THE FRONTIERS COLLECTION

P.-M. Binder · K. Smith (Eds.)

# THE LANGUAGE PHENOMENON

## Human Communication from Milliseconds to Millennia

Springer

# THE FRONTIERS COLLECTION

For further volumes:
http://www.springer.com/series/5342

# THE FRONTIERS COLLECTION

The books in this collection are devoted to challenging and open problems at the forefront of modern science, including related philosophical debates. In contrast to typical research monographs, however, they strive to present their topics in a manner accessible also to scientifically literate non-specialists wishing to gain insight into the deeper implications and fascinating questions involved. Taken as a whole, the series reflects the need for a fundamental and interdisciplinary approach to modern science. Furthermore, it is intended to encourage active scientists in all areas to ponder over important and perhaps controversial issues beyond their own speciality. Extending from quantum physics and relativity to entropy, consciousness and complex systems—the Frontiers Collection will inspire readers to push back the frontiers of their own knowledge.

For a full list of published titles, please see back of book or springer.com/series/5342

P.-M. Binder · K. Smith
Editors

# THE LANGUAGE PHENOMENON

## Human Communication from Milliseconds to Millennia

Springer

*Editors*
P.-M. Binder
Natural Sciences Division
University of Hawaii
Hilo, Hawaii
USA

K. Smith
School of Philosophy, Psychology
and Language Sciences
University of Edinburgh,
Dugald Stewart Building
Edinburgh, Scotland
UK

ISSN 1612-3018
ISBN 978-3-642-36085-5 ISBN 978-3-642-36086-2 (eBook)
DOI 10.1007/978-3-642-36086-2
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013931134

© Springer-Verlag Berlin Heidelberg 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Preface

You have in your hands what we believe is a significantly different book about language. It is not a technical book that tries to teach you an academic discipline. It is not a book that tries to condense the knowledge and insights of a single writer for the benefit of the general reader. Instead, it is a collection of chapters written by leading experts on practically all facets of language: its processing in the brain, language learning and evolution, the influence of genes and biology on language, and its place in human ecology. The level of rigor is halfway between academic writing and writing for the general public. We use timescales as an organizing principle, starting from processes that take a small fraction of a second all the way to others that span about a million years. Yet, our goal is not just to describe language but also to locate many of its mysteries precisely at the seams between such processes. We hope you enjoy the results and emerge with a more complete perspective on what language is.

We have many people to thank. First of all, the scholars who generously shared their knowledge and patiently gave their time in writing the chapters of this book. Secondly, those who have influenced our careers or supported us as this book was starting to take shape. P. M. B. thanks Martine Cattarelli and Yumiko Ohara for encouragement, and Brett Larive for initiating contact between him and the publisher. K. S. thanks Simon Kirby and Andrew Smith for very helpful advice. Thirdly, we acknowledge the support of our families, and mildly apologize for the time this project took away from them. Thanks to Shalila, Penelope, Becky, Jessie, and Sam for being part of our lives.

We would like to dedicate this book to all those who spend their lives studying language, especially those who concentrate on saving endangered languages. Their work in preserving a small but very important part of what makes us human deserves all our admiration. Among these, P. M. B. would like to recognize the people at Centro Colombiano de Estudios de Lenguas Aborígenes at Universidad de Los Andes, Bogotá, Colombia, and at the Ka Haka ʻUla O Keʻelikōlani College of Hawaiian Language at the University of Hawaii, Hilo.

Finally, we would like to acknowledge our Publishing Editor at Springer, Angela Lahee, for her outstanding competence, patience, and optimism. Angela, it has been a pleasure working with you.

Hilo, Hawaii, United States P.-M. Binder
Edinburgh, Scotland K. Smith

# Contents

# Chapter 1
# Introduction

**K. Smith and P.-M. Binder**

## 1.1 Preamble

Language is a highly complex, socially learned system of communication which allows us to express a limitless range of concepts and convey very fine-grained and subtle distinctions: anything you can think, you can transmit to others using language. The capacity for language is unique to humans: no other species has a communication system that works like language or that has anything close to its communicative potential. Indeed, language seems to underpin many of humanity's most striking achievements: it's hard to imagine how we would have such complex social structures and technologies, were it not for our ability to use language to exchange information and coordinate our activities. These technical achievements have taken human civilization beyond our own planet, both through manned space travel and unmanned probes that have reached the edge of our solar system. No other species on Earth has such achievements of boast of, nor language to boast in.

Despite the manifest specialness of language, its central role in explaining some of the key achievements of our species, the long history of the academic study of language (dating at least back to the classical philosophy of ancient Greece in the 5th century

---

**Phenomenon:** A fact or event of scientific interest susceptible to scientific description and explanation (Merriam-Webster Dictionary).

**Phenomenon:** A fact or occurrence, the cause or explanation of which is in question. A very notable or extraordinary thing; a highly exceptional or unaccountable fact or occurrence; a thing, person, or animal remarkable for some unusual quality; a prodigy (Oxford English Dictionary).

---

K. Smith (✉)
School of Philosophy, Psychology and Language Sciences, University of Edinburgh,
Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK
e-mail: kenny@ling.ed.ac.uk

P.-M. Binder
Natural Sciences Division,University of Hawaii, 200 West Kawili Street,
Hilo, HI 96720-4091, USA
e-mail: pbinder@hawaii.edu

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_1,

BC), and the constant fascination that language holds for non-academics, language remains a phenomenon: something presumably describable and explainable, and yet shrouded in mystery. Why is this? Why haven't we figured out a long time ago how language works, and why it works the way it does, and why only humans have a communication system which works like this? While great progress has been made in moving towards addressing these questions in various scientific disciplines (linguistics of course, but also various branches of psychology, biology, anthropology, and more recently neuroscience and genetics, to name but a few), the perspective of this book is that language poses such a challenge, and requires efforts from such a broad array of disciplines, because it takes place across multiple timescales, and is shaped by processes which operate across such scales. Language is produced by the articulatory apparatus and processed by the brain via operations which operate at scales of milliseconds to seconds. Such utterances form parts of conversations which last a few seconds, minutes, or even hours. These daily linguistic experiences form the input to processes of language learning in children and the modification of patterns of language use in adults, processes which operate across days, weeks, months and years. Languages themselves change over the course of tens, hundreds and thousands of years: new words or phrases are adopted, pronunciation changes, even the systematic structure of the language and the basic principles along which it is organised might change. Language change and evolution are a consequence of the processes which operate at the level of individual lifetimes: the ways in which languages change are driven by the ways in which children learn language and adults use language. Superimposed on these faster processes, our biology (including, presumably, any biological component of our capacity for language) evolves over tens or hundreds of millennia, or even longer. Evolution tinkers with the lowest-level processes involved in language: the physical equipment used for speech, and the neural substrates on which language depends. Therefore, the cycle of interaction loops back on itself: the longest timescale we might imagine studying language on, that of biological evolution of the capacity of our species for language, ultimately becomes the study of the equipment involved in language at the shortest timescales.

This book tracks language across these various timescales, from the very short timescale of sentence processing in the brains of individuals to the very long scales of language change and the biological evolution of our species, with an emphasis on how processes at one level interact with those at other levels. Understanding how each of these individual processes takes place is in itself a challenge, and constitutes the subject matter of the disciplines outlined above. However, a central theme of this book is that, in order to understand language, we also need to understand these interactions across scales: how processes occurring at one timescale can shape, or be shaped by, processes occurring at another timescale. The chapters in this book describe language at many levels, and also attempt to explain how these various levels tie together, to offer an integrated account of how language is shaped by the processes that make it happen.

In this introductory chapter we'll set the scene for the rest of the book. We begin in the next Section by contrasting our approach, focusing on interactions between processes, with a classic alternative: abstracting away from those interactions, to focus on language as a mental object in the heads of individuals. What we'll suggest here is not that this is an inherently bad approach, but rather that at least some aspects of how language works are more naturally and succinctly explained by taking the interactions into account. Then, in Sect. 1.3, we'll provide a short roadmap of the rest of the book, explaining how we think the various processes that shape language fit together, and how the chapters of this book are organized to explain this to you.

## 1.2 How Should We Study Language?

### *1.2.1 Studying Language from the Perspective of the Individual: Chomsky and the Cognitive Revolution*

All scientific enquiry involves abstraction and simplification: if you want to know how something works, then a well-established way to analyze it is to isolate the object of study from the surrounding environment as much as you possibly can, and if possible break it down into its component parts and figure out how they work individually. For instance, if studying the human body, you might want to start off by studying healthy adults (rather than, say, children or diseased individuals, whose bodies might pose additional challenges, either because they are damaged or because they are still developing), and you might want to study the component parts separately (how does the heart work? how do the lungs work?). Implicit or explicit in this approach is the acknowledgement that at some point we'll have to understand how the various components interact (e.g. how the heart and lungs combine to distribute oxygenated blood around the body) and how more complex cases function (how do the lungs develop? what happens in a heart attack?), but the starting point is to idealize, to isolate, and to decompose.

This approach has proved to be enormously profitable when applied to the study of language. Modern linguistics, and indeed the field of cognitive science, owes much of its existence and form to the work of Noam Chomsky (e.g. Chomsky 1965, 1986, 1995). Chomsky's approach to language, in turn arguing against a preceding school of thought which emphasized the importance of studying observable linguistic behaviour (see a more detailed discussion in Müller, this volume), was to study language as a *mental object*: a thing which exists in the heads of people who know a language (see Jackendoff 2002 for an excellent introduction). As Chomsky pointed out, speakers can produce and understand utterances they have never encountered before: language isn't just a memorized collection of sentences, or even a list specifying which words can follow which words, but a far richer and more complex mental system for producing and understanding sentences according to the rules of a language—a so-called generative system. To borrow an example from Wonnacott (Chap. 4 of this book), speakers of English will be able to guess that a sentence like *John gorped me the ball* involves some kind of transfer of a ball from John to the speaker, despite never having heard the novel verb *gorped* or that complete sentence before. In order to pull off this trick, English speakers must have an internal system of rules (or something equivalent) to draw on to allow them to handle novel utterances, rather than just a stored set of words and sentences which they have heard others use and which they can parrot back. This mental grammar, which exists in the head of a person who knows a language, was the object that Chomsky wanted to study.

Chomsky therefore used the isolate-and-decompose strategy to focus on language at the level of an individual mind, rather than as a system operating in populations featuring a mish-mash of speakers with different accents and dialects, operating in a noisy environment full of distractions and interruptions, attempting to use their language to achieve their communicative goals. In his own words:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of language in actual performance... To study actual linguistic performance [i.e. observable linguistic behavior, rather than mental grammar], we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one. In this respect, study of language is no different from empirical investigation of other complex phenomena. . . .The problem for the linguist. . . is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer . . . [h]ence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior (Chomsky 1965, pp. 3–4).

This focus on grammar as a mental entity allowed enormous progress to be made in characterizing the structure of languages. More broadly, Chomsky's arguments generalized to other aspects of human behavior, and spawned the cognitive sciences: if there were underlying mental rules governing language, which linguists could study by focusing on idealized individuals, why not study other aspects of the mind in the same way, e.g. the ability to form memories, identify categories, or apply reasoning to solve problems?

### *1.2.2 The Role of Interactions in Understanding Language*

Despite the powerful appeal of this individual-centered approach to studying language, and its influence far beyond linguistics, the focus on the idealized speaker-hearer in the homogenous speech community is clearly more troublesome for those branches of linguistics which study variation in languages. Perhaps the most obvious feature of language to a non-linguist is that it varies within populations: not everyone sounds the same, and you can often guess where someone comes from (and perhaps something about their socio-economic status) from the way they talk. Cultural commentators also frequently notice, and typically lament, the fact that language varies over time: Deutscher (2005) amusingly runs through two millennia of authors bemoaning the fall from perfection of their language, going back to Cicero comparing the Latin of 46 BC unfavorably with the Latin spoken 100 years earlier. The observation that languages are heterogenous, both at a given time point and across time, is central to the subfields of sociolinguistics and historical linguistics. Indeed, one of the earliest developments in linguistics was the systematic attempt to study the processes by which languages change over time, and to reconstruct the ancestral languages from which modern languages descend. In keeping with the theme of this book, a classic approach to the study of language change has been to explain the ways in which such change occurs over time as a response to low-level pressures operating on the production and perception of speech (see Box 1).

**Box 1: Assimilation and Sound Change (examples from Trask (2007))**

Producing speech involves moving various parts of the articulatory apparatus—the tongue, the lips, the velum and the vocal folds—in rapid sequence. This is hard work, and there is a natural tendency on the part of speakers to minimize the movements they must perform when speaking. This can result in various changes to the way in which words are pronounced: for instance, sounds can change to be more similar to nearby sounds, reducing the amount of movement required between sounds: this process as known as *assimilation* (speakers can do other things to minimize articulatory effort, including 'weakening' or even dropping sounds entirely). We can see traces of these assimilatory processes in language change, as the normal pronunciation (featuring assimilation) becomes a standard part of the written language. For example, the Latin word for 'night', *nocte*, has become *notte* in Italian: the relatively awkward 'ct' sequence has assimilated to an easier-to-pronounce 'tt'. Systematic weakening and deletion of sounds also shows up in the historical record: in general, actions of individual speakers when speaking result in the language changing to embody those actions.

Chomsky's idealisation of a homogenous speech community is clearly not useful when studying the social function of linguistic variation, or the heterogeneity of language over time. There are other cases where, it has been argued, the focus on language in the minds of individuals obscures the best possible explanations for how language works, and why it works the way it does. In addition to developing the mentalistic approach to studying language, Chomsky is well-known for proposing that language learning requires some innate and richly structured knowledge of how languages work, sometimes called Universal Grammar (e.g. Chomsky 1965, 1986, although in more recent years he has downplayed this idea, e.g. Chomsky 2005). For instance, rather than having to figure out from scratch the rules underpinning the linguistic behaviour they are exposed to, children might be born 'knowing' that there are a limited set of possibilities, and simply have to identify which of these options applies to the language they are learning: the so-called "principles and parameters" theory (see Baker 2001 for an accessible and ambitious version of this theory).

Chomsky claimed that the strongest evidence for the existence of this innate linguistic knowledge was that children seem to learn facts about language for which there is no good evidence in the linguistic input they receive: for instance, children might know that *Is the boy who is happy singing?* is a perfectly normal question in English, whereas *Is the boy who happy is singing?* is not, despite not having encountered questions of this level of complexity (involving a relative clause *who is singing* which modifies the subject noun *boy*) before. In other words, when presented with the need to make guesses during language acquisition, children seem to have a knack for making the *right* guess. Chomsky's interpretation for this happy knack is that children are born knowing the correct answers: 'learning' of these features of language is an illusion.

However, others have argued that this interpretation of children's abilities as language learners is a consequence of the focus on language at the level of the individual. If you consider an isolated language learner attempting to learn an arbitrary language, then it is indeed surprising if they seem to magically know how that language should work. However, language learners don't exist in isolation, and they don't attempt to learn any logically possible language: they exist in populations, and the language they are trying to learn is a language that is used by other members of their population who also learned their language in the same circumstances. Therefore, as a consequence of their existence and transmission in populations, languages will tend to have precisely those features that language learners expect them to have:

> The structure of a language is under intense selection because in its reproduction from generation to generation, it must pass through a narrow bottleneck: children's minds … Language operations that can be learned quickly and easily by children will tend to get passed on to the next generation more effectively and intact than those that are difficult to learn. So, languages should change through history in ways that tend to conform to children's expectations (Deacon 1997, p. 110).

> Human children appear pre-adapted to guess the rules of syntax correctly, precisely because languages evolve so as to embody in their syntax the most frequently guessed patterns (Deacon 1997, p. 122).

This pressure on languages to be learnable, resulting from the transmission of language in populations, is sometimes called *cultural selection for learnability* (Brighton et al. 2005). Of course there remains debate about exactly what expectations children have, how strong these expectations are, and whether they are expectations about the world in general or linguistic systems in particular (Kirby 1999; Christiansen and Kirby 2003; Kirby et al. 2004; Christiansen and Chater 2008). But the point remains that the study of language beyond the level of the individual allows a natural and elegant explanation for the good fit between how languages work and how children expect them to work, which is perhaps puzzling or even shocking under other approaches. Similar arguments have been made for other aspects of language. For example, in Chap. 3, Garrod and Pickering argue that dialogue is one such area: considered from the level of individuals, the actual use of language in conversation seems like an amazingly hard problem; however, when we consider the detail of how interaction in a conversation actually works, and in particular how it facilitates development of shared linguistic norms, it becomes apparent why dialogue in fact feels easy, rather than extremely hard. These moves in the language sciences are paralleled in the cognitive sciences more generally, where there has been a movement towards considering how the messy details of the environments in which cognition takes place might radically change what we think cognition is (e.g. Clark 1998).

## 1.3 The Content and Structure of this Book

The previous section concluded with an example of how a process operating at one level (learning) could impact on another process (language change): learning causes languages to change and to become more learnable, Of course, the timescales/processes of learning by individuals and the cultural/historical timescale/process of language change are not the only possible timescales and processes at which language might work.

Language is a mental object: as with any mental operation you conduct, language is ultimately instantiated in your brain. Understanding how the brain responds to linguistic stimuli—how incoming linguistic stimuli are processed on a millisecond-to-millisecond and second-to-second basis—forms the shortest timeframe at which linguistic processes can be studied (speech production and perception of course work at similar timescales). In Chap. 2, Müller examines the operation of language in the brain, and how this might influence the functional structure of the brain over the timescale of a lifespan. In the quest to abstract and reduce, much research on the neuroscience of language has sought to abstract away from individual variation in how language is processed in the brains of different individuals, and identify "language areas"—parts of the brain that seem to be reliably involved in language processing in multiple experimental participants. Müller argues that differences between individuals are also an inevitable consequence of the extreme plasticity of the brain: language areas are then an emergent property of the low-level wiring of the brain, which causes language to gravitate to those areas of the brain which are particularly suited to do language processing, i.e. have the right neural resources and the right patterns of connectivity.

Sentences are typically produced in conversations, which last a few seconds or minutes or even hours: the study of dialogue therefore occupies the next natural level of analysis, discussed in Garrod and Pickering's Chap. 3. They review the experimental evidence which shows that people *align* linguistically as they interact: as a consequence of talking with someone, you tend to use the same words, the same linguistic constructions and even come to share the same mental representation of the world as that person, and they in turn tune themselves to you. Alignment is ultimately a consequence of a basic property of the wiring of the brain, in particular the ubiquity of automatic perception-action links (some neurons fire when you see someone performing some action, or perform it yourself). These patterns of linguistic behaviour which form during conversations can become entrenched or routinized: initially temporary conversational conventions can become a stable component of people's linguistic knowledge. Garrod & Pickering emphasise the implications of this view for language at longer timescales: language learning can be conceived of as routinization on a grand scale, and routinized linguistic behaviours can become permanent features of a language.

Single conversations are one-off events in a lifetime spent communicating, spanning years and tens of years. One important linguistic process which operates at this timescale is the process of language learning or language acquisition: children learn to speak the language(s) of their speech community. In Chap. 4, Wonnacott looks in detail at this process, in particular focusing on statistical learning accounts of language acquisition. Statistical learning theories emphasise the richness of the linguistic data children are presented with, and the various ways in which they can exploit that richness. The fundamental conundrum at the heart of language acquisition is the need to explain how language learners can *generalise* from a finite set of linguistic data to an infinitely expressive language, without *overgeneralising*, or extrapolating too far from the available data. Wonnacott reviews some of the theories about how children achieve this feat, emphasising the assumptions that these theories make about the capacities, biases and predispositions of language learners. Explaining what these capacities, biases and predispositions are – explaining what the human capacity for language *actually consists of*—is a major goal of modern linguistics.

In Chap. 5, Croft picks up on the theme introduced by Garrod and Pickering, and explores the consequences of language use for language change. Croft argues that the properties of language can best be understood in terms of the kinds of things language is used for. In particular, Croft emphasises the role of language as a coordination device to enable us to undertake joint action (i.e. to cooperate), and as a mechanism for the verbalisation of experience. We vary how we express ourselves depending on what we want to communicate and who we want to communicate it to, and in turn flexibly interpret the utterances of others depending on the context in which communication takes place. These processes of flexible production and interpretation generate a pool of linguistic variants (Croft calls them *linguemes*) which interact and compete: for example, prestigious variants, or variants of prestigious speakers, might spread more rapidly than alternative formulations. Croft therefore argues that language change, operating at very long timescales, can be seen as an evolutionary process whose engine is the short timescale, moment-to-moment choices of language users.

The next two chapters follow Croft in studying the long-term change and evolution of linguistic systems. In Chap. 6, like Croft, Kirby argues that processes operating at the level of individual lifetimes will impact on long-term patterns of language change, and ultimately on the fundamental structure of languages. In contrast to Croft's emphasis on language use, Kirby argues that some of the fundamental structural properties of language (compositional syntax, phonemic coding, a content-function split in the lexicon) can be explained as long-term consequences of pressures inherent in language *learning*. The development of these fundamental structural properties of language are instances of more general patterns of evolution: in particular, they parallel the major transitions in the evolution of life, in that they involve the loss of independent replication and the division of labour between interacting replicators. In Chap. 7, Dediu reviews our current state of knowledge of the genetics of language. Genetics is relevant to the study of language at at least two timescales. At the scale of the lifespan of individuals, as Dediu shows, there is good evidence that there is a genetic component to many aspects of language, as evidenced for example by studies of twins, studies of the pedigrees of language pathologies, or more recent techniques of trawling the whole genome in search of genetic markers which seem to correlate with linguistic properties of individuals. However, as Dediu goes on to show, the relationship between languages and genes is relevant at longer timescales too. Firstly, languages and genes share their history to some extent: neighbouring populations can influence one another linguistically (through language contact) and genetically (by interbreeding), which inevitably leads to some historical entwinement between language and the genes. However, Dediu also argues for a second kind of causal relationship between genetics and language: based on an unexpected correlation between genetic markers and the linguistic feature of tone, he argues that genetic differences between populations may lead to linguistic differences, as languages adapt to the different biases and predispositions of learners in different populations.

New physical characteristics and behavioral repertoires of species themselves evolve at the level of tens of thousands, hundreds of thousands, or millions of years. Zuidema adopts this evolutionary perspective in Chap. 8, in order to answer the question of what the human capacity for language actually is, and where it comes from. Zuidema lays out the types of evidence that we can use to answer such questions: from linguistics, but also comparative biology, evolutionary theory, and archaeology. The physiological adaptations in humans for language may actually be relatively minor: we have a somewhat unusual vocal tract, a fairly mundane ear. However, Zuidema does pinpoint two unusual features of the human brain: we possess the neural machinery for vocal learning (see also Oudeyer, this volume), and the less-well-understood machinery for hierarchical compositionality (the ability to build complex, meaningful utterances from a set of component parts). Zuidema sketches a set of conditions under which these biological building blocks might evolve, and how these basic capacities might lead to the other features of language that we find so striking. Again, in keeping with the theme of the book, the consideration of language at these long timescales makes reference to language as it operates on shorter timescales (e.g., how learning drives language change), and suggests particular perspectives on the articulatory, acoustic and neural machinery which influences how language works at the shortest times. Language is like a Mobius strip, where following language at increasingly long timescales leads us back to a consideration of the machinery used by language at the shortest timescales possible.

In Chap. 9, Oudeyer presents a detailed elaboration of the basic position that we have sketched in this introductory chapter: that language is a complex system whose properties emerge from the interaction of multiple processes operating across time scales, where understanding the behaviour of those component processes alone is not enough to understand how language works. As an example, Oudeyer considers the emergence of *phonemic coding* in language, the construction of words using a combinatorial system of basic contrastive units, such as speech sounds. Oudeyer shows, via computational simulation, that the interaction of individuals with only very basic properties (the capacity for vocal imitation, requiring only plastic connections between the neural circuitry involved in perceiving and producing speech) results in the emergence of a system of phonemic coding of the sort we see in language: phonemic coding is not necessarily wired into any individual, but emerges over time from the interactions between individuals.

Finally, by way of a rather sobering epilogue, in Chap. 10 Romaine outlines a further consequence of the interconnectedness of language: language endangerment and death. Languages evolve in specific environments to meet the communicative needs of their users, as discussed by Croft, and perhaps to indirectly reflect local topography (regions with mostly isolated valleys or islands may develop more distinct languages). However, when languages come into contact with each other, particularly in situations of unequal status or power, the individual-level choices of language users, in particular the choice of whether to use or not use a lower-status language, can have dire consequences for language systems. Constructing interventions to reverse language decline or stave off language death is challenging because of the interactive nature of language: Romaine reviews modelling techniques which are beginning to be used to understand the interactions involved in language contact and death, in the hope that the loss of the planet's linguistic diversity can be slowed.

In Chap. 11, we wrap up the book, reviewing what has gone before and looking to what we think will be the future of the language sciences, focussing on the interactive nature of language and the challenges (and opportunities) this poses for language scientists. To get a detailed feel for the interactions that shape language, and how they can be studied, read on.

## References

Baker, M.: The Atoms of Language: The Mind's Hidden Rules of Grammar. Basic Books, New York (2001)

Brighton, H., Kirby, S., Smith, K.: Cultural selection for learnability: three principles underlying the view that language adapts to be learnable. In: Tallerman, M. (ed.) Language Origins: Perspectives on Evolution, pp. 291–309. Oxford University Press, Oxford (2005)

Chomsky, N.: Aspects of the Theory of Syntax. MIT Press, Cambridge (1965)

Chomsky, N.: Knowledge of Language: Its Nature, Origin and Use. Praeger, London (1986)

Chomsky, N.: The Minimalist Program. MIT Press, Cambridge (1995)

Chomsky, N.: Three factors in language design. Linguist. Inqu. **36**, 1–22 (2005)

Christiansen, M.H., Chater, N.: Language as shaped by the brain. Behav. Brain Sci. **31**, 489–509 (2008)

Christiansen, M.H., Kirby, S. (eds.): Language Evolution. Oxford University Press, Oxford (2003)

Clark, A.: Being There: Putting Brain, Body, and World Together Again. Bradford Books, Cambridge (1998)

Deacon, T.: The Symbolic Species. Penguin, London (1997)
Deutscher, G.: The unfolding of Language. Metropolitan, New York (2005)
Jackendoff, R.: Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press, Oxford (2002)
Kirby, S.: Function, Selection and Innateness: The Emergence of Language Universals. Oxford University Press, Oxford (1999)
Kirby, S., Smith, K., Brighton, H.: From UG to universals: linguistic adaptation through iterated learning. Stud. Lang. **28**, 587–607 (2004)
Trask, R.L.: Trask's Historical Linguistics. 2nd edn. Hodder Education, London (2007) (revised by Robert McColl Millar)

# Chapter 2
# Neurobiology: Language By, In, Through and Across the Brain

**Ralph-Axel Müller**

Language has come to be commonly understood as something accomplished *by the brain*. Much scientific investigation has consequently focused on looking for language *in the brain*. Although it may sound intuitive, this approach suggests that language is an object located inside another object. This spatial metaphor has generated important insights into the brain sites important for language, from nineteenth century studies of brain-damaged patients to more recent and refined evidence from brain imaging. However, with this refinement has come the realization that brains are not truly maps, and that language is not simply the inhabitant of a localized part of the brain. On the time scale of lifetime development, this is supported by evidence of plasticity in the brain's organization for language, prompted by maturation, experience, or localized damage. On the smaller time scale of an individual language process *through the brain,* the participation of many brain regions can be observed. Some of these regions are known for their role in perceptual, motor, and other seemingly non-linguistic functions. It remains uncertain whether any brain regions are exclusively dedicated to language. As examples, the classical language areas of Broca and Wernicke participate in many functions, such as imitation, action understanding, perception of biological motion that are not specifically linguistic. This suggests that language acquisition is built upon and embedded in precursor and component functions that emerge during the first years of life from networks distributed *across the brain*.

## 2.1 By the Brain: Language as a Biological Object

In 1959, the young linguist Noam Chomsky reviewed B. F. Skinner's book *Verbal Behavior* (Skinner 1957), in which behaviorist principles of conditioning and reinforcement were applied to the question of language acquisition in children.

---

R.-A. Müller (✉)
Brain Development Imaging Laboratory, Department of Psychology, San Diego State University, San Diego, CA 92120, USA
e-mail: rmueller@mail.sdsu.edu

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_2,

Chomsky's review (1959) is remembered for debunking behaviorist linguistics in an uncompromising way. As an alternative, Chomsky put forth a 'generative grammar' in an attempt to transform linguistics into a science on a par with mathematics or physics (Chomsky 1965, 1972a). This approach proposes a set of rules and principles capable of generating an infinite number of sentences, and dramatically changed the concept of language in the second half of the twentieth century. Chomsky's views about the biological basis of these grammatical principles were equally revolutionary. In the mentioned review Chomsky accounts for children's ability to develop mental grammars through innate tendencies, which endow human beings with specific types of learning. He then tied these thoughts to the view of mind and body as separate and independent realms, as proposed earlier by Descartes (Chomsky 1966).[1] From the contemporary point of view, Cartesian innatism (the belief that the mind is born with ideas) and mind-body dualism may be considered strange bedfellows of linguistics as an exact science. This appears to be reflected in the ultimately ambivalent attitude towards a biological explanation of language that has characterized Chomsky's work and that of many fellow generative grammarians. While the presumed genetic bases of universal grammar have been a mantra for decades in generative linguistics, actual efforts to identify relevant genes remained modest for a long time (see Chap. 7 for recent advances).

Chomsky's views not only opposed behaviorist teachings, which he directly attacked, but also the tradition of Ferdinand de Saussure, which had dominated early twentieth century linguistics. While Saussure's structural linguistics foreshadowed Chomsky (Newmeyer 1986) by viewing language ("langue") as a system that could be subjected to exact scientific study ("un object de nature concrète"), he considered this object of language as a social entity, "exterior to the individual" (Saussure 1915/1972). To Saussure, the core object of linguistics was social, whereas the study of "parole", the "grammatical system that exists in every brain", was considered secondary. Chomsky reversed these priorities, declaring the language ability (which he called "universal grammar") to be the core object of linguistics—a biological entity that matured in similar ways as other bodily organs (Chomsky 1976, 1980). Pragmatic and sociological approaches to language, such as Searle's theory of speech acts (Searle 1969), were dismissed with disdain: "You can also collect butterflies and make many observations. If you like butterflies, that's fine; but such work must not be confounded with research…" (Chomsky and Ronat 1979).

Given Chomsky's strong views on the biological and innate nature of language, generative grammarians showed surprisingly little initiative to bridge the gap between linguistic theory and neuroscience or genetics. Lenneberg's *Biological Foundations of Language* Lenneberg (1967) attempted to link principles of generative grammar with biological, neurological, and evolutionary evidence. However, Lenneberg's conclusion, according to which human language abilities reflected "an adaptation of a ubiquitous process (among vertebrates) of categorization and extraction of similari-

---

[1] Chomsky's forays into dualistic rationalism were blended with and at least partly motivated by his political views, which eclectically (and bizarrely) equated empiricism with US imperialism of the Vietnam era and rationalism with progressive enlightenment (Chomsky 1972b).

ties", was hard to reconcile with Chomsky's strong views of the autonomy of language with respect to nonverbal cognition (Chomsky 1957, 1976). Lenneberg furthermore dismissed the idea of "genes for language", pointing out that species-specific changes in genes during hominid evolution may have resulted in protracted infancy, enhanced brain plasticity, and "a peculiar ontogenetic phase of an optimal confluence of various abilities… [and] thus a critical period for language acquisition" (ibid.; see also Sect. 2.2.5). As will become clear in Sect. 2.4, these ideas are surprisingly modern and still capture the gist of neurobiological evidence that has accumulated in the past four decades.

## 2.2 In the Brain: Language as a Spatial Location

### *2.2.1 History*

Chomsky's critique of Skinner came in the context of a mid-twentieth century "cognitive revolution" in linguistics, philosophy, psychology, and artificial intelligence. This revolution aimed to replace behaviorist views of the mind as a "black box" with cognitive models of the mind (Gardner 1987). Yet, the brain was surely not new territory at the time. On the contrary, neurology had an extensive track record in the study of brain-behavior relationships. Overcoming empirically dubious early attempts (such as medieval ventricular theory or phrenology in the late eighteenth century; Clarke and Dewhurst 1972), modern neurology had accumulated a large literature relating localized brain damage to specific types of cognitive-behavioral impairment. As neuroimaging techniques identifying local damage in living patients were yet unavailable, the link was usually established after death. Broca's (1861) landmark case of Leborgne, also called "tan-tan" for his repetitive utterances, serves as a slightly ironic example. First, Leborgne conveniently died only 6 days after his case had been declared a litmus test of the theory of functional localization at the Parisian 'Société d'Anthropologie' in 1861 (Harrington 1987). Secondly, while results from his autopsy prompted the notion of the left inferior frontal gyrus ("Broca's area") as a language area critically involved in speech production, the actual damage in Leborgne's brain was far more extensive, including underlying white matter, and parts of the frontal and parietal lobes (Dronkers et al. 2007; Signoret et al. 1984). Nonetheless, 150 years later the inferior frontal cortex is still considered a core region for language processing or simply a "language area", for valid (though not necessarily compelling) reasons, as will become clear below (brain regions are shown in Fig. 2.1). Regardless of its empirical accuracy in the specific case of Leborgne, the basic logic of the localizationist approach is apparent: The damaged brain is equated to the healthy minus the injured part, and similarly the impaired mind of a lesion patient is equal to its healthy counterpart minus the function attributed to the site of injury. For example, if Leborgne's loss of speech production resulted from damage to Broca's area, the function of this region in the healthy brain was inferred to be speech production. This “patho-normal inference” (Müller 1996) survives in modern cognitive neuropsychology (see Shallice 1988 for a thorough review).

The case of Leborgne attracted great interest in 1861 because it was considered a test case in the ongoing debate between those who believed in functional brain localization, such as Jean-Baptiste Bouillaud, and others who considered the brain a unitary organ, without distinct functional subdivisions, such as Pierre Flourens.[2] Critics of the localizationist program have made contributions to the debate about brain-behavior relationships that remain of interest today. The British neurologist John Hughlings-Jackson (1878) goes beyond the traditional emphasis on the lesion patient’s “negative condition” (e.g., loss of expressive speech) and highlights the importance of the “positive condition”. Jackson describes the example of a paraphasia, such as the production of the word “table” instead of “chair”. In Jackson’s view, this utterance is not “the direct result of the disease”, as typically assumed. Instead it is “the best speech under the circumstances… owing to activity of healthy… nervous elements” (ibid.: 316)”.

