for both language directions. Also one can see the proportion of fixations towards distractors. It is obvious from the plot that with the spoken word onset, viewers started to orient their attention towards the written word which was a cohort of the translation equivalent much more than the distractors. Although there were three distractors, such attention bias was stronger towards this target word.

The table below gives quantitative data on theproportion of fixations for each time window (Table 1). Time windows were taken for each 200 ms for the entire time course. For most time windows the proportion of fixations to the target was

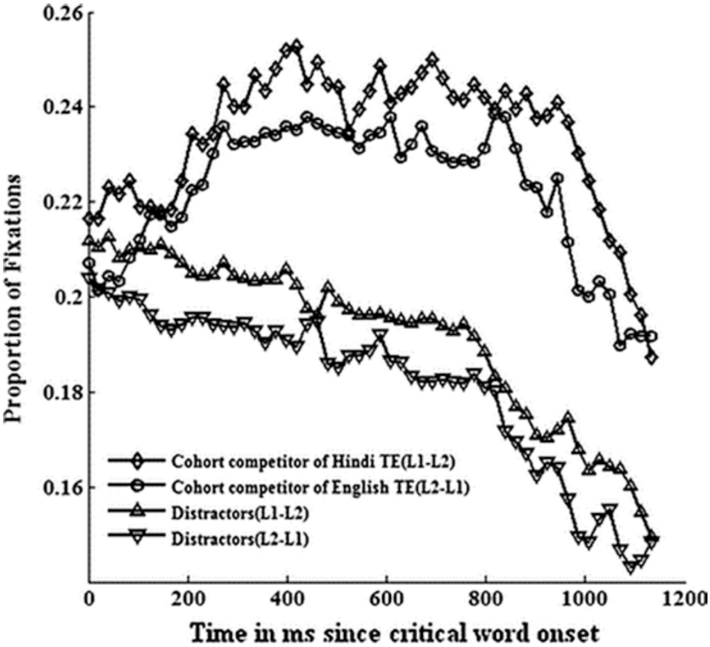


Fig. 3 Proportion of fixations to competitors and distractors when spoken word was in L1 and L2

Table 1 Comparison of mean fixation ratio for the two language conditions

	Ratio		Statistics
Bins (ms)	(L1−L2 direction)	(L2−L1 direction)	(t_1 df = 78, t_2 df = 74)
Baseline	0.49	0.49	$t_1 = 0.114$, $p = .91$; $t_2 = 0.07$, $p = .05$
200–400	0.50	0.52	$t_1 = 0.76$, $p = .44$; $t_2 = 0.71$, $p = .477$
400–600	0.52	0.54	$t_1 = 0.89$, $p = .37$; $t_2 = 0.62$, $p = .51$
600–800	0.53	0.54	$t_1 = 0.54$, $p = .58$; $t_2 = 0.19$, $p = .91$
800–1000	0.54	0.55	$t_1 = 0.35$, $p = .72$; $t_2 = 0.51$, $p = .60$

statistically significant compared to the distractors. Therefore we can conclude that when listeners were presented with spoken words in either of the languages, they could activate the translation equivalent and more importantly the orthographic representation of the cohort. Therefore they oriented their attention towards this particular word more readily and this bias continued for the entire time course. This was the first study with Hindi-English bilinguals that provided strong evidence for a language non-selective activation of orthography in bi-scriptal bilinguals while they listened to spoken words. Earlier studies were with monolinguals.

Theoretical Implications

In this chapter apart from some introduction to orthographic processing in Hindi, I also presented data from one eye tracking experiment where the issues were the interface between spoken phonology and script. The data were cross-modal and eye tracking was used in a novel paradigm. In this section, I will attempt to provide some theoretical implications and also what further could be explored. As for bilingualism, the data provided strong evidence of online translation activation for Hindi-English bilinguals in a visual world task. It showed that even highly proficient second language users activate translations when listening to spoken words in any language. We did not find much difference in the magnitude of activation between the two language directions. This suggests that language proficiency and bilingual status had no influence on orthographic activation. Such data are very important for cross-linguistic and cross-cultural comparison of bilingual data.

Although we presented full words in both English and Hindi, we had not controlled for many factors such as orthographic frequency or neighborhood size. It is well known that orthographic bigrams are very important for phonology retrieval and also for reading. Since English and Hindi are different orthographies, their bigrams are different. Importantly, if processing style is different (i.e., English is phonemic and Hindi is an Akshara), then it is not clear how phonological activation from one language and its mapping into another language with a distinct orthography works. Thus from these data, we cannot infer if such bilingual processing is restricted to orthographic processing style of one language or both at least when bilinguals have to translate. Therefore, beyond suggesting that bilingual spoken word processing is language non-selective, such data can also suggest how orthographic information is processed in the cross-modal scenario.

While some have studied how the Brahmi-derived scripts are processed using simple reaction time-based method or phonological awareness batteries, it is important to study them using techniques that offer a higher temporal resolution. The issue as to whetherthe Hindi orthography is processed as an Akshara or at times phonemically cannot be examined with a simple priming or pen and paper task. More appropriate will be to use a word reading task and eye tracking as has been done for many other languages already. For example, while we can convince ourselves that in the word ‘kitab’ readers process the vowel later after the consonant although visually it appears first, we need to also know whether readers look at them in that order while they derive the phonological representation from the orthography.

It may well turn out that fluent readers may not even fixate on such constituents yet derive proper phonological representations and meaning. The study that I presented did not involve visual word decoding. However, the method is powerful enough to settle many issues that currently researchers have been raising on the Indian scripts. Another most important implication is in the area of reading acquisition in children. Particularly, we know relatively little about how bilingual children acquire such distinct orthographic systems such as Hindi and English who become bi-literate. Little is known how eye movement control develops in these two scripts as they adopt and learn both. These questions have not been examined in these scripts much. While the theoretical debate on the issue of Akshara as a unit currently has legitimate empirical support, for such claims to be fully accepted, we need to make use of multiple methodologies.