Although Jackson’s statement may appear trivial, it reflects a crucial change in perspective that contrasts with localizationism, as described above. Rather than equating the effect of brain damage with a simple subtraction of a specific function, Jackson considered it as a regression into a more primitive and automated state that reflected an earlier stage of evolution. More generally, Jackson’s emphasis of the positive condition following brain damage underscores the importance of plasticity, which complicates any inference based on subtractions. These implications were further elaborated by Henry Head, who (1920) defined Jackson’s “positive effects” as follows: “A negative lesion produces positive effects by releasing activities, normally held under control by the functions of the affected level” (ibid.: 805)”. In order to fully understand cognitive processes in a lesion patient, it is therefore necessary to accept that brain damage triggers functional *reorganization*. Head’s account of plasticity remains a crucial insight today, because it implies that lesion effects are not simply subtractive. Later, Head (1926) adopted the term *diaschisis* from Constantin von Monakow to describe alterations in the brain far away from the site of the damage. Diaschisis contradicts the idea of strictly local effects of brain damage.[3]

Although skepticism over subtractive arguments in the works of Jackson, Head, Marie (1906) and Goldstein (1948) raised many deep and important issues, it did not put an end to the idea of localization. One consideration in favor of localizationism has been its strength in generating clear and falsifiable hypotheses. As argued by Popper (1965), true science resembles trial and error, or “conjectures and refutations”.

---

[2] Interestingly, this scientific debate had political and ideological undertones in mid-nineteenth century France of the second empire, as holistic views were considered a conservative and theologically warranted reflection of the unity of the soul, whereas localizationism was associated with antimonarchic and anticlerical left-wing views at the time (Harrington 1987).

[3] Modern neurology has provided evidence in support of diaschisis. For example, in patients with damage in cerebral cortex, reduced glucose metabolism is often seen far away from the site of damage, in the opposite hemisphere of the cerebellum. This phenomenon, called ‘crossed cerebellar diaschisis’, is attributed to contralateral (“crossed”) connections between forebrain and cerebellum (Gold and Lauritzen 2002; Pantano et al. 1986).

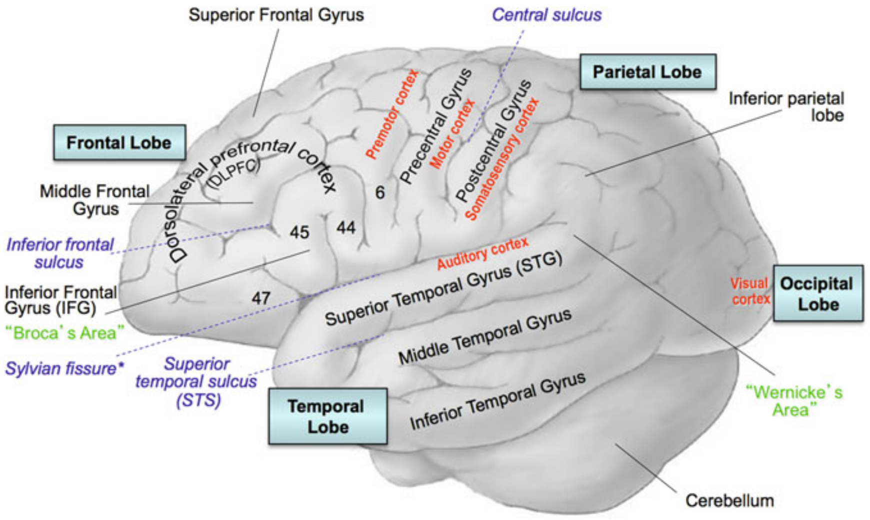


**Fig. 2.1** Outline of left hemisphere with anatomical terms used in text

Conversely, any school of thinking that does not generate falsifiable hypotheses may be considered unscientific, as it is impossible to produce empirical evidence that could refute it. The latter may apply to many of the insights provided by critics of localizationism. Hughlings-Jackson (1878), for example, emphasized that conclusions on brain-behavior relationships were ultimately limited to each individual patient: "There is… no single well-defined entity—loss of speech or aphasia—and thus… such a question as, 'Can an aphasic make a will?' cannot be answered any more than the question, 'Will a piece of string reach across the room?'… The question should be, 'Can this or that aphasic person make a will?' " (ibid.: 314). In Goldstein's work (1948), this focus on the individual appears to become an explicit denial of theory construction: "Experience showed that I must free myself from any definitive theory and investigate patients in a way as unbiased from any of them as possible".

Localizationist approaches, on the other hand, provided explicit models and predictions. A prime example is Ludwig Lichtheim's 'house model', which transparently predicts seven different types of language impairment (aphasia), based on lesion sites hitting either of the classical language areas (Broca's or Wernicke's), or connections between these and other nodes of a simple network (Fig. 2.1). Modern classifications of aphasia (e.g., Broca's, Wernicke's, transcortical sensory and motor, conduction, global) still resemble Lichtheim's model (e.g., Benson and Ardila 1996). A review by Norman Geschwind in *Science* (1970) was most instrumental in reintroducing Broca's, Wernicke's, and Lichtheim's localizationist approach to the general community of cognitive psychology and neuroscience; this review praised the concept of functionally specialized centers for language as "one of the greatest achievements of the last half of the nineteenth century".

### *2.2.2 Localizationism Afresh: Functional Neuroimaging*

While this revitalized localizationism did not share much scientific heritage with Chomsky's claim of the autonomy of syntax, there is some common ground. For an interesting example, Grodzinsky (1990) reviews Geschwind's "appealing" model, which was based on Lichtheim's house diagram (Fig. 2.1), and argues against it because it does not implement adequate theoretical concepts. The point is well-taken. How can a neurological model localize language, if the latter is not even fully understood, but construed in simplistic terms of speech comprehension, production etc.? The inverse approach, which prioritizes linguistic models and then looks for corresponding neural substrates, however, faces a similar problem.

How do we know whether a linguistic theory that adequately models empirical facts about language behavior actually relates to the way the brain is organized? In an influential article in *Behavioral and Brain Sciences,* Grodzinsky (2000) illustrates the issue. Stating that Broca's area "is more specialized than previously thought", he explains that this specialization pertains to grammatical relations within sentences[4] and that "syntactic abilities… are represented entirely and exclusively in the left hemisphere" (ibid.: 1). An empirical evaluation of these statements will follow further below. The justified critique of Geschwind's approach encounters a complementary problem: superimposing a theoretical concept (transformation) onto brain tissue without adequate understanding of brain function. While there is no perfect resolution to the tension between linguistic theory-driven 'top-down' and neuroscience data-driven 'bottom-up' approaches, it appears that extreme positions may hamper progress and only careful consideration of both can advance cognitive neurolinguistics.[5] Moreover, neuroscience is not a discipline of unassorted data, but is itself shaped by theoretical models, which may be usefully applied in cognitive neurolinguistics, as I will discuss later.

The views on the brain organization for syntax described above reflect a modern version of localizationism often called *modularity*. This concept was most influentially propagated by Fodor (1983), who portrayed cognitive modules as independent ("encapsulated") systems that are "innate" (genetically specified) and supported by localized brain structures. How compatible are such views with evidence from cognitive neuroscience? Functional imaging techniques, such as PET and fMRI (Box 1),

---

[4] In more technical verbatim terms: Broca's area "handles only intrasentential dependency relations" (Grodzinsky 2000: 21) and "is the neural home to receptive mechanisms involved in the computation of the relation between transformationally moved phrasal constituents and their extraction sites…" (ibid.: 1).

[5] Corresponding issues have been debated in artificial intelligence for decades. Computational models informed by cognitive theories may approximate the behavior of a human being, but their internal organization in no way resembles functional brain organization. The cognitive model-driven approach in AI (e.g., Newell and Simon 1963) was countered in the 1980s by parallel-distributed ("connectionist") models (McClelland and Rumelhart 1986). These were inspired 'bottom-up' by simplified features of neurons and their connections and aspired to demonstrate complex input-output behaviors mimicking human cognition as emergent properties of such very simple neural networks.

have greatly enhanced the investigation of brain-behavior relationships in recent decades. Note that in conventional applications these techniques have a localizationist bias, as they are typically used to generate "maps" of cognitive functions on the brain (similar in principle to phrenological maps or those of classical localizationism). Some early neuroimaging results in the study of syntax indeed appeared consistent with a modular view. Several PET studies by Caplan and colleagues (1998, 1999; Stromswold et al. 1996) implemented comparisons of sentences at different levels of complexity. Both identified activity in left inferior frontal area 44, presenting prominent illustrations with exclusive activity in this portion of Broca's area (Fig. 2.2a). Closer inspection, however, shows that stronger activations were identified in other parts of the brain, but were deemed irrelevant because they occured elsewhere.

**Box 1**. The two most important functional brain imaging techniques are positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). In **PET**, a small amount of radioactive tracer is injected into the bloodstream and biological measures, such as blood flow or glucose metabolism, can be taken based on the distribution of the tracer in the brain, which is detected by the PET scanner. **fMRI** does not involve radioactivity and is therefore more broadly used (even in children). This technique takes advantage of the relative increase of oxygen in the blood during regional brain activation, which results in slight changes in magnetic resonance that can be detected with an MRI scanner. This blood oxygen level dependent (**BOLD**) effect indirectly reflects local neuronal activity. A more recent development is **event-related fMRI**, in which the BOLD effect is detected for single trials of a task (rather than long blocks of trials), resulting in improved temporal resolution (see Box 2).

In a PET study by Caplan and colleagues ( 1999) for example, participants made plausibility judgments for more complex cleft object sentences ("It was the juice that the child enjoyed") and less complex subject cleft sentences ("It was the child that enjoyed the juice"). Comparison of the two conditions yielded significant activation solely in Broca's area, as depicted by a small and lonely activation blob in area 44 (Fig. 2.2a). The presentation is, however, misleading, as at least two additional sites with stronger activation were found in other parts of the brain, but were considered "non-significant", as they were outside the region of hypothesized activity. The wizardry requires explanation. Functional imaging studies (PET or fMRI) face severe issues of multiple comparisons, as statistics are typically performed separately for each volume element ("voxel"). A whole brain imaging study may include 100,000 voxels or more.[6] While technical solutions can mitigate the problem (for example,

---

[6] Since conventional fMRI statistics test for activation effects separately in each voxel, the large number of voxels creates a need to correct for multiple comparisons. For example, a given voxel shows an effect that is significant at the 95 % confidence level (i.e., the probability of obtaining the finding by chance alone is $\leq 5\,\%$). If comparisons are performed in 10 voxels, the probability

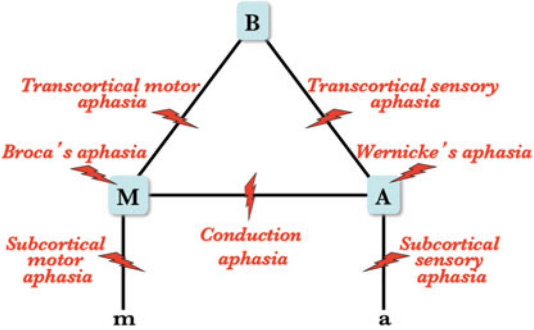


**Fig. 2.2** Lichtheim's "house diagram" of the language system. *Letters* in *blue boxes* stand for brain centers: *M* motoric language center (Broca's area); *A* auditory language center (Wernicke's area); *B* conceptual center ("Begriffszentrum"). *Small letters* at the *bottom* stand for motor and auditory periphery. Exactly seven distinct types of aphasia are predicted from this model, based on lesion sites (indicated by *lightening icons*) hitting either one of the two language centers (*M*, *A*) or connections between *M* and *A*, with the periphery, or with the conceptual center

based on the fact that true activation tends to occur in clusters; Forman et al. 1995), a more dramatic shortcut is provided by planned comparisons in regions of interest.[7] Thus in the cited studies by Caplan and colleagues, activity related to syntactic complexity was predicted for Broca's area, for which no correction was performed, but not for any of the remaining brain regions, where signal changes therefore remained below the significance threshold due to multiple comparison correction. A similar "demonstration" of exclusive and modular activation can in principle be construed for just about any task paradigm and any moderately activated part of the brain, based on an *a priori* hypothesis, especially if the study is underpowered due to a small number of participants.

Whether the activation pattern for a given language task appears localized or widely distributed therefore depends relatively little on the task itself, but largely on methodological specifics.[8] It is thus important to keep in mind that functional imaging evidence from PET or fMRI itself cannot provide direct evidence on how localized a given functional component may be. Ultimately, only reasonable interpretation of

---

of obtaining the finding by chance increases to 50 %. A correction for multiple comparisons is therefore necessary. In its simplest form (the "Bonferroni correction"), the *p*-value is multiplied by the number of comparisons. Such a correction would require that an effect in a single voxel reaches a *p*-value of 0.0000005 in order to "survive" a Bonferroni correction for 100,000 comparisons (assuming that an fMRI study includes this number of brain voxels).

[7] In technical terms, the region of interest approach drastically reduces the need to correct for multiple comparisons, as only a small part of the brain is considered, thus increasing the probability of a statistically significant finding.

[8] Among these are statistical power (number of participants), analysis type (whole-brain vs. region of interest), signal to noise ratio (reflecting head motion and other sources of noise), and the investigator's selection of significance thresholds and method of multiple comparison correction.

converging results from many studies, with careful consideration of methodological detail, can provide clues.

### 2.2.3 Localizing Syntax and Semantics

A number of neuroimaging studies have directly compared effects of syntactic and semantic task components. Friederici et al. (2003) found that processing of syntactic violations (e.g., “The blouse was on ironed”) and semantic violations (e.g., “The thunderstorm was ironed”) was associated with activation in large overlapping regions of left perisylvian cortex, the area surrounding the Sylvian fissure (Fig. 2.2b). Specificity, however, was seen in the *degree* of regional activation for the two types of condition. Some regions (the anterior superior temporal gyrus [STG] and the basal ganglia) showed significantly more activation for syntactic violations; others (the middle portion of the STG and the insula, a brain region in the depth of the Sylvian fissure) showed more activation in both the left and right hemispheres) showed more activation for semantic violations. Yet other regions, such as the posterior STG and inferior frontal gyrus (IFG), were activated during both conditions and also when sentences without violations were presented.

These and the findings by Caplan and colleagues described earlier illustrate that interpreting neuroscientific evidence as support for modular localization or holistic organization may be futile, as purely methodological details related to data acquisition, processing, and statistical analysis determine whether imaging results “look” modular, distributed, or holistic. An important additional consideration concerns the specifics of task paradigms. Imaging studies may be presented as comparisons of syntactic and semantic processes, but in reality each study will use a set of specific stimuli and task instructions, which may affect the results and may limit the extent to which findings can be generalized. For example, Kuperberg et al. (2003) used a conceptually similar approach by presenting sentences with syntactic and semantic anomalies, which however occurred in sentences of greater length than those used by Friederici et al. (2003). These and other design differences may have contributed to a rather different pattern of results, with portions of left IFG showing strongest activity for semantic anomalies, but modest activity for syntactic anomalies (which was actually weaker than for sentences without anomalies) in the study by Kuperberg and colleagues.

Note further that the use of sentences with violations implies that the type of violation specifically taxes the corresponding language component (e.g., syntax, semantics). In violation paradigms, the timing of a violation can be pinpointed by the presentation of a single word that creates the violation, which is advantageous in techniques with high temporal resolution (Box 2). It may be reasonable to assume that presentation of syntactically anomalous sentences puts a heavy load on syntactic decoding, but this process may differ from syntactic decoding of sentences without anomalies, as typically encountered in everyday language use.

**Box 2**. Functional imaging techniques can be distinguished in regard to two dimensions of resolution. **Spatial resolution** relates to a technique's ability to pinpoint where in the brain activity occurs. In colloquial terms, images with high spatial resolution are sharp, those with low spatial resolution are blurry. In more technical terms, spatial resolution is largely determined by the size of voxels (3-dimensional pixels). **Temporal resolution** refers to a technique's ability to pinpoint when activation occurs. For example, electrophysiological techniques, such as EEG and MEG (Box 3), have high temporal resolution, with measurements in the millisecond range. PET and fMRI have lower resolution, in the range of seconds to minutes. In fMRI, temporal resolution is largely determined by the repetition time (TR), which is the interval between time points (roughly, how long it takes to acquire a complete 3D image of the brain), typically between 1 and 4 s.

Other groups have therefore used task paradigms, in which syntactic complexity was manipulated (rather than violated). Keller et al. (2001) implemented such a paradigm and additionally used high-frequency (common) or low-frequency (rare) nouns in syntactically simple and complex sentences. They found that effects of syntactic complexity and word frequency interacted in several brain areas (left IFG, STG, and inferior parietal lobe). Greater activation for syntactically complex sentences in these regions was found *only* if sentences included low frequency nouns (e.g., "The pundit that the regent attacked admitted the gaffe at the conclave"), but not for those with high-frequency nouns (e.g., "The writer that the king attacked admitted the mistake at the meeting"). While these results are not easily reconciled with findings from other studies that implemented similar manipulations of syntactic complexity, including the studies by Caplan and colleagues mentioned above, they may partly relate to limited temporal resolution. For example, in the study by Keller et al. (2001), the repetition time (Box 2) was 3 s. The authors correctly caution that any brain site, such as left IFG, may be involved in different aspects of a process at different points in time. Specialization may thus be temporal as well as spatial. Although PET and fMRI have intrinsically limited temporal resolution due to the sluggishness of blood flow increases that accompany neuronal activation, temporal resolution of event-related fMRI (Box 1) can be improved to about 1 s. For example, Ni et al. (2000) were able to identify distinct sequences of activity in response to syntactic or semantic anomalies, modeling blood oxygenation level dependent (BOLD) effects with short latency (a delay time of 4–5 s) and with longer latencies (up to 10 s). Effects for syntactic anomalies occurred in left IFG at mid-latency (about 6 s), while those in right IFG and basal ganglia were seen at long latency.

Whereas fMRI is limited in its capacity to resolve the temporal sequence of language processing (as will be further discussed in Sect. 2.3.2), attempts have been made to increase the spatial resolution of functional results. In particular, a number of studies have explored functional subdivisions within Broca's area. Dapretto and Bookheimer (1999) used a task that required either primarily syntactic or semantic operations.[9] Direct comparison yielded stronger effects for syntactic processing in a superior portion of IFG (areas 44/45), whereas stronger effects for the semantic condition were seen in a more inferior portion (area 47). Activation in area 47 has been observed in other semantic studies (Booth et al. 2002; Poldrack et al. 1999; Uchiyama et al. 2008). However, not all types of task conform to this simple model of subdivision. In Friederici et al. (2000a), participants made explicit syntactic (noun or function word?)[10] and semantic judgments (abstract or concrete noun?). While syntax-related activation in the inferior tip of area 44 was expected given the above findings, semantic judgment was unexpectedly associated with more *superior* IFG activation in area 45. This again suggests that simply characterizing a task as "syntactic" or "semantic" may be insufficient and potentially misleading since details of stimulus presentation, comparison condition, and task may affect observed activation patterns.

This raises a further question. Should activity that is localized to a specific site within a "language area", such as left IFG, be attributed to specifically *linguistic* components of a task [e.g., syntactic transformations and the movement of constituents, as suggested by Grodzinsky (2000)], or rather to components that are shared across functional domains? Broca's area is located close to premotor cortex and dorsolateral prefrontal cortex (DLPFC), and reasonable candidates of such shared components are therefore the planning motor of complex motor sequences and working memory. Sustained activity, which reflects working memory maintenance, has been observed in animal studies of neurons in DLPFC (Miller and Cohen 2001). Human imaging studies support the crucial role of DLPFC, including IFG, in working memory (Duncan and Owen 2000). Possibly related to this, Fiebach et al. (2005) observed increased activity in Broca's area associated with greater demand on syntactic working memory, as opposed to syntactic complexity. However, this leaves open the question of specificity. Does syntactic decoding involve a type of working memory that is exclusively specialized for syntactic operations, as opposed to one that is shared with non-syntactic maintenance (e.g., word lists or letter strings)? This question requires some linguistic detail and readers solely interested in the conceptual issues may move on to the next section directly.

Uchiyama and colleagues (2008) directly compared working memory in sentence and non-sentence tasks (matched for working memory load) and found greater activity in parts of left IFG (areas 44 and 45) for the sentence task. A second experiment using garden-path sentences[11] identified activation in area 44 overlapping with that from the first experiment. The authors considered these overlapping activations to

---

[9] Participants had to judge whether sentences were the same or different. For some sentences, this required syntactic processing (e.g., "The policeman arrested the thief" is the same as "The thief was arrested by the policeman"), whereas in other sentence pairs it required semantic processing ("East of the city is the lake" is different from "East of the city is the river").

[10] Function words (e.g., "if" or "by") are primarily grammatical with relatively little meaning, whereas content words (e.g., nouns such as "house" or "love") are meaningful (semantically rich).

[11] Garden-path sentences have misleading syntactic structure. A classic example is the sentence "The horse raced past the barn fell", where the final word prompts reanalysis of the initially expected syntactic structure ("The horse raced past the barn").

reflect verbal working memory. Santi and Grodzinsky (2007) used two types of sentences with filler-gap dependencies, requiring either Binding ("… *the mailman* who [gap] burnt *himself*") or Movement ("…*the woman* who Kate burnt *[gap]*"). In addition, they varied working memory load, i.e., the number of noun phrases between the dependent elements (indicated by italics in the examples above). For example, the sentence "Kate loves *the woman* who THE MAILMAN and THE MOTHER of JIM pinched *[gap]*" requires Movement between two elements (indicated by italics) separated by three noun phrases (indicated by small capitals). Participants performed grammaticality judgments during fMRI scanning. Interestingly, a site in IFG (area 45) showed an interaction effect: Activation increased linearly with working memory load for sentences in the Movement, but not in the Binding condition. The authors interpret this finding as support for a 'specificist' view, according to which "Broca's area is specific to the working memory needs of syntactic Movement rather than general dependency relations" (ibid.: 16). Not considered in this study is the possible impact of other nonverbal task components, such as inhibition. In the example sentence "Kate loves *the woman* who the mailman and the mother of Jim pinched *[gap]*", working memory maintenance during syntactic decoding may interact with inhibition of the object "the woman" in the final gap position. Indeed, activation in area 45 of left IFG has been reported in previous studies of response inhibition (Collette et al. 2001; Rubia et al. 2001). Jonides et al. (1998) found that inhibitory activation in left IFG interacted with working memory and was specifically enhanced when words had to be rejected after being maintained in working memory, possibly similar to the sentence requiring Movement cited above ("Kate loves…"). The peak activation effect in the study by Jonides et al. occurred within a few millimeters of the peak observed by Santi and Grodzinsky.

### 2.2.4 Terminology and Methodology

The debates about functional specificity and neural localization within the language system (syntax vs. semantics) and between language and other systems (e.g., language versus memory or motor planning) highlight open questions regarding the metaphor of "language in the brain". First, there is the unresolved issue of what exactly those mentalistic concepts are that cognitive neuroscience is supposed to find or locate in the brain. Churchland's eliminative materialism[12] (1986) was instrumental in debunking the idea that traditional psychological concepts could be directly related to the brain, since "folk psychology may be irreducible with respect to neuroscience—irreducible because dead wrong" (ibid.: 384). Cognitive psychologists and linguists are not immune from this criticism simply because they have invented a highly sophisticated terminology to model human behavior. The assurance that theories are "cognitive" or "mentalistic" surely helped overcome the lim-

---

[12] Churchland's eliminative materialism claims that psychology will be eventually replaced by neuroscience (1986).

itations of behaviorist black-box theories, which had ruled out any consideration of mind/brain (Gardner 1987), as mentioned earlier. However, as long as models are constrained solely by behavioral data, the constructs of a cognitive model may still remain inadequate for neuroscientific investigation. The example above from Santi and Grodzinsky (2007) showed that importing the linguistic concept of Movement into a neuroimaging experiment resulted in puzzling findings, arguably because Movement was solely considered as a syntactic operation, ignoring its potential relation to inhibition—a cognitive concept that may translate more directly into the language of neuroscience (Chambers et al. 2009).

As a general conclusion, results from imaging and other cognitive neuroscientific techniques require careful consideration of the methodological fine print. Summarized interpretations will typically resort to cognitive terminology ("syntax", "working memory"), whereas the actual results depend on the precise implementation of tasks, on the details of data acquisition and processing, and on sample size and statistical significance thresholds. Each well-designed experiment will contribute a small piece to the puzzle (for example towards an understanding of the functional organization of Broca's area), but pieces may be mislabeled.

The previous sections have shown that language is surely "in the brain", in the sense that linguistic behavior requires brain function. More specifically, there is overwhelming evidence that certain parts of the brain play crucial roles in language-related functions, in particular the regions of the left hemisphere traditionally known as Broca's and Wernicke's areas. However, beyond this bland statement things become less transparent. Imaging and lesion findings have pinpointed a number of additional regions, such as the anterior superior temporal gyrus in syntactic processes (Friederici and Kotz 2003; Friederici et al. 2003), the superior temporal sulcus (STS) in sentence comprehension (Devauchelle et al. 2008; Redcay 2008), and the cerebellum in speech production (Ackermann 2008), language comprehension (Baillieux et al. 2008), and possibly morphosyntactic[13] abilities (Justus 2004). It is obvious that these brain regions are not exclusively involved in language processing. For example, portions of the STS are considered crucial for the perception of biological motion (Puce and Perrett 2003), as well as social cognition and audiovisual integration (Hein and Knight 2008; Redcay 2008); the cerebellum participates in numerous nonverbal domains, including executive function, working memory, and attention (Baillieux et al. 2008). Some functional links are intuitively appealing; for example, those between biological motion perception, audiovisual integration and speech perception in STS (Redcay 2008), or those between action understanding, imitation, and language in Broca's area (Nishitani et al. 2005). However, except for rare studies in neurosurgery patients (Duffau 2008; Ojemann 1991), neuroscience techniques avail-

---

[13] The grammatical structure of sentences is not fully determined by word order, but also by insertion of small and relatively meaningless grammatical words (such as prepositions), changes in word endings etc. This is reflected in the term "morphosyntax". For example, interpretation of the sentence "The Rottweiler *was* bit*ten by* the mailman" is not possible based on word order alone, but requires decoding of morphological cues (passive voice indicated by the auxiliary verb "be" and the ending of the verb "bite", which makes it a participle) and the preposition "by" indicating that "mailman" is the agent of the sentence, i.e. the one who (unexpectedly) does the biting.

able for studies of living humans are limited in their spatial resolution (see Box 2) and are therefore unable to answer questions about the functional specialization of individual neurons or small assemblies of neurons ("minicolumns").[14]

## *2.2.5 Variability and Plasticity*

As a tentative conclusion, the existing literature predominantly suggests an integrative organization, with distributed networks for language that partly share neuronal resources with other functional domains. Historically, the schools of localizationist and holistic thinking were thus both partly justified because local functional specialization is accompanied by distributed network organization. However, even if precise evidence were available regarding brain loci specialized in language functions (e.g., some neurons in STS showing increased firing rates only in response to visual stimuli of biological motion, others only in response to speech stimuli), a fundamental question would remain: If such a finding is true for person A, will it equally apply to person B? The question makes little sense when it comes to individual neurons because there is simply no way to identify "the identical neuron" in two human brains. On a more macroscopic level of brain regions and their functional specialization, however, the question is reasonable and can be experimentally tested. Strangely, neuroimaging studies of individual variability are rare. Variability within groups (typically healthy "normal" young adults) is conventionally considered noise. Anatomical variability of the brain is minimized through 'spatial normalization', after which (ideally) each image element (voxel) falls onto exactly the same spot in the brain in each individual. If a voxel shows activation for a given task in nine participants out of a group of ten, lack of activity in the tenth individual reduces significance of the group finding, but is otherwise shrugged off as an uninterpretable nuisance.

This attitude is based on a convenient fiction (the universality of functional brain organization), which may hamper an adequate understanding of neurofunctional organization. In the few language studies of healthy adults that compared activation findings across individuals, substantial variability was found. Herholz and colleagues (1996) studied seven adults during covert verb generation and found dramatic differences in magnitude and localization of activation in left IFG and in overall patterns of cerebral blood flow changes. However, task compliance could not be monitored in this study because participants were producing words only in their heads (covertly). A second study by Xiong et al. (2000) instead used overt verb generation in a larger sample of 20 young adults. Interestingly, highest consistency of activation was found in motor regions. Areas 44 and 47 in left IFG showed expected groupwise activation, but were not activated in four of the participants. Handedness could affect these find-

---

[14] A minicolumn includes about 100 neurons with strong excitatory interconnections. These neurons are vertically arranged, giving the impression of a 'column'. Minicolumns are considered basic functional units of mammalian cortex (Mountcastle 1997).

ings, but was not reported by Xiong and colleagues. Seghier and colleagues (2004) included 30 exclusively right-handed participants in an fMRI study of phonological rhyme judgments and semantic category decisions. Although Broca's area (areas 44, 45, and 47) showed robust groupwise activation and was activated in most participants (88 and 96 % for phonological and semantic tasks, respectively), the individual variability of peak activation loci was high in the left IFG for both tasks.

Quantifying individual variability in neuroimaging is not entirely straightforward, and methodological details such as smoothing (image blurring) and spatial normalization (warping to a standard anatomy) will affect results. However, ample evidence is available from other techniques to appreciate that variability reflects basic principles of functional differentiation (rather than measurement error). Drastic manipulations of brain development in animal models have shown *cross-modal plasticity* in differentiation of the cerebral cortex (O'Leary and Nakagawa 2002; Sur and Leamey 2001). For example, 'rewiring' of connections in deep subcortical regions can transform temporal *auditory* cortex into *visual* cortex, if temporal cortex receives input from visual nuclei in the thalamus during critical developmental periods (Sur et al. 1990). Cortical tissue can also be transplanted during early development and will assume the functional organization of the surrounding 'host' region. Developing occipital cortex will thus assume somatosensory specialization (such as tactile perception), if it is transplanted into the postcentral region and receives input from somatosensory nuclei from the thalamus (Schlaggar and O'Leary 1991). Human imaging studies suggest that similar principles of cross-modal plasticity are at work in people with early loss or congenital absence of one sensory modality. Temporal 'auditory' cortex assumes visual functions in deafness (Finney et al. 2001), and occipital 'visual' cortex takes over tactile functions in congenitally blind people, especially during Braille reading (Sadato et al. 2002).

How is this related to individual variability of language networks? Evidence of cross-modal plasticity can be considered the 'tip of the iceberg' of experience-driven plasticity that is at work in every typically developing brain. Probably the strongest set of evidence for experience-based changes in healthy brain organization comes from studies in professional musicians. Imaging studies have shown effects of musical experience on functional organization in motor (Elbert et al. 1995; Hund-Georgiadis and von Cramon 1999) and auditory cortex (Pantev et al. 1998), as well as on brain anatomy in these regions (Amunts et al. 1997; Hyde et al. 2009; Schlaug et al. 2005). One study using magnetic resonance spectroscopy (Aydin et al. 2005) found that neuronal numbers and function in the left superior temporal gyrus were significantly higher in professional musicians than in non-musicians. It is especially compelling that in all of these studies functional and anatomical plasticity was related to the age of inception of musical practice, which suggests that plasticity reflects the amount of lifetime experience.

Another manifestation of plasticity, i.e., plasticity following brain damage, had been proposed by nineteenth century neurologists, such as Hughlings-Jackson and Head, as discussed in Sect. 2.2.1. With modern neuroimaging techniques, functional reorganization for language can now be experimentally demonstrated. PET and fMRI studies have shown that early-onset damage to the left hemisphere often results in

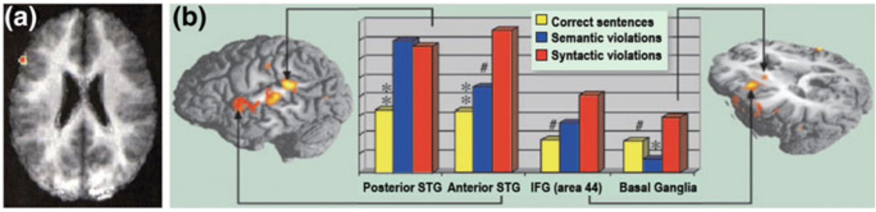


**Fig. 2.3** Results from two fMRI studies of syntactic processing, as described in main text. (**a**) Caplan et al. (1999) find only a single and small activation in area 44 of the left hemisphere, based on an *a priori* hypothesis. (**b**) Friederici et al. (2003) observe effects related to syntactic processing in distributed left perisylvian regions. *Bars* in the *center* show relative activity in several left-hemisphere regions (posterior and anterior STG; frontal operculum; basal ganglia) *color-coded* by condition. Pronounced effects for the syntax condition are seen in anterior superior temporal gyrus, inferior frontal operculum, and basal ganglia, but these regions also show some degree of activation for the other two conditions

reorganization of language to the right hemisphere, if the right hemisphere remains intact (Guzzetta et al. 2008; Lidzba et al. 2006; Liégeois et al. 2008; Müller et al. 1999; Staudt et al. 2002; Fig. 2.3).

Greater plasticity in children than in adults can be largely attributed to developmental changes in synaptic density. Early in postnatal development, the number of synapses is extremely high (Bourgeois 2001). During child development, synaptic pruning dramatically reduces the number of synapses (Chugani et al. 1987). The early period of synaptic ‘abundance’ corresponds to greater plasticity, most likely because synaptic pruning is still in process and can be partly suspended, resulting in the emergence of alternative brain networks (Müller 2004; Müller and Courchesne 2000). For example, connections between Broca’s and Wernicke’s area in the left hemisphere may be lost due to early injury. Corresponding connections in the right hemisphere, which in typical development will be mostly pruned away, may now be maintained and strengthened, resulting in atypical right hemisphere dominance for language. The period of intense plasticity roughly corresponds to the ‘critical period for language acquisition’ (Lenneberg 1967), reflected in the easy and apparently automatic acquisition of language(s) in children compared to more effortful and limited acquisition of new languages in adolescents and adults.

In conclusion, individual variability of neural networks for language is largely a reflection of brain plasticity. Functional localization (e.g., the existence of Broca’s and Wernicke’s areas) is roughly universal, but a more detailed and microscopic view tends to show variability that is partly experience-driven, partly genetic (see Müller 2009 for review; Thompson et al. 2001). The role of experience, i.e., brain activity in response to and interaction with the environment, and the impact of gene-based brain maturation illustrates that specialized language networks are the product of a dynamic developmental process, rather than a set of stationary locations.

## 2.3 Through the Brain: Language as a Distributed Process

### *2.3.1 Ontogeny*

Studies in infants suggest some degree of early specialization related to language. Near-infrared spectroscopy (NIRS; see Box 3) has recently become a method of choice in the study of infants (Minagawa-Kawai et al. 2008). Differential responses to spoken language in left temporal lobe and visual stimulation in occipital cortex have been reported for infants 6–9 months of age (Bortfeld et al. 2007). Minagawa-Kawai and colleagues (2007) found that changes in vowel length across phonemic category (compared to within-category changes)[15] were associated with greater blood flow response in temporal cortex in 6–7 months old infants. However, a lateralized response (greater in the left than in the right hemisphere) was only seen after 13 months of age.

A similar pattern of change has been observed in event-related potential (ERP) studies (Box 3), which combine EEG with stimulus presentation. Cheour and colleagues used ERP mismatch negativity[16] and showed that newborns were able to learn phonological discriminations during sleep (Cheour et al. 2002) and that phonemic distinctions of the native language emerge in infants before age 12 months (Cheour et al. 1998). Whereas basic phonemic abilities thus seem present early in infancy, word acquisition begins slightly later. Mills et al. (1997) found that words understood by 13–17-month-old children were associated with widespread and bilateral ERP changes (when compared to unknown words). In 20-month-old children, ERP differences between known and unknown words were seen primarily in temporo-parietal regions of the left hemisphere, suggesting increasing localization and lateralization of word meanings.

---

[15] For an example, consider the length of the vowel /i/ in the word “fit”, which is around 60 ms. With length of > 100 ms, a phonemic boundary will be crossed and the vowel will be perceived as a long /iː/, resulting in the word “feet”. On the other hand, a change in vowel length from 100 to 150 ms will not be perceived as a different phoneme (and the perceived word will remain “feet”). Whereas acoustic length of the vowel can thus vary in graded fashion, some relatively small changes in vowel length around the category boundary will be perceived as distinct phonemes that can change the meaning of a word.

[16] The mismatch negativity (MMN) is an EEG response that follows about 150–250 ms after a deviant stimulus (e.g., a high tone among a series of low tones, or an omitted click in an otherwise regular sequence of clicks; Cheour et al. 2000). It can be reliably measured even in infants and therefore provides a technique for demonstrating perceptual discrimination at an age when verbal responses cannot be obtained. If an infant shows an MMN to stimulus A in a series of presentations of stimulus B, one can infer that the infant brain is able to distinguish A from B.

**Box 3. Near-infrared spectroscopy** (NIRS) is an optical imaging technique detecting changes in blood oxygenation through the skull. It is non-invasive and relatively insensitive to motion artifacts, which makes it ideal for the study of infants and small children. However, spatial resolution (see Box 2) is low. Two other techniques suitable for the study of small children detect neuronal activity directly (rather than the blood flow changes associated with it). **Event-related potentials** (ERPs) are an adaptation of traditional electroencephalography (EEG) and use microelectrodes on the scalp to detect small changes in electrical currents triggered by a stimulus or task. **Magnetoencephalography** (MEG) relies on the same electromagnetic effects of neuronal "activity, but detects changes in magnetic fields outside the skull in a shielded environment."

However, neither NIRS nor ERP provide sufficient spatial resolution to pinpoint developmental changes in language organization. FMRI studies of infants remain rare, given the motion sensitivity of this technique. Studying infants 2–3 months of age, Dehaene-Lambertz and colleagues (2002) could show activation in STG in response to speech, which was stronger in the left than in the right hemisphere for both forward and backward speech. In a follow-up study (Dehaene-Lambertz et al. 2006), activation in the STG during presentation of sentences was found to be bilateral in 3-months olds. Interestingly, response in Broca's area was delayed (occurring several seconds after STG activation) and enhanced when sentences were repeated verbatim (compared to novel sentences). The authors speculate that this finding may reflect "a sentence learning mechanisms" already at work in the left IFG, which they relate to this region's status as an "integrative node common to perception and production" and the participation of IFG in emerging mirror neuron networks crucial for imitation and action understanding (Nishitani et al. 2005). I will return to this theme in Sect. 2.4.

Imaging and electrophysiological studies in infants suggest *language readiness* of left perisylvian regions (those surrounding the Sylvian fissure, see Fig. 2.4). This is not equivalent to fully established *functional specialization*, but indicates that STG in the left hemisphere tends to have an early advantage in processing complex and fast-changing auditory stimuli (Zaehle et al. 2004). It is less clear, though, whether language readiness in left STG and IFG reflects what Greenough et al. (1987) call *experience-expectant* properties, which primarily apply to evolutionarily old sensorimotor systems.[17] Chomsky's metaphor of language "growing" in the child similar to a bodily organ (Chomsky and Ronat 1979: 83–84) would imply that plasticity in these regions is experience-expectant. This appears to be supported by studies in congenitally deaf users of sign language showing that left perisylvian cortex is

---

[17] Greenough et al. (1987, p. 540) define experience-expectant plasticity as "designed to utilize the sort of environmental information that is ubiquitous and has been so throughout much of the evolutionary history of a species." An example is the development of ocular dominance columns, which are small vertically organized units in visual cortex that respond to information coming from one eye only.

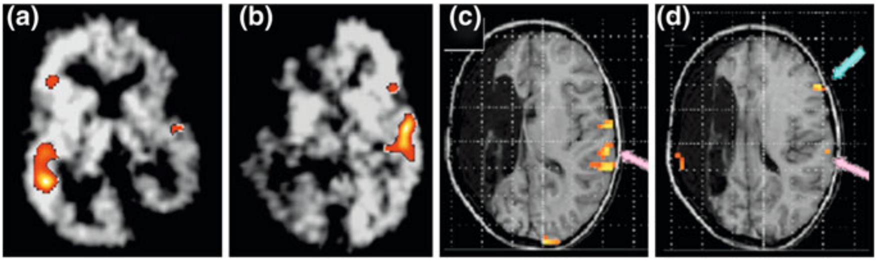


**Fig. 2.4** Clinical studies showing effects of developmental plasticity. (**a**)–(**b**) PET studies showing activation for listening to sentences in two children with Sturge-Weber syndrome, a congenital condition that results in progressive calcification of one hemisphere, leaving the other intact. (**a**) A child with calcification and shrinkage of the right hemisphere shows a typical pattern of left hemisphere activation in superior and middle temporal gyri and inferior frontal lobe. (**b**) A second patient, with damage to the left hemisphere shows almost a mirror image of activity, in inferior frontal and superior/middle temporal gyri of the *right* hemisphere (see Müller et al. 1998 for complete data). The capacity for language to reorganize into the right hemisphere early in life is further illustrated in (**c**)–(**d**) by the example of a patient with left hemispherectomy (removal of the left hemisphere. Activations during overt word generation are seen in right IFG (*blue arrow*) and right superior and middle gyri (*magenta arrows*). (**c**)–(**d**) adapted from Liégeois et al. (2008)

crucial for language acquisition even when one typical sensory modality of language learning (hearing) is missing (Campbell et al. 2008). Conversely, the finding of language reorganization into the right hemisphere in children with early left hemisphere damage, as described in Sect. 2.2.5, may suggest that plasticity at work in emerging language systems is to some extent *experience-dependent*, i.e., partially dependent on environmental interaction.

While regional proto-specialization or ‘readiness’ for language can thus be observed surprisingly early in the infant brain, imaging studies in older children have shown substantial changes in language networks over time. One finding is the increase in lateralization of language-related activity with age (Brauer and Friederici 2007; Holland et al. 2001; Ressel et al. 2008). Holland and colleagues (2001) found that leftward asymmetry of whole-brain activity during covert verb generation correlated significantly with age in 7–18 year-old children, although a follow-up in a larger sample suggests that age-dependent lateralization is subtle and task-dependent (Holland et al. 2007). Some studies have found age-dependent increase of activation in Broca’s area for verb generation (Holland et al. 2001; Szaflarski et al. 2006) and phonological rhyme judgment (Booth et al. 2004). However, this intuitively appealing finding is not always replicated. For example, Brauer and Friederici (2007) used syntactic and semantic violation tasks (as described earlier) and found left IFG activation in children for all types of sentences (including those without violations), whereas activation in adults was confined to area 44 and only observed for sentences with violations. Conversely, Schlaggar, Petersen, and colleagues (Brown et al. 2005; Schlaggar et al. 2002) identified positive age-dependent effects in left frontal cortex using several overt word generation tasks, which require spoken production of words and are preferable to covert generation (without articulated speech) because

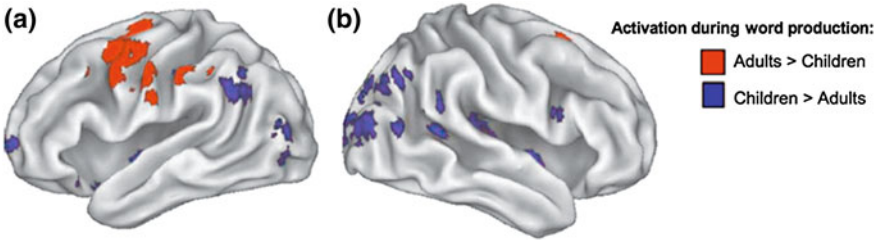


**Fig. 2.5** Brain regions showing age-dependent differences in activation during word production, in an fMRI study of children and adults between 7 and 32 years of age. *Red clusters*, mostly in motor and premotor regions of the left hemisphere (**a**), show greater activation in adults than in children. Conversely, greater activation in children (*blue clusters*) is seen mostly in posterior parts of the brain, in both the left (**a**) and the right hemispheres (**b**). Adapted from Brown et al. (2005)

performance can be monitored (Fig. 2.5). The site of age-dependent increase in these studies was not the IFG, however, but premotor area 6, possibly related to motor planning. The overall pattern in these studies suggests greater activity in children on the perceptual side (receiving a word stimulus) and greater activity in adults on the response side (articulating a word). It is unclear whether greater activity in visual cortex, which was very pronounced in children up to age 10 years, but leveled off in older children and adults, represents simply different attention-related strategies of dealing with a verbal stimulus-response task (young children being more input-oriented), or whether it reflects a stronger perceptual dependence of word representations in children.

This brief review[18] suggests substantial developmental changes in language organization during childhood. However, as in previous discussions of language organization in adults, simple generalized conclusions (such as increasing leftward asymmetry or increasing activity in left IFG) are thwarted by complex results and dependence of findings on task designs and methodological detail. For example, Brauer and Friederici (2007), as mentioned, could not find evidence for greater left IFG activity in adults than in children. However, when further inspecting activation in Broca's area in a follow-up study (Brauer et al. 2008), they found that BOLD effects (Box 1) were delayed in children around age 6 years, compared to young adults. Such delay was not seen in auditory cortex in the left temporal lobe, suggesting that it might not be simply explained by a slow blood flow response, but by delayed neuronal activity in Broca's area. This study highlights dynamic changes in language processes on a smaller time scale, i.e., changes observable during a single language-related process, to which I will turn now.

---

[18] For a more comprehensive review of the literature on developmental change in language networks see Müller and Palmer (2008).

### 2.3.2 Microgeny

*Microgeny* relates to the unfolding of a cognitive process that typically occurs on a scale of milliseconds to seconds. The term will be used here in analogy to ontogeny (and phylogeny), as a time scale of neurocognitive change, and not as a theory assuming parallels between time scales, as proposed by Brown (1988).

Given the low temporal resolution of PET and fMRI, the technique of choice for microgenetic studies of language has been ERP (Box 3). At least four distinct electrophysiological components have been attributed to different language subprocesses (reviewed by Friederici and Kotz 2003). An early negative component (N100) corresponding to phonological analysis is followed by an early left anterior negativity (ELAN) at about 150–200 ms after stimulus presentation, which reflects first-pass syntactic decoding (cf. also Neville et al. 1991). This is followed at about 400 ms by a semantic component (N400), and finally in some cases of reanalysis or repair (e.g., for sentences with complex structure or violations) by a positive component (P600). While these ERP components can be indirectly related to imaging and lesion findings for clues on anatomical sources, magnetoencephalography (MEG; Box 3) provides an improved combination of high temporal resolution and better spatial resolution, as signals are less distorted by the skull in MEG than in EEG (Wheless et al. 2004).

Magnetic fields corresponding to the N100 and associated with phonological processing can expectedly be localized to sources from the posterior STG, near the primary auditory cortex (as reviewed in Salmelin 2007). MEG components corresponding to the syntactic ELAN have been identified less consistently (Knösche et al. 1999; Kubota et al. 2005; Kwon et al. 2005). This may relate to differences between EEG and MEG. Signals predominantly originate from neurons at the top ("crest") of a gyrus in ERP, but from neurons buried in the sulci in MEG. As a consequence, ERP and MEG data may not always coincide. Friederici et al. (2000b) therefore used fMRI to localize brain activity that was the source of magnetic fields and were thus able to identify an MEG component that peaked around 150 ms after presentation of a syntactic violation (e.g., "The fish was in *caught*"). Interestingly, the sources were located in STG (just anterior to the primary auditory cortex) and IFG of *both* hemispheres. This is surprising given the overwhelming lesion and imaging evidence of left-hemisphere lateralization for syntax (discussed in previous sections). However, a few imaging studies have also reported right hemisphere effects related to syntactic complexity (Just et al. 1996), syntactic violations (Newman et al. 2001), and repair[19] of such violations (Meyer et al. 2000). Relatively early morphosyntactic effects (with a latency of around 200 ms) have also been shown in frontal and tempo-

---

[19] Meyer and colleagues (2000) instructed some of their participants to repair sentence violations. For example, when presented with the sentence "The spy was in the caught", participants would not only indicate that the sentence was ungrammatical, but also covertly fix the violation, generating "The spy was caught" in their minds (without speaking). Activation for this repair condition was particularly strong in right IFG and STG.

ral regions for grammatical gender decisions in an MEG study with German speakers (Härle et al. 2002).[20]

Maess and colleagues (2006) used MEG for more accurate localization of the N400 that has been reported for semantic violations in numerous ERP studies over the past 30 years (Kutas and Hillyard 1984). Corresponding magnetic evoked fields (MEFs) were found in distributed frontal and temporal regions (areas 45 and 47 in IFG, STG, and inferior temporal gyrus). While effects were more robust in the left hemisphere, most corresponding (homotopic) regions in the right hemisphere also showed effects of semantic violation. MEFs with long latencies ($\geq 400$ ms) have also been found for sentence and word stimuli in meaningful contexts, without violations (reviewed in Salmelin 2007).

The findings described above seem to suggest that syntactic processes happen generally before semantic ones. However, this simplification is not entirely correct. Some MEG studies have detected semantic MEFs at earlier stages. Shtyrov and Pulvermüller 2007 used word pairs, such as "ball-kick". MEFs related to semantic mismatch (e.g., "ball-eat") peaked in left temporal cortex at $\sim$115 ms latency, followed by a weaker MEF in the inferior frontal lobe 16 ms later. On the other hand, both semantic *and syntactic* anomalies have also been found to be associated with prolonged effects, often occurring at $>$500 ms latency. Kwon et al. (2005) used Korean stimuli with sentence-final morphosyntactic or semantic violations. Both violation types were associated with MEFs in left temporal lobe peaking at a latency of 600 ms.

## 2.4 *Across the Brain*: Language Connected

Overall, the neuroscientific evidence on language suggests a picture that differs from what Chomsky may have imagined when he declared language to be a gene-based biological object. Although it cannot be entirely ruled out that a neural architecture uniquely specialized for language and genes that specify this architecture might yet be discovered (cf. Fisher and Marcus 2006), the prospect of such innately predetermined autonomy (or modularity) of language is faint. The findings reviewed above show that several brain regions are quite consistently involved, mostly in the left perisylvian cortex. Some of them, such as left IFG and STG, correspond to Broca's and Wernicke's areas that have been considered 'language areas' since early studies of aphasic patients. However, the closer one inspects the functional characteristics of these 'language areas' from a cognitive neuroscience perspective (rather than a narrow neurolinguistic view), the more these areas appear to assume 'multiple personalities'. Broca's area is called by the name of "syntax" in some studies (Caplan

---

[20] Unlike English, German morphosyntax (cf. Footnote 13) relies heavily on grammatical gender since every noun has one of three genders (feminine, masculine, neuter) and articles and pronouns change in agreement with the gender of the corresponding noun. For example, in "Die Tüte riss und ihr Inhalt fiel heraus" ("The bag tore and its contents fell down"), the article "die" and the pronoun "ihr" have the feminine form, in agreement with the grammatical gender of "Tüte".

et al. 1999; Heim et al. 2003), "semantics" in another (Homae et al. 2002), and "phonology" in yet another (Fiez et al. 1993). Then the plot thickens and we hear the same persona called "imitation" (Heiser et al. 2003), "motor preparation" (Krams et al. 1998), "planning" (Fincham et al. 2002) and "imagery" (Binkofski et al. 2000), "action understanding" (Buccino et al. 2004; Hamzei et al. 2003), "visuomotor coordination" (Müller and Basho 2004), "sequence learning" (Haslinger et al. 2002), "tonal discrimination" (Müller et al. 2001), "artificial grammar learning" (Bahlmann et al. 2008), "working memory" (Nystrom et al. 1998), "rule shifting" (Konishi 1998), "response selection" (Thompson-Schill et al. 1997), "response inhibition" (Collette et al. 2001) and so on. As there is no technique allowing neuroscientists to probe for functional preferences of individual neurons in the healthy human brain, it remains theoretically possible that each of these specializations is entirely separate from linguistic specializations, and that the left IFG consists of a large array of functionally discrete modules. One may invoke the known columnar organization of cortex (Mountcastle 1997; cf. Sect. 2.3) and speculate that neighboring columns might have distinct and unrelated functions.

However, neuroscientists who have learned their trade beyond the confines of linguistics will tend to find such conjectures far-fetched. One reason is that cortical connections are predominantly local (Braitenberg 1991) and that neighboring sites therefore tend to participate in similar (though subtly specialized) functions. While one may cling to historical ideas of innate specialization, autonomy of language and modularity, the neuroscience of 'language areas' provides a unique opportunity for a deeper understanding of how children develop the ability to understand and produce words and string them together in phrases and sentences. Imaging and other neuroscience techniques provide a complex picture of the brain regions involved in language. By learning what else these regions are also involved in, we obtain clues as to what the cognitive and sensorimotor components and precursor functions may be that support and guide language acquisition (for detailed review see Müller 2009).

Our understanding of how relatively elementary functional specializations come together in support of more complex higher cognitive functions has been boosted by recent developments in the study of white matter, which contains axons connecting different parts of the brain. In particular, diffusion tensor imaging, an MRI technique that can be used for mapping out fiber tracts in the brain (Le Bihan 2003), has been applied to language-related connections. Traditional notions of the main pathway connecting Broca's and Wernicke's areas, known as the arcuate fasciculus, had to be modified recently based on evidence from diffusion tensor imaging (Catani et al. 2005; Fig. 2.6).

The study of connectivity is crucial because it helps overcome the idea of intrinsic local specialization that seems tied to the localizationist tradition. Each cortical neuron is connected to thousands or tens of thousands of other neurons (Abeles 1991). The functional "specialization" of each neuron can therefore be defined by its connectivity pattern, rather than as an intrinsic property (Passingham et al. 2002). This is

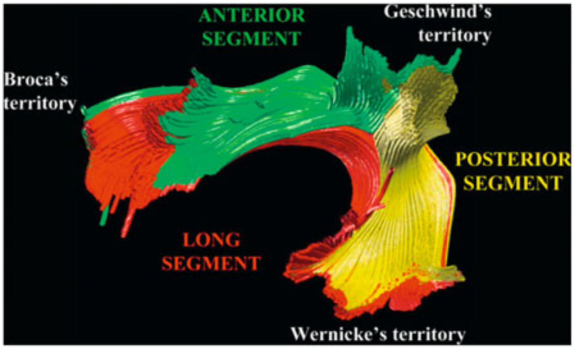


**Fig. 2.6** Reconstruction of white matter fibers corresponding connecting Broca's and Wernicke's areas, by means of diffusion tensor imaging (DTI). This study, which was performed in vivo in eleven healthy right-handed men, shows that there are two pathways, a direct pathway connecting Broca's and Wernicke's area, and an indirect pathway via the inferior parietal lobe ("Geschwind's territory"). From Catani et al. (2005)

a principle of simple Hebbian logic,[21] as synaptic connections are strengthened from simultaneous firing of pre- and postsynaptic neurons, and thus simultaneous firing will reflect shared sensorimotor or cognitive representations (Kandel et al. 2000). Applying such neuroscientific truisms to a 'language area', such as Broca's area, we obtain a glimpse of this region's true colors by considering the complex connectivity of IFG (Anwander et al. 2007; Friederici et al. 2009). This includes connectivity with inferior parietal and lateral temporal regions in posterior perisylvian cortex via the arcuate fasciculus (Fig. 2.6), as well as connectivity along the dorsal stream, which is crucial for visuospatial processing and visuomotor coordination (Goodale and Westwood 2004); the mirror neuron system (Rizzolatti and Craighero 2004), considered to be involved in imitation, action understanding, and possibly some aspects of social cognition (Pineda 2008; see Chap. 3); and the ventral stream (Rilling et al. 2008; Saur et al. 2008), known to provide meaningful interpretation of visual and auditory stimuli (Grill-Spector and Malach 2004). The best available hypothesis on why Broca's area is also a 'language area' is that this terrain of cortex has the ideal connectivity patterns bringing together information that a child needs to acquire language. A similar argument has been made for posterior perisylvian cortex, such as STS (Redcay 2008).

---

[21] Donald Hebb (1949) hypothesized that a synapse is strengthened as a result of synchronous firing of the two neurons the synapse connects. This idea, popularized in the phrase "What fires together wires together", has proven basically accurate.

## 2.5 Final Note

In the localizationist tradition, which has been boosted by the advent of modern techniques of functional "mapping" (such as fMRI), language tends to be regarded as a location (or a set of locations) in the brain. However, functional brain organization is characterized by plasticity throughout life. Experience and activity are associated with continuous change, which may be anywhere between subtle and dramatic (as in professional musicians or the congenitally blind). Plasticity is particularly pronounced during child development and in response to injury. On a smaller time scale, links between language and the brain are also characterized by constant change. A language process (e.g., comprehending an utterance and responding to it) involves a large number of brain regions far beyond the classic language areas of Broca and Wernicke. Anatomical evidence and imaging of fiber tracts suggests that many of these regions are tightly connected, allowing them to cooperate functionally. In view of the functional characteristics of regions participating in language networks, it appears likely that language emerges from processes shared with numerous functional domains traditionally considered nonverbal. Electrophysiological evidence further suggests that the profile of regional activity changes on a millisecond by millisecond basis. Today it is not fully understood what the 'neural code' of language processing may be. Single neurons are unlikely to possess complex intrinsic representational capacities. For example, there will be no neuron (or small set of neurons) fully representing the meaning of the word "cat". However, it is possible that the complex connectivity of neurons relates more directly to their representational capacity. The synchronous activity of a distributed set of well-connected neurons in several areas of the brain (e.g., IFG and visual, auditory, somatosensory, and premotor cortices) may more fully represent the meaning of the word "cat" (what it looks like, what sounds it makes, what it feels like, how one can interact with it; Martin 2007; McClelland and Rogers 2003; Hwangetal 2009). If this model is true, the question remains of how exactly distributed sets of neurons are transiently tied together to form a representation or cognitive state. Animal studies suggest that synchronous firing patterns play an important role, especially phase-locked oscillations in the gamma band (around 40 Hz; Singer 2001), or in even higher frequency domains (Canolty et al. 2007). However, correlated distributed activity within language networks can also be seen in very low frequency domains (below 0.1 Hz, Hampson et al. 2002) and it remains possible that many different frequency bands of neuronal communication contribute in a complex hierarchical organization (Lakatos et al. 2008) to language representations and processing.

**Acknowledgments** Preparation of this review was supported by the National Institutes of Health, grants R01-NS43999, R01-DC006155 and R01-MH081023. Thanks to Patricia Shih for help in editing the manuscript.

## References

Abeles, M.: Corticonics. Cambridge University Press, Cambridge (1991)

Ackermann, H.: Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives. Trends Neurosci. **31**, 265–272 (2008)

Amunts, K., Schlaug, G., Jäncke, L., Steinmetz, H., Schleicher, A., Dabringhaus, A., Zilles, K.: Motor cortex and hand motor skills: structural compliance in the human brain. Hum. Brain Mapp. **5**, 206–215 (1997)

Anwander, A., Tittgemeyer, M., von Cramon, D.Y., Friederici, A.D., Knosche, T.R.: Connectivity-based parcellation of Broca's area. Cereb. Cortex **17**, 816–825 (2007)

Aydin, K., Ciftci, K., Terzibasioglu, E., Ozkan, M., Demirtas, A., Sencer, S., Minareci, O.: Quantitative proton MR spectroscopic findings of cortical reorganization in the auditory cortex of musicians. Am. J. Neuroradiol. **26**, 128–136 (2005)

Bahlmann, J., Schubotz, R.I., Friederici, A.D.: Hierarchical artificial grammar processing engages Broca's area. Neuroimage **42**, 525–534 (2008)

Baillieux, H., De Smet, H.J., Paquier, P.F., De Deyn, P.P., Marien, P.: Cerebellar neurocognition: insights into the bottom of the brain. Clin. Neurol. Neurosurg. **110**, 763–773 (2008)

Benson, D.F., Ardila, A.: Aphasia. Oxford University Press, New York (1996)

Binkofski, F., Amunts, K., Stephan, K.M., Posse, S., Schormann, T., Freund, H.J., Zilles, K., Seitz, R.J.: Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. Hum. Brain Mapp. **11**, 273–285 (2000)

Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M.: Modality independence of word comprehension. Hum. Brain Mapp. **16**, 251–261 (2002)

Booth, J.R., Burman, D.D., Meyer, J.R., Gitelman, D.R., Parrish, T.B., Mesulam, M.M.: Development of brain mechanisms for processing orthographic and phonologic representations. J. Cogn. Neurosci. **16**, 1234–1249 (2004)

Bortfeld, H., Wruck, E., Boas, D.A.: Assessing infants' cortical response to speech using near-infrared spectroscopy. Neuroimage **34**, 407–415 (2007)

Brauer, J., Friederici, A.D.: Functional neural networks of semantic and syntactic processes in the developing brain. J. Cogn. Neurosci. **19**, 1609–1623 (2007)

Bourgeois, J.P., Synaptogenesis in the neocortex of the newborn: The ultimate frontier for individuation? In: Nelson, C.A., Luciana, M. (Eds.), Handbook of Developmental Cognitive neuroscience. MIT Press, Cambridge (MA), pp. 23–34 (2001)

Braitenberg, V., Schüz, A., Anatomy of Cortex. Statistics and Geometry. Springer, Berlin (1991)

Brauer, J., Neumann, J., Friederici, A.D.: Temporal dynamics of perisylvian activation during language processing in children and adults. Neuroimage **41**, 1484–1492 (2008)

Broca, P.: Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bulletins et Mémoires de la Société Anatomique de Paris **36**, 330–357 (1861)

Brown, J.W.: Introduction: Microgenetic theory. In: Brown, J.W. (ed.) The Life of the Mind, pp. 1–26. Lawrence Erlbaum, Hillsdale (1988)

Brown, T.T., Lugar, H.M., Coalson, R.S., Miezin, F.M., Petersen, S.E., Schlaggar, B.L.: Developmental changes in human cerebral functional organization for word generation. Cereb. Cortex **15**, 275–290 (2005)

Buccino, G., Binkofski, F., Riggio, L.: The mirror neuron system and action recognition. Brain Lang. **89**, 370–376 (2004)

Campbell, R., MacSweeney, M., Waters, D.: Sign language and the brain: a review. J. Deaf Stud. Deaf Educ. **13**, 3–20 (2008)

Canolty, R.T., Soltani, M., Dalal, S.S., Edwards, E., Dronkers, N.F., Nagarajan, S.S., Kirsch, H.E., Barbaro, N.M., Knight, R.T.: Spatiotemporal dynamics of word processing in the human brain. Front. Neurosci. **1**, 185–196 (2007)

Caplan, D., Alpert, N., Waters, G.: Effects of syntactic structure and propositional number on patterns of regional cerebral blood flow. J. Cogn. Neurosci. **10**, 541–552 (1998)

Caplan, D., Alpert, N., Waters, G.: PET studies of syntactic processing with auditory sentence presentation. Neuroimage **9**, 343–351 (1999)

Catani, M., Jones, D.K., ffytche, D.H.: Perisylvian language networks of the human brain. Ann. Neurol. **57**, 8–16 (2005)

Chambers, C.D., Garavan, H., Bellgrove, M.A.: Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. Neurosci. Biobehav. Rev. **33**, 631–646 (2009)

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Naatanen, R.: Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. **1**, 351–353 (1998)

Cheour, M., Leppanen, P.H., Kraus, N.: Mismatch negativity (MMN) as a tool for investigating auditory discrimination and sensory memory in infants and children. Clin. Neurophysiol. **111**, 4–16 (2000)

Cheour, M., Martynova, O., Naatanen, R., Erkkola, R., Sillanpaa, M., Kero, P., Raz, A., Kaipio, M.L., Hiltunen, J., Aaltonen, O., Savela, J., Hamalainen, H.: Speech sounds learned by sleeping newborns. Nature **415**, 599–600 (2002)

Chomsky, N.: Syntactic Structures. Mouton, s'-Gravenhage (1957)

Chomsky, N.: Review of B.F. Skinner "Verbal Behavior". Language **35**, 26–58 (1959)

Chomsky, N.: Aspects of the Theory of Syntax. MIT Press, Cambridge (1965)

Chomsky, N.: Cartesian Linguistics. Harper & Row, New York (1966)

Chomsky, N.: Language and Mind. Harcourt Brace Jovanovich, New York (1972a)

Chomsky, N.: Psychology and ideology. Cognition **1**, 11–46 (1972b)

Chomsky, N.: Reflections on Language. Temple Smith, London (1976)

Chomsky, N.: Rules and Representations. Columbia University Press, New York (1980)

Chomsky, N., Ronat, M.: Language and Responsibility. Pantheon, Hassocks (1979)

Chugani, H.T., Phelps, M.E., Mazziotta, J.C.: Positron emission tomography study of human brain functional development. Ann. Neurol. **22**, 487–497 (1987)

Churchland, P.S.: Neurophilosophy. MIT Press, Cambridge (1986)

Clarke, E., Dewhurst, K.: An Illustrated History of Brain Function. Sanford, Oxford (1972)

Collette, F., Van der Linden, M., Delfiore, G., Degueldre, C., Luxen, A., Salmon, E.: The functional anatomy of inhibition processes investigated with the Hayling task. Neuroimage **14**, 258–267 (2001)

Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L.: Functional neuroimaging of speech perception in infants. Science **298**, 2013–2015 (2002)

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., Dehaene, S.: Functional organization of perisylvian activation during presentation of sentences in preverbal infants. Proc. Natl. Acad. Sci. U.S.A **103**, 14240–14245 (2006)

Devauchelle, A.D., Oppenheim, C., Rizzi, L., Dehaene, S., Pallier, C.: Sentence syntax and content in the human temporal lobe: an fMRI adaptation study in auditory and visual modalities. J. Cogn. Neurosci. **21**, 1000–1012 (2008)

Dronkers, N.F., Plaisant, O., Iba-Zizen, M.T., Cabanis, E.A.: Paul Broca's historic cases: high resolution MR imaging of the brains of Leborgne and Lelong. Brain **130**, 1432–1441 (2007)

Duffau, H.: The anatomo-functional connectivity of language revisited. New insights provided by electrostimulation and tractography. Neuropsychologia **46**, 927–934 (2008)

Duncan, J., Owen, A.M.: Common regions of the human frontal lobe recruited by diverse cognitive demands. Trends Neurosci. **23**, 475–483 (2000)

Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., Taub, E.: Increased cortical representation of the fingers of the left hand in string players. Science **270**, 305–307 (1995)

Fiebach, C.J., Schlesewsky, M., Lohmann, G., Von Cramon, D.Y., Friederici, A.D.: Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. Hum. Brain Mapp. **24**, 79–91 (2005)

Fiez, J.A., Raichle, M.E., Tallal, P., Petersen, S.E.: Activation of left frontal area near Broca's area during auditory detection and phonological access tasks. J. Cereb. Blood Flow Metab. **13**(Suppl. 1), S504 (1993)

Fincham, J.M., Carter, C.S., van Veen, V., Stenger, V.A., Anderson, J.R.: Neural mechanisms of planning: a computational analysis using event-related fMRI. Proc. Natl. Acad. Sci. U.S.A **99**, 3346–3351 (2002)

Finney, E.M., Fine, I., Dobkins, K.R.: Visual stimuli activate auditory cortex in the deaf. Nat. Neurosci. **4**, 1171–1173 (2001)

Fisher, S.E., Marcus, G.F.: The eloquent ape: genes, brains and the evolution of language. Nat. Rev. Genet. **7**, 9–20 (2006)

Fodor, J.A.: The Modularity of Mind. MIT Press, Cambridge (1983)

Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C.: Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. **33**, 636–647 (1995)

Friederici, A.D., Kotz, S.A.: The brain basis of syntactic processes: functional imaging and lesion studies. Neuroimage **20**(Suppl 1), S8–17 (2003)

Friederici, A.D., Opitz, B., von Cramon, D.Y.: Segregating semantic and syntactic aspects of processing in the human brain: an fMRI investigation of different word types. Cereb. Cortex **10**, 698–705 (2000a)

Friederici, A.D., Wang, Y., Herrmann, C.S., Maess, B., Oertel, U.: Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study. Hum. Brain Mapp. **11**, 1–11 (2000b)

Friederici, A.D., Ruschemeyer, S.A., Hahne, A., Fiebach, C.J.: The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb. Cortex **13**, 170–177 (2003)