In sum, the chapter discussed data where spoken word phonology activated orthography. One limitation of the paradigm could be that written words were presented as such and this may have helped the participants to orient their eyes towards the word which was in some way related to the spoken words. Even then given the difference between Hindi and English at the orthographic level, the data strongly indicate some kind of orthographic convergence in such fluent bi-literates. They may be activating the orthographic representation of words in the other language. This means, even if one of the orthographies is Akshara based and another alphabetic, they are activated together in the bilingual mind. These issues should be explored more in different populations in order to understand how orthography is processed during language processing.

References

Abu-Rabia, S., & Taha, H. (2005). Reading in Arabic orthography. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy*. Amsterdam: Lawrence Erlbaum Associates.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.

Barr, D. J. (2008). Analyzing 'visual world'eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*(4), 457–474.

Bhatia, D., Prasad, S. G., Sake, K., & Mishra, R. K. (2017). Task irrelevant external cues can influence language selection in voluntary object naming: Evidence from Hindi-English bilinguals. *PLoS One, 12*(1), e0169284.

Chéreau, C., Gaskell, M. G., & Dumay, N. (2007). Reading spoken words: Orthographic effects in auditory priming. *Cognition, 102*(3), 341–360.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area. *Brain, 123*(2), 291–307.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84–107.

Das, T., Kumar, U., Bapi, R. S., Padakannaya, P., & Singh, N. C. (2009). Neural representation of an alphasyllabary-The story of Devanagari. *Current Science, 97*(7), 1033–1038.

De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language, 80*(3), 617–626.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Daal, V. H., Polyzoe, N., et al. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly, 39*(4), 438–468.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., & Mascetti, G. G. (2000). Visual-spatial attention in developmental dyslexia. *Cortex, 36*(1), 109–123.

Guo, T., & Peng, D. (2006). Event-related potential evidence for parallel activation of two languages in bilingual speech production. *NeuroReport, 17*(17), 1757–1760.

Gupta, A. (2004). Reading difficulties of Hindi-speaking children with developmental dyslexia. *Reading and Writing, 17*(1–2), 79–99.

Huang, X., Jing, J., Zou, X. B., Wang, M. L., Li, X. H., & Lin, A. H. (2008). Eye movements characteristics of Chinese dyslexic children in picture searching. *Chinese Medical Journal, 121*(17), 1617–1621.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience, 31*(1), 80–93.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482.

Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica, 137*(2), 138–150.

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia, 10*(3), 196–214.

Karanth, P. (1992). Developmental dyslexia in bilingual-biliterates. *Reading and Writing, 4*(3), 297–306.

Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology, 94*, 67–84.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science, 30*(3), 481–529.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*(2), 149–174.

Leong, C. K. (1986). What does accessing a morphemic script tell us about reading and reading disorders in an alphabetic script? *Annals of Dyslexia, 36*(1), 82–102.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*(02), 97–115.

Miles, E. (2000). Dyslexia may show a different face in different languages. *Dyslexia, 6*(3), 193–201.

Mishra, R. K. (2009). Interface of language and visual attention: Evidence from production and comprehension. *Progress in Brain Research, 176*, 277–292.

Mishra, R. K., Huettig, F., & Olivers, C. N. (2012). Automaticity and conscious decisions during language-mediated eye gaze in the visual world. *Progress in Brain Research*, 212.

Mishra, R. K., & Singh, N. (2014). Language non-selective activation of orthography during spoken word processing in Hindi–English sequential bilinguals: an eye tracking visual world study. *Reading and Writing, 27*(1), 129–151.

Mishra, R. K., & Singh, N. (2016). The influence of second language proficiency on bilingual parallel language activation in Hindi–English bilinguals. *Journal of Cognitive Psychology, 28*(4), 396–411.

Nag, S. (2011). The akshara languages: What do they tell us about children's literacy learning? In *Language-cognition: State of the art* (pp. 291–310). Munich, Germany: Lincom Europa.

Nehru, R. (2001). *Distorted grapheme representation: A new hypothesis to explain dyslexic reading errors*. Poster presented at the Fifth British Dyslexia Association international conference, York, UK.

Padakannaya, P., Pandey, A., Saligram, D., & Ranga Rao, S. (2016). Visual-orthographic complexity of Akshara and eye movements in reading: A study in Kannada alphasyllabary. *Writing Systems Research, 8*(1), 32–43.

Padakannaya, P., Pandey, A., Saligram, D., & Ranga Rao, S. (2016). Visual-orthographic complexity of Akshara and eye movements in reading: A study in Kannada alphasyllabary. *Writing Systems Research, 8*(1), 32–43.
Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain Research, 1188*, 132–138.
Purushothama, G. (1990). Reading vowels in Kannada script. *Journal of Learning Disabilities, 23*(3), 198–200.
Ramaa, S. (2000). Two decades of research on learning disabilities in India. *Dyslexia, 6*(4), 268–283.
Rao, C., & Singh, N. C. (2015). Visuospatial complexity modulates reading in the brain. *Brain and Language, 141*, 50–61.
Rao, C., Vaid, J., & Chen, H. C. (2017). The processing cost for reading misaligned words is script-specific: Evidence from Hindi and Kannada/Hindi readers. *Journal of Cultural Cognitive Science, 1*(1), 39–48.
Rao, C., Vaid, J., Srinivasan, N., & Chen, H. C. (2011). Orthographic characteristics speed Hindi word naming but slow Urdu naming: Evidence from Hindi/Urdu biliterates. *Reading and Writing, 24*(6), 679–695.
Rayner, K. (1985). Do faulty eye movements cause dyslexia? *Developmental Neuropsychology, 1*(1), 3–15.
Roychoudhuri, K. S., Prasad, S. G., & Mishra, R. K. (2016). Iconic native culture cues inhibit second language production in a non-immigrant population: Evidence from Bengali-English bilinguals. *Frontiers in Psychology, 7*, 1516.
Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1108.
Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 546.
Singh, N., & Mishra, R. K. (2015). Unintentional activation of translation equivalents in bilinguals leads to attention capture in a cross-modal visual task. *PLoS One, 10*(3), e0120131.
Spivey, M. (2008). *The continuity of mind*. Oxford, UK: Oxford University Press.
Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281–284.
Sunderman, G. L., & Priya, K. (2012). Translation recognition in highly proficient Hindi–English bilinguals: The influence of different scripts but connectable phonologies. *Language and Cognitive Processes, 27*(9), 1265–1285.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.
Vaid, J., & Gupta, A. (2002). Exploring word recognition in semi-alphabetic script: The case of Devanagari. *Brain and Language, 81*, 679–690.
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*(1), 1–25.
Weekes, B. S. (2005). Acquired disorders of reading and writing: Cross-script comparisons. *Behavioural Neurology, 16*(2–3), 51–57.