Friederici, A.D., Makuuchi, M., Bahlmann, J.: The role of the posterior superior temporal cortex in sentence comprehension. Neuroreport **20**, 563–568 (2009)

Gardner, H.: The Mind's New Science, 2nd edn. Basic Books, New York (1987)

Geschwind, N.: The organization of language and the brain. Science **170**, 940–944 (1970)

Gold, L., Lauritzen, M.: Neuronal deactivation explains decreased cerebellar blood flow in response to focal cerebral ischemia or suppressed neocortical function. Proc. Natl. Acad. Sci. USA **99**, 7699–7704 (2002)

Goldstein, K.: Language and Language Disturbances. Grune & Stratton, New York (1948)

Goodale, M.A., Westwood, D.A.: An evolving view of duplex vision: separate but interacting cortical pathways for perception and action. Curr. Opin. Neurobiol. **14**, 203–211 (2004)

Greenough, W.T., Black, J.E., Bailey, C.H.: Experience and brain development. Child Dev. **58**, 539–559 (1987)

Grill-Spector, K., Malach, R.: The human visual cortex. Annu. Rev. Neurosci. **27**, 649–677 (2004)

Grodzinsky, Y.: Theoretical Perspectives on Language Deficits. MIT Press, Cambridge (1990)

Grodzinsky, Y.: The neurology of syntax: language use without Broca's area. Behav. Brain Sci. **23**, 1–71 (2000)

Guzzetta, A., Pecini, C., Biagi, L., Tosetti, M., Brizzolara, D., Chilosi, A., Cipriani, P., Petacchi, E., Cioni, G.: Language organisation in left perinatal stroke. Neuropediatrics **39**, 157–163 (2008)

Hampson, M., Peterson, B.S., Skudlarski, P., Gatenby, J.C., Gore, J.C.: Detection of functional connectivity using temporal correlations in MR images. Hum. Brain Mapp. **15**, 247–262 (2002)

Hamzei, F., Rijntjes, M., Dettmers, C., Glauche, V., Weiller, C., Buchel, C.: The human action recognition system and its relationship to Broca's area: an fMRI study. Neuroimage **19**, 637–644 (2003)

Härle, M., Dobe, C., Cohen, R., Rockstroh, B.: Brain activity during syntactic and semantic processing-a magnetoencephalographic study. Brain Topogr. **15**, 3–11 (2002)

Harrington, A.: Medicine, Mind, and the Double Brain. Princeton University Press, Princeton (1987)

Haslinger, B., Erhard, P., Weilke, F., Ceballos-Baumann, A.O., Bartenstein, P., Grafin von Einsiedel, H., Schwaiger, M., Conrad, B., Boecker, H.: The role of lateral premotor-cerebellar-parietal circuits in motor sequence control: a parametric fMRI study. Cogn. Brain Res. **13**, 159–168 (2002)

Head, H.: Studies in Neurology. Oxford University Press, London (1920)

Head, H.: Aphasia and Kindred Disorders of Speech. Cambridge University Press, Cambridge (1926)

Hebb, D.O.: The Organization of Behavior. John Wiley & Sons, New York (1949)

Heim, S., Opitz, B., Friederici, A.D.: Distributed cortical networks for syntax processing: Broca's area as the common denominator. Brain Lang. **85**, 402–408 (2003)

Hein, G., Knight, R.T.: Superior temporal sulcus-It's my area: or is it? J. Cogn. Neurosci. **20**, 2125–2136 (2008)

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., Mazziotta, J.C.: The essential role of Broca's area in imitation. Eur. J. Neurosci. **17**, 1123–1128 (2003)

Herholz, K., Thiel, A., Wienhard, K., Pietrzyk, U., von Stockhausen, H.-M., Karbe, H., Kessler, J., Bruckbauer, T., Halber, M., Heiss, W.D.: Individual functional anatomy of verb generation. Neuroimage **3**, 185–194 (1996)

Holland, S.K., Plante, E., Weber Byars, A., Strawsburg, R.H., Schmithorst, V.J., Ball Jr, W.S.: Normal fMRI brain activation patterns in children performing a verb generation task. Neuroimage **14**, 837–843 (2001)

Holland, S.K., Vannest, J., Mecoli, M., Jacola, L.M., Tillema, J.M., Karunanayaka, P.R., Schmithorst, V.J., Yuan, W., Plante, E., Byars, A.W.: Functional MRI of language lateralization during development in children. Int. J. Audiol. **46**, 533–551 (2007)

Homae, F., Hashimoto, R., Nakajima, K., Miyashita, Y., Sakai, K.L.: From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. Neuroimage **16**, 883–900 (2002)

Hughlings-Jackson, J.: On affections of speech from disease of the brain. Brain **1**, 304–330 (1878)

Hund-Georgiadis, M., von Cramon, D.Y.: Motor-learning-related changes in piano players and non-musicians revealed by functional magnetic-resonance signals. Exp. Brain Res. **125**, 417–425 (1999)

Hwang, K., Palmer, E.D., Basho, S., Zadra, J.R., Müller, R.-A.: Category specific activations during word generation reflect experiential sensorimotor modalities. Neuroimage **48**, 717–725 (2009)

Hyde, K.L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A.C., Schlaug, G.: Musical training shapes structural brain development. J. Neurosci. **29**, 3019–3025 (2009)

Jonides, J., Smith, E.E., Marshuetz, C., Koeppe, R.A., Reuter-Lorenz, P.A.: Inhibition in verbal working memory revealed by brain activation. Proc. Natl. Acad. Sci. U.S.A **95**, 8410–8413 (1998)

Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R.: Brain activation modulated by sentence comprehension. Science **274**, 114–116 (1996)

Justus, T.: The cerebellum and English grammatical morphology: evidence from production, comprehension, and grammaticality judgments. J. Cogn. Neurosci. **16**, 1115–1130 (2004)

Kandel, E.R., Jessell, T.M., Sanes, J.R.: Sensory experience and the fine tuning of synaptic connections. In: Kandel, E.R., Schwartz, J.H., Jessell, T.M. (eds.) Principles of Neural Science, pp. 1115–1130. Elsevier, New York (2000)

Keller, T.A., Carpenter, P.A., Just, M.A.: The neural bases of sentence comprehension: a fMRI examination of syntactic and lexical processing. Cereb. Cortex **11**, 223–237 (2001)

Knösche, T.R., Maess, B., Friederici, A.D.: Processing of syntactic information monitored by brain surface current density mapping based on MEG. Brain Topogr. **12**, 75–87 (1999)

Konishi, S., Nakajima, K., Uchida, I., Kameyama, M., Nakahara, K., Sekihara, K., Miyashita, Y.: Transient activation of inferior prefrontal cortex during cognitive set shifting. Nat. Neurosci. **1**, 80–84 (1998)

Krams, M., Rushworth, M.F., Deiber, M.P., Frackowiak, R.S., Passingham, R.E.: The preparation, execution and suppression of copied movements in the human brain. Exp. Brain Res. **120**, 386–398 (1998)

Kubota, M., Inouchi, M., Ferrari, P., Roberts, T.P.: Human magnetoencephalographic evidence of early syntactic responses to c-selection violations of English infinitives and gerunds by L1 and L2 speakers. Neurosci. Lett. **384**, 300–304 (2005)

Kuperberg, G.R., Holcomb, P.J., Sitnikova, T., Greve, D., Dale, A.M., Caplan, D.: Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. J. Cogn. Neurosci. **15**, 272–293 (2003)

Kutas, M., Hillyard, S.A.: Brain potentials during reading reflect word expectancy and semantic association. Nature **307**, 161–163 (1984)

Kwon, H., Kuriki, S., Kim, J.M., Lee, Y.H., Kim, K., Nam, K.: MEG study on neural activities associated with syntactic and semantic violations in spoken Korean sentences. Neurosci. Res. **51**, 349–357 (2005)

Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E.: Entrainment of neuronal oscillations as a mechanism of attentional selection. Science **320**, 110–113 (2008)

Le Bihan, D.: Looking into the functional architecture of the brain with diffusion MRI. Nat. Rev. Neurosci. **4**, 469–480 (2003)

Lenneberg, E.H.: Biological Foundations of Language. Wiley, New York (1967)

Lidzba, K., Staudt, M., Wilke, M., Grodd, W., Krageloh-Mann, I.: Lesion-induced right-hemispheric language and organization of nonverbal functions. Neuroreport **17**, 929–933 (2006)

Liégeois, F., Connelly, A., Baldeweg, T., Vargha-Khadem, F.: Speaking with a single cerebral hemisphere: fMRI language organization after hemispherectomy in childhood. Brain Lang. **106**, 195–203 (2008)

Maess, B., Herrmann, C.S., Hahne, A., Nakamura, A., Friederici, A.D.: Localizing the distributed language network responsible for the N400 measured by MEG during auditory sentence processing. Brain Res. **1096**, 163–172 (2006)

Marie, P.: Revision de la question de l'aphasie. La Semaine Médicale **26**, 241–247 (1906)

Martin, A.: The representation of object concepts in the brain. Annu. Rev. Psychol. **58**, 25–45 (2007)

McClelland, J., Rumelhart, D.E. (eds.): Parallel Distributed Processing. MIT Press, Cambridge (Mass.) (1986)

McClelland, J.L., Rogers, T.T.: The parallel distributed processing approach to semantic cognition. Nat. Rev. Neurosci. **4**, 310–322 (2003)

Meyer, M., Friederici, A.D., von Cramon, D.Y.: Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. Brain Res. Cogn. Brain Res. **9**, 19–33 (2000)

Miller, E.K., Cohen, J.D.: An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. **24**, 167–202 (2001)

Mills, D.L., Coffey-Corina, S.A., Neville, H.J.: Language comprehension and cerebral specialization from 13 to 20 months. Dev. Neuropsychol. **13**, 397–445 (1997)

Minagawa-Kawai, Y., Mori, K., Naoi, N., Kojima, S.: Neural attunement processes in infants during the acquisition of a language-specific phonemic contrast. J. Neurosci. **27**, 315–321 (2007)

Minagawa-Kawai, Y., Mori, K., Hebden, J.C., Dupoux, E.: Optical imaging of infants' neurocognitive development: recent advances and perspectives. Dev. Neurobiol. **68**, 712–728 (2008)

Mountcastle, V.B.: The columnar organization of the neocortex. Brain **120**(Pt 4), 701–722 (1997)

Müller, R.-A.: Innateness, autonomy, universality? Neurobiological approaches to language. Behav. Brain Sci. **19**, 611–631 (1996)

Müller, R.-A.: Genes, language disorders, and developmental archaeology: what role can neuroimaging play? In: Rice, M., Warren, S. (eds.) Developmental Language Disorders: From Phenotypes to Etiologies, pp. 291–328. Erlbaum, Mahwah (2004)

Müller, R.-A.: Language universals in the brain: how linguistic are they? In: Christiansen, M.H., Collins, C., Edelman, S. (eds.) Language Universals, pp. 224–252. Oxford University Press, Oxford (2009)

Müller, R.-A., Basho, S.: Are nonlinguistic functions in "Broca's area" prerequisites for language acquisition? FMRI findings from an ontogenetic viewpoint. Brain Lang. **89**, 329–336 (2004)

Müller, R.-A., Courchesne, E.: The duplicity of plasticity: a conceptual approach to the study of early lesion and developmental disorders. In: Ernst, M., Rumsey, J. (eds.) The Foundation and Future of Functional Neuroimaging in Child Psychiatry, pp. 335–365. Cambridge University Press, New York (2000)

Müller, R.-A., Palmer, E.D.: Language and neurophysiological development. In: Rickheit, G., Strohner, H. (eds.) Communicative Competence. Mouton de Gruyter, Berlin (2008)

Müller, R.-A., Rothermel, R.D., Behen, M.E., Muzik, O., Mangner, T.J., Chakraborty, P.K., Chugani, H.T.: Brain organization of language after early unilateral lesion: a PET study. Brain Lang. **62**, 422–451 (1998)

Müller, R.-A., Rothermel, R.D., Behen, M.E., Muzik, O., Chakraborty, P.K., Chugani, H.T.: Language organization in patients with early and late left hemisphere lesion: a PET study. Neuropsychologia **37**, 545–557 (1999)

Müller, R.-A., Kleinhans, N., Courchesne, E.: Broca's area and the discrimination of frequency transitions: a functional MRI study. Brain Lang. **76**, 70–76 (2001)

Neville, H.J., Nicol, J.L., Barss, A., Forster, K.I., Garrett, M.F.: Syntactically based sentence processing classes: evidence from event-related brain potentials. J. Cogn. Neurosci. **3**, 151–165 (1991)

Newell, A., Simon, H.A.: GPS, a program that simulates human thought. In: Feigenbaum, E., Feldman, J. (eds.) Computers and Thought, pp. 279–293. McGraw-Hill, New York (1963)

Newman, A.J., Pancheva, R., Ozawa, K., Neville, H.J., Ullman, M.T.: An event-related fMRI study of syntactic and semantic violations. J. Psycholinguist. Res. **30**, 339–364 (2001)

Newmeyer, F.J.: Linguistic Theory in America, 2nd edn. Academic Press, Orlando (1986)

Ni, W., Constable, R.T., Mencl, W.E., Pugh, K.R., Fulbright, R.K., Shaywitz, S.E., Shaywitz, B.A., Gore, J.C., Shankweiler, D.: An event-related neuroimaging study distinguishing form and content in sentence processing. J. Cognitive Neurosci. **12**, 120–133 (2000)

Nishitani, N., Schurmann, M., Amunts, K., Hari, R.: Broca's region: from action to language. Physiology (Bethesda) **20**, 60–69 (2005)

Nystrom, L.E., Delgado, M.R., Sabb, F.W., Noll, D.C., Choen, J.D.: Dynamics of fMRI: Broca's area activation reflects independent effects of duration and intensity of working memory processes. Neuroimage **7**, S7 (1998)

O'Leary, D.D., Nakagawa, Y.: Patterning centers, regulatory genes and extrinsic mechanisms controlling arealization of the neocortex. Curr. Opin. Neurobiol. **12**, 14–25 (2002)

Ojemann, G.A.: Cortical organization of language. J. Neurosci. **11**, 2281–2287 (1991)

Pantano, P., Baron, J.C., Samson, Y., Bousser, M.G., Desrouesne, C., Comar, D.: Crossed cerebellar diaschisis. Brain **109**, 677–694 (1986)

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L.E., Hoke, M.: Increased auditory cortical representation in musicians. Nature **392**, 811–814 (1998)

Passingham, R.E., Stephan, K.E., Kotter, R.: The anatomical basis of functional localization in the cortex. Nat. Rev. Neurosci. **3**, 606–616 (2002)

Pineda, J.A.: Sensorimotor cortex as a critical component of an 'extended' mirror neuron system: Does it solve the development, correspondence, and control problems in mirroring? Behav. Brain Funct. **4**, 47 (2008)

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., Gabrieli, J.D.: Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. Neuroimage **10**, 15–35 (1999)

Popper, K.R.: Conjectures and Refutations: The Growth of Scientific Knowledge. Routledge & Kegan Paul, Neuw York (1965)

Puce, A., Perrett, D.: Electrophysiology and brain imaging of biological motion. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **358**, 435–445 (2003)

Redcay, E.: The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. Neurosci. Biobehav. Rev. **32**, 123–142 (2008)

Ressel, V., Wilke, M., Lidzba, K., Lutzenberger, W., Krageloh-Mann, I.: Increases in language lateralization in normal children as observed using magnetoencephalography. Brain Lang. **106**, 167–176 (2008)

Rilling, J.K., Glasser, M.F., Preuss, T.M., Ma, X., Zhao, T., Hu, X., Behrens, T.E.: The evolution of the arcuate fasciculus revealed with comparative DTI. Nat. Neurosci. **11**, 426–428 (2008)

Rizzolatti, G., Craighero, L.: The mirror-neuron system. Annu. Rev. Neurosci. **27**, 169–192 (2004)

Rubia, K., Russell, T., Overmeyer, S., Brammer, M.J., Bullmore, E.T., Sharma, T., Simmons, A., Williams, S.C., Giampietro, V., Andrew, C.M., Taylor, E.: Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. Neuroimage **13**, 250–261 (2001)

Sadato, N., Okada, T., Honda, M., Yonekura, Y.: Critical period for cross-modal plasticity in blind humans: a functional MRI study. Neuroimage **16**, 389–400 (2002)

Salmelin, R.: Clinical neurophysiology of language: the MEG approach. Clin. Neurophysiol. **118**, 237–254 (2007)

Santi, A., Grodzinsky, Y.: Working memory and syntax interact in Broca's area. Neuroimage **37**, 8–17 (2007)

Saur, D., Kreher, B.W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C.: Ventral and dorsal pathways for language. Proc. Natl. Acad. Sci. U.S.A **105**, 18035–18040 (2008)

Saussure, F., Cours de Linguistique Générale. Payot, Paris (1915/1972)

Schlaggar, B.L., O'Leary, D.: Potential of visual cortex to develop an array of functional units unique to somatosensory cortex. Science **252**, 1556–1560 (1991)

Schlaggar, B.L., Brown, T.T., Lugar, H.M., Visscher, K.M., Miezin, F.M., Petersen, S.E.: Functional neuroanatomical differences between adults and school-age children in the processing of single words. Science **296**, 1476–1479 (2002)

Schlaug, G., Norton, A., Overy, K., Winner, E.: Effects of music training on the child's brain and cognitive development. Ann. N.Y. Acad. Sci. **1060**, 219–230 (2005)

Searle, J.R.: Speech Acts. An Essay in the Philosophy of Language. Cambridge University Press, Cambridge (1969)

Seghier, M.L., Lazeyras, F., Pegna, A.J., Annoni, J.M., Zimine, I., Mayer, E., Michel, C.M., Khateb, A.: Variability of fMRI activation during a phonological and semantic language task in healthy subjects. Hum. Brain Mapp. **23**, 140–155 (2004)

Shallice, T.: From Neuropsychology to Mental Structure. Cambridge University Press, Cambridge (1988)

Shtyrov, Y., Pulvermuller, F.: Early MEG activation dynamics in the left temporal and inferior frontal cortex reflect semantic context integration. J. Cogn. Neurosci. **19**, 1633–1642 (2007)

Signoret, J.L., Castaigne, P., Lhermitte, F., Abelanet, R., Lavorel, P.: Rediscovery of Leborgne's brain: anatomical description with CT scan. Brain Lang. **22**, 303–319 (1984)

Singer, W.: Consciousness and the binding problem. Ann. N.Y. Acad. Sci. **929**, 123–146 (2001)

Skinner, B.F.: Verbal Behavior. Appleton-Century-Crofts, New York (1957)

Staudt, M., Lidzba, K., Grodd, W., Wildgruber, D., Erb, M., Krageloh-Mann, I.: Right-hemispheric organization of language following early left-sided brain lesions: functional MRI topography. Neuroimage **16**, 954–967 (2002)

Stromswold, K., Caplan, D., Alpert, N., Rauch, S.: Localization of syntactic comprehension by positron emission tomography. Brain Lang. **52**, 452–473 (1996)

Sur, M., Leamey, C.A.: Development and plasticity of cortical areas and networks. Nat. Rev. Neurosci. **2**, 251–262 (2001)

Sur, M., Pallas, S.L., Roe, A.W.: Cross-modal plasticity in cortical development: differentiation and specification of sensory neocortex. Trends Neurosci. **13**, 227–233 (1990)

Szaflarski, J.P., Schmithorst, V.J., Altaye, M., Byars, A.W., Ret, J., Plante, E., Holland, S.K.: A longitudinal functional magnetic resonance imaging study of language development in children 5 to 11 years old. Ann. Neurol. **59**, 796–807 (2006)

Thompson, P.M., Cannon, T.D., Narr, K.L., van Erp, T., Poutanen, V.P., Huttunen, M., Lonnqvist, J., Standertskjold-Nordenstam, C.G., Kaprio, J., Khaledy, M., Dail, R., Zoumalan, C.I., Toga, A.W.: Genetic influences on brain structure. Nat. Neurosci. **4**, 1253–1258 (2001)

Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., Farah, M.J.: Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc. Natl. Acad. Sci. U.S.A **94**, 14792–14797 (1997)

Uchiyama, Y., Toyoda, H., Honda, M., Yoshida, H., Kochiyama, T., Ebe, K., Sadato, N.: Functional segregation of the inferior frontal gyrus for syntactic processes: a functional magnetic-resonance imaging study. Neurosci. Res. **61**, 309–318 (2008)

Wheless, J.W., Castillo, E., Maggio, V., Kim, H.L., Breier, J.I., Simos, P.G., Papanicolaou, A.C.: Magnetoencephalography (MEG) and magnetic source imaging (MSI). Neurologist **10**, 138–153 (2004)

Xiong, J., Rao, S., Jerabek, P., Zamarripa, F., Woldorff, M., Lancaster, J., Fox, P.T.: Intersubject variability in cortical activations during a complex language task. Neuroimage **12**, 326–339 (2000)

Zaehle, T., Wüstenberg, T., Meyer, M., Jäncke, L.: Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study. Eur. J. Neurosci. **20**, 2447–2456 (2004)

# Chapter 3
# Dialogue: Interactive Alignment and Its Implications for Language Learning and Language Change

**Simon Garrod and Martin J. Pickering**

**Abstract** This chapter discusses language processing during conversation. In particular, it considers why taking part in a conversation is more straightforward than speaking or listening in isolation. We argue that conversation is easy because speakers and listeners automatically align with each other at different linguistic levels (e.g., sound, grammar, meaning) which leads to alignment at the level of interpretation. This alignment process is reflected in the repetitiveness of dialogue at different levels and occurs both on the basis of local mechanisms of priming and more global mechanisms of routinization. We argue that the latter process may tell us something about both acquisition of language and historical processes of language change.

## 3.1 Introduction

It is generally accepted that the most basic use of language is in conversation or dialogue. Everyone who speaks can converse, whereas the ability to give a speech or even the ability to listen to one is difficult to acquire. Yet dialogue has never taken priority in the language sciences. Theoretical linguists analyze isolated sentences of the kind found in monologue. Until recently, this was also true for computational linguists. In turn, psycholinguists concentrate exclusively either on processes of language production or on processes of language comprehension without considering the relationship between the two.

By contrast, this chapter deals with dialogue processes and attempts to explain why interactive language use is so easy compared to speaking or listening on your own. It is not just that dialogue is basic. We argue that it may also tell us something about

---

S. Garrod (✉)
Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8Q, UK
e-mail: Simon.Garrod@glasgow.ac.uk

M. J. Pickering
School of Philosophy, Psychology and Language Sciences, University of Edinburgh,
7 George Square, Edinburgh EH8 9JZ, UK

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_3,

language learning and language change. Language is acquired through conversation. Furthermore, even adult conversationalists adapt their language to that of their partner. Hence, during dialogue language learning takes place all the time. In this respect there is a kind of continuity between childhood language acquisition and adult conversation. Continuous speaker adaptation may also help explain why languages are always historically changing. In this way language processes occurring in short-lived interactions may tell us something about language over a larger time-scale.

## 3.2 The Challenge of Conversation

Conversation involves an extremely complicated set of processes in which participants have to interweave their activities with precise timing, and yet it is a skill that all speakers seem very good at (Garrod and Pickering 2004). To understand how remarkable this is, consider this transcript of a dinner-party conversation (Tannen 1984), with brackets indicating overlapping speech and numbers indicating noticeable pauses in seconds:

1—**A:** I shook hands with Rubenstein once? [and his hand
2—**B:** [Yeah we did together
3—**A:** That’s right. we were together. wasn’t it incredible?
4—**B:** *(laughing)* oh it was like a cushion.
5—**C:** What’s this?
6—**A:** [I *(0.5)* we shook] hands with Rubenstein.
7—**B:** [Rubenstein’s hands].
8—**D:** and he had –?
9—**A:** his hands –
10—**D:** short stubby hands?
11—**A:** they were like *(0.5)* [jelly. they were like — *(1.0)*
12—**B:** [a famous concert pianist
13—**A:** they were like *(0.5)* putty. *(0.5)*
14—**D:** [really?
15—**A:** [just completely soft and [limp
16—**B:** [mush
17—**A:** just mush. it was as though there was [no bone
18—**B:** [and warm.
19—**D:** and short stubby fingers?
20—**A:** short stubby fingers but just *(0.5)* totally covered with
21—**B:** fat.
22—**A:** fat

This conversation differs greatly from formal prose (such as the rest of this chapter). In particular, the speakers regularly produce elliptical and fragmentary utterances that would make little sense on their own (e.g., 7, 10, 12, 16, 17, 20). It is jointly constructed by all four speakers, and involves a great deal of interruption, overlapping speech, and disfluency. However, the participants appear to be satisfied with the conversation. They seem to understand what everyone says, as do non-participants such as ourselves. How can this be?

The more we think about conversations such as this, the more remarkable they appear. The interlocutors cannot be sure what contributions their partners are going to make, so they cannot securely plan far in advance. They have to construct their utterances so that they are appropriate for their listeners at that particular point, and therefore must pay constant attention to any feedback (e.g., whether a particular term is understood). For example, B's interruption at (2) causes A to abandon (1) and produce the appropriate response (3) on the fly. They have to decide whether to contribute to a conversation and if so precisely when they should do so, and they may have to decide who to address. In addition, they have to constantly switch between speaking and listening, even though task-switching is often difficult.

So why is dialogue so easy? We believe that the key to this question is found in its repetitiveness. Notice how the participants reuse each others words and expressions. For instance, consider the various repetitions of *hands* and *Rubenstein* in the conversation above (6, 7, 9, 10). Our central argument is that such repetitiveness is mirrored in the participants' minds, so that they are replicating each other's mental states and not merely their form of words. This is the core to what we shall call interactive alignment.

## 3.3 Interactive Alignment During Conversation

One argument for why conversation is so easy is that conversational partners tend to become aligned at different levels of linguistic representation and therefore find it easier to perform this joint activity than the individual activities of speaking or listening (Garrod and Pickering 2009). Pickering and Garrod (2004) explain the process of alignment in more detail in terms of their interactive-alignment account. According to this account, conversation is successful to the extent that participants come to understand the relevant aspects of what they are talking about in the same way as each other. More specifically, they construct mental representations or models of the situation under discussion, and successful conversation occurs when these models become aligned. Such alignment largely occurs as a result of the tendency for conversational partners to repeat each other's choices at many different linguistic levels, such as words and grammar (e.g., Branigan et al. 2000; Brennan and Clark 1996; Garrod and Anderson 1987). This is a form of imitation. Essentially, conversational partners prime each other to speak about things in the same way, and people who speak about things in the same way are more likely to think about them in the same way as well (Box 1). In this way the language processing system makes a virtue out of what appears to be a vice, by coupling together speaking and listening processes.

The interactive alignment account has three implications for language processing in dialogue. First, it implies that there is *parity of representations* used in speaking and listening. Second, it depends on the idea that alignment processes operating at different levels (words, structure, meaning) interact in such a way that increased alignment at one level leads to increased alignment at other levels. Finally, it assumes that these alignment processes are based on imitation which is largely automatic.

**Box 1: Evidence for linguistic imitation at many levels**

Evidence for imitation is found in many language experiments. Interlocutors become aligned at many different linguistic levels simultaneously, almost invariably without any explicit negotiation. At the level of the situation model, interlocutors align on spatial reference frames: if one speaker refers to objects egocentrically (e.g., ‘on the left’ to mean on the speaker’s left), then the other speaker tends to use an egocentric perspective as well (Watson et al. 2004). More generally, they align on a characterization of the domain, for instance using coordinate systems (e.g., A4, D3) or figural descriptions (e.g., T-shape, right indicator) to refer to positions in a maze (Garrod and Anderson 1987; Garrod and Doherty 1994, see Box 4). They also repeat each other’s referring expressions, even when they are unnecessarily specific (Brennan and Clark 1996). Imitation also occurs for grammar, with speakers repeating the syntactic structure used by their interlocutors for cards describing events (Branigan et al. 2000, see Box 2 for details) or objects (Cleland and Pickering 2003), and repeating syntax or closed-class lexical items in question-answering (Levelt and Kelter 1982). They even repeat syntax between languages, for example when one interlocutor speaks English and the other speaks Spanish (Hartsuiker et al. 2004). Finally, there is evidence for alignment of sound representations (Pardo 2006), and of accent and speech rate (Giles et al. 1992).

### *3.3.1 Parity of Representations*

A critical aspect of the alignment model is what we term parity of representations—the same representations are constructed during speaking and listening. In other words, language involves use of a common code for representing your own actions (your speech) and your partner’s actions (his or her speech). This explains why linguistic repetition occurred in experiments such as Branigan et al. (2000), who had participants take turns to describe and match picture cards, and found that they tended to use the form of utterance just used by their partner (Box 2). For example, they tended to use a “prepositional object” form such as *the pirate giving the book to the swimmer* following another prepositional object sentence but a “double object” form such as *the pirate giving the swimmer the book* following another double object sentence (though both sentences have essentially the same meaning).

In such cases, the same grammatical representation is activated during speaking and listening.

Though the common-coding assumption may appear to follow from the reasonable claim that language users do not call upon different knowledge of language when speaking and listening, it is important to realize that traditional psycholinguistic theories of production and comprehension have largely developed in isolation from each other (see Fodor et al. 1974; Harley 2007). For example, theories of lexical representation during production (e.g., Levelt et al. 1999) are not used in theories of word recognition. Historically, this separation of the study of comprehension and production goes back to the idea that language can be thought of as a code. On this view, communication involves two distinct processes: encoding a message (language production) or decoding a signal to reveal the message (language comprehension). If one accepts such an account then it makes sense to study the production (encoding) process and the comprehension (decoding) process as distinct activities. However, this approach is not appropriate for understanding communication in dialogue (Garrod and Pickering 2007). During dialogue, production and comprehension processes become inextricably linked. Speakers need to interpret feedback from their addressees while speaking and addressees need to prepare appropriate responses (e.g., spoken feedback or subsequent responses to queries) while listening to the speaker. The most straightforward way of accounting for this interplay between production and comprehension processes is to assume close parity of linguistic representations underlying production and comprehension processes.

### *3.3.2 Percolation Between Levels of Alignment*

Another important aspect of the interactive alignment account is that alignment at one level affects alignment at other levels. For example alignment of syntactic structure is enhanced by repetition of words, with participants being even more likely to say *The cowboy handing the banana to the burglar* after hearing *The chef handing the jug to the swimmer* than after *The chef giving the jug to the swimmer* (Branigan et al. 2000). Thus, alignment at one level (in this case, lexical alignment) enhances alignment at another level (in this case, grammatical alignment). Similarly, people are more likely to use an unusual form like *the sheep that's red* (rather than *the red sheep*) after they have just heard *the goat that's red* than after they heard *the door that's red* (Cleland and Pickering 2003). This is because the meaning of *sheep* is related to the meaning of *goat* but not *door*. So alignment at the semantic level increases syntactic alignment. Furthermore, alignment of words leads to alignment of situation models—people who describe things the same way tend to think about them in the same way too (Markman and Makin 1998). This means that alignment of low-level structure can eventually affect alignment at the crucial level of speakers' situation models, the hallmark of successful communication.

**Box 2: Confederate dialogue experiment to test for syntactic priming (Branigan et al. 2000)**

A naïve participant and a confederate sat on opposite sides of a table with a divider between them. They take turns to describe cards to each other and to select the appropriate card from an array. For example, the confederate described a card as *The chef giving the jug to the swimmer*. After the participant selected the matching card, she tended to describe her next card as *The cowboy handing the banana to the burglar*. But if the confederate had described the card as *The chef giving the swimmer the jug*, the participant tended to say *The cowboy handing the burglar the banana*. Such repetition of syntactic form occurred on about 4 trials out of 6 when the confederate and the participant used different verbs. But when they both described cards with the same verb (e.g., *handing*), repetition occurred on about 5 trials out of 6 (Fig. 3.1).

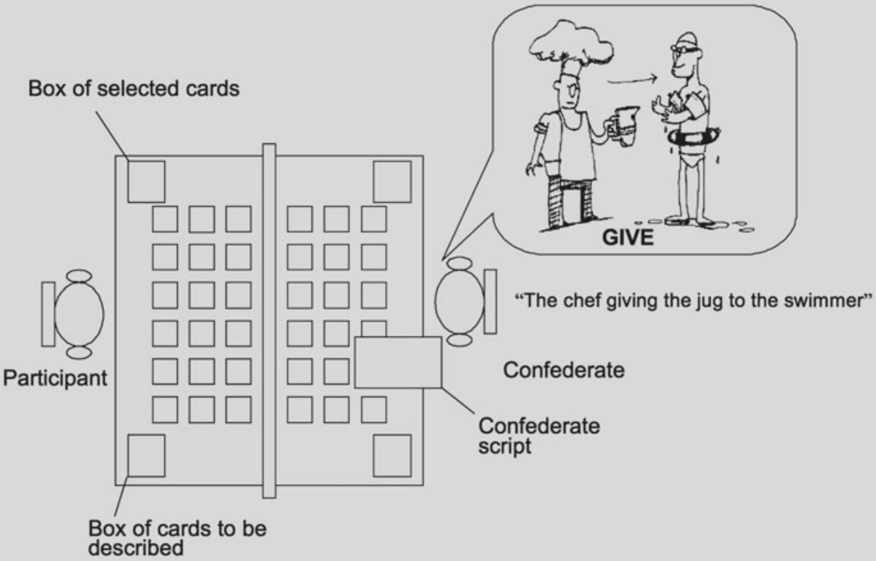
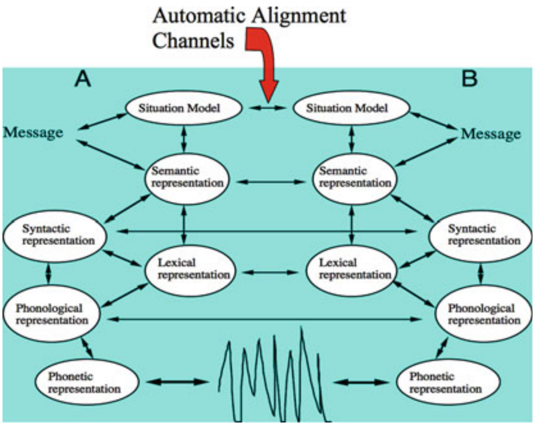


**Fig.3.1.** Schematic representation of the experimental set-up for Branigan et al. (2000) confederate scripted syntactic priming experiment

### 3.3.3 *Automatic Channels of Alignment*

An important property of interactive alignment is that it is automatic in the sense that speakers are not aware of the process and that it does not appear effortful. Such automatic imitation or mimicry occurs in social situations more generally. Thus, Dijksterhuis and Bargh (2001) argued that many social behaviours are automatically

**Fig. 3.2** Automatic channels of alignment for participants *A* and *B* during a conversation



triggered by perception of action in others (Box 3). We propose that the automatic alignment channels linking different levels of linguistic representation operate in essentially the same fashion (see Fig. 3.2). In other words, conversationalists do not need to decide to interpret the different levels of linguistic representation carried by alignment channels for them to influence alignment (Garrod and Pickering 2006). This is because the alignment channels reflect priming rather than interpretation. In addition there are aspects of automatic non-linguistic imitation that can facilitate alignment at linguistic levels (Garrod and Pickering 2009). For example, when speakers and listeners align their gaze to look at the same thing this can facilitate alignment of interpretation (Richardson and Dale 2005; Richardson et al. 2007).