Language and Literacy Practices that Influence Bilingual and Bi-Literate Acquisition in L1 Kannada and L2 English in Bangalore, India

Sunaina Shenoy and Richard K. Wagner

Abstract This study was conducted in Bangalore, India, to measure the influence of language and literacy practices on students' bilingual and bi-literate competencies in L1 Kannada and L2 English. Shenoy S (Psychol Stud 61(3):126–136, 2016) modified the Clinical Evaluation of Language Fundamentals (CELF5) screening tool in English and Kannada to provide teachers with an opportunity to better understand the language profiles of their students. Based on this screening tool, more than 50% of the students from a low-income school were identified as being below criterion on both their L1 and L2, compared to 13% of students from a high-income school. To identify the language and literacy practices that could account for bilingual test scores, informal measures were developed to collect data from parents and teachers. Ninety-three parents completed a questionnaire on language and literacy practices at home, and regression analyses depicted that students' dominant home language explained their bilingual test scores in both low-SES and high-SES schools. Ten teachers were interviewed, and qualitative analyses depicted that teachers from low-income schools code-switched between English and Kannada approximately 50% of the time compared to teachers from high-income schools, who followed a 100% English-immersion model.

Keywords Bangalore · Bilingualism · Bi-literacy · English immersion · Home language · L2 English · L1 Kannada · Socio-economic status

This chapter addresses a complex intersection of language and literacy practices at home and school that influence students' bilingual and bi-literate competencies in

S. Shenoy (✉)
Department of Psychology and Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

Department of Special Education, University of New Mexico, Albuquerque, NM, USA
e-mail: shenoy@unm.edu

R. K. Wagner
Department of Psychology and Florida Center for Reading Research, Florida State University, Tallahassee, FL, USA

R. M. Joshi, C. McBride (eds.), *Handbook of Literacy in Akshara Orthography*,
Literacy Studies 17, https://doi.org/10.1007/978-3-030-05977-4_19

their first language Kannada and their second language English. It was set in the context of schools in Bangalore, India that serve students whose first language is Kannada, but follow an English immersion program. Their home language and literacy practices also vary as a result of socio-economic status, with some students having more continuity between home and school languages than others. Navigating two language systems is a difficult proposition for most students, but it becomes more trying when students are not afforded bilingual programs and/or phonics instruction, and/or have limited literacy support at home. The purpose of this chapter is to consider literacy acquisition in two languages through the lens of the larger ecological systems at play that strongly influence the level of competence in both languages.

Effect of Bilingualism on Developing Bi-Literate Competencies

Reyes and Azuara (2008) defined bi-literacy as the process of developing reading and writing competencies in two languages. It is important to distinguish between sequential and simultaneous bilinguals. Cummins (1995) observed that children who represent the former group typically learn one language at home and one in school, and children who represent the latter group learn two languages concurrently, usually before age 3. Moreover, although sequential bilinguals have the added advantage of linguistic transfer of skills from their first language (L1) to a second language (L2), simultaneous bilinguals cannot use L1 to influence L2 because both languages are learned together and support each other (Cummins, 1995). Reyes and Azuara (2008) found that children who are bilingual develop their own theories and concepts of language and literacy from an early age through conversations and contextual cues. Thus, the socio-cultural model functions as a key to literacy, but students need continuous support in both languages from caregivers to become bi-literate (Reyes & Azuara, 2008). Based on this process, Collier and Thomas (1989) observed that bilingual and bi-literacy programs are based on two assumptions: (a) reading is a meaning-constructing process and needs to be in a language that the child understands, and (b) literacy skills acquired in the first language transfer to the second language.

Another common question is related to the likelihood of language interference when children are taught two languages. Cummins (1991) developed the *common underlying proficiency hypothesis*, which states that there are two kinds of literacy skills, namely, foundational level literacy skills (involving intentionality of print, letter-sound correspondences, conventions of print) and surface-level literacy skills (involving relations between two languages and the larger literacy context). He stressed that linguistic transfer is not restricted to linguistic abilities alone but also impacts academic skills, literacy development, subject knowledge, and learning skills that can transfer from L1 to L2. For example, if a child is exposed to both

English and Chinese, the surface-level skills such as different writing systems in both languages are understood as two separate systems in the brain, and one does not affect the other. However, foundational level skills such as literacy development in one language have a positive effect on the other, such that, for example, the child's interactions with reading and writing in both English and Chinese serve to develop basic understanding of literacy. Collier and Thomas (1989) found cognitive facilitation, rather than interference, for bilinguals, which can be parlayed into long-term academic success.

Bi-Literacy Acquisition in Kannada vs. English

Kannada is a Dravidian language spoken by roughly 50 million people in Karnataka in Southern India and is characterized by Devanagiri script, which is used as an example in this chapter to illuminate how literacy acquisition might vary for speakers of Kannada compared to speakers of English. While alphabetic scripts, such as English, have distinct graphemic elements that correspond to individual phonemes, syllabaries have graphic units that correspond to syllables. Alpha-syllabaries, like Korean Hangul, are a combination of both as their primary graphic unit is a syllable, and they have distinct subunits that correspond to phonemes (Daniels, 2000). However, Devanagari derived orthographies may belong to a different category more similar to Abugida.