### Box 3: Automatic perception-action links during social interactions

Automatic perception–action links are well documented in the neurophysiological literature (e.g., motor imitation arising from the firing of mirror neurons in monkey premotor cortex; Rizzolatti and Craighero 2004) and in the psychological literature (Hommel et al. 2001). There is evidence for automatic links in controlling facial expressions, movements and gestures, and speech. For example, when observing another person experiencing a painful injury and wincing, observers imitate the wince in their own expression (Bavelas 1986). Similarly, participants will mimic postures such as foot shaking and nose rubbing carried out by a person with whom they are conversing (Chartrand and Bargh 1999) and when they repeat another's speech they adopt the other's tone of voice as well (Neumann and Strack 2000). Finally, conversational partners align their posture (Shockley et al. 2003).

## 3.4 Alignment and Routinization

The interactive alignment account gives a basic mechanism for alignment of understanding during dialogue. But also it may have a bearing on both acquisition of language and the process of language change. To understand this, we need to consider the interactive alignment process in more detail. In particular, we need to consider how it works on two time scales. First, there is alignment based on short-term co-activation of representations at various linguistic levels. This comes about through priming, whereby there is a boost in activation of relevant representations (e.g., for words or for syntactic structures) following exposure to their corresponding forms. Second, there is longer term alignment arising from the repeated co-activation of different representations. This longer term process we refer to as *routinization*.

As we have noted already real conversation is extremely repetitive, and the comparison with carefully crafted monologue (as in texts) is very striking indeed (Tannen 1989). Pickering and Garrod (2004) argued that expressions that are repeated become routines for the purposes of the dialogue. By routine we mean an expression that is "fixed" to a relatively large extent. Extreme examples include repetitive conversational patterns such as *How do you do?* and *Thank you very much*. Many examples are idioms, such as *kick the bucket* (where all the words are fixed) or *keep (lose) one's cool* (where some words are fixed but others can vary). However, many common expressions such as *I love you* have literal interpretations.

Groups of people may develop particular types of routine, perhaps in order to aid their fluency. Kuiper (1996) described the fixed language used by auctioneers and sportscasters. For example, radio horseracing commentators produce highly repetitive and stylized speech which is quite remarkably fluent. He argued that the commentators achieve this by storing routines, which can consist of entirely fixed expressions (e.g., *they are coming round the bend*) or expressions with an empty slot that has to be filled (e.g., *X is in the lead*), in long-term memory, and then accessing these routines, as a whole, when needed. Processing load is thereby greatly reduced in comparison to non-routine production. Of course, this reduction in load is only possible because particular routines are stored; and these routines are stored because the commentators repeatedly produce the same small set of expressions in their career.

Most discussion of routines refers to the long-term development of fixed expressions that come to behave like words (e.g., Aijmer 1996; Kuiper 1996; Nunberg et al. 1994; Bybee 2006). But routines may also be established for the purposes of a particular interchange. If one speaker starts to use an expression and gives it a particular meaning, the other will most likely follow suit. In other words, routines are set up 'on the fly' during conversation. We propose that the use of routines contributes greatly to the fluency of conversation. For example, Pickering and Garrod (2004) give the example *the previous administration*, which can take on a specific meaning (referring to a particular political body) as part of a conversation, and where other interpretations of the individual words (e.g., *administration* meaning work) or of the expression as a whole (e.g., referring to a different political body) are not considered. The establishment of this form of words and meaning as a routine has the effect that interlocutors access it without seriously considering alternatives. In production, they do not make a difficult choice between using the word *administration* or its near-synonym *government*; and in comprehension, they do not consider (non-routinized) interpretations of the words (e.g., of *administration*). After the conversation is over, however, the interlocutors may ‘drop’ this routine and return to their ‘standard’ use of the words. Conversational routines can be elicited experimentally. Consider the brief transcript of an interaction (shown in Box 4) in which A and B are trying to establish their respective positions in a maze. In particular, the expression *right indicator* takes on a specific meaning (referring to a particular configuration within mazes). Once the players have fixed on this expression and interpretation, they do not describe the configuration in alternative ways. Although we can be less certain of what happens during comprehension, the responses to references to *right indicator* strongly suggest that they also understand the expression in its special sense.

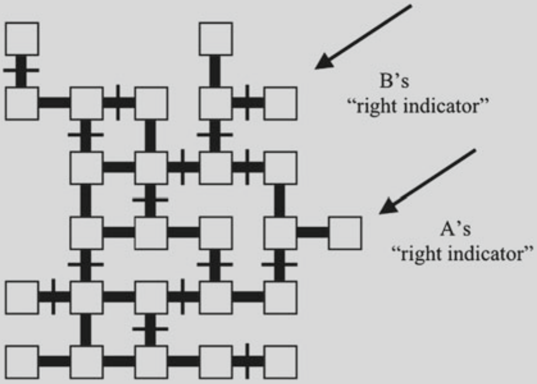
Pickering and Garrod (2005) drew a distinction between short-term interactive alignment and routinization. Interactive alignment involves the priming of particular levels of representation and the links between those levels. Producing or comprehending any utterance leads to the activation of those representations, but their activation gradually decays. However, when interactive alignment leads to sufficiently strong activation of the links between the levels, routinization occurs. Routinization involves the setting down of new memory traces associated with a particular expression, so that the expression becomes lexicalised. A formal approach compatible with this is found in Jackendoff (2002), who argues that lexical entries consist of linked components concerned with meaning, sound structure (phonology), and syntax. For example, the word *indicator* would consist of a sound representation (in phonemes) linked to a syntactic representation (Noun) linked to a conceptual representation (POINTING DEVICE). This scheme can be extended to account for complex lexicalisations such as *right indicator* or *kick the bucket*.

Pickering and Garrod (2005) argued that routines are not simply recovered from long-term memory as complete chunks (in contrast to Kuiper 1996, for example). They enumerated various reasons to suspect that producing routines involves some compositional processes. First, it can straightforwardly explain how people produce semi-productive routines with a variable element, as in *take X to task*, where *X* can be any noun phrase referring to a person or people. Second, the structure of non-idiomatic sentences can be primed by idiomatic sentences in production (Konopka and Bock 2009). Third, it is consistent with the production of idiom blends like *That’s the way the cookie bounces* (Cutting and Bock 1997). There is also evidence for syntactic processing of routines in comprehension. For example, syntactically appropriate continuations to phrases are responded to faster than syntactically inappropriate ones when the phrase is likely to be the beginning of an idiom (e.g., *kick the*) (Peterson et al. 2001). We now consider the implications of routinization for language acquisition and language change.

### Box 4: History of a conversational routine

Below is an extract from a maze-game dialogue taken from Garrod and Anderson (1987), and which relates to the figure below. When B says *It's like a right indicator* (11), the expression *right indicator* is not a routine, but is composed of two expressions whose interpretations are relatively standard, and whose meaning involves normal processes of meaning composition. So, B accesses the standard meanings of the words *right* and *indicator* and creates a phrase with the standard meaning. Importantly, however, B does not simply use *right indicator* to refer to any object that can be referred to as a right indicator, but instead uses it to refer to a particular type of object that occurs within this maze. A accepts this description with *yes* (12), presumably meaning that he has understood B's utterance correctly. He then interprets A's utterance at this stage using the normal processes of meaning corresponding to the compositional processes that A has used in production. The expression *right indicator* now keeps recurring, and is used to refer to positions in the maze. Whereas initially it was used as part of a simile [*it's like a right indicator* in (11)], subsequently it is used referentially [*that right indicator you've got* in (15)].

8—**A:** You know the extreme right, there's one box.
9—**B:** Yeah right, the extreme right it's sticking out like a sore thumb.
10—**A:** That's where I am.
11—**B:** It's like a right indicator.
12—**A:** Yes, and where are you?
13—**B:** Well I'm er: that right indicator you've got.
14—**A:** Yes.
15—**B:** The right indicator above that.
16—**A:** Yes.
17—**B:** Now if you go along there. You know where the right indicator above yours is?
18—**A:** Yes.
19—**B:** If you go along to the left: I'm in that box which is like: one, two boxes down O.K.?



How does such routinization occur? Pickering and Garrod (2005) proposed that the activation of *right* and *indicator* plus the specific meaning that *right indicator* has acquired leads to the activation of the phonological representation and syntactic representation together with the activation of the specific meaning ("right-hand-protrusion-on-maze"). Therefore the links among the phonology, syntax and semantics are activated (as specified in the interactive alignment model). That increases the likelihood that the interlocutors are going to subsequently use *right indicator* with that specific meaning. But in addition to this basic interactive-alignment process, the activation of the links "suggest" the positing of a new long-term association, essentially that *right indicator* can have the meaning "right-hand-protrusion-on-maze". We propose that when activation is strong enough, a new lexical entry (similar to a word) is constructed and stored in memory as a routine.

### *3.4.1 Routinization and Language Learning*

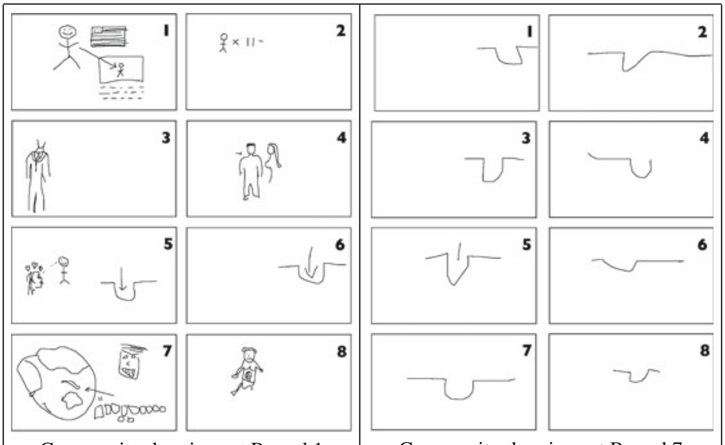
So far we have focused on the establishment of temporary routines for the purpose of a particular interchange. This appears to be an important and almost entirely neglected aspect of language use. But routines need not be 'dropped' once the conversation is over. When this happens, the new lexical entry remains in the speaker's lexicon.

In fact, experimental evidence suggests that routines do extend beyond the particular interchange. Garrod and Doherty (1994) had people play the maze game (see Box 4) with different partners. When all members of a group played with each other (e.g., A with B, C with D, then A with C, B with D, then A with D, B with C), they converged on description schemes (consisting of both fixed and semi-productive routines) to a much greater extent than when participants played with members of a different group on each interchange (e.g., A with B, C with D, A with C, A with E, B with F). In other words, interlocutors who formed a ‘network’ converged to a much greater extent than those who did not (and indeed converged more than those who played repeatedly with the same partner). This shows that they converged on description schemes that lasted beyond one interchange, and hence that the routinization of the schemes persisted. (Interestingly, this same convergence can be demonstrated for non-linguistic graphical communication among groups of communicators, see Box 5).

### Box 5: Group convergence during graphical communication

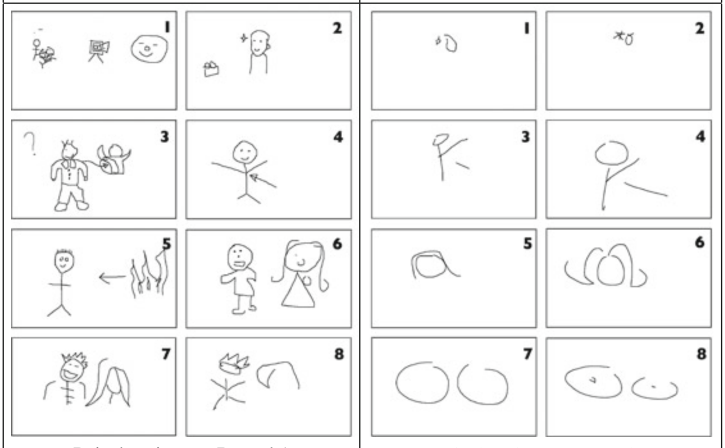
Garrod et al. (2007) developed a non-linguistic communication task to study the emergence of novel graphical signs. The task was a laboratory version of the popular parlour game ‘Pictionary’. Participants would take turns to draw pictures of concepts drawn from a list in such a way that their partner could identify the concept from the same list. The process was then repeated over a number of blocks (within each block participants communicated 12 items from a list of 16). In the original version of the task Garrod et al. (2007) found that with repetition the drawings for the same item became increasingly simple and abstract and the 2 participants would end up depicting a given concept in the same way as each other (see bottom right panel in Fig. 3.3). Fay et al. (2010) developed a community version of this experiment similar to Garrod and Doherty (1994)’s community maze game study. Groups of 8 players carried out the ‘Pictionary’ task in successive pairs involving 7 rounds of play. Each round consisted of 6 blocks of trials with a new partner drawn from the same group. In this way, by the end of the experiment each member of the group had interacted graphically with each other member. The top panel of Fig. 3.3 shows the drawings from 1 group of players for 1 item (Brad Pitt). On the top left of the figure are drawings taken from the beginning of the first round for each of the original pairs and on the right top panel are drawings taken from the beginning of the final round. Whereas the original drawings are complex and varied, the final drawings are simple and homogenous. This suggests that interactive communication with members of a closed community leads to the evolution of a common representation whether it be a linguistic or a non-linguistic one.

Garrod and Doherty (1994) showed that interlocutors who did not come from the same community failed to converge. In terms of our current proposal, this occurred because of a clash between routinization and priming: one participant’s routinized lexical entries may not match with the priming that occurs as a result of the other participant using a different lexical entry. In other words, if A has routinized a particular expression with partner B (e.g., *right indicator*) and now encounters partner C from a different community, then A’s routines will not correspond to B’s routines (e.g., B might have routinized *T on its side* for the same maze configuration). As a consequence after encountering a number of different partners from different com-



Community drawings at Round 1

Community drawings at Round 7



Pair drawings at Round 1

Pair drawings at Round 7

**Fig. 3.3** Drawings of ‘Brad Pitt’ elicited by the ‘Pictionary’ task (Fay et al. 2010, see Box 5). The *top left* panel shows drawings from community pairs in the first round (1 and 2, 3 and 4, etc. before the community has been established), the *top right* panel shows drawings from the same individuals in the final round. The *bottom* panels show drawings from matched isolate pairs in the first and final rounds of the task

munities each interlocutor’s tendency to use different routines will get in the way of the short-term priming process.

This suggests that the establishment of routines can be equated with the processes that take place during language learning. In particular, the process by which children set down representations for novel words and expressions may be akin to routinization. However, we need to explain why routinization might lead to large-scale vocabulary acquisition, when it clearly extends adults’ store of expressions to a much more limited extent.

Of course, children encounter new words much more often than adults. But in addition, young children are much more set up to accept novel pairings between form and meaning (and grammar, though we ignore this here) than adults. In other words, the links between the components of linguistic representations are particularly strong. This can be seen in the strong tendency children have to avoid synonyms (e.g., Clark 1993). For example, if a young child refers to particular footwear as *boots* she will tend not to accept the term *shoes* to refer to the same objects. This is compatible with a particularly strong link being set up between the word and a particular meaning. Garrod and Clark (1993) found that children (aged 7–8 years) playing the maze game (Box 4) would converge on referring expressions and description schemes to refer to maze positions to at least as great an extent as adults. But they were much less happy than adults to abandon those referring schemes when it became clear that they were leading to misunderstanding. Garrod and Clark interpreted this result as showing that the natural tendency for the children is to converge (as predicted by interactive alignment) and it is only as they mature that they are able to inhibit this tendency when required to do so.

Such commitment to particular form-meaning pairings is efficient both for processing and acquisition. For processing, it means that the space of alternatives that the child has to consider is rapidly reduced. But it has the difficulty that it reduces the ability of the child to express a wider range of concepts (assuming that synonyms can have slight differences in meaning, or can have differences imposed for particular interchanges) and to comprehend the full range of meanings that a speaker expresses. These problems do not of course matter so much if the speaker (the "parent") is aware of the child's limitations, and (for instance) employs a limited vocabulary. For acquisition, if novel lexical items follow from the fixation of form-meaning pairings, then children will establish new routines more easily than adults. If a child hears *right indicator* being used to refer to a bit sticking out from a maze, then she will establish the link between *right indicator* and its meaning in such a way that she will be unable to accept another term to refer to the same thing. We have argued that this occurs in adults too, but the assumption is that adults can abandon such conventions more straightforwardly than children. This means that adults' conversation is more flexible than children's, but that the establishment of novel items is more straightforward for children.

### 3.4.2 Routinization and Language Change

Moving to a larger time-scale languages undergo historical change. Expressions come into the language and drop out of it and may change as a consequence of usage (Labov 1994, see also Croft, this volume). Can interactive alignment and routinization tell us anything about this process?

A key issue in the study of language change is explaining how changes in the language can spread within and across generations of speakers. Kirby (1999) refers to this as the *problem of linkage*. In biological evolution, linkage occurs through the inheritance of genes from one generation to the next. The traditional linguistic analogy is to explain linkage through the passing down of a language from one generation to the next during its acquisition. It is then assumed that language change is determined by constraints (which Kirby 1999, calls the linguistic bottleneck) that apply to the language learning mechanism (see also Kirby, this volume). However, interactive alignment and routinization offer an alternative linkage mechanism associated with language use. In the same way that experimental communities of speakers establish their own routines over the course of repeated interactions, so real communities of speakers can establish and maintain routines as well. Hence, one kind of language variation is found in what Clark (1996) calls *communal lexicons*—particular sets of expressions associated with different communities. For example, skiers talk of *piste*, physicists of *quarks*, statisticians have a special interpretation of *significance* and *normal distribution*. This kind of variation would be expected if each community establishes its own routines.

As we have argued, routines can be considered lexicalisations, bits of language stored and accessed directly from memory. One important topic in the study of language change is the emergence and maintenance of simple and complex lexicalisations. Take for example, the process of grammaticalization in which lexical elements increasingly take on grammatical functions. A good example of this is the evolution in English of the complex future auxillary *going to* from the simple lexical verb of motion *going*, which may even become reduced to the simple *gonna* (Hopper and Traugott 1993). This historical process follows a similar pattern to that of routinization in dialogue. Initially, an expression takes on a contextually determined interpretation (in this case with reference to a future action presumably involving motion). This expression-meaning mapping then becomes fixed and eventually generalizes to other analogous future actions that do not involve motion. As soon as it becomes fixed in this way it becomes a routine which can be reduced like any other lexicalisation with repeated usage (e.g., becoming the simple lexical item *gonna*). The important distinction between this account of language change and the more traditional acquisition-based account is that the evolutionary process arises from usage rather than constraints on learning, because the linkage is through interactive alignment and routinization. For a more detailed discussion of how frequency of usage relates to processes of grammaticalization we refer the reader to Bybee (2006).

Another evolutionary phenomenon in English concerns the steady loss of irregular verb forms. Here the problem is somewhat different from that of the *going to* auxillary. Over the years irregular past tense verbs such as *mown* have been replaced by their regular counterparts in English (*mowed*). Interestingly, this regularization process is sensitive to the frequency of use of the verb, with recent research suggesting that verbs regularize at a rate that is inversely proportional to the square root of their usage frequency (Lieberman et al. 2007). How can this be explained? If we consider irregular expressions as lexicalised routines, this may help to explain the circumstances in which they are lost. On our account speakers use routines because they can be accessed directly from memory, thereby bypassing the complex decisions of non-routine language production. However, this is only beneficial if the routine is readily accessible (see Wonnacott, this volume, for discussion of dual-route models of production). In other words, if accessing the routine (e.g., *mown*) takes longer than formulating the full form (e.g., MOW + -ED), or if speakers fail to access it at all on occasion, then it will fall out of use to be replaced by the non-routine regular form. Again Bybee (2006) gives a detailed account of how the process of regularization can be explained in terms of the probability of retrieving stored representations.

## 3.5 Summary and Conclusions

We began the chapter with the observation that taking part in a conversation is more straightforward than speaking or listening in isolation, despite the apparent complexity of the process. We went on to explain this paradox in relation to an account of dialogue processing called interactive alignment. Interactive alignment arises from automatic priming processes that link production with comprehension and *vice versa*. The essential notion is that people prime each other to use similar expressions at many linguistic levels simultaneously. This kind of alignment of speaking leads in turn to alignment at the level of deeper representations including the situation model adopted by the conversational partners. Because such alignment of situation models is the hallmark of successful communication, the interactive alignment process, operating during dialogue, greatly facilitates communication.

Interactive alignment also enables conversational partners to adapt to each other. Such adaptation happens both at a local level with speakers and listeners adopting each others' grammar and meaning in consecutive utterances and over longer time-scales. The longer term alignment occurs through a process of routinization with speakers and listeners creating routines or partially frozen expressions. We argued that this longer term alignment may be a central mechanism both for the acquisition of language and processes of historical language change.

## References

Aijmer, K.: Conversational Routines in English: Convention and Creativity. Longman, London (1996)

Bavelas, J.B.: "I show how you feel": motor mimicry as a communicative act. J. Pers. Soc. Psychol. **50**, 322–329 (1986)

Branigan, H.P., Pickering, M.J., Cleland, A.A.: Syntactic coordination in dialogue. Cognition **75**, B13–B25 (2000)

Brennan, S.E., Clark, H.H.: Conceptual pacts and lexical choice in conversation. J. Exp. Psychol.: Learn. Mem. Cogn. **22**, 1482–1493 (1996)

Bybee, J.: From usage to grammar: the mind's response to repetition. Language **82**, 711–733 (2006)

Chartrand, T.L., Bargh, J.A.: The chameleon effect: the perception-behavior link and social interaction. J. Pers. Soc. Psychol. **76**, 893–910 (1999)

Clark, E.V.: The Lexicon in Acquisition. Cambridge University Press, Cambridge (1993)

Clark, H.H.: Using Language. Cambridge University Press, Cambridge (1996)

Cleland, A.A., Pickering, M.J.: The use of lexical and syntactic information in language production: evidence from the priming of noun phrase structure. J. Mem. Lang. **49**, 214–230 (2003)

Cutting, J.C., Bock, J.K.: That's the way the cookie bounces: syntactic and semantic components of experimentally elicited idiom blends. Mem. Cogn. **25**, 57–71 (1997)

Dijksterhuis, A., Bargh, J.A.: The perception-behavior expressway: automatic effects of social perception on social behavior. In: Zanna, M.P. (ed.) Advances in Experimental Socialpsychology, vol. 33, pp. 1–40. Academic Press, San Diego (2001)

Fay, N., Garrod, S., Roberts, L., Swoboda, N.: The interactive evolution of human communication systems. Cogn. Sci. **34**, 351–386 (2010)

Fodor, J.A., Bever, T.G., Garrett, M.F.: The Psychology of Language. McGraw Hill, New York (1974)

Garrod, S., Anderson, A.: Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. Cognition **27**, 181–218 (1987)

Garrod, S., Doherty, G.: Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. Cognition **53**, 181–215 (1994)

Garrod, S., Fay, N., Lee, J., Oberlander, J., MacLeod, T.: Foundations of representation: where might graphical symbol systems come from? Cogn. Sci. **31**, 961–987 (2007)

Garrod, S., Pickering, M.J.: Why is conversation so easy? Trends Cogn. Sci. **8**, 8–11 (2004)

Garrod, S., Pickering, M.J.: Automaticity of language production in monologue and dialogue. In: Wheeldon, L., Meyer, A., Krott, A. (eds.) Automaticity and Control in Language Processing, pp. 1–20. Psychology Press, Hove (2006)

Garrod, S., Pickering, M.J.: Alignment in dialogue. In: Gaskell, G. (ed.) Handbook of Psycholinguistics, pp. 443–451. Oxford University Press, Oxford (2007)

Garrod, S., Pickering, M.J.: Joint action, interactive alignment, and dialogue. Top. Cogn. Sci. **1**, 292–304 (2009)

Garrod, S.C., Clark, A.: The development of dialogue co-ordination skills in schoolchildren. Lang. Cogn. Process. **8**, 101–126 (1993)

Giles, H., Coupland, N., Coupland, J.: Accommodation theory: communication, context and consequences. In: Giles, H., Coupland, J., Coupland, N. (eds.) Contexts of Accommodation, pp. 1–68. Cambridge University Press, Cambridge (1992)

Harley, T.: The Psychology of Language: From Data to Theory, 3rd edn. Psychology Press, Hove (2007)

Hartsuiker, R.J., Pickering, M.J., Veltkamp, E.: Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish/English bilinguals. Psychol. Sci. **15**, 409–414 (2004)

Hommel, B., Müsseler, J., Aschersleben, G., Prinz, W.: The theory of event coding (TEC): a framework for perception and action planning. Behav. Brain Sci. **24**, 849–937 (2001)

Hopper, P.J., Traugott, E.C.: Grammaticalization, 2nd edn. Cambridge University Press, Cambridge (1993)

Jackendoff, R.: Foundations of Language: Brain, Meaning, Grammar. Evolution. Oxford University Press, Oxford (2002)

Kirby, S.: Function, Selection and Innateness: The Emergence of Language Universals. Oxford University Press, Oxford (1999)

Konopka, A.E., Bock, K.: Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. Cogn. Psychol. **58**, 68–101 (2009)

Kuiper, K.: Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters. Erlbaum, Mahwah (1996)

Labov, W.: Principles of Linguistic Change, Vol. 1: Internal Factors. Basil Blackwell, Oxford (1994)

Levelt, W.J.M., Kelter, S.: Surface form and memory in question answering. Cogn. Psychol. **14**, 78–106 (1982)

Levelt, W.J.M., Roelofs, A., Meyer, A.S.: A theory of lexical access in speech production. Behav. Brain Sci. **22**, 1–75 (1999)

Lieberman, E., Michel, J.-B., Jackson, J., Tang, T., Nowak, M.A.: Quantifying the evolutionary dynamics of language. Nature **449**, 713–716 (2007)

Markman, A.B., Makin, V.S.: Referential communication and category acquisition. J. Exp. Psychol. Gen. **127**, 331–354 (1998)

Neumann, R., Strack, F.: "Mood contagion": the automatic transfer of mood between persons. J. Pers. Soc. Psychol. **79**, 211–223 (2000)

Nunberg, G., Sag, I.A., Wasow, T.: Idioms. Language **70**, 491–538 (1994)

Pardo, J.S.: On phonetic convergence during conversational interaction. J. Acoust. Soc. Am. **119**, 2382–2393 (2006)

Peterson, R.R., Burgess, C., Dell, G.S., Eberhard, K.: Disassociation between syntactic and semantic processing during idiom comprehension. J. Exp. Psychol. Learn. Mem. Cogn. **27**, 1223–1237 (2001)

Pickering, M.J., Garrod, S.: Toward a mechanistic psychology of dialogue. Behav. Brain Sci. **27**, 169–225 (2004)

Pickering, M.J., Garrod, S.: Establishing and using routines during dialogue: Implications for psychology and linguistics. In: Cutler, A. (ed.) Twenty-First Century Psycholinguistics: Four Cornerstones, pp. 85–102. Lawrence Erlbaum Associates, London (2005)

Richardson, D.C., Dale, R.: Looking to understand: the coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. Cogn. Sci. **29**, 1045–1060 (2005)

Richardson, D.C., Dale, R., Kirkham, N.Z.: The art of conversation is coordination: common ground and the coupling of eye movements during dialogue. Psychol. Sci. **18**, 407–413 (2007)

Rizzolatti, G., Craighero, L.: The mirror-neuron system. Annu. Rev. Neurosci. **27**, 169–192 (2004)

Shockley, K., Santana, M.V., Fowler, C.A.: Mutual interpersonal postural constraints are involved in cooperative conversation. J. Exp. Psychol. Hum. Percept. Perform. **29**, 326–332 (2003)

Tannen, D.: Conversational Style: Analyzing Talk Among Friends. Ablex, Norwood (1984)

Tannen, D.: Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse. Cambridge University Press, Cambridge (1989)

Watson, M., Pickering, M.J., Branigan, H.P.: Alignment of reference frames in dialogue. In: Forbus, K., Gentner, D., Regier, T. (eds.) Proceedings of the 26th Annual Conference of the Cognitive Science Society, pp. 1434–1439. Lawrence Erlbaum Associates, Mahwah (2004)

# Chapter 4
# Learning: Statistical Mechanisms in Language Acquisition

**Elizabeth Wonnacott**

**Abstract** The grammatical structure of human languages is extremely complex, yet children master this complexity with apparent ease. One explanation is that we come to the task of acquisition equipped with knowledge about the possible grammatical structures of human languages—so-called "Universal Grammar". An alternative is that grammatical patterns are abstracted from the input via a process of identifying reoccurring patterns and using that information to form grammatical generalizations. This statistical learning hypothesis receives support from computational research, which has revealed that even low level statistics based on adjacent word co-occurrences yield grammatically relevant information. Moreover, even as adults, our knowledge and usage of grammatical patterns is often graded and probabilistic, and in ways which directly reflect the statistical makeup of the language we experience. The current chapter explores such evidence and concludes that statistical learning mechanisms play a critical role in acquisition, whilst acknowledging holes in our current knowledge, particularly with respect to the learning of 'higher level' syntactic behaviours. Throughout, I emphasize that although a statistical approach is traditionally associated with a strongly empiricist position, specific accounts make specific claims about the nature of the learner, both in terms of learning mechanisms and the information that is primitive to the learning system. In particular, working models which construct grammatical generalizations often assume inbuilt semantic abstractions.

## 4.1 Introduction

Speaking at least one language is a ubiquitous human ability. Wherever humans are discovered, whatever else they are doing, they are talking. Conversing in our mother-tongue feels so effortless that it is rarely regarded as a skill or accomplishment, yet

---

E. Wonnacott (✉)
Department of Psychology, University of Warwick, Conventry CV4 7AL, UK
e-mail: E.A.Wonnacott@warwick.ac.uk

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_4,

we know that this behaviour relies on a highly complex body of knowledge about the structure of that language. This knowledge is sometimes called ‘grammatical’, though it is important to realize that we are not talking about the sorts of prescriptive grammatical rules which (depending on the decade) we may have been taught at school. Rather, the type of grammatical knowledge which is the concern of Cognitive Science is what psychologists call *implicit* knowledge, i.e. knowledge which is subconscious and largely inaccessible to the speaker.

Consider an example at the level of *morphology*—the constraints governing how meaningful strings of sounds, *morphemes*, may combine to form words. An English speaker implicitly knows that the regular past tense ending, the one we write as ‘ed’, differs according to the sound of the final consonant of the verb stem: if the verb ends in a /t/ or /d/ sound, the past tense is pronounced as /əd/ (e.g. *loaded*); if it ends in a voiceless sound (i.e. one produced without the vocal chords vibrating) it is pronounced as /t/ (e.g. *liked*); if it ends in a voiced sound (i.e. one produced with the vocal chords vibrating) it is pronounced /d/ (e.g. *loved*). This knowledge goes beyond a memory for the forms of individual verbs, since we are able to produce appropriate past forms for new verbs—try using the verbs *wid*, *wuf* and *wug* in the past tense (Berko 1958). As we will see, there is considerable debate about how such generalizations should be characterized. The point here is simply that such patterns must somehow be incorporated into our implicit knowledge of English “grammar”.

For an example at the level of syntax (the constraints governing how words combine into higher level structures), consider the following English sentence:

(1) a. *Jack threw Henry the ball*

Our understanding of this sentence includes not only the meanings of the individual words within it, but also the semantic roles imposed by their structural positions: *Jack* was the agent of the throwing action; *the ball* was the transferred object; *Henry* was the recipient. The following sentences, though composed of different words, have the same formal structure and exemplify the same semantic relationships:

(1) b. *Oliver sent William the parcel*
c. *Poppy gave Charlie her book*
d. *Jasmine told Jessica the news*

(Note that, as in the last example, the ‘transfer’ action may be metaphorical rather than physical). This relationship between an abstract structure and a semantic event is known as a *construction*. As with morphology, a new words test can reveal our implicit understanding of this generalized knowledge: given an appropriate context, we can spontaneously produce and understand the construction with new verbs as in 1e (Gropen et al. 1989):

(1) e. *He gorped me the ball*

Again, this suggests that a mental grammar of English must contain knowledge of the relationship between a general *X Verb Y Z* pattern and the semantic information it conveys. Interestingly, however, the grammar must also contain information which *prevents* us from applying the construction in certain circumstances, in particular there are a number of English verbs which *can't* be used in this construction, as in the ungrammatical, though perfectly comprehensible English sentences 1f–g:

(1) f. * *Oliver* **explained** *William the news*
g. * *Jack* **carried** *Henry the ball*

Thus an account of language acquisition has to explain *both* how we acquire the generalizations and the exceptions to those generalizations.