Kannada is a diglossic language with the formal variety of the language differing from the colloquial one (Karanth, 2006). It has 43 phonemes, which are represented by 50 syllables in the kagunitha or syllable matrix (Karanth, 2006). The letter-graphemes or *aksharas* do not have separate names; their names reflect the sounds they represent, with an almost one-to-one correspondence between sound and *akshara.* According to Nishanimut, Johnston, Joshi, Thomas, and Padakannaya (2013), Kannada *aksharas* can therefore be used as an intermediary stage for learning English letter sounds, and form a compelling reason for transitioning from the home language to English instruction in school. It is important to note, however, that Kannada is taught using an alphabetic approach which involves a process of rote memorizing the individual akshara to sound correspondences, though learning to read the syllables represented by the *aksharas* and blending sounds might be a better approach (Nishanimut et al., 2013). Similarly, English in schools is taught using an alphabet-spelling method, which involves rote memorizing the spellings of words (Gupta, 2014), though a systematic phonics approach is shown to be more effective (National Reading Panel, National Institute of Child Health and Human Development (NICHD), 2000).

How do the approaches to teaching Kannada and English affect the identification of reading disabilities? The prevalence of reading disabilities (RD) in India is a complex problem that is complicated by a multitude of variables, such as income and resources of schools, home language backgrounds, fluency in first and second languages, and English immersion programs. On the one hand, there is a lack of awareness among teachers, and on the other, there is an underlying belief that RD is largely a by-product of children navigating deep orthographies compared to transparent ones like Kannada (Karanth, 2006). The initial identification of RD among Indian children appears to be more common to urban schools where English is the medium of instruction. According to Ramaa, Miles, and Lalithamma (1993), 10.36% of students from a sample of Kannada speakers were identified as poor readers and 2.51% were identified as having a reading disability, which is lower than what typically is reported in Western countries.

English Language Instruction in Indian Schools

Most private schools in India follow an English immersion model with limited or no bilingual support. One reason for the increase in private school education is the poor quality of education afforded by the government schools. According to the Annual Status of Education Report, India (2012), 80% of Indian schools are government schools, but because of the poor quality of education, 27% of Indian children are privately educated. Another reason for private schooling is the belief among Indian parents that if they can afford it, private education programs are a better option for their children compared to government education programs. This is especially true in urban centers, where more than 50% of children (27 million) attend private schools (Annual Status of Education Report, India, 2012). The private schools typically follow an international, national, or state-level standardized curriculum, and the medium of instruction in these schools is usually English (Kurrien, 2005). In contrast, government schools typically follow a state-level curriculum and the medium of instruction is usually in the State language, even though English and one other language is introduced to meet the three-language formula of the central government.

English is an integral part of the education system in India because it is one of the two official languages of the country, along with Hindi (National Council of Educational Research and Training, 2011). Given the linguistic diversity of the country and numerous languages that are spoken in different states, English is also used as the national mode of communication and the unifying link language (National Council of Educational Research and Training, 2011). It is the language of economics and business, and is viewed as a requirement for economic and social mobility (Ramanathan & Bruning, 2003). It is estimated that 90 million children in India are currently being formally schooled in English (Kalia, 2007). According to Ramanathan and Atkinson (1999), a key assumption has been that the inner circle of countries (Britain, US, Canada and Australia) with native speakers of the language sets English standards for countries in the outer circle (e.g., India, and parts of Africa), where English is used non-natively but extensively and has been given official language status. Unfortunately, English and the privileges associated with it remain inaccessible to those who are from a lower socioeconomic population in

India, with the Indian middle-class assuming a position of power through its access to English (p. 212).

Schools in India follow a three-language formula (Aggarwal, 1991) that is endorsed by the National Curriculum Framework 2005 (Ramachandran et al., 2005). The first language is the medium of instruction; the second is required to be taught at least by Grade 5, and the third between Grades 6–10. Hindi and English have to be introduced as two of these three languages (Saini, 2000), and the third one is typically the state language. Nag-Arulmani (2000) found that English-teaching in India is dependent on: (a) the teacher's English language proficiency, and (b) the students' exposure to English outside of school. Kurrien (2005) identified four types of schools:

(a) English-medium middle-high cost private schools, where teachers are proficient in English, but students have varying levels of exposure to English in their environment; including as a home or first language;
(b) English- medium low cost private schools, where teachers have limited proficiency in English and students have limited exposure to English in their environment, but parents aspire towards upward mobility via English;
(c) Government-aided regional schools, where teachers use English and other regional languages, with students from a variety of backgrounds; and
(d) Government regional schools run by district and municipal education authorities, where teachers are least proficient in English, and students have the least exposure to English in their environments.

There could be wide variations within these schools in terms of learning opportunities, number of students in the classroom, class libraries, culture, management, which eventually lead to varying levels of oral-language and reading proficiency of these students (Nag-Arulmani, 2003).

Language and Literacy Practices in Indian Homes

Parenting practices are the strongest predictor of early childhood literacy skills (Early Child Care Research Network, 2003). The most common home literacy practice, shared book reading, provides a rich source of information and opportunity for children to learn context-specific language (Pellegrini, Brody, & Sigel, 1985). Print exposure and shared book reading foster vocabulary development in pre-school aged children (Cunningham & Stanovich, 1991; Senechal & Cornell, 1993). While the above studies focused on monolingual children, more recent studies have found similar results with bilingual children as well. Chow, McBride-Chang, and Cheung (2010) observed that dialogic reading enhanced phonological awareness in both English and Chinese.

Kalia (2007) conducted a study in two pre-schools in Bangalore, and found that book reading at home was associated with bilingual children's oral language, narrative and literacy development in L2 English. Some of the languages parents spoke at home were Kannada, Malayalam, Urdu, Hindi, Tamil, Telugu, Bangla, Gujarati and Oriya. These students came from middle-income backgrounds and their parents were high school or college graduates who read to them in English. Parents' report of frequency of library visits was significantly associated with children's scores for complex syntax, and their receptive vocabulary scores were significantly correlated with their scores on concepts of print, blending, elision and the complexity of narrative they produced (Kalia, 2007). Though this is true for students from middle-high income backgrounds in India, we could not find a study that looked at literacy practices in low-income homes. Thus, the present study offers some insight into that area, as well as how it corresponds to more students from low-income homes being identified as having learning disabilities.

Sreekanth (2011) conducted a study to measure parents' involvement in the education of their children. He reported that most of the parents in his sample were committed to attending parent-teacher meetings, supported their children with their homework and were proactive with the teachers to focus on the welfare of their children. But apart from expected roles, parents usually do not question the authority of the teachers and do not have an understanding of alternative education programs, teaching and learning styles beyond what the school expects of them. Moreover, most parents are determined to have their children succeed on school-based exams and believe that the primary goal of education is focused on academic content knowledge.

Context of Present Study and Research Questions

Linguistic Context

All students in the sample spoke Kannada at home and their schools followed an English immersion model. The students' fluency rates in English varied from beginning and early intermediate to advanced depending on the level of English language support they had at home.

Language Support at School

In both English-medium middle-high income private schools and low-income private schools in Bangalore, India, the first language was English, the second language was Hindi and the third was the state language, which in the case of this study was Kannada. These schools followed an English-immersion model with no bilingual support, and second and third languages are taught much like Spanish is taught in the US.

Study Aims and Research Questions

This study focused on language acquisition in both Kannada and English as a precursor to literacy, and was measured across low and middle-high-income schools. It had three over-arching goals: (a) to examine language and literacy practices at home that impacted second language acquisition in school, (b) to examine language markers in L1 and L2 that could potentially be part of a screener for later reading disabilities, and (c) to examine the efficacy of English immersion programs in low-income and middle-high-income schools.

The following were the specific research questions that were explored in the study.

1. How do the language and literacy practices at home contribute to students' L2 English acquisition in low-income vs. middle-high income schools?
2. How do the indicators of early language problems vary as a function of assessing L1 Kannada vs. L2 English?
3. Are English immersion programs effective in both low-income and middle-high income schools?

Method

Participants

Students The sample for the current study consisted of 104 students: Sixty-four students attended a low-income private school, and 40 students attended a middle-high income private school. Of these students, 62 were female and 42 were male. The students were selected to participate in the study if their home language was Kannada (which is primarily spoken in the state of Karnataka, India), and their medium of instruction in school was English. All of them were from grades 2-5 and were between 7 and 10 years old. The school sites were located in a large urban city, Bangalore, in the state of Karnataka in South India. All the programs followed an English immersion model with no L1 bilingual support.

Parents Of the 104 parents who consented to their children being part of the bilingual assessment study, 93 of them also completed a parent questionnaire. Fifty-four of these parents came from a low-income background and 39 came from a middle-high income background. They identified as belonging to the Dravidian linguistic group, and were either dominant in Kannada or bilingual in both English and Kannada.

Teachers Ten teachers, five from the low-income school and five from the high-income school participated in teacher interviews. All the teachers were female, working with students in Grades 2–5, following an English-immersion model in their respective schools.

Measures

Modified CELF-5 Screening Tool in English and Kannada An adaptation of The *Clinical Evaluation of Language Fundamentals 5 Screening Test* (Semel, Wiig, & Secord, 2013) was utilized for the study. The tool was adapted into Kannada and English, contents referring to India. It consisted of the following subtests: word structure, word classes, following directions, sentence recall, sentence assembly, and semantic relationships. These items were developed to assess language skills that have been shown to be problematic for and/or indicative of individuals with language disorders. It was rendered culturally appropriate for students in an Indian context by ensuring that the language used and picture prompts were grounded in artifacts and experiences that are relevant to and typical of the culture of the region. Some of the items were changed from American English to reflect Indian English use, and the picture prompts were changed to be more context-specific, but they still tested the same language skill. For example, item 8: subjective pronoun, under the word structure subtest, had the words "hot dog" and "hamburger," which were changed to "sandwich" and "burger" which are more familiar terms in Indian English. Similarly, in the case of item 10, under the word classes subtest, the word "marker" was replaced with "sketch pen," again a term that the students would be more familiar with. The only item that had to be dropped was question 37, which uses the phrase "A quarter past three" because that is not a common way in which time is expressed either in Indian English or Kannada. All other items on the test remained the same and were translated the same way into Kannada.

A pilot study was conducted to develop age and grade appropriate items as well as to establish criterion scores. The researchers consulted with a panel of three bilingual psychologists and five bilingual teachers who rendered both versions of the tool age and grade appropriate. The tests were then administered on a sample of 50 students in Bangalore aged 7–10 years (Grades 2–5) and averages were established for each age group and for both Kannada and English versions of the test. These scores served as criterion scores to determine if students were at/above criterion or below criterion on both tests. It was further piloted on a group of 10 students in the non-clinical population and a group of 8 students who were previously identified as having a learning disability. Ninety percent of the former group were identified as "above criterion in L1 or L2 tests" and 100% of the students previously identified as having a learning disability were identified as "below criterion on both L1 and L2 tests" and in need of further language assessment.

Reliability and Validity Cronbach's alphas for the six subtests were calculated to measure internal consistency for the Indian English and Kannada versions of the test. For the Indian English test, the word structure subtest consisted of 9 items

(α = .72), the word classes subtest consisted of 5 items (α = .76), the following directions subtest consisted of 5 items (α = .72), the sentence recall subtest consisted of 7 items (α = .75), the sentence assembly subtest consisted of 6 items (α = .82) and the semantic relationships subtest consisted of 7 items (α = .77). For the Kannada test, the word structure subtest consisted of 9 items (α = .78), the word classes subtest consisted of 5 items (α = .86), the following directions subtest consisted of 5 items (α = .83), the sentence recall subtest consisted of 7 items (α = .69), the sentence assembly subtest consisted of 6 items (α = .77) and the semantic relationships subtest consisted of 7 items (α = .73). Moreover, the content validities of both versions of the test were rendered grade and age-appropriate by a panel of three bilingual psychologists and five bilingual teachers.