One final example will further illustrate the type of abstract structures which play a role in our use and understanding of language. Consider the sentence:

(2) *Put the block in the box on the table*

Your school-learned grammar might allow you to identify the following underlying linguistic categories or "parts of speech" (where 'Det' stands for Determiner and 'Prep' for Preposition).

| Put | the | block | in | the | box | on | the | table |
|---|---|---|---|---|---|---|---|---|
| Verb | Det | Noun | Prep | Det | Noun | Prep | Det | Noun |

These abstract categorizations also feature in our implicit knowledge. That is, we know which particular set of English words can fill, for example, the *Preposition* slots above. Note that this is not simply a question of knowing the word's meaning, since categorization is partially arbitrary (consider that the equivalent for a word which is a preposition in one language may be a verb in another and vice versa: for example in Chinese the instrument reading of the English preposition *with*—as in *eat with chopsticks*—is the verb[1] *yong*: DeLancey 2005). Moreover, our grammatical understanding of this sentence goes beyond an unstructured string of categories. Rather, we recognize that substrings of words may be grouped, and that this grouping affects our interpretation of the sentence. For example, the above sentence may be described using two different structural organizations, which can be shown with schematic bracketing—where NP stands for "noun phrase" and PP "prepositional phrase":

$$\textit{Put}\ {}_{\mathrm{NP}}[\textit{the block}]\ {}_{\mathrm{PP}}[\textit{in}\ {}_{\mathrm{NP}}\ [\textit{the box}\ {}_{\mathrm{PP}}[\ \textit{on}\ {}_{\mathrm{NP}}[\textit{the table}]]]]$$
$$\textit{Put}\ {}_{\mathrm{NP}}[\textit{the block}\ {}_{\mathrm{PP}}[\textit{in}\ {}_{\mathrm{NP}}[\textit{the box}]]]\ {}_{\mathrm{PP}}[\textit{on}\ {}_{\mathrm{NP}}\ [\textit{the table}]]$$

The first structure imposes an interpretation in which a block is placed in a box which is situated on a table. The second imposes an interpretation where the block was initially reposing in a box and is then is moved to the table. Note that the structures labeled "PP" and "NP" are embedded within each other—yielding what linguists

---

[1] More accurately, this word is usually categorized as a "co-verb". Li and Thompson (1974) (cited in DeLancey 2005) argue that co-verbs are graded in how syntactically "verb like" they are.

refer to as *hierarchical phrase structure*. This organizing principle is central to our understanding of syntactic phenomena. For example, in English the relationship between a *statement* and *yes–no question* is that the entire *NP* which is the *subject*[2] of the verb inverts with the auxiliary—as in the following examples:

[*The boy who is happy]*$_{\mathrm{NP}}$ **is** *singing*?
**Is** *[the boy who is happy]*$_{\mathrm{NP}}$ *singing*?

The purpose of the above examples was to give the English speaking reader an insight into his or her implicit knowledge of the language. This, of course, only scrapes the surface of the intricacies of English grammar, and similar complexity underlies all human languages.[3] The topic of this chapter is how structural patterns of different levels of abstraction are acquired by native speakers. Perhaps the most remarkable feature of this learning is that, in normal circumstances, it occurs in early childhood: a good bulk of the grammatical system is in place by the age of four, meaning that the average child is in some sense a competent grammarian before she can brush her own teeth. It is clear that this is not a result of explicit teaching. Few parents or teachers are aware of the types of patterns discussed above—and I doubt that any would relish the prospect of explaining the relevant concepts to a young child. Of course adult speakers do have an *intuitive* knowledge of the grammatical patterns of their native language, and so will be aware when their children produce utterances which are un-adult-like. However, studies have repeatedly shown that children receive very little explicit correction for grammatical errors (Braine 1971; Brown and Hanlon 1970; Newport et al. 1977).

Somehow, then, small children extract grammatical patterns via exposure to the language they hear around them *without explicit instruction*. Moreover the outcome of learning is very consistent, i.e. native speakers largely agree in their grammatical intuitions.[4] This makes learning a native language quite different from some types of human learning, such as learning how to grow crops or do mathematics, but rather similar to others, such as learning to perceive scenes in terms of discrete objects with particular locations and depths. In contrast to visual learning, however, language learning is a species specific behavior. No other animal communication system even approaches human language in its complexity. Attempts to teach human

---

[2] "Subject-hood" is itself defined in terms of the position that the NP holds within the hierarchical structure.

[3] The nineteenth century assumption that non-Western languages are more grammatically primitive is long discredited. This is not to say that particular languages may lack particular grammatical devices. To take an extreme example, Pirahã, a language spoken by a tribe of around a hundred people in a remote area of the Amazon, has been reported (controversially—e.g. Nevins et al. 2009) to lack certain grammatical structures previously thought to be universal. Nevertheless, Everett points out that Pirahã employs a highly complex, intricate grammatical system: "No one should draw the conclusion from this paper that the Pirahã language is in any way 'primitive'. It has the most complex verbal morphology I am aware of and a strikingly complex prosodic system." (footnote in Everett 2005).

[4] Languages may have different dialects, but there is internal agreement for speakers of that dialect.

language to other primates showed that these animals had little propensity to acquire the grammatical structure of human languages, despite intensive training regimes (Terrace et al. 1979; Seidenberg and Pettito 1979). In contrast, there is evidence that children begin learning the patterns of their native language from the first months of life (Aslin et al. 1998) and spontaneously produce their own utterances from about one year of age. Strikingly, the latter has also been found to be true even for children who are not exposed to any language. This is seen in deaf children who do not have exposure to a signed language. Being deaf, they do not acquire spoken language, but instead create their own gestural communication systems, dubbed ‘home-sign’. Although more simple than mature languages, these systems have been found to have several properties in common with other human languages (but lacking in other species’ communication systems), including use of discrete symbols to indicate fixed meanings (i.e. words) and, as we shall see later, the use of certain grammatical devices (see Goldin-Meadow 2003, for a review).

All of this indicates that children are born with a biological predisposition for language learning. But what is the nature of this predisposition? In the 1960s Chomsky famously proposed that it takes the form of an innate ‘Universal Grammar’ (henceforth UG), i.e. that children are born with innate knowledge about the possible grammatical organization and structure of human languages (Chomsky 1965). This radical theory revolutionized the scientific study of language which, at the time, was primarily conducted according to the principles of Behaviorism, a paradigm which rejected a role for mental structures in psychological theory. Chomsky pointed out the inadequacy of this approach for understanding human language: any account of linguistic behaviour must allow for the mental structures which underlie the utterances we produce and understand. He also argued that the simple associative learning mechanisms of behaviourist learning theory were inadequate to account for the abstraction of the necessary linguistic structures. Thus innate UG was proposed to act as a ‘blueprint’ for acquisition. According to this account of learning, the child’s task is not to create structure, but rather to identify which of a set of known structures match the sample of language she hears around her. The theory received apparent support from the fact that linguistic structures frequently recur across the languages of the world, even in ‘unrelated’ languages whose speakers have little or no contact (Greenberg 1963). One explanation is that languages are constructed from a single grammatical template with parameters which can set differentially for different languages.

It would be hard to overstate the influence of the UG hypothesis in Linguistics and Cognitive Science: the existence of some form of UG became an underlying premise of the main stream *Generative Linguistics* paradigm in the 1960s, and remains so to this day (although it is explicitly rejected by other brands of Linguistics: Langlacker 1987; Lakoff 1987; Bybee 1985). Nevertheless, the concept has been through many permutations over the years, even for researchers working within the Chomskian framework (for some current approaches see Chomsky 1995; Hauser et al. 2002; Crain and Pietroski 2006, and for a very different UG perspective, Jackendoff 2002). Some researchers use the term UG more generally, to include whatever structures and processes, language specific or otherwise, we bring to the task of language learning (for example, see Goldin-Meadow 2005). However the argument that all languages follow from, and are thus learnable from, an innate template of specifically *grammatical* knowledge has become increasingly untenable. For example, one claim about UG (e.g. Pinker 1984) has been that the categories which occur in the world's languages are drawn from a fixed set. However several researchers have argued that cross-linguistic evidence does not support this claim. Although certain categories can be identified across many languages (e.g. nouns and verbs, adjectives and prepositions/postpositions) this identification relies largely upon knowing the semantic properties of the words in the category. However, comparing *across* languages these classes may be syntactically quite different. For example, in some languages "verbs" (i.e. the class of words referring to actions) are marked for tense and action, but in other languages that property is associated with "nouns" (i.e. the class of words referring to things). In fact, Croft (2001) argues that categories across languages are so varied that they are essentially language specific (see also Evans and Levinson 2009). An alternative explanation for the fact that linguistic categories which are very similar—both in terms of semantics and grammatical behavior—do frequently reoccur across languages is that they comprise a 'good solution' for building a communication system within the confines of human conceptual biases and broader cognition. They thus emerge in the process of language change (see Christiansen and Chater 2008 for a general account of this type, and also Kirby and Oudeyer, this volume).

If children are not "pre-equipped" with grammatical knowledge, they must instead be endowed with learning mechanisms which abstract that information from their input. In recent years, many researchers have argued that this depends on a process of *statistical learning* (Elman 1990; Newport and Aslin 2000; Rumelhart and McClelland 1986; Seidenberg 1997), that is, an ability to identify reoccurring relationships between elements of the input, and make appropriate generalizations from probabilistic patterns. A growing body of evidence suggests that young children come to the task of learning with an ability to track probabilistic patterns. For example, Saffran et al. (1996) demonstrated that 8 month old infants are sensitive to syllable co-occurrence probabilities. Such information provides a useful cue for identifying word boundaries—an important part of acquisition since, in spoken language, unlike in written language, there are no 'gaps' between words. For example, in the sequence of syllables *pre-ty-ba-by* one cue to the fact that *pre* and *ty* form a 'unit', while *ty* and *ba* do not, is that across the whole language *pre* is followed *ty* about 80 % of the time, but *ty* is followed by *ba* only about 0.03 % of the time. Saffran et al.'s experiments demonstrated that 8-month-olds who were exposed to a stream of nonsense syllables could distinguish between those syllables which had frequently co-occurred in the string ("words") versus those which had infrequently co-occurred (i.e. "part-words" which crossed words boundaries).[5] Computational work has revealed that very similar statistics are relevant to *grammatical* learning. For example, Mintz et al.

---

[5] A variety of techniques exist for assessing whether pre-verbal infants distinguish different types of stimuli. Saffran et al. (1996) used *preferential listening* where infants indicate their interest in some aural stimuli by looking at a light which they associate with that stimuli. Longer looking times are taken to indicate greater interest in the stimuli. Saffran et al. (1996) found that, after exposure

(2002) conducted computational analyses over samples of speech (English) spoken to particular children (from the CHILDES database, see MacWhinney 2000). Their analysis treated each utterance in the input set as a string of (meaningless) words and tracked how often particular words co-occurred adjacently. Importantly analyses were conducted over very large samples (15,000–20,000 utterances in each corpus).[6] Clustering techniques were then applied to this data and revealed that there was sufficient information to separate words into the English categories 'noun' and 'verb' with good accuracy. Finally, further evidence that language learning involves tracking co-occurrence statistics comes from the abundant evidence that such probabilistic knowledge plays a role in real time language understanding. For example, many studies have shown that when we encounter a verb we predict what type of construction is likely to follow that verb on the basis of our past experience. For example, English speakers expect the verb *find* to be used in a transitive construction with a direct object, which is the construction with which it is most likely to occur across the language. Our sensitivity to this probabilistic information shows up when we read a sentence in which this expectation is violated as in 3 where 'found' is followed by a sentence complement.

(3) *The chef found the recipe would require using fresh basil*

The reader's 'surprise' can be captured using various psycholinguistic techniques (such as monitoring hesitation in eye-movements at the word *would*). Importantly the same 'surprise' does *not* occur for verbs which are likely to be followed by a sentence complement (e.g. *claim*: Trueswell et al. 1993; see also Garnsey et al. 1997; Snedeker and Trueswell 2004; Trueswell and Kim 1998). The point is that if language *processing* relies on knowledge of statistical likelihoods, that same information must somehow be accumulated as a part of language *learning* (see also Wonnacott et al. 2008).

In the remainder of this chapter, I will consider the statistical learning hypothesis with respect to the acquisition of certain aspects of Morphology and Syntax. The aim is to illustrate domains in which a statistical learning approach has been applied and explore the strengths and weaknesses of current accounts. Two overarching themes emerge. The first is that both our knowledge of grammatical patterns, and the ways in which we use and process them reflect the probabilistic nature of the input to which we are exposed. The second is that a statistical account of language acquisition is far from a "blank slate" theory of learning.[7] In fact, any such account must specify

---

(Footnote 5 continued)
to the nonsense syllable stream, infants showing longer looking times for *part-word* test items than for *word* test items (the stimuli were played repeatedly until the infant looked away from the light). The interpretation is that they found the part-words to be more *novel* and therefore more interesting.

[6] This under-estimates, rather than over-estimates, the quantity of language to which a child is likely to be exposed. Hart and Risely (1995) estimate that working class children hear an average of 6 million words per year.

[7] Both of these themes have been emphasized by other researchers. See Newport and Aslin (2000) for a statistical learning approach which strongly emphasizes the importance of innate constraints on learning. See Elman et al. (1996) for a connectionist approach to the issue of "innateness" in terms of

both the sources of information that are accessible to the learner (i.e. the *primitives* to the learning system), and the ways in which these different sources are combined and evaluated to yield generalization.

## 4.2 Statistical Learning in Morphology

*Words* may have internal structure. That is, we can identify meaningful parts, known as *morphemes*, which re-occur across different words in systematic ways. In the introduction I used the example of the English *"ed"* past tense form to illustrate our implicit knowledge of morphology. Linguists have traditionally described such patterns using "rules" which capture the relationships between stem morphemes, inflectional (i.e. grammatical) morphemes and composed forms. For example, Fig. 4.1 shows a formulation of the English past tense in terms of a conditionalized rule

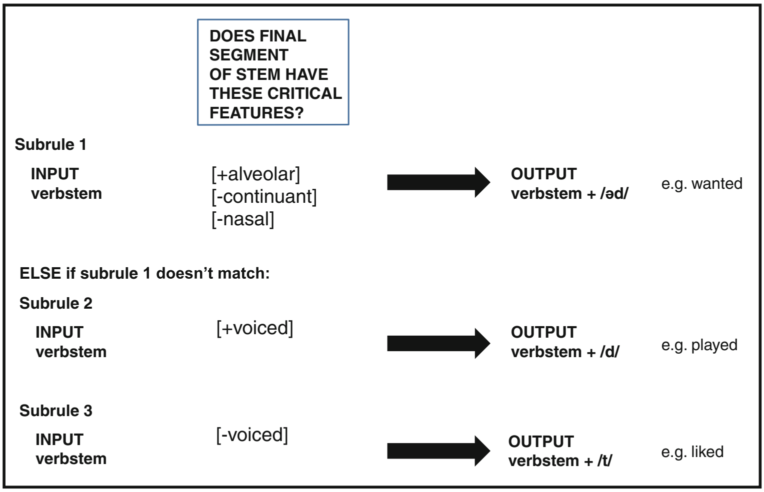


**Fig. 4.1** Rule for forming the English regular past tense, taking the stem as input. The properties in square brackets are distinctive features (roughly following Chomsky & Halle, 1968) which pick out a set of speech sounds. [+alveolar] means "produced via contact between the tongue and the alveolar ridge", [-nasal] means "air escapes from the mouth" (and not the nose, which is the case for e.g. /n/) and [-continuant] means "production involves a complete closure completely blocking airflow". This combination thus picks out the alveolar oral stops (i.e. /t/ and /d/). [+voiced] means "produced with the vocal chord vibrating", and picks out all the voiced speech sounds (i.e. vowels and consonants such as /g/ and /v/). [-voiced] picks out the non-voiced consonants such as /k/ and /f/. The sub-rules must be ordered so that subrules 2 and 3 only apply if the word does not end in /t/ or /d/.

---

(Footnote 7 continued)
the architectural make up of networks in different domains; See Seidenberg (1997) for a discussion of the relationship between statistical effects in language learning and language processing.

which takes the stem morpheme (like, love, load etc.) as ‘input’, concatenates it with an appropriate form of the inflectional morpheme (/t/, /d/ or /*əd*/), in accordance with the phonology (i.e. sounds) of the stem, and provides the appropriate past tense form (*likeT*, *lovD*, *loadID* etc.) as ‘output’. This rule not only provides a neat description of a widespread linguistic phenomenon (97 % of English verbs form their past tense form in this manner), but also appears to capture actual human *behavior*—when given a new verb stem (e.g. *wug*, *wuf*, *wid*) as input, we can produce past tense output (*wugD*, *wufT*, *widID*) according to the pattern.

Similar “new words” tests will generally reveal a variety of such productive rules for a given language, not only for verbs but also for other parts of speech (e.g. plural marking and case marking on nouns, agreement features on adjectives, etc).[8] From about 3–4 years of age, children have also been shown to productively inflect new words (Akhtar and Tomasello 1997; Berko 1958). This behavior indicates that the child has abstracted these regular morphological patterns from her experience of words in the input, and has somehow incorporated this knowledge as a productive part of the mental grammar. The question for theories of acquisition is *how* such learning occurs. One apparently straightforward account could be that that the child comes equipped with a learning algorithm which explicitly seeks out linguistic rules like those in Fig. 4.1. This process must involve both identifying the various reoccurring morphemes across a language, and extrapolating and mentally representing the rules which combine these morphemes into complex forms. According to this story, a speaker has no need to store forms like *“walkT”* and *“likT”* in memory, only morphemes like *walk*, *like* and *T*, since once the rule has been acquired complex forms can be routinely composed and decomposed “on-line” during language production and comprehension (e.g. Prasada and Pinker 1993).[9]

An obvious problem with the theory as described above is that there is no account of how we learn exceptional or *irregular* forms such as the past tense forms *went* and *ate*. Such forms appear to have to be stored as wholes, with some additional mechanism which ‘blocks’ the application of morphological rules where there is a stored exception (e.g. the stored form *went* blocks the formation of *goed*: Marcus et al. 1995). However a little further probing reveals that a system of wholly productive rules and unproductive exceptions is insufficient. For example, try applying the new words test to the stem *ping*. You may come up with *pingD*, in line with Fig. 4.1, but you might also hit on *pang* (Bybee and Moder 1983; Prasada and Pinker 1993). This behavior is clearly related to the existence of verbs like *sing*, *ring*, *spring* with their past tense forms *sang*, *rang*, *sprang*. The critical point is that the underlying pattern appears again to have some productivity, i.e. English speakers can access some generalized process which converts *ing* → *ang*. In fact, for the English past tense there are a number of such semi-productive patterns, so we might also get past

---

[8] Languages may make much more extensive use of productive morphology than English. For example, in many languages (e.g. many of the Eskimoan languages) entire nouns may be attached to the verb-stem as dependent morphemes, rather than appearing as separate words within the sentence (a phenomenon known as “noun incorporation”).

[9] Some more recent versions of this theory allow that at least some regular forms also be stored as whole forms (Pinker 1999; Pinker and Ullman 2002).

tense forms *med* for the stem *meed* (in line with the *eed* → *ed* pattern in *bleed/bled*, *feed/fed* etc) and *prew* for *prow*, (in line with the *ow*→ *ew* pattern in *blow/blew*, *grow/grew* etc). Although some early theories (e.g. Chomsky and Halle 1968; Halle and Mohanan 1985) attempted to capture these types of regularities in terms of rules also (for example rules which altered the vowel in the stem in particular contexts), it turns out to be very difficult to identify a precise set of phonological patterns which trigger particular past tense forms (see Bybee & Slobin 1982; Pinker 1999).

More fundamentally, characterizing these semi-productive regularities in terms of clear cut rules ignores an important finding: the extent to which the patterns are extended to new words depends on our experience of how consistently they apply across the language. For example, the probability of producing *pang* as the past tense of *ping* will be affected by the number of verb stems which are phonologically similar to *ping* and have past tense forms similar to *pang*. This statistic is known as 'type frequency'. Exactly how similar the new verb is to familiar verbs which use the pattern is also important. The effects of these statistics may also be seen in the patterns of errors made by young children. The majority of morphological errors are so called overgeneralization errors which arise from over-applying the regular pattern (e.g. saying *goed, gived* etc), but other patterns may also be over-generalized. For example, incorrect forms such as *brung* and *brang* (Xu and Pinker, 1995; Bybee & Slobin, 1982) may result from relatively high frequency of the relevant patterns across past tense forms in the language. The frequency of individual verbs ('token' frequency) is also important—children are less likely to use the wrong pattern with a more frequent word (e.g. *sleeped* is a less common error than *weeped*).

Generalization errors are not restricted to child language but are also made by adults, particularly the overgeneralization of high frequency patterns to low frequency words (these are also the items that are most likely to change their morphological behavior over *generations* of speakers and understanding the interaction of type and token frequency is critical to understanding the process of language change: Bybee 1985). In fact, such similarity-based *graded* productivity turns out to be rife in natural language morphology (see Bybee 1995; Hay and Baayen 2005). From the perspective of language acquisition, our theories must therefore include an explanation of how learners come to track different statistics, such as type and token frequencies and how these become integrated into the productive morphological system which emerges. One class of statistical learning system which has been extensively studied in this domain are *connectionist models*, also known as *neural networks* since their architecture is inspired by the fact that neural circuitry is comprised of networks of interconnected units (neurons) which learn by adjusting the connections between those units (synapses).[10] These models are able to extract probabilistic patterns in the course of learning mappings between sets of input and output nodes. Such a model was first applied to morphological learning by Rumelhart and McClelland (1986) who presented a landmark model of the acquisition of English past tense. This model (and many subsequent models, e.g. Plunkett and Marchman 1991; Hare

---

[10] Although connectionist models are neurally inspired, there is no claim that they constitute a biologically plausible model of neural circuitry.

et al. 1995; Daugherty and Seidenberg 1992)[11] learned to map a set of input nodes representing the sound patterns in the verb-stem to a set of output nodes representing the related past tense form. Different input/output nodes represent different aspects of the phonology of the stem or past tense form, meaning that representations were *distributed* (for example, the representations of *sing* and *ring* would have shared components, i.e. some subset of nodes would be activated for both). These models can be "trained" to map a set of stems to past tense forms (e.g. given *sing* they generate *sang*), importantly, without having any explicit rule formation process—after sufficient training the models may also generalise appropriately to new words, i.e. given *wug* generate *wugD* and given *ping* generate *pingD* or possibly *pang*. The models also make errors, particularly before they are fully trained, and, as for human children and adults, errors are affected by token and type frequency statistics. In the models, this is a direct consequence of the statistical nature of the learning process: as particular words are frequently encountered, the mappings between the stem phonology and past tense phonology are proportionally strengthened (token frequency) but, since words have *distributed representations*, aspects of those mappings which frequently re-occur across words are also strengthened (type frequency). Errors are therefore likely if a verb is low frequency but its stem is consistent with an alternative high-frequency pattern. These models thus capture the probabilistic effects of *phonological similarity*.

Some more recent connectionist models have reconstrued the learning problem so that rather than mapping directly between different phonological forms, the network's central 'task' is to map *phonological* and *semantic* representations, i.e. the sounds of words to their meanings (e.g. Joanisse and Seidenberg 1999; Plaut and Gonnerman 2000). So, for example, for past tense the phonological string *walkT* might map to a semantic representation including the information *WALK-PAST-THIRDPERSON* and the form *walkS* to a semantic representation including *WALK-PRESENT-THIRDPERSON*. Links between different forms of the same verb can thus emerge from shared semantic and shared phonological representations. Models of this ilk also have the potential to capture effects of *semantic* similarity (i.e. when clusters of words with similar meanings show similar morphological behavior, as is quite common in morphological systems across the world's languages: Wierzbicka 1988). Moreover, mapping the sounds of words to their meanings provides a more natural model of the child's actual task during language acquisition.[12] Note that in this view "morphemes" such as */s/ = THIRD-PERSON-PRESENT* are *emergent* rep-

---

[11] Later models had more complex architectures, including layers of hidden units between the input and output units, and used different learning algorithms.

[12] All connectionist models require an error signal to drive learning. The models learn by predicting outputs for given inputs (early on predictions are random guesses), receiving feedback as to what the correct response should be, and then updating the "weights" (which drive the predictions) accordingly. For models which map between phonology and semantics, we are envisioning a child who implicitly compares the sound she would have expected for a given meaning with the one she is hearing, and the meaning she would have expected for a given form with the one that is currently implied.

resentations which arise when pieces of form and meaning are *repeatedly* associated (e.g. Gonnerman et al. 2007).

In short, connectionist models have proved useful in elucidating the origins of graded productivity and the probabilistic usage of morphological patterns, for example, explaining effects of token and type frequency. However whether this type of account is appropriate remains highly controversial. In particular, there is concern that the generalizations which networks acquire only approximate bona fide "rules". One theory claims that these are necessary to account for *regular* morphological processes, although generalization seen with irregulars may be accounted for by the storage of those forms in a connectionist-style associative memory system (a so-called 'dual-route' account, Prasada and Pinker 1993; Pinker and Ullman 2002). Debate has therefore centered around whether regular and irregular forms show differences in processing. For example, some studies found effects of token frequency (e.g. Prasada et al. 1990) and phonological similarity (e.g. Prasada and Pinker 1993) for irregulars but *not* for regulars. However both of these effects have since been found for regulars (token frequency: Schreuder et al. 1999; Hare et al. 2001; phonological similarity: Albright and Hayes 2003). From a statistical learning perspective, the fact that graded, statistical effects are harder to detect with regulars, so that the patterns therefore appear more 'rule-like' in their application, results from the statistical properties of the input. Regular patterns generally have a much higher type frequency than the alternatives,[13] resulting in a strong drive to apply those patterns across the board. This tends to overwhelm any factors concerning the properties of particular words.

More recently, arguments have focused on *neurological* evidence which suggests that producing and comprehending regular and irregular forms involves different brain areas. This comes both from brain imaging studies (e.g. participants hear/read a word and we see which brain areas are activated; e.g. Joanisse and Seidenberg 2005) and from studies of individuals who suffer damage to different areas of the brain (i.e. damage to one area of the brain may affect the production and comprehension of regulars, damage to another the processing of irregulars; Marslen-Wilson and Tyler 2007). However, the interpretation of these differences remains controversial. One explanation is that they actually arise from differences in the extent to which producing and comprehending regular versus irregular forms relies on *semantic* and *phonological* representations. For example, the task of producing the idiosyncratic *took* when presented with *take* strongly relies on identifying the particular word. Thus semantic representations play a role (Joanisse and Seidenberg 1999). Also, regular forms tend to be more phonologically complex than irregulars (for example they are more likely to end in consonant clusters as in the /spt/ in *claspT*: McClelland and Patterson 2003). Joanisse and Seidenberg (1999) showed that when a connectionist

---

[13] Some researchers have argued that the most frequent form is not always the one that acts as the regular rule (e.g. Marcus et al. 1995). However in such cases the *variety* of types may be important. Plunkett and Nakisa (1997) demonstrate that a pattern which is not the most frequent can become the most productive in a connectionist model provided that the set of words to which the pattern applies are more dissimilar to each other than is the case for the sets of words associated with alternative patterns. Capturing such variability relies on the use of models with a more complex architecture, including a layer of hidden units between input and output mappings.

model with separate banks of semantic and phonological units had been trained up on the English past tense, the production of *irregulars* and *regulars* could be differentially affected by knocking out semantic and phonological areas respectively.

An alternative account of the neurological evidence is that producing and comprehending regularly inflected words does actually involve the assembly/decomposition of complex words from/to their component morphemes. Some direct evidence for decomposition has been presented for comprehension. Post et al. (2008) argue that any word, including a new word, which potentially matches the output of the schema in Fig. 4.1, may be automatically decomposed. For example, the word *nomd* (presented aurally) is a potential past tense which can be formed from *nom*+/d/ (try reading 'nommed' aloud). Evidence that such forms are actually decomposed comes from an experiment in which listeners had to say if pairs of words were the "same" or "different". Participants took longer to say that pairs like *nomd–nom* were different than pairs like *nomt–nom*. This is interpreted as evidence that *nomd* is decomposed into *nom*+/d/ (note that *nomt* does not fit with the schema in Fig. 4.1 since /t/ should follow a voiceless consonant and /m/ is voiced).[14] Thus, at least for comprehension, there is some evidence for the storage of separable morphemes (like /t/, /d/ and /əd/) and for the usage of "rules" to access these morphemes where appropriate. It remains to be seen whether connectionist-style models where rules and morphemes are *emergent* forms, with *graded* representations, can capture this type of data.

In summary, the statistical make up of the input language has important consequences for the emergent morphological system, and probabilistic patterns may be seen even for very rule-like systems. Any model of morphological learning must account for this, as well as accounting for situations in which processing is indeed very rule-like. Connectionist models have been important in elucidating the origins of statistical phenomena, but it remains to be seen whether they can account for the full range of behavioral evidence, or whether statistical learning systems with different architecture assumptions are necessary. One piece of evidence suggests there may be some further constraints or biases in the statistical learning system. This comes from the study of a child whose language input contained *inconsistent* patterns of morphological usage (Singleton and Newport 2004). This child was deaf and his only input was the sign language used by his parents who were imperfect users of that language, having themselves not been exposed to the language until adulthood (this is typical of so called late-learners of a language; Newport 1990). Surprisingly, the child's own language in many ways surpassed that of his parents. Most relevant here is that there were situations where the child's parents erratically used multiple complex morphological forms (a little like randomly using all three of, say, *sleepT*, *sleepD* and *slept*) but the child himself did not replicate this probabilistic usage and instead boosted the frequency of the most frequent form and eliminated the others. Thus the child did not replicate the probabilistic patterns of the input, but in a sense

---

[14] The critical factor appears to be whether past-tense forms are potentially decomposable, rather than whether the relationship between stem and past tense is regular. For example, *slept* is traditionally irregular but is nevertheless decomposable into *slep* + /t/ (note that this fits Fig. 4.1 as /p/ is voiceless) and it seems to be processed akin to regulars rather than irregulars (Joanisse and Seidenberg 2005).

'sharpened' these patterns to make the system more consistent and rule-like (see also Newport and Aslin 2000). To my knowledge, this type of language change within a single learner has not yet been addressed within the connectionist literature.

Whatever the adequacy of statistical models of morphology, one final point is worth emphasizing. As I said in the introduction, statistical models are far from blank slate learners. This is particularly clear when considering working computational models. Models instantiate hypotheses, not only as to the architecture of the learning system, but also as to the information available to that system. For example, models of English past tense may assume features such as 'voiced', 'voiceless' and 'alveolar' as primitives in the phonological representations of words. If the model maps phonology to semantics it will have (at least) PRESENT and PAST as primitives. This does not preclude an account in which these features are themselves learned (though this of course opens the question of how *that* learning occurs), but where the behavior of the model depends on a particular set of primitives this makes the strong claim that by the time that morphological learning occurs, such features are available as candidates for mapping.

## 4.3 Statistical Learning of Syntax

While *morphology* governs how words are formed from smaller meaningful parts (morphemes), *syntax* is the system which governs how those words combine to form phrases and sentences. It is syntax which provides the massive productivity and expressivity of human language. Following early (and extremely influential) arguments made in Chomsky (1957), many researchers rejected the notion that statistical learning mechanisms could appropriately abstract syntactic knowledge. For example, Pinker (1987) argued that a learning procedure which simply attended to how words are distributed within sentences could easily be led astray. As an example, he suggested that a distributional analysis of the sentences in 4a–c could lead to the incorrect generalization in 4d:

(4) a. *John ate fish.*
b. *John ate rabbits.*
c. *John can fish.*
d. **John can rabbits.*

More recently however, access to fast computing has allowed researchers to explore how a distributional learner would fare if given access to very large amounts of linguistic input. We saw in the introduction that distributional computational analyses which cluster words on the basis of adjacent co-occurrence statistics can distinguish English "nouns" and "verbs" with good accuracy, provided they are applied to a sizable corpus of sentences, as opposed to just three utterances (Mintz et al. 2002). The potential error in 4d—which is the result of a mis-categorizing the word *rabbits*—is avoided because words like *fish* get clustered with both nouns and verbs and words like *rabbit* do not, since *rabbit* shares many more distributional characteristics with words used primarily as nouns (Mintz et al. 2002). Further research has shown that distributional statistics can divide words into a more comprehensive set of categories, and these correspond fairly well to the types of syntactic categories identified by linguists (adjectives, prepositions etc.; Mintz 2003; see also Finch and Chater 1994). An inherent advantage of a statistical approach is that it has the potential to capture the situation where category membership is graded rather than absolute, and where words appear to act like partial members of more than one category (e.g. the English word *near* appears to be a blend between an adjective and a preposition, Manning and Schutze 2001). Related statistical analyses may also capture some information about permissible and impermissible sequences of categories (Elman 1990; Church 1988).

Such computational research has played an important role in demonstrating that the input holds a good deal of information for a learner trying to build a syntactic system, providing that that learner is equipped with mechanisms which can tap into sequential patterns. On the other hand, we know that human syntactic knowledge cannot be captured by a grammar which generates unstructured sequences of categories. How far can we take a statistical approach to syntax learning? In the remainder of this section I consider this problem with respect to three topics: linking syntactic structure to abstract semantics, avoiding overgeneralization and acquiring hierarchical phrase structure.

### *4.3.1 Linking Formal Structure and Meaning*

The types of statistical analyses discussed may yield a formal system for generating possible word strings, but the strings themselves convey no further information. In contrast, the *raison d'etre* of natural language syntax is to provide a means of systematically encoding a structured *message*. For example, we have seen that the *X verb Y Z* structure (as in *Jack threw Henry the ball*), conveys a *transfer* event and further indicates the roles which the entities denoted by the noun phrases *X*, *Y* and *Z* play in that event (the so called "thematic" roles which linguists label *agent*, *recipient* and *theme*). This is an example of a *construction* i.e. a systematic mapping between a formal pattern (here the positions that words and phrases can occupy within the utterance) and a semantic pattern. Many researchers have focused on the acquisition of constructions, and particularly constructions centered around verbs (verb-argument structure constructions), as a core component of syntax acquisition (e.g. Tomasello 2000; Gleitman et al. 2005).

As always for theories of language acquisition, theoretical debate concerning the learning of constructions has focused on whether the necessary structures can be gleaned via exposure to the input, or whether children come equipped with relevant innate knowledge. For example, Tomasello (2000) argues against a UG approach on the basis of evidence that young children's grammars (before they reach about 3 years) are not 'adult-like'. In particular, he claims that for verb argument structure constructions such as the *X verb Y Z* structure, there is no evidence that children know the link between the formal structures and abstract thematic relations like *agent* and *theme*. This is because, unlike adults and older children, young children are unable to use the structure with new verbs, and their usage of the structures in everyday speech is generally limited to one or two verbs. On the basis of such data, Tomasello (2000) proposed the "verb island hypothesis": early on children use structures which revolve around specific verbs, such as *"X want Y"* where X = *person-that-wants*, Y = *thing-wanted*. According to this theory, abstract constructions only emerge once children have acquired multiple related verb-specific structures and notice the relationships across them. Interestingly, however, experiments using preferential looking[15] have revealed that children may have some knowledge of the relationship between word order and abstract thematic roles at a much earlier age than they are able to demonstrate in production. For example, one study found that 21-month-olds who heard a transitive sentence containing a new verb (such as *Rabbit is blicking Monkey*) whilst viewing two scenes with a novel action—one with correct noun assignments, one with the roles reversed (e.g. correct: *RABBIT–JUMPS-ON–MONKEY*, reversed: *MONKEY–JUMPS-ON–RABBIT*)—tended to look longer at the correct scene. Some researchers have argued that this early evidence of abstract knowledge indicates that learning is not entirely input-driven and that the child "contributes" some structure to the learning process (Gertner et al. 2006; see also Fisher 2002).

What is the role of statistical learning in these accounts? In fact, the verb-island hypothesis relies on statistical learning mechanisms: the child must have an ability to form generalizations once a 'critical mass' of related structures has been accumulated, just as we saw that repeated patterns could lead to generalization in morphology. Here, however, the ability to generalize also relies on the child's ability to notice (subconsciously of course) the abstract semantic relationships which hold across sentences. That is, she must be able to identify that in *Henry kisses mummy* and *Poppy drinks milk* the roles played by *Henry–Poppy* and *mummy–milk* in the kissing and drinking events are analogous. One way that a statistical learning model might capture this type of learning is to include semantic representations alongside 'word string' representations of input sentences and some models have taken this approach (St. John and McClelland 1990; Miikkulainen 1996; Chang et al. 2006). One such model, presented by Chang et al. (2006), not only proved able to learn abstract constructions, but also captured some of the developmental data discussed above. The model included an SRN (Serial Recurrent Network) which is a type of connectionist statistical learning system which learns by sequentially predicting upcoming words and learning from incorrect predictions. This type of model has been shown to be able to abstract grammatically relevant information from word sequences (Elman 1990). Critically, in the Chang et al. (2006) model, each sentence was also coupled with a structured semantic representation including—amongst other things—thematic roles like *agent* and *theme*. This semantic information also fed into the prediction process. In line with the developmental data discussed above, the model showed evidence of verb-island effects early in learning. Specifically, it showed different degrees of accuracy when using the same structure with different verbs, and its ability to produce sentences with new verbs only gradually developed. However, results from

---

[15] This is a standard methodology for assessing infant preferences for a particular visual stimuli.

preferential looking experiments were also replicated: given a sentence containing a new verb, the model could identify which of two semantic representations was correct *before* it would be able to correctly produce that sentence itself.[16] Eventually, like older children and adults, the model also passed the new verbs test in production, indicating that abstract structures had been learned.

The Chang et al. (2006) model provides a good illustration of how an input-driven, statistical learning explanation may still embody strong claims about what is 'built in' to the learner. On the one hand, the model is able to acquire abstract syntactic representations without access to innate *syntactic* knowledge of the type envisioned by some UG accounts (e.g. Pinker 1989; Radford 1988). On the other hand, it does assume that the learner has access to abstract *semantic* structures, which in the model are given rather than learned. In this way it is in line with some approaches which emphasize the structure innately contributed by the learner (Gertner et al. 2006; Goldin-Meadow 2003). In particular, the model comes pre-equipped with abstract representations like *agent*, *recipient* and *theme*. Its behaviour thus demonstrates that input-driven 'verb-island' effects still arise, given the task of matching such representations up to particular words in the input.

Assuming that thematic roles are innate is a strong hypothesis, but it is supported by some independent evidence. This comes from the study of *home-sign* systems—the self-created language systems of deaf children, to which I alluded in the introduction. These language systems—which are formed without linguistic input—have repeatedly been found to encode a fixed sets of abstract thematic roles (including *agent*, *recipient* and *theme*, Goldin-Meadow 2003). This is consistent with a hypothesis in which such representations are inbuilt. It also demonstrates an unlearned bias to desire to communicate *this specific type of information*—a bias also inherent in the Chang et al. (2006) model.

In other ways, the Chang et al model is relatively unconstrained. For example, the link between word order and thematic role emerges during learning, given the serial processing nature of the SRN architecture and the semantic representations. This contrasts with approaches which assume innate links between thematic role and word-order (and also between thematic-role and morphological case-marking—the other cross-linguistically common device; Pinker 1984; Jackendoff 2002). It has been shown that the model can equally learn languages in which the same information is marked via morphology or some combination of morphology and word order (Chang 2009).[17]

In short, there is evidence that a statistical learning system can acquire essential links between *syntax* and *semantics*, provided it has access to structured semantic

---

[16] The data from the Gertner et al. (2006) experiments were not specifically modelled in Chang et al. (2006) but the result is generally consistent with the model's account.

[17] Ultimately we need an account of language learning and language change which explains why word order and case marking are so prevalent as means of encoding thematic information. However from the perspective of learning, the account must also be sufficiently flexible to explain the learning of other additional or alternative devices. For example, sign languages may also employ the modality specific device of directing signs with the signing space (e.g. moving a GIVE gesture towards a particular person to indicate that they are the recipient).

representations over which it can generalize. Of course current models are far from acquiring the full range of constructions for any language. Nevertheless, this type of modeling work is likely to play a central role in future research into syntactic learning.

### *4.3.2 Avoiding Overgeneralization*

A classic criticism of input-driven theories of acquisition is that unconstrained learning may lead to an over-generalized grammar (e.g. Baker 1979; Pinker 1989). This can again be illustrated with respect to verb-argument structures. We saw in the introduction that not all combinations of verbs and argument structures are grammatical in the adult grammar, even where that combination would seem semantically plausible (sentences 1f and 1g above). However, once they are able to generalize constructions to new verbs, children may start to spontaneously use known verbs in constructions in which they have not encountered them. This may result in overgeneralization as in 5a and 5b (from Gropen et al. 1989):

(5) a. *Carry me the ball.*
b. *Don't say me that!*

The theoretical problem is how the child eventually learns that such combinations of verbs and structures are incorrect, given that they cannot rely on overt correction from caregivers.[18] In other words, if children are able to generalize verbs to new structures, but they don't get corrected when they use them with incorrect structures, how do they eventually learn that this generalization is actually ungrammatical?

This is the classic problem of *no negative evidence* and it applies whenever there is a plausible but ultimately incorrect linguistic generalization. One possible solution to the paradox is that apparent "exceptions" to generalizations are not arbitrary but are in fact conditioned. Overgeneralization will therefore cease once the child has identified the correct conditioning factors, perhaps with the help of innate knowledge of what such factors might be. For example Pinker (1989) proposes that the argument structures of verbs is in fact conditioned by subtle semantic factors which are not apparent at first glance. However attempts to come up with sets of absolute conditions have generally been found to be unsuccessful (Bowerman 1988; Braine and Brooks 1995; Goldberg 1995) and arbitrary exceptions remain.

Although the problem of no-negative evidence is often presented as evidence against input-driven accounts, many researchers have argued that the solution may lie in the statistical nature of language learning and usage. The first step is to relax

---

[18] Since I have found that people outside of this discipline (particularly middle class academically minded parents, accustomed to explicitly correcting their children's grammar) have difficulty accepting this point, it is worth highlighting. To further see that parental correction does not account for our knowledge of verb syntax, consider that many of the verbs which are ungrammatical in this construction are Latinate verbs (e.g. *donate*). It seems unlikely that such verbs are widely used (and therefore corrected) in childhood, yet we all known their syntactic restrictions.

the criteria on what is learned. If the end state of learning is a grammar which can determine "grammaticality" with an absolute yes-no judgment, there is indeed a learnability paradox. If instead we permit a grammar which allows *varying degrees of certainty*, "grammaticality" may be determined via probabilistic inference. In fact, at least for verb-argument structure constraints, there is evidence that judgments are graded in just this way. For example, Theakston (2004) asked both children and adults to rate "ungrammatical" sentences in which verbs occurred in the wrong structures. She found that such sentences received higher ratings when the verbs were of low frequency. For example, children judged *"He arrived me to school"* to be better than *"He came me to school"* (*come* occurs with higher frequency than *arrive*). Even adults, who of course have more familiarity with all verbs than children, nevertheless gave higher ratings when the verbs were very low frequency (for example preferring *"He vanished the rabbit"* to *"He disappeared the rabbit"*). In other words, the more a particular verb has been encountered in a particular set of structures, the less likely speakers are to extend that verb to a new structure (Braine and Brooks 1995). As we saw in the introduction, the idea that we track how often different verbs occur in different structures is further supported by data from sentence processing.

From this statistical perspective, determining 'grammaticality' is a question of weighing up the evidence in the input. It is logically true that a child can never know that the verb *come* may not one day show up in a transitive structure—but their wide experience of that verb appearing in other structures *but not the transitive* can make them pretty certain. With less frequent, or entirely novel, verbs, it makes sense to assume that more general patterns may apply. Note that this tendency for more over-generalization with low frequency items is exactly what we saw with morphology, pointing to common statistical inference processes. Partial conditioning factors such as verb semantics can now be considered cues which play a role in this statistical inference. For example, if other verbs similar to verb X occur in structure Y, that provides evidence that verb X may also do so. There is evidence that children and adults are influenced by both semantic and phonological similarity in just this way (Braine and Brooks 1995; Ambridge et al. 2008; Brooks and Tomasello 1999; Gropen et al. 1989). Finally statistical patterns at a 'higher level' may also play a role. If there is evidence that a construction is very "open", learners are more likely to generalize using that construction than if there is evidence that the construction's usage is restricted to particular words (Goldberg 2005; Wonnacott et al. 2008).

The picture that emerges is one in which multiple sources of information can influence a judgment of grammaticality. The problem of how to evaluate and combine probabilistic cues is in fact well known in cognitive science more generally. For example, it is seen in the problem of combining visual cues to give percepts of depth and localization. Recent approaches to cognition have emphasized the use of *Bayesian statistical inference* to estimate cue reliability from correlations in the input (Jacobs 2002; Chater et al. 2006). This type of statistical inference is to some extent implicit in the connectionist approach discussed previously, however Bayesian models differ in making the formation and evaluation of hypotheses explicit. Both connectionist and Bayesian approaches have been applied to the problem of constraining overgeneralization (Connectionist: Allen and Seidenberg 1999; Chang 2002; Bayesian: Perfors et al. 2010; Dowman 2000; Onnis et al. 2002)

An important question for future research is whether there are further constraints on the process of restricting generalization. For example, if a verb does not appear in structure X, does its frequent appearance in *any* other structure count as evidence against its future appearance in X? Some researchers have argued not (e.g. Goldberg 2005). Identifying such constraints may be important in understanding how native language learners end up with a set of grammatical intuitions that are so similar.

### *4.3.3 Hierarchical Phrase Structure*

Syntactic analyses of very different and unrelated languages have repeated revealed that sentences are composed from phrases, which may themselves be composed from smaller phrases, and so on. This is *hierarchical phrase structure*. We saw examples in the introduction with the two structures underlying the ambiguous *"Put the block on the box on the table"*. Any theory of language acquisition must account for how children are able to acquire grammars which generate these types of structures. Within the Chomskyan tradition, this principle of syntactic organization constitutes part of Universal Grammar. In other words, children are supposed to come to acquisition assuming that utterances are composed of phrases and expecting syntactic relationships to operate over phrases rather than single word categories (the principle of *structure dependence*, Chomsky 1968). However this approach has assumed that constituency is universal, and this is controversial. Evans and Levinson (2009) argue that there are many languages which show no evidence of constituency since they have "free" word order and words which are semantically grouped are not necessarily contiguous within a sentence. It is interesting that these languages do nevertheless have a means of encoding a hierarchical message: elements of distinct levels of structure may be grouped using multiple levels of morphological case marking (i.e. word endings). This suggests the possibility that it may be the structured nature of conceptual representations which is "universal" to human language[19], rather than a particular means of encoding that information.

Nevertheless, the ability to learn hierarchical phrasal structures poses an important challenge for statistical models of acquisition (Chomsky 1957). Within the connectionist tradition there has been an attempt to demonstrate that models can capture behaviours which appear to rely on phrasal constituency. For example, the ability to compose indefinitely long complex noun phrases means that agreement relations may hold over several words—as in the subject-verb agreement in 6 (note that only the highest levels of phrasal structure are shown):

(6) *[The* **boy** *[who chases dogs which chase cats ]]*$_{\text{NP}}$ **runs** *fast*

---

[19] Although Everett (2005) controversially claims that Pirahã lacks the ability to encode *recursion*, a particular type of hierarchical structure whereby the same phrase may be embedded within a phrase of the same type.

Elman (1993) probed whether an SRN was sensitive to such long distance dependencies. As discussed above, SRNs learn by predicting upcoming words in a sentence (input units represent the current word, output units represent the next word—and the difference between this prediction and what the next word turns out to be provides the error signal which drives learning). SRNs also have *hidden units* between the input and output which feed into the prediction, and, critically, a set of 'context' units, which carry a copy of the previous state of the hidden units. Since these serve as additional inputs to the hidden units, the current activation of these units is affected by both their current and previous activation, which was in turn affected by the previous activation, and so forth. Thus, although the predictions of SRN models are most strongly dependent upon the previous word, there is also a rapidly diminishing memory for the earlier sentence context. Elman showed that an SRN which was trained on sentences from a pseudo-English grammar learned to reject sentences like **The* **boys** *who chase dogs which chase cats* **runs** *fast*, demonstrating that it was able to hold information about agreement over long distances (in fact the network only succeeded when it was first trained on simple sentences such as *The boy runs* and *The boys run*; however, this is controversial since it was not replicated in a later study, Rohde and Plaut 2003). Another study (Lewis and Elman 2001) showed that an SRN could learn that question forms such as 7a were acceptable whilst forms such as 7b were not.

(7) a. Is the man who is coming here?
b. * Is the man who coming is here?

This ability appears to rely on an ability to recognize the noun phrase in 7c:

(7) c. [The man who is coming]$_{\text{NP}}$ is here.

Does the SRN succeed in these tasks by learning something about hierarchical structure? It is certainly clear that the type of structure that is acquired is not equivalent to that which can be implemented in a symbolic processor. For example, a symbolic system has no limits on the depth of embedding which it can process. In contrast, processing in the SRN may rapidly breakdown, particularly given a certain type of embedding known as "center embedding" (Christiansen and Chater 1999). However, this is not necessarily a shortcoming of the models, since human processing may also break down in these circumstances (try figuring out *"A man that a woman that a child that a cat that I heard saw knows loves beer"*).

Nevertheless, it is not clear that SRNs do extract phrasal structure. Steedman (2002) argues that the SRN models approximate the class of Finite State Markov Machines. This means that they treat essentially represent sentences as an unstructured string of categories. The form of a word may thus depend upon the previous set of words up to some length (a so called "*n-gram*"—although unlike fixed n-gram models, SRNs can potentially learn what length n-gram is most relevant). Such a system may prove able to track fairly long distance relationships, but will never be able to represent structures of the type necessary to disambiguate such sentences as *Put the box on the table on the shelf*. In addition, the interpretation that the network presented by Lewis and Elman (2001) had learned something about an NP constituent is challenged by recent work showing that the relevant sentences can be differentiated by a learner sensitive only to relationships between adjacent words (Reali and Christiansen 2005). This same statistic *cannot* deal with equivalent question formation in other languages (Kam et al. 2008) or with a variety of other linguistic phenomena which rely on internal sentence structure.

One obvious limitation of the studies discussed above is that the models were asked to learn syntactic patterns without access to semantic structure. Yet phrase structure is a means of representing conceptual groupings—for example an entire noun phrase serves to picks out a particular entity or set of entities. Still it is interesting that connectionist models which do attempt to link syntactic and semantic structure have tended to employ additional specialized mechanisms for dealing with the encapsulated interpretation of embedded structures (Miikkulainen 1996; Pollack 1988; though see Bryant and Miikkulainen 2001). These systems still differ from symbolic systems in showing plausible memory degradation for center embedding, as discussed above. However to date such work has only dealt with fairly basic linguistic phenomena. It remains to be seen how statistical approaches will scale up to deal with the full complexity of natural language syntax, and the types of learning architectures necessary to capture these behaviors.

## 4.4 Concluding Remarks

In this chapter, I have presented evidence that statistical learning processes play an important role in language acquisition. We have seen that statistical models are necessary to explain graded, probabilistic effects in morphological systems (even those that appear very rule-like), although it is currently unclear whether and how current models will scale up to capture all of the human data. We also saw evidence that a statistical system can learn abstract relationships between syntactic form and semantic structure (at least if given access to the requisite semantic representations). Further, an ability to track and evaluate probabilistic evidence may explain how learners avoid rampant overgeneralization and converge on highly similar grammatical intuitions. However, it is important to emphasize that we are still a long way from possessing a full account of statistical grammar learning. In particular, accounts of many ‘higher level’ syntactic behaviors are lacking, particularly those which require access to hierarchical structure.

Another phenomenon which statistical learning theory must address, and one which I have neglected in this chapter, is the fact that acquisition is generally more successful when it begins in early childhood. This has been shown to be the case even when controlling for years of exposure and external factors such as ‘motivation’ (Johnson and Newport 1989; Newport 1990). These studies reveal that although the ability to learn language is not entirely lost, the grammatical system acquired by late-learners is characterized by grammatical inconsistency and probabilistic use of incorrect forms. This suggests that there may be important differences in the statistical learning process that takes place at different ages. Newport (1990) suggests that these stem from constraints placed on the system by children's limited memory capacity, which restrict the input to the statistical learning system in the early stages of learning (Hudson Kam and Newport 2009; Elman 1993—though see Rohde and Plaut 2003). Another possibility is that there may be differences in the way child and adult learners weigh and combine different probabilistic sources. Exploring these possibilities may further illuminate the mechanisms of native language learning and why it is so consistently successful.

Despite holes in our current knowledge, it seems clear that statistical learning mechanisms play a critical role in human language. Since I began this chapter by emphasizing our biological "predisposition" for language, it is worth considering again how this approach fits into the long-standing nature-nature controversy. Traditionally, statistical learning has been associated with an empiricist approach to language which deemphasizes the contribution of the learner. In contrast, I have emphasized that a full statistical learning account must specify (a) what statistical computations the system can calculate, and how information is integrated (b) the nature of the representations (formal and semantic) over which these analyses occur. In fact, statistical learning theories, and in particular working computational models, actually force us to make quite precise claims about the type of information that is primitive to the learning system. Of course it is always possible that representations which are primitive in one linguistic domain may ultimately be derived from lower-level primitives—but this leads to testable hypotheses about *that* learning process and how the derived representations feed into higher level processes.

Another contentious issue within the acquisition literature is the extent to which language learning rests on language-specific versus domain-general processes (see also Müller, this volume). I think that this division arises primarily from reactions for and against the "strong" UG view, which certainly makes claims about linguistic specificity. However, if our goal is to understand the cognitive processes and structures which allow human language, the focus on whether these are shared by other cognitive systems appears less important. For example, we have seen that children's self-created gestural systems communicate "thematic role" information. Many researchers have pointed out that such conceptual information might not be specific to the linguistic system (e.g. McClelland and Bybee 2007). In fact, Goldin-Meadow (2005) does not dispute this point. Nevertheless, as she points out, the fact that the children communicate this particular set of conceptual structures, and that these also show up across human languages, is surely important in understanding what children bring to language learning. Similarly, the types of hierarchical relations seen in human language may also be evident in other cognitive systems such as motor planning (Rosenbaum et al. 1983). However, recognizing that human language (and no other communication system) conveys messages which are hierarchically structured is surely critical to understanding the nature of our endowment for language.

Other more general cognitive developments are undoubtedly vital for language learning. Another topic which I have neglected in this chapter is the contribution of more general social cognition. Tomasello in particular has argued that human language learning rests on more general social adaptations, and reports various ways in which human social interactions differ from those of primates (Tomasello et al.

2005). In particular, he has emphasized the human ability to comprehend *intention*, which is critical in inferring the message conveyed by a linguistic utterance.

In short, it is likely that our 'specialization' for language relies on a variety of different cognitive abilities. Each of these may also play a role in other cognitive behaviors, and be shared to some extent by other species. Our goal is to understand how these come together to give us the—uniquely human—Language Phenomenon.

**Acknowledgments** Many thanks to the following people for helpful discussions and/or comments on earlier drafts of the chapter: Adele Goldberg, Franklin Chang, Joanne Taylor, Jennifer Thomson and Edward Longhurst.

## References

Akhtar, N., Tomasello, M.: Young children's productivity with word order and verb morphology. Dev. Psychol. **33**, 952–965 (1997)

Albright, A., Hayes, B.: Rules vs. analogy in English past tenses: a computational/experimental study. Cognition **90**, 119–161 (2003)

Allen, J., Seidenberg, M.S.: The emergence of grammaticality in connectionist networks. In: MacWhinney, B. (ed.) Emergentist Approaches to Language: Proceedings of the 28th Carnegie Symposium on Cognition, pp. 115–151. Lawrence Earlbaum Associates, Hillsdale (1999)

Ambridge, B., Pine, J.M., Rowland, C.F., Young, C.R.: The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument-structure overgeneralization errors. Cognition **106**, 87–129 (2008)

Aslin, R.N., Jusczyk, P., Pisoni, D.B.: Speech and auditory processing during infancy: constraints on and precursors to language. In: Kuhn, D., Siegler, R. (eds.) Handbook of Child Psychology: Cognition, Perception, and Language, pp. 147–254. Wiley, New York (1998)

Baker, C.L.: Syntactic theory and the projection problem. Linguist. Inquiry **10**, 533–581 (1979)

Berko, J.: The child's learning of english morphology. Word **14**, 150–177 (1958)

Bowerman, M.: The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In: Hawkins, J.A. (ed.) Explaining Language Universals, pp. 73–101. Basil Blackwell, New York (1988)

Braine, M.D.S.: On Two Types of Models of the Internalization of Grammars. In: Slobin, D.I. (ed.) The Ontogenesis of Grammar: A Theoretical Symposium, pp. 153–186. Academic Press, New York (1971)

Braine, M.D.S., Brooks, P.J.: Verb argument strucure and the problem of avoiding an overgeneral grammar. In: Tomasello, M., Merriman, W.E. (eds.) Beyond Names for Things: Young Children's Acquisition of Verbs, pp. 352–376. Lawrence Earlbaum Associates, Hillsdale (1995)

Brooks, P.J., Tomasello, M.: Young children learn to produce passives with nonce verbs. Dev. Psychol. **35**, 29–44 (1999)

Brown, R., Hanlon, C.: Derivational complexity and order of acquisition in child speech. In: Hayes, J.R. (ed.) Cognition and the Development of Language, pp. 11–53. Wiley, New York (1970)

Bryant, B. D. & Miikkulainen, R. (2001). From word stream to gestalt: A direct semantic parse for complex sentences. Technical Report TR-AI98-274, Department of Computer Sciences, The University of Texas at Austin.

Bybee, J.L.: Morphology: A Study of the Relation between Meaning and Form. John Benjamins, Amsterdam (1985)

Bybee, J.L.: Diachronic and typological properties of morphology and their implications for representation. In: Feldman, L. (ed.) Morphological Aspects of Language Processing, pp. 225–246. Lawrence Earlbaum Associates, Hillsdale (1995)

Bybee, J.L., Moder, C.L.: Morphological classes as natural categories. Language **59**, 251–270 (1983)

Bybee, J.L., Slobin, D.I.: Rules and schemas in the development and use of the English past tense. Language **58**, 265–289 (1982)

Chang, F.: Symbolically speaking: a connectionist model of sentence production. Cogn. Sci. **26**, 609–651 (2002)

Chang, F.: Learning to order words: a connectionist model of heavy NP shift and accessibility effects in Japanese and English. J. Mem. Lang. **61**, 374–397 (2009)

Chang, F., Dell, G.S., Bock, K.: Becoming syntactic. Psychol. Rev. **113**, 234–272 (2006)

Chater, N., Tenenbaum, J.B., Yuille, A.: Probabilistic models of cognition: conceptual foundations. Trends Cogn. Sci. **10**, 287–291 (2006)

Chomsky, N.: Syntactic Structures. Mouton, The Hague (1957)

Chomsky, N.: Aspects of the Theory of Syntax. MIT Press, Cambridge (1965)

Chomsky, N.: Language and Mind. Harcourt, Brace and World, New York (1968)

Chomsky, N.: The Minimalist Program. MIT Press, Cambridge (1995)

Chomsky, N., Halle, M.: The Sound Pattern of English. Harper and Row, New York (1968)

Christiansen, M.H., Chater, N.: Toward a connectionist model of recursion in human linguistic performance. Cogn. Sci. **23**, 157–205 (1999)

Christiansen, M.H., Chater, N.: Language as shaped by the brain. Behav. Brain Sci. **31**, 489–509 (2008)

Church, K.W.: A stochastic parts program and noun phrase parser for unrestricted text. In Proceedings of the 2nd Conference on Applied Natural Language Processing, (pp. 136–143). ACL (1988)

Crain, S., Pietroski, P.: Is generative grammar deceptively simple or simply deceptive? Lingua **116**, 64–68 (2006)

Croft, W.: Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford University Press, Oxford (2001)

Daugherty, K., Seidenberg, M.S.: Rules or Connections? The Past Tense Re-Visited. Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society, pp. 259–264. Lawrence Earlbaum Associates, Hillsdale (1992)

DeLancey, S.: Adpositions as a Non-Universal Category. In: Frajzyngier, Z., Hodges, L., Rood, D.S. (eds.) Linguistic Diversity and Language Theories, pp. 185–202. John Benjamins, Amsterdam (2005)

Dowman, M.: Addressing the learnability of verb subcategorizations with Bayesian inference. In: Gleitman, L.R., Joshi, A.K. (eds.) Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates, Mahwah (2000)

Elman, J.L.: Finding structure in time. Cogn. Sci. **14**, 179–211 (1990)

Elman, J.L.: Learning and development in neural networks: the importance of starting small. Cognition **48**, 71–99 (1993)

Elman, J.L., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K.: Rethinking Innateness: A Connectionist Perspective on Development. MIT Press, Cambridge (1996)

Evans, N., Levinson, S.C.: The myth of language universals: Language diversity and its importance for cognitive science. Behav. Brain Sci. **32**, 429–492 (2009)

Everett, D.: Cultural constraints on grammar and cognition in Pirahã. Curr. Anthropol. **46**, 621–646 (2005)

Finch, S., Chater, N.: Distributional bootstrapping: from word class to proto-sentence. In: Eiselt, A.R.K. (ed.) Proceedings of the 16th Annual Conference of the Cognitive Science Society, pp. 301–306. Lawrence Earlbaum Associates, Hillsdale (1994)

Fisher, C.: The role of abstract syntactic knowledge in language acquisition: a reply to Tomasello (2000). Cognition **82**, 259–278 (2002)

Garnsey, S.M., Pearlmutter, N.J., Meyers, E., Lotocky, M.A.: The contribution of verb bias to the comprehension of temporarily ambiguous sentences. J. Mem. Lang. **37**, 58–93 (1997)

Gertner, Y., Fisher, C., Eisengart, J.: Learning words and rules: Abstract knowledge of word order in early sentence comprehension. Psychol. Sci. **17**, 648–691 (2006)

Gleitman, L.R., Cassidy, K., Nappa, R., Papafragou, A., Trueswell, J.C.: Hard words. Lang. Learn. Dev. **1**(1), 23–64 (2005)

Goldberg, A.E.: A Construction Grammar Approach to Argument Structure. The University of Chicago Press, Chicago (1995)

Goldberg, A.E.: Constructions At Work: The Nature of Generalization in Language. Oxford University Press, Oxford (2005)

Goldin-Meadow, S.: The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language. Psychology Press, New York (2003)

Goldin-Meadow, S.: What language creation in the manual modality tells us about the foundations of language. Linguist. Rev. **22**, 199–225 (2005)

Gonnerman, L.M., Seidenberg, M.S., Andersen, E.S.: Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. J. Exp. Psychol. Gen. **136**, 323–345 (2007)

Greenberg, J.H. (ed.): Universals of Language. MIT Press, Cambridge (1963)

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., Wilson, R.: The learnability and acquisition of the dative alternation in English. Language **65**, 203–255 (1989)

Halle, M., Mohanan, K.P.: Segmental phonology of modern English. Linguist. Inquiry **16**, 57–116 (1985)

Hare, M., Elman, J.L., Daugherty, K.G.: Default generalization in connectionist networks. Lang. Cogn. Process. **10**, 601–630 (1995)

Hare, M.L., Ford, M., Marslen-Wilson, W.D.: Ambiguity and Frequency Effects in Regular Verb Inflection. In: Bybee, J., VHopper, P. (eds.) Frequency and the Emergence of Linguistic Structure, vol. 45, pp. 181–200. of Typographical Studies in LanguageJohn Benjamins, Amsterdam (2001)

Hart, B., Risely, J.: Meaningful Differences in the Everyday Experience of Young Children. Brookes Publishing Co, Baltimore (1995)

Hauser, M.D., Chomsky, N., Fitch, W.T.: The faculty of language: what is it, who has it and how did it evolve? Science **298**, 1569–1579 (2002)

Hay, J.B., Baayen, R.H.: Shifting paradigms: gradient structure in morphology. Trends Cogn. Sci. **9**, 342–348 (2005)

Hudson Kam, C.L., Newport, E.L.: Getting it right by getting it wrong: When learners change languages. Cogn. Psychol. **59**, 30–66 (2009)

Jackendoff, R.: Foundations of Language: Brain, Meaning, Grammar, Evolution. Oxford University Press, Oxford (2002)

Jacobs, R.A.: What determines visual cue reliability? Trends Cogn. Sci. **6**, 345–350 (2002)

Joanisse, M.F., Seidenberg, M.S.: Impairments in verb morphology following brain injury: a connectionist model. Proc. Nat. Acad. Sci. U.S.A **96**, 7592–7597 (1999)

Joanisse, M.F., Seidenberg, M.S.: Imaging the past: neural activation in frontal and temporal regions during regular and irregular past tense processing. Cogn. Affect. Behav. Neurosci. **5**, 282–296 (2005)

Johnson, J.S., Newport, E.L.: Critical period effects in second language learning: The influence of maturational state on the acquisition of english as a second language. Cogn. Psychol. **21**, 60–99 (1989)

Kam, X.-N., Stoyneshka, I., Tornyova, L., Fodor, J., Sakas, W.: Bigrams and the richness of the stimulus. Cogn. Sci. **32**, 771–787 (2008)

Lakoff, G.: Women, Fire, and Dangerous Things: What Categories Reveal About the Mind. University of Chicago Press, Chicago (1987)

Langlacker, R.W.: Foundations of Cognitive Grammar: Volume 1, Theoretical Prerequisites. Stanford University Press, Stanford (1987)

Lewis, J.D., Elman, J.L.: Learnability and the statistical structure of language: poverty of stimulus arguments revisited. In: Proceedings of the 26th Annual Boston University Conference on Language Development (2001)

Li, C., Thompson, S.A.: Coverbs in Mandarin Chinese: verbs or prepositions? J. Chin. Linguist. **2**, 257–278 (1974)

MacWhinney, B.: The CHILDES project: Tools for Analyzing Talk. 3rd edn, Lawrence Erlbaum Associates, Mahwah (2000)

Manning, C.D., Schutze, H.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (2001)

Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R., Pinker, S.: German inflection: the exception that proves the rule. Cogn. Psychol. **29**, 189–256 (1995)

Marslen-Wilson, W.D., Tyler, L.: Morphology, language and the brain: the decompositional substrate for language comprehension. Philos. Trans. R. Soc. B Biol. Sci. **362**, 823–836 (2007)

McClelland, J.L., Bybee, J.: Gradience of gradience: a reply to Jackendoff. Linguist. Rev. **24**, 437–455 (2007)

McClelland, J.L., Patterson, K.: Differentiation and integration in human language: a reply to Marslen-Wilson and Tyler. Trends Cogn. Sci. **7**, 63–64 (2003)

Miikkulainen, R.: Subsymbolic case-role analysis of sentences with embedded clauses. Cogn. Sci. **20**, 47–73 (1996)

Mintz, T., Newport, E.L., Bever, T.: The distributional structure of grammatical categories in speech to young children. Cogn. Sci. **26**, 393–424 (2002)

Mintz, T.H.: Frequent frames as a cue for grammatical categories in child directed speech. Cognition **90**, 91–117 (2003)

Nevins, I., Pesetsky, D., Rodrigues, C.: Pirahã exceptionality: a reassessment. Language **85**, 355–404 (2009)

Newport, E.L.: Maturational constraints on language learning. Cogn. Sci. **14**, 11–28 (1990)

Newport, E.L., Aslin, R.: Innately constrained learning: blending old and new approaches to language acquisition. In: Howell, S.C., Fish, S.A., Keith-Lucas, T. (eds.) Proceedings of the 24th Annual Boston University Conference on Language Development. Cascadilla Press, Somerville (2000)

Newport, E.L., Gleitman, H., Gleitman, L.: Mother, I'd rather do it myself: some effects and non-effects of maternal speech style. In: Snow, C., Ferguson, C.A. (eds.) Talking to Children: Language Input and Acquisition. Cambridge University Press, Cambridge (1977)

Onnis, L., Roberts, M., Chater, N.: Simplicity: A Cure for Overregularizations in Language Acquisition? Proceedings of the 24th Conference of the Cognitive Science Society, pp. 720–725. Lawrence Earlbaum Associates, Mahwah (2002)

Perfors, A., Tenenbaum, J.B., Wonnacott, E.: Variability, negative evidence, and the acquisition of verb argument constructions. J. Child Lang. **37**, 607–642 (2010)

Pinker, S.: Language Learnability and Language Development. Harvard University Press, Cambridge (1984)

Pinker, S.: The bootstrapping problem in language acquisition. In: MacWhinney, B. (ed.) Mechanisms of Language Acquisition. Lawrence Earlbaum Associates, Hillsdale (1987)

Pinker, S.: Learnability and Cognition: The Acquisition of Argument Structure. Learning, development and conceptual change. MIT Press, Cambridge (1989)