The test is not a diagnostic tool designed to provide an in-depth diagnosis of speech/language disability or the degree of impairment of speech or language abilities. Rather, it is used to identify students who are "at risk" for a language disorder, and need to be referred for further language assessment. It helps in measuring whether the students' language abilities appear to be adequate for their age. The total score attained by the student is compared to a research-based criterion score appropriate for the student's age and certain recommendations are made. Typically, these recommendations include conducting a diagnostic test and conducting informal assessments, such as teacher and parent interviews as well as classroom observations.

Parent Questionnaire This questionnaire served as a home language survey, and consisted of 10 questions targeted at getting more information on the demographics of the family, language of dominance at home, and home literacy practices in both L1 and L2. It was developed in both English and Kannada in order to give parents the option of answering in their dominant language. The researchers believe that this was one of the main reasons the parents felt comfortable answering the questionnaire and consenting to their children being part of the study. All forms sent home from school are usually in English only, limiting access in their L1. Since our forms were in L1 and L2, we had 93/104 (90%) parents send back their completed questionnaires, improving our retention rate. The goal was to get more information on language and literacy practices at home that could explain a variance in test scores between students from low income and middle-high income backgrounds.

Teacher Interview This assessment consisted of 24 questions in total. These included 8 questions regarding demographic information such as gender, age, educational qualifications, number of years teaching and number of students in classrooms; 5 questions regarding dominant language use; 3 open-ended questions on culture of the school and teaching style; and 8 open-ended questions regarding the special education referral process, resources and accommodations used in classrooms to address the needs of students with disabilities. Specific questions targeted pedagogical practices and the current policies in place to address concerns regarding second language acquisition in the classroom. This tool was developed and administered in English because all the schools visited were schools that followed an English-immersion model and the medium of instruction was English. For the purposes of this chapter, only the questions regarding dominant language use in the classroom were coded and will be reported in the following section.

Results

Modified CELF 5 Screening Test in English and Kannada

The Modified CELF-5 Screening Test in English and Kannada was used to establish a classification system of "At/Above criterion" and "Below criterion" on both L1 and L2. Table 1 presents the classification of students across low-income and middle-high income schools. Whereas 18 students (28.13%) were above criterion on the L1 Kannada test and below criterion on the L2 English test in the low-income school, 28 students (70%) were above criterion on the L2 English test and below criterion on the L1 Kannada test in the middle-high income school. Although the majority of students from the middle-high income schools were either proficient in L2 English or bilingual in English and Kannada, the majority of their peers in the low-income school were either below criterion on both L1 and L2 tests or proficient in L2 Kannada. Moreover, 5 students (12.5%) from the middle-high income school were below criterion on both L1 and L2 tests, compared to 34 students (53.12%) of the sample from the low-income school, which is quite an alarming statistic.

Specific Differences in Subtest Scores Between Kannada and English Looking across the two languages, average percentages of correct responses on the subtest scores on the Modified CELF 5 Screening Test were recorded. Table 2 represents these percentage scores. The word structure subtest measured grammatical components of the language like plurals; word classes measured categorizing words that were semantically similar; following directions was aimed at listening comprehension; sentence recall focused on short-term memory; sentence assembly measured re-organizing jumbled sentences to make a coherent sentence, and semantic relationships measured understanding connecting words like between, after. There was

Table 1 Classification of students on L1 Kannada and L2 English scores on the modified CELF-5 screening test

	At/Above criterion on both tests	At/Above criterion on Kannada test; Below criterion on English test	At/Above criterion on English test; Below criterion on Kannada test	Below criterion on both tests
Low-income school N = 64	10 (15.63%)	18 (28.13%)	2 (3.12%)	34 (53.12%)
Middle-high income school N = 40	7 (17.50%)	0	28 (70%)	5 (12.5%)

Table 2 Percentage of correct responses on subtest scores on L1 Kannada and L2 English on the modified CELF 5 screening test

	Low-income school (N = 64)		Middle-high income school (N = 40)	
	English	Kannada	English	Kannada
Word structure	54.11	67.33	83.55	62.33
Word classes (Age 5–8)	52.60	59.20	66.66	59.00
Following directions	30.20	47.88	50.40	45.40
Sentence recall	31.42	44.64	67.85	24.57
Sentence assembly	34.94	32.66	73.66	24.50
Semantic relationships	40.00	48.57	75.85	55.57
Word classes (Age 9–21)	26.00	40.00	48.40	35.60

an overlap in Kannada and English scores, with students from both low-income and middle-high income schools doing poorly on the word classes subtest in English and the sentence assembly subtest in Kannada. The best scores were on the word structure subtest in both English and Kannada across the two schools.

Dominant Home Language Use

In terms of dominant home language, 51 parents (94.44%) from the low-income sample identified Kannada as their dominant language for speaking, listening, reading, and writing activities at home, and 56.48% of family members reported that they had trouble understanding and speaking English. In contrast to this, 28 parents (71.80%) from the middle-high income sample identified English as their dominant home language. They mentioned that although they were fluent in both languages while speaking and listening, their dominant language for reading and writing was English. While some parents mentioned that they had trouble understanding and speaking Kannada (28.80%), some said they had the same trouble with English (20.02%).

Middle-High Income Homes Multiple regression analysis was used to test if the dominant home language use in terms of speaking, listening, reading and writing, significantly explained students' bilingual test scores. The results of the regression indicated that two variables explained 29.06% of the variance ($R^2 = .29$, $F(7, 31) = 1.81$, $p < .05$). Speaking both English and Kannada at home significantly accounted for students' Kannada test scores ($\beta = -6.28$, $p < .05$) and students' Kannada scores significantly accounted for their English scores ($\beta = 0.43$, $p < .05$). Students who came from bilingual homes did significantly better on both English and Kannada versions of the test.

Low-Income Homes Multiple regression analysis was used to test if the dominant home language use in terms of speaking, listening, reading and writing, significantly explained students' bilingual test scores. The results of the regression on English scores indicated that three variables explained 27.39% of the variance ($R^2 = .27$, $F(7, 46) = 2.48$, $p < .05$). Speaking ($\beta = 3.44$, $p < .05$), listening ($\beta = -3.72$, $p < .05$) and reading ($\beta = -5.61$, $p < .05$) in English at home significantly accounted for students' English test scores. The results of the regression on Kannada scores indicated that one variable explained 15% of the variance ($R^2 = .15$, $F(7, 46) = 1.21$, $p < .05$). Listening in Kannada at home significantly accounted for students' Kannada test scores ($\beta = 9.02$, $p < .05$). Thus, exposure to English at home in domains of speaking, listening and reading had a significant effect on improving students' English scores on the bilingual test. Since a majority of parents mentioned that their dominant language at home was Kannada, this could explain some of the variance in scores between low-cost and middle-high cost schools.

Reading Frequency at Home

Figure 1 illustrates the frequency of reading to children in low-income versus middle-high income homes. A majority of 30 parents (55.55%) from the low-income schools reported that they read to their children in Kannada approximately once a month at home. In contrast to this, a majority of 26 parents (66.66%) from the middle-high income schools reported that they read to their children in English at least once a week at home. There was no significant effect of reading frequency on students' bilingual test scores across both income categories.

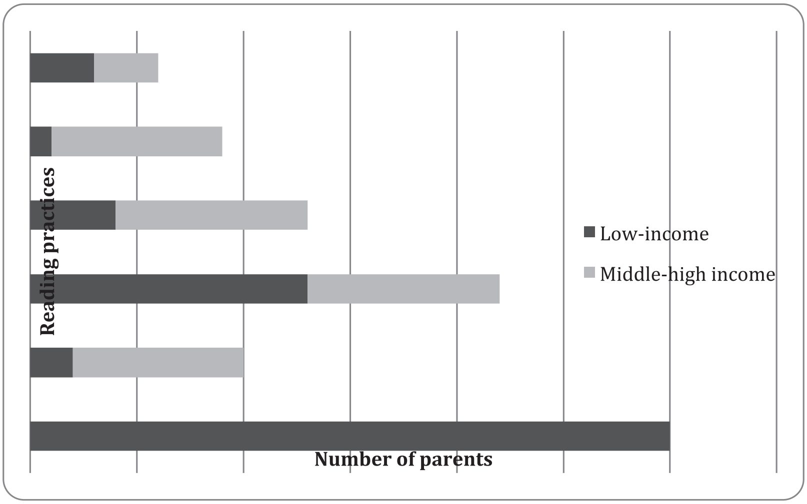


Fig. 1 Bar graph comparing the frequency of reading to children in low-income versus middle-high income homes

Dominant Language Practices in School

The dominant languages of instruction in the low-income classrooms were both English and Kannada, whereas the dominant language of instruction in the middle-high income classrooms was only English. In the low-income school, all the teachers agreed that students in the lower grades between K-5 would benefit from Kannada instruction, but they followed an English-immersion model prescribed by the school board and alternated between the two languages in their classrooms, especially to explain harder concepts. One second-grade teacher said, “They (the students) need English in the future, so it is better to start young. But I use English 50% of the time and Kannada 50% of the time in my classroom so students understand.”

In contrast to this, teachers in the middle-high income schools agreed that their students would not benefit from native-language instruction especially since they came from different home language backgrounds and English served as a link language. One fourth-grade teacher responded, “They are from Bangalore, so they are fluent in English!” Another teacher said, “Most students start English medium from Kindergarten, so they pick it up along the way.”

Discussion

The Modified CELF 5 in English and Kannada was used to develop a classification system to measure language proficiency in both languages as well as linguistic markers that could potentially lead to later literacy problems. In the middle-high income school, 12.5% of students were classified “below criterion on both L1 and L2” which suggests that they might be at-risk for literacy acquisition later on. This could be a more plausible statistic in identifying and diagnosing learning disabilities later on. In comparison, 53.12% of students from the low-income school were identified as “below criterion on both L1 and L2,” which suggests that parsing out language differences from disorders in the low-income school becomes a harder task, and it is imperative to look at other factors such as home language and literacy practices, school instructional practices, and other skills to rule out the exclusionary clauses before students are recommended for a complete diagnostic evaluation.

Parents whose children were enrolled in low-income schools mentioned that on average they (a) had high school degrees, (b) earned below 10,000 Rupees per month, (c) spoke Kannada as their home language, and (d) typically read to their children once a month or less. In contrast, parents whose children were enrolled in middle-high income schools mentioned that on an average they (a) had college degrees, (b) earned between 31,000 and 70,000 Rupees a month, (c) spoke English as the dominant home language, and (d) typically read to their children at least once a week. Regression analyses were conducted on bilingual test scores in English and Kannada and two home environment factors, namely dominant home language and reading frequency. While dominant home language use accounted for students' bilingual test scores in both low-income and middle-high income schools, reading frequency was not. Socio-economic status (SES) played a significant role in home language and literacy practices across low-income and middle-high income populations. This in turn significantly explained scores on the English (L2) assessment, and could serve to explain why mastery of a first language may be required to learn the second language.

Another important finding from the study was that it was not enough to test students in their L1 and L2, but there is a need to ground assessment tools within a cultural context. For example, the students from the low-income school did not have any exposure to English outside of school, whereas the students from middle-high income schools were exposed to language and literacy activities in both English and Kannada at home. This was reflected in their performance scores, with students from low-income schools demonstrating overall low scores on both language assessments, probably because of the nature of their distinct language acquisition patterns across home and school contexts. Compared to this, students from middle-high income schools demonstrated overall higher scores on both language assessments, especially with regard to L2 English. Cummins (1991) proposed a *Developmental Interdependence Hypothesis*, which states that proficiency in L1 is required to develop proficiency in L2 and a common underlying proficiency between L1 and L2 facilitates the transfer of cognitive skills in addition to linguistic skills. The reading and writing experiences established in one language are passed on to the second language as children navigate between the two to understand literacy development. The development of bilingualism seems to have a positive linguistic and cognitive effect for most children. They develop meta-linguistic awareness in both languages, seamlessly navigate two language and literacy environments, and acquire two sets of vocabulary words and an expanded lexicon to describe everything in their learning contexts. Tapping on heritage language resources helps bilingual students value their L1 in relation to their instructional L2, helps parents of bilingual students be more involved in school-based projects, and promotes cross-linguistic transfer between the two languages.

Qualitative analyses of the teacher interviews emphasized the differences in language of instruction between low-income and middle-high income schools; whereas teachers in the former school alternated between English and Kannada, teachers in the latter schools used English only. While an English immersion program has been effective with students from a higher socio-economic backgrounds where they have more exposure to their L2 English both at home and school (Kurrien, 2005), it has not proved to be the best option with students from lower socio-economic backgrounds where there is limited or less proficient exposure to their L2 in school and no English L2 support at home. This study makes a case for examining the efficacy of English immersion programs in low-income schools, and recommends a bilingual program instead.

In conclusion, the purpose of this study was to expose the language and literacy practices that influenced students' bilingual and bi-literate competencies in L1

Kannada and L2 English. Results indicated that the dominant language practices at home and school contributed to students from low-income backgrounds losing out on both their heritage language, Kannada and their language of instruction, English.

References

Aggarwal, S. (1991). *Three language formula: An educational problem*. New Delhi, India: Gyan Publishing House.

Annual status of education report (rural). (2012). *ASER centre, New Delhi*.

Chow, B. W. Y., McBride-Chang, C., & Cheung, H. (2010). Parent–child reading in English as a second language: Effects on language and literacy development of Chinese kindergarteners. *Journal of Research in Reading, 33*(3), 284–301.

Collier, V. P., & Thomas, W. P. (1989). How quickly can immigrants become proficient in school English. *Journal of Educational Issues of Language Minority Students, 5*(1), 26–38.

Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. In *Language Processing in Bilingual Children* (pp. 70–89). Cambridge, UK: Cambridge University Press.

Cummins, J. (1995). Discursive power in educational policy and practice for culturally diverse students. In *Discourse and power in educational organizations* (pp. 191–209). Cresskill, NJ: Hampton Press.

Cunningham, A. E., & Stanovich, K. E. (1991). Tracking the unique effects of print exposure in children: Associations with vocabulary, general knowledge, and spelling. *Journal of Educational Psychology, 83*(2), 264.

Daniels, P. T. (2000). On writing syllables: Three episodes of script transfer. *Studies in the Linguistic Sciences, 30*(1), 73–86.

Early Child Care Research Network. (2003). Does amount of time spent in child care predict socioemotional adjustment during the transition to kindergarten. *Child Development, 74*(4), 976–1005.

Gupta, R. (2014). Change in teaching practices: Case of phonics instruction in India. *Procedia-Social and Behavioral Sciences, 116*, 3911–3915.

Kalia, V. (2007). Assessing the role of book reading practices in Indian bilingual children's English language and literacy development. *Early Childhood Education Journal, 35*(2), 149–153.

Karanth, P. (2006). The Kagunita of Kannada—Learning to read and write an Indian alphasyllabary. In *Handbook of orthography and literacy* (pp. 389–404). Mahwah, NJ: L. Erlbaum Associates.

Kurrien, J. (2005). Notes for the meeting of the national focus group on teaching of English, and note on tntroduction of English at the primary stage. *Ms., NFG-English*.

Nag-Arulmani, S. (2000). *Types and manifestations of learning difficulties in Indian classrooms*. In Paper presented at the first Orientation Programme for Schoolteachers, National Institute for Public Co-operation and Child Development (NIPCCD), Bangalore, India.

Nag-Arulmani, S. (2003). Reading difficulties in Indian languages. In *Dyslexia in different languages: Cross-linguistic comparisons* (pp. 235–254). London, UK: Whurr Publ.

National Council of Educational Research and Training. (2011). *National curriculum framework 2005* (No. id: 1138).

National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.

Nishanimut, S. P., Johnston, R. S., Joshi, R. M., Thomas, P. J., & Padakannaya, P. (2013). Effect of synthetic phonics instruction on literacy skills in an ESL setting. *Learning and Individual Differences, 27*, 47–53.

Pellegrini, A. D., Brody, G. H., & Sigel, I. E. (1985). Parents' book-reading habits with their children. *Journal of Educational Psychology, 77*(3), 332–340.

Ramaa, S., Miles, T. R., & Lalithamma, M. S. (1993). Dyslexia: Symbol processing difficulty in the Kannada language. *Reading and Writing, 5*(1), 29–41.

Ramanathan, H., & Bruning, M. D. (2003). Reflection on teaching oral English skills in India: A research report. *Journal of the International Society for Teacher Education, 7*(1), 48–55.

Ramanathan, V., & Atkinson, D. (1999). Individualism, academic writing, and ESL writers. *Journal of Second Language Writing, 8*(1), 45–75.

Ramachandran, V., Pal, M., Jain, S., Shekar, S., & Sharma, J. (2005). *Teacher motivation in India* (pp. 96–103). Discussion Paper. Bangalore: Azim Premji Foundation.

Reyes, I., & Azuara, P. (2008). Emergent biliteracy in young Mexican immigrant children. *Reading Research Quarterly, 43*(4), 374–398.

Saini, A. (2000). Literacy and empowerment: An Indian scenario. *Childhood Education, 76*(6), 381–384.

Senechal, M., & Cornell, E. H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly, 28*(4), 360–374.

Semel, E., Wiig, E., & Secord, W. A. (2013). *Clinical evaluation of language fundamentals* (5th ed.). San Antonio, TX: Pearson.

Shenoy, S. (2016). Efficacy of a bilingual screening tool in L1 Kannada and L2 English to differentiate between language differences and disorders in English language learners. *Psychological Studies, 61*(3), 126–136.

Sreekanth, Y. (2011). Parents involvement in the education of their children: Indicators of level of involvement. *International Journal about Parents in Education, 5*(1), 36–45.