Pinker, S.: Words and Rules. Basic Books, New York (1999)

Pinker, S., Ullman, M.: The past and future of the past tense. Trends Cogn. Sci. **6**, 456–463 (2002)

Plaut, D.C., Gonnerman, L.M.: Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? Lang. Cogn. Process. **15**, 445–485 (2000)

Plunkett, K., Marchman, V.: U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. Cognition **38**, 43–102 (1991)

Plunkett, K., Nakisa, R.C.: A connectionist model of the Arabic plural system. Lang. Cogn. Process. **12**, 807–836 (1997)

Pollack, J.B.: Recursive auto-associative memory: Devising compositional distributed representations. Proceedings of the 10th Annual Conference of the Cognitive Science Society, pp. 33–39. Erlbaum, Hillsdale (1988)

Post, B., Marslen-Wilson, W.D., Randall, B., Tyler, L.K.: The processing of English regular inflections: phonological cues to morphological structure. Cognition **109**, 1–17 (2008)

Prasada, S., Pinker, S.: Generalizations of regular and irregular morphology. Lang. Cogn. Process. **8**, 1–56 (1993)

Prasada, S., Pinker, S., Snyder, W.: Some evidence that irregular forms are retrieved from memory but regular forms are rule generated. Paper presented at the Psychonomic Society meeting (1990)

Radford, A.: Transformational Grammar: A First Course. Cambridge University Press, Cambridge (1988)

Reali, F., Christiansen, M.H.: Uncovering the richness of the stimulus: structure dependence and indirect statistical evidence. Cogn. Sci. **29**, 1007–1028 (2005)

Rohde, D.L.T., Plaut, D.C.: Less is less in language acquisition. In: Quinlan, P. (ed.) Connectionist Modeling of Cognitive Development, pp. 189–231. Psychology Press, Hove (2003)

Rosenbaum, D.A., Kenny, S., Derr, M.A.: Hierarchical control of rapid movement sequences. J. Exp. Psychol. Hum. Percept. Perform. **9**, 86–102 (1983)

Rumelhart, D., McClelland, J.: On Learning the past tenses of English verbs. In: McClelland, J., Rumelhart, D.: The PDP Research Group (eds.) Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 2, MIT Press, Cambridge (1986)

Saffran, J.R., Aslin, R.N., Newport, E.L.: Statistical Learning by 8-month-old Infants. Science **274**, 1926–1928 (1996)

Schreuder, R., de Jong, N., Krott, A., Baayen, H.: Rules and rote: beyond the linguistic either-or fallacy. Behav. Brain Sci. **22**, 1038–1039 (1999)

Seidenberg, M.S.: Language acquisition and use: learning and applying probabilistic constraints. Science **275**, 1599–1604 (1997)

Seidenberg, M.S., Pettito, L.A.: Signing behavior in apes: a critical review. Cognition **7**, 177–215 (1979)

Singleton, J.L., Newport, E.L.: When learners surpass their models: the acquisition of American sign language from inconsistent input. Cogn. Psychol. **49**, 370–407 (2004)

Snedeker, J., Trueswell, J.C.: The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. Cogn. Psychol. **49**, 238–299 (2004)

St. John, M.F., McClelland, J.L.,: Learning and applying contextual constraints in sentence comprehension. Artif. Intell. **46**, 217–257 (1990)

Steedman, M.: Connectionist and symbolic representations of language. In: Encyclopedia of Cognitive Science, Macmillan (2002)

Terrace, H.S., Petitto, L.A., Sanders, R.J., Bever, T.: Can an ape create a sentence? Science **206**, 891–902 (1979)

Theakston, A.L.: The role of entrenchment in constraining children's verb argument structure overgeneralisations: a grammaticality judgment study. Cogn. Dev. **19**, 15–34 (2004)

Tomasello, M.: Do young children have adult syntactic competence? Cognition **74**, 209–253 (2000)

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H.: Understanding and sharing intentions: the origins of cultural cognition. Behav. Brain Sci. **28**, 675–735 (2005)

Trueswell, J.C., Kim, A.E.: How to prune a garden-path by nipping it in the bud: Fast-priming of verb argument structures. J. Mem. Lang. **39**, 102–123 (1998)

Trueswell, J.C., Tanenhaus, M.K., Kello, C.: Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. J. Exp. Psychol. Learn. Mem. Cogn. **19**, 528–553 (1993)

Wierzbicka, A.: The Semantics of Grammar. John Benjamins, Amsterdam (1988)

Wonnacott, E., Newport, E.L., Tanenhaus, M.K.: Acquiring and processing verb argument structure: distributional learning in a miniature language. Cogn. Psychol. **56**, 165–209 (2008)

Xu, F., Pinker, S.: Weird past tense forms. J. Child Lang. **22**, 531–556 (1995)

# Chapter 5
# Evolution: Language Use and the Evolution of Languages

**William Croft**

**Abstract** Language change can be understood as an evolutionary process. Language change occurs at two different timescales, corresponding to the two steps of the evolutionary process. The first timescale is very short, namely, the production of an utterance: this is where linguistic structures are replicated and language variation is generated. The second timescale is (or can be) very long, namely, the propagation of linguistic variants in the speech community: this is where certain variants are selected over others. At both timescales, the evolutionary process is driven by social interaction and the role language plays in it. An understanding of social interaction at the micro-level—face-to-face interactions—and at the macro-level—the structure of speech communities—gives us the basis for understanding the generation and propagation of language structures, and understanding the nature of language itself.

## 5.1 Time Scales and Evolutionary Processes in Language Change

Language evolution, in the broadest sense of the term, appears to occur at different time scales. It is more precise to say that language evolution involves different types of processes occurring at different time scales, which produce an overall effect of a dynamic evolving system. What matters most is not the time scale but the type of evolutionary process that is taking place.

The process to be described in this chapter is language change: the process by which sounds change in their pronunciation and their status in the linguistic system, words change in both form and meaning (and are created or lost), and grammatical constructions change in their structure and function. This is a process that can, and usually does, take place at a time scale longer than a single speaker's lifetime: historical linguists observe an arc of continuous change in a particular linguistic element or structure than spans centuries or even longer.

---

W. Croft (✉)
Department of Linguistics, University of New Mexico, Albuquerque, NM 87131, USA
e-mail: wcroft@unm.edu

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_5,

But the topic of this chapter is the relationship between language change—the evolution of languages over centuries or millennia—and language use. Language use takes place when two speakers interact in conversation (see also Garrod and Pickering, this volume). Conversations do not take centuries or millennia (even if they sometimes feel like they do). In fact, conversations take place in a matter of minutes or even seconds—a time scale that is much smaller than that of the human lifetime. So the topic of this chapter is also a process that takes place at a different time scale from that of language change as it is normally understood; a time scale that is orders of magnitude shorter than a speaker's lifetime.

### 5.1.1 Evolutionary Processes and Evolutionary Theories

How can this paradox be resolved? It can be resolved by recognizing that language change is an evolutionary process. Evolutionary processes can be described independently of any particular domain in which such processes occur. Evolutionary processes are processes in which change happens by replication of new entities, such as the replication of linguistic structures every time we produce an utterance. Evolutionary processes have another property: they take place in two steps. The first step is the generation of variation. Variation is generated in the process of *replication* of replicators. The second process is *selection*, in which the variants are differentially replicated so that some variants spread and others go extinct. These two processes and their interplay form the basis of Darwinian evolution. Both processes are evident in cultural change, and language change in particular. In language, novel sounds, words and constructions, and novel uses of sounds, words and constructions, are generated all the time. The variability of language in use is a hallmark of human communication. But over time some variants are propagated in a speech community, while others fade from use and go extinct. In other words, some variants are selected over others.

These basic observations have led biologists since Darwin to observe parallels between biological evolution and cultural change, including language change. It has also led some biologists and philosophers of biology to propose generalized theories of evolution to subsume biological evolution, cultural change and language change. It has also led linguists, archeologists, anthropologists and other scholars of cultural change to adopt concepts of biological evolution. However, this interdisciplinary interaction has been limited by two factors. Scholars of cultural change, including historical linguists, have generally treated cultural change and language change as analogous to biological evolution, rather than as instances of a generalized theory of evolutionary change. Analogies can provide intellectual inspiration, but they are opportunistic and unreliable. One cannot determine which concepts should carry over from biological evolution to cultural change and language change. If there is a disanalogy between biological evolution and language change, one cannot determine whether it is a difference between the two domains that is irrelevant, or a profound problem in applying evolutionary thinking to language change or cultural change.

On the other hand, biologists do not entirely agree on what the properties of a general theory of evolution are that apply to biological evolution, let alone to cultural change and language change. The fundamental problem is the "unit of evolution": what is the unit in biology (and in culture and language) that evolves? The next three sections discuss this debate. Then Sect. 1.5 turns to a less contentious part of Darwinian evolutionary theory, population thinking, which is equally important in evolutionary theories of language change.

### *5.1.2 Replication and Evolution in Biology, Culture and Language*

Evolutionary processes are most familiar to us, and probably best understood, in biological evolution. In the neo-Darwinian synthesis of Darwinian evolution and population genetics (Mayr 1982), the replication process is biological reproduction. Reproduction generates variation, either through mutation or (in sexually reproducing organisms) recombination of genetic material. Selection occurs across generations, in which a biological character is propagated or goes extinct. Both of these processes are probabilistic, a point I will return to below.

The generation of variation and the selection of variants occur at different timescales. The generation of variation occurs in the act of reproduction, which is short even if gestation and organismic development are included (see Sect. 1.4; human beings are somewhat unusual in the animal world for the lengthy times of both gestation and development). The selection process takes place over many generations, and therefore may be very long. The lifetime of the organism defines the time scale of both processes: reproduction including development must be no longer than a lifetime, while selection is defined in generations which are in turn defined by an organism's lifetime, or at least time to reproductive maturity. Nevertheless, these are two different time scales, because generation of variation takes place in a single replication, while selection takes place over very many generations of replications. Even so, lifetimes and therefore rates of evolution vary greatly in the biological world. While human beings take around 15 years to become sexually mature, and may take twenty or more years longer to actually reproduce, viruses and bacteria and even insects reproduce at extremely fast rates from a human being's point of view.

Applying these evolutionary concepts to cultural change and language change requires us to identify what it is that reproduces in each of these domains. In the neo-Darwinian theory of evolution, the entity that ultimately reproduces is the genome; the genome contains the genes that partly determine the structure and behavior of the organism (its phenotype). The first and most influential application of evolutionary theory to cultural change starts with a generalization of the gene as a replicator (Dawkins 1976). Dawkins identifies several abstract properties of replicators (see Box 1). These properties define two essential properties of replicators. The first is that replicators form lineages through successive replications, lineages that can persist indefinitely and spread more or less widely. The second is that the structure of the replicates (the "copies", though we will see that this is a too restrictive view of replication) is at least partly maintained through the lineage. That is, structural traits of the replicator are inherited by its daughters. However, replication is not always perfect, and variation arises. The generation of variation is of course a fundamental requirement for evolutionary change to take place.

**Box 1: Replicators (Dawkins 1976)**

*Replicator*: an abstraction over the neo-Darwinian concept of gene, with the following sets of properties.

*Lineage formation:*

(i) A replicated unit can itself be replicated, forming lineages of replicators
(ii) Replicators have longevity, that is, they can survive long enough to replicate
(iii) Replicators have differing degrees of fecundity, that is, they replicate at different rates or with differing degrees of success

*Inheritance:*

(iv) Replicates of a replicator possess much of the structure of the original replicator
(v) Replication is mostly faithful
(vi) But replication can produced heritable variation

*Meme:* a cultural replicator

Dawkins also proposes that cultural replicators exist; he calls them memes, and he spawned an approach to cultural change called memetics. Memeticists (including Dawkins) have proposed three different types of cultural replicators: concepts, behaviors and artifacts. Some cultural replicators appear to be a combination of these three. For example, replication in language change occurs in language use: every time we open our mouths, we replicate sounds, words and constructions. Language use is an instance of behavior. But linguists transcribe such utterances for analysis, and language can also be written: these are artifacts. And language behavior also appears to involve concepts, namely a speaker's knowledge about her language.

Theories of cultural replication suffer from a number of problems. Concepts are the most popular replicators among memeticists (Dawkins 1982; Blackmore 1999; Aunger 2002). But is unclear how concepts are replicators. Their structure is invisible: the best candidates are neural structures, but these differ between individuals. A problem with all three types of replicators is that they are replicated very indirectly. Concepts are replicated from one person to another only via behaviors and artifacts, which are publicly observable. Artifacts and behaviors appear to be better candidates for replicators, since their structure is clear; but they can only be replicated via the employment of the knowledge of the users, namely concepts.

Another problem with memetics is that Dawkins embeds his theory of replicators into a genocentric theory of evolution, in which genes control all aspects of the process of biological evolution and organisms and the environment play a highly subordinated role. The translation of Dawkins' genocentrism into memetics as a model of cultural evolution in which "selfish" concepts drive cultural change has been widely criticized as implausible, not recognizing the important role that human beings play in cultural evolution.

### 5.1.3 The General Analysis of Selection and the Theory of Utterance Selection

Hull (1988) presents a generalized theory of evolutionary processes that builds on Dawkins' abstract concept of replicator but rejects his genocentrism and memocentrism. Hull proposes a second abstract concept or role in the evolutionary process, the interactor, and develops a General Analysis of Selection (GAS) based on causal interactions between replicator, interactor, and environment (see Box 2). Selection (differential replication of variant replicators) results from environmental interaction. Hull argues that the entity filling the role of interactor is not necessarily the same entity that fills the role of replicator. In biology, the organism functions as an interactor because its interaction with its environment causes differential replication. More generally, Hull argues that different entities in the biological hierarchy function as interactors, from the gene itself to the cell, the organism, and possibly also species.

**Box 2: The General Analysis of Selection (Hull 1988)**

*Evolution*: a process that takes place through replication (e.g. biological evolution)
*Replication*: a process in which a replicator (see Box) preserves most of its structure
*Interactor*: an entity, not necessarily the same as the replicator, whose interaction with its environment causes differential replication of replicators
*Selection*: the process by which an interactor's interaction with its environment causes the differential replication of the relevant replicators

GAS is useful for cultural evolution because the distinct evolutionary roles of interactor and replicator fit well the distinct roles played by human beings and their cultural products respectively. Croft (2000) applies GAS to language change, with speakers as interactors and linguistic behaviors as replicators. He calls his application the Theory of Utterance Selection (see Box 3). When a historical linguist talks about language change, s/he describes change in sounds, words and constructions. But those linguistic entities do not exist in the abstract (although many linguists appear to talk as if they do). The concrete linguistic entities are the tokens of those entities that occur in language use, in discourse. The word *salmon* exists as a token produced in a linguistic utterance. Each occurrence of *salmon* is a replication of a prior occurrence of *salmon* in another utterance. The speaker, of course, brings about the replication of the word, using her knowledge of her prior conversations in which that word was used, by others or by herself. This replication process takes place on a very small time scale, that of language production. This time scale is far smaller than a speaker's lifetime: thousands or millions of replications of a word may occur in a single speaker's lifetime. Hence language change can, and does, happen within a single speaker's lifetime, though it may also take place over hundreds of years. (Another advantage of GAS is that it delinks interaction and replication: these two processes do not have to proceed in lockstep, and in cultural evolution, they do not.) Variation is generated in that process, in ways that will be described in later sections of this chapter. Selection of variants, which is described after the fact as a change in the form and/or meaning of the word, is a process that can take many generations of the replication of the word, which may—or may not—extend beyond the lifetime of an individual speaker.

**Box 3: The Theory of Utterance Selection** (**Croft 2000**)

*Replicator*: a *lingueme*—a token of linguistic structure in an utterance
*Interactor*: a *speaker*, interacting with other speakers and with the experiences she wishes to communicate
*Variation*: generated in the process of language use, for communicative and processing reasons
*Selection*: differential replication of linguemes in language use, for social reasons
*Lineages*: sound lineages (sound changes), word lineages (etymologies) and construction lineages (grammaticalization chains)

In other words, there is a linguistic evolutionary process that is different from that of a speaker's biological lifetime, and in fact takes place at both a much shorter timescale (a conversation) and a much longer timescale (hundreds or thousands of years) than a human lifetime. Actually, human biological evolution takes place at a time scale far greater than the entire history of human language. What is important is that language change is a different evolutionary process: the evolution of utterances.

Thus speakers are simultaneously the interactors in the evolutionary process of language change, and as biological organisms they are also the interactors in the evolutionary process of biological evolution. Hence it is clear that coevolution has taken place, between the biological and linguistic or cultural evolutionary processes. The evolution of the language capacity has been a coevolutionary process from the beginning, from the point at which linguistic behavior can be accurately described as a replication process. Linguistic behavior involves replication when it results from imitative learning, hence the (co)evolution of language presupposes, among other things, the evolution of the imitative learning capacity among our human ancestors.

Nevertheless, GAS leaves important issues in cultural evolution and biological evolution unresolved. Hull allows for artifacts, behaviors and concepts to be replicators; but the indirectness of replication, and the ineffability of the structure of concepts, remain issues in GAS just as they are in memetics. In biological evolution, Hull retains a neo-Darwinian position: organisms that reproduce sexually cannot be replicators, because they are replicated indirectly (via development from the gamete). Instead, genes code for the phenotype of the organism. This view of biological evolution has been challenged recently; this challenge and its consequences for language change will be described in the next section.

### *5.1.4 Developmental Systems Theory and Its Relevance to Language Change*

Neo-Darwinism ignores development, and the role of the environment in development. Instead it assigns a unique role to genes as coding the structure of organisms, although it acknowledges that environmental factors influence development. Developmental Systems Theory (DST) argues that the contributions of the genome and the environment to development cannot be separated. Current models of information in biology do not distinguish the role of the genome and the environment in development.

DST incorporates the developmental process into biological theory by proposing that the replicator is a process, namely the life cycle of the developing organism. This process includes interactions with all resources, both the genetic (and other) contribution from the parent organism and all environmental contributions, including coevolving organisms (see Box 4). Selection is the result of the life cycle's interaction with its resources.

**Box 4: Developmental Systems Theory (Griffiths and Gray 1994, 2001)**

*Parity thesis:* the informational contributions of the genome and the environment (in traditional terms) to the structure and behavior of the phenotype of the organism during development are equal; in particular, the informational contribution of the genome is not privileged over that of the environment
*Life cycle:* a process of development that includes the interactions of the organism with all of its resources through its lifetime; the life cycle is the replicator in Developmental Systems Theory
*Resources:* all of the entities with which an organism interacts in its lifecycle, including both its genome and all of the entities in its environment at all levels
*Selection:* a process by which the life cycle-resource interaction causes differential replication of life-cycles

DST was developed to deal with biological issues and, unlike Dawkins' and Hull's theories, is not a generalized theory of evolution. Suggestions to generalize DST will be found in Croft (In preparation). Cultural replicators can best be thought of as life cycles of cultural acts, from human intention to their manifestation as behavior (and artifacts) to its social effect. This solves the conundrum of what counts as a cultural replicator. It requires one to accept that replication can be quite indirect, but there appears to be no alternative in cultural evolution. The cultural act's interaction with its resources is a complex adaptive system, which is the cultural parallel to a biological developmental system. Cultural acts coevolve just as life cycles do. Even so, one entity (the *reproducer*, roughly in Griesemer (2000) sense) directly (re)produces the cultural act, and provides the distinct role that human beings play in cultural evolution.

### *5.1.5 Population Thinking in Biology and Language*

A less contentious part of evolutionary theory in biology is the central role of populations in evolution (see Box 5). Species have been defined in terms of essential properties. But individual members of species are highly variable, making it difficult if not impossible to devise an essentialist definition that distinguishes one species from another. Finally, a species evolves, and therefore can lose any "essential" property that might be proposed for it.

Population thinking defines species as historical entities, namely a population of individuals that exist in a bounded region of space and time. Populations are defined by the interactions among individuals, rather than by inherent properties. More precisely, populations are defined by the absence of interactions between individuals who are members of distinct populations (although this is relative: populations have porous boundaries).

**Box 5: Population thinking (Mayr 1982)**

*Population*: a spatiotemporally bounded set of entities, that is, bounded in space and bounded in time by its beginning and its end (by extinction or splitting); a historical entity
*Isolation*: a population is defined by an interaction among its members (e.g. sexual reproduction, communicative intercourse) which is mostly or entirely lacking between a member of the population and an entity outside the population
*Isolating mechanism*: a process that gives rise to distinct populations, and maintains their distinctness

Population thinking requires a substantial recasting of theoretical concepts in language and language change (see Box 6). Most sociolinguists would accept a definition of a speech community as a population of communicatively (relatively) isolated speakers. A language, in the formal grammar sense, is commonly defined in essentialist terms: a set of possible sentences generated by a fixed set of grammatical rules. In an evolutionary framework for language change, a language is defined as a population of actual utterances produced by the members of the speech community. Those utterances contain tokens of linguistic structures, called *linguemes* in Croft (2000). The population of linguemes can be called the *lingueme pool*; the lingueme pool is what is sampled and analyzed by sociolinguists. Sociolinguistics has generally used population thinking, but population thinking is relatively new to the analysis of language structure, being found mainly in usage-based and exemplar-based approaches to phonology and grammar (see next section).

**Box 6: Populations in language change (Croft 2000)**

*Speech community*: a population of speakers who communicate with each other and are relatively communicatively isolated from other speakers
*Language*: a population of utterances produced by a speech community
*Lingueme pool*: a population of linguemes found in the language of a speech community

### 5.1.6 *Evolutionary Linguistics and Other Approaches to Linguistics*

There is a long-held position that language change is crucially linked to human lifetimes. This is the hypothesis that language change occurs through the change of generations from parents to children, hence the locus of language change is the process by which children acquire language. However, there is no empirical evidence that this is how language change takes place (see Croft 2000, Sect. 3.2; Bybee 2010, Sect. 6.6). "Errors" that a child produces in language acquisition are not of the same type as observed language changes. At any rate, children "unlearn" errors by around the age of seven; in fact, children are so good at it that it remains a puzzle exactly how children learn the ambient language so well. Finally, children, especially under the age of seven, are not the agents propagating language change; the agents are adolescents or older (Labov 2001), and the children are following their lead.

The alternative hypothesis, also of considerable antiquity, is that language change emerges from language use. This hypothesis is based on the position that language change involves a different evolutionary process than biological evolution, where change occurs via biological replication (reproduction). The evolutionary process of language change is logically independent of the change of biological generations of speakers, and in fact can take place within the lifetime of a single speaker. More and more evidence indicates that a speaker's grammatical knowledge is actually the result of language use (e.g. Bybee 2001, 2006). This is obviously the case in the fact that the forms and conventions of a particular language have to be learned via exposure to use of that language in discourse (cf. Garrod and Pickering, this volume). But it is also true that a speaker's grammatical knowledge continues to be shaped by language use, most clearly in the role of frequency of use in structuring a speaker's knowledge.

Structuralist and generative theories of language have focused on the analysis of grammatical structures and the relationships among different grammatical structures that define a linguistic system. Two major guiding assumptions were that the linguistic system is most fruitfully analyzed as a static or synchronic system, and that it is also most fruitfully analyzed as a self-contained system, autonomous from the cognitive and social matrix of language use.

Much progress has been made in our understanding of grammatical structure, but a number of researchers argued that the major guiding assumptions of structuralism and generative theory leave out essential properties of human language. Language behavior is actually highly variable. The variation is itself systematic, governed by social factors among other things. A model of the language system that incorporates language variation will allow for language change. This critique of the static linguistic system gave rise to the linguistic subfield of variationist sociolinguistics, also known as sociohistorical linguistics (the classic work here is Weinreich et al. 1968; see also Labov 1994, 2001).

Language behavior is not autonomous but a component of a more general phenomenon of social interactional behavior. One consequence of this observation is that what is communicated in a linguistic interaction is more than what is conventionally meant by the actual utterance. Conversely, properties of grammatical structures can best be understood in terms of how language is used and what it is used for in social interaction. This critique of the autonomy of the linguistic system emerged from different intellectual traditions. It gave rise to the linguistic subfield of pragmatics, with contributions from philosophy (e.g. Austin 1962; Searle 1969; Lewis 1969, cited below) and also from sociology and anthropology. It also gave rise to the linguistic approach called cognitive linguistics, which focuses on the role of general cognitive processes in shaping language structure (e.g. Lakoff 1987; Langacker 1987).

The evolutionary framework provides a way to embed language structure in its larger context. The evolutionary framework gives us a theory of structured entities that vary through replication, and a systematic relationship between the acts of language use (the replicators) and the knowledge and behavior of language speakers (the interactors). To flesh out an evolutionary framework for understanding language and language change, one must integrate richer models of language use that provide plausible mechanisms for the realization of linguistic structure, the generation of variation in linguistic structure, and the selection of structural variants in an evolving population of linguistic utterances.

## 5.2 Language Use: Social Interaction and Verbalization

Language plays an essential role in social interaction. However, the production of utterances, or even the communication of experiences—often described as the "function" of language—is only a small part of the process, and certainly not the chief goal of language use. A complete understanding of language use will emerge only by understanding the role of language in social interaction. Much of the research on this topic has been conducted outside the field of linguistics, in philosophy, psychology and sociology.

### *5.2.1 Social Interaction and Language Use*

Social interaction, including any accompanying language use, involves joint action between two or more individuals (Clark 1996; Tomasello 2008; see also Garrod and Pickering, this volume). In a joint action, each individual performs her or his individual action, but each individual's action is intended to combine in a cooperative way with the other individual's action in order to produce a joint action. The combination of individual actions that have this property are called a shared cooperative activity by Bratman (1992, see Box 7).

**Box 7: Shared Cooperative Activity (Bratman 1992)**

(a) Each individual participant intends to perform the joint action. That is, each participant's intention is not directed simply towards his/her individual action, but towards the joint action that is carried out by both participants together.
(b) Each participant intends to perform the joint action in accordance with and because of each one's meshing subplans. That is, each participant's individual actions are intended to mesh with the other participant's actions in order to successfully achieve the joint action.
(c) Neither participant is coercing the other.
(d) Each participant has a commitment to mutual support. That is, each one will help the other to carry out the subplans; each participant is thus responsible for more than just execution of their own subplan.
(e) All of (a)–(d) are *common ground*, or shared knowledge between the individuals.
(f) There must be mutual responsiveness in action on the part of the participants. That is, the participants will *coordinate* their individual actions as they are executed in order to ensure that they mesh with each other and hence the joint action will be successfully carried out (to the best of their abilities).

Where does language fit into social interaction? Language plays a crucial role in coordination of individual actions. Without coordination, joint actions would generally be unsuccessful. The achievement of a joint action thus poses coordination problems between the participants (Lewis 1969). The means used to solve the coordination problem on a particular occasion is a coordination device. Human beings have a number of coordination devices available to them. The simplest is *joint attention* to jointly salient properties of the environment (Tomasello 1999). For example, if I point to a bird and you follow my gesture, we have established joint attention to the bird. This cognitive ability also appears to be largely restricted to human beings. Another coordination device that Lewis describes is *precedent*: if two participants succeeded in coordinating their actions, by luck or joint attention or both, then doing their component individual actions in the same way the next time they want to carry out that joint action will likely be a device for successful coordination.

But by far the most powerful and effective coordination device is for the participants to communicate with each other. By communicating their mental states with a much greater degree of precision than via joint attention, the participants tremendously facilitate the execution of any joint action. The participants can communicate much more specifically about the joint action, their individual actions, and the ways by which they can coordinate their individual actions.

However, communication is itself a joint action. We thus have two steps in the process of social interaction: the joint action intended by the participants, and the communicative action used by the participants to coordinate their joint action. The joint action of communication is that the two participants, who we will call the signaler and recognizer (adopting terms from Clark 1996), must converge on a recognition of the signaler's intention by the recognizer (this is the Gricean definition of [intentional] meaning; Grice 1948/1989). The joint action of communication poses coordination problems of its own. The essential problem for the joint action of communication is that the participants cannot read each other's minds. Language is the primary coordination device used to solve the coordination problem of communication, which is in turn used to solve the coordination problem for joint actions in general. Indeed, that is the ultimate purpose of language: to solve the coordination problem for joint actions, ranging from the everyday to the profound (Clark 1999). This fact is essential for understanding the structure of discourse and the linguistic expressions used in it, as (Clark 1992, 1996) has shown for many aspects of conversational interaction, and it also accounts for many fundamental properties of linguistic structure.

There are now three steps in the process of social interaction: the joint action; communication to coordinate the joint action; and the production of some signal to coordinate communication. Linguistic utterances are the most significant type of signal for human beings. Linguistic utterances are a special type of coordination device, namely, a convention (see Box 8).

**Box 8: Convention (Lewis 1969, reformulated by Clark 1996)**

(i) A regularity in behavior,
(ii) that is partly arbitrary (that is, we could have equally chosen an alternative regularity of behavior),
(iii) that is common ground in the community,
(iv) as a coordination device,
(v) for a recurrent coordination problem.

Conventions are coordination devices for *recurrent* coordination problems: a convention can emerge when members of the community have shared knowledge that a certain repeated behavior can act among them as a coordination device for the recurrent coordination problem. This definition of convention is general: it applies to conventions such as shaking hands (or kissing on the cheek) for greeting, or driving on the right (left) side of the road. The definition also applies straightforwardly to language: a string of sounds (i.e. a word such as *eagle*) emerges as a convention when it becomes a regularly used means to solve the recurrent coordination problem of referring to a specific experience that is to be communicated. A convention therefore is a symbol, including a linguistic symbol.

### *5.2.2 Language Use and the Verbalization of Experience*

Linguistic utterances are frequently more complex than single words, of course. They are combinations of words and morphemes that constitute a coherent whole. The structure of utterances solves the problem that in fact, the sorts of experiences that human beings wish to communicate are in fact not recurrent; every experience taken as a whole is unique. The conceptual solution is to decompose the experience to be verbalized into parts that are recurrent across experience (such as eagles), and then recombine the recurring parts into the experience that the speaker currently intends to verbalize. This decomposition-recombination process is remarkably similar across languages, presumably reflecting cognitive commonalities across humans, though one must not overlook the high degree of diversity in the parsing of specific types of experiences. Following Chafe (1977), it appears that conceptualizers divide the unique whole experience into smaller chunks of the same conceptual type, which broadly correspond to what are verbalized as clauses. Then a conceptualizer breaks each chunk into pieces of different types: relatively persistent entities, corresponding to participants that are verbalized as argument phrases, and a relative transitory state of affairs linking the participants together, verbalized as the predicate (sometimes in combination with adverbial satellites of the predicate). The parts are then identifiable as recurrent categories of participants and situation types, and are then verbalized as content words (basically nouns, verbs, and adjectives).

The categorized parts of the experience are then particularized and the decomposed experience then recombined: both processes use grammatical forms and constructions to reconstitute the unique whole of the experience (Croft 2007). In verbalizing an experience such as *Her dog barked*, the chunk has been propositionalized and categorized into Bark(Dog), to use a logical notation. But the actual linguistic utterance (in a discourse context) re-particularizes the categories to the specific dog (identified by the possessive pronoun *her* modifying *dog*) and the specific barking event (identified by the past tense suffix on *bark*); the grammar of the clause links the participant (her dog) to its role in the barking situation. Finally, in a longer utterance consisting of multiple clauses, various grammatical devices for discourse cohesion link together the chunks of experience that verbalized as clauses into the whole experience that the conceptualizer intended to verbalize. Thus, both content words and grammatical forms and constructions can be interpreted in terms of their role in the process of verbalizing one's experience. It should be mentioned that this approach to grammatical structure and its relation to verbalization is a novel one in linguistics, but I believe it is the most plausible way to account for universal patterns of grammatical structure across languages.

### *5.2.3 The Action Ladder of Language Use*

However, this is still not the end of the process. The conceptualization of experience for verbalization described in the preceding two paragraphs is basically a cognitive process. For it to become joint, that is, shared between the interlocutors, the verbalization process must occur in a jointly perceivable medium. For spoken languages, this is sound; for sign languages and for written language, it is sight. The fourth and final step is the production of perceptible sounds (or gestures or images) that the listener can attend to. Joint attention on the actual physical utterance is achieved; the perceivable utterance, plus shared knowledge of linguistic conventions, allows the hearer to identify the speaker's conceptualization of the experience; that conceptualization allows the intended meaning of the speaker to be recognized by the hearer; and this communicative act assists the hearer in carrying out the joint action that the interlocutors are trying to achieve. This joint action may be simply to increase shared knowledge, as in many declarative utterances (e.g. *I bought bananas this afternoon*), or to achieve some other sort of social goal, as with other types of utterances (e.g. *Hand me the pliers*).

Language is therefore a joint action that operates simultaneously at four levels, which (Clark 1996) calls an *action ladder* (see Box 9).

**Box 9: The action ladder of speaking** (**Clark 1996**)

(4) *proposing* and *taking up* a joint project (joint action);
(3) *signaling* and it *recognizing* the communicative intention;
(2) *formulating* and *identifying* the proposition;
(1) *producing* and *attending* to the utterance.

The higher numbered levels are dependent on the lower numbered levels; the individual actions of the interlocutors are given in italics. The highest level, (4), corresponds to the illocutionary act in speech act theory (Austin 1962); the next level corresponds to Gricean meaning or the informative act (Clark 1992); the following level corresponds to the propositional act (Searle 1969), and the lowest level corresponds to the utterance act (Austin 1962; Searle 1969). Each level enables the level(s) above it, and succeeds only if the level(s) below has been successfully achieved (e.g. one cannot recognize the communicative intention if one did not pay attention to the utterance produced).

### *5.2.4 Language and Biological Evolution*

The model of language use described here allows us to situate linguistic utterances in the context of human social behavior. It also shows how language, while distinctive, is not utterly unique. Language is one kind of convention, and also one type of complex (multi-part) action. Convention is one type of signalling device. A signalling device is one means for communication. Communication is one type of coordination device for joint action. And language is one cog in the machine of social interaction, albeit a crucial one for achieving the sort of rich and complex social interactions that have given rise to human culture.

One consequence of the fact that language is only a part of a larger phenomenon, human social interaction, is that the biological evolution of human language can only be understood as part of the evolution of the entire suite of social cognitive abilities that have been described in this section. The social cognitive abilities necessary for shared cooperative activity appear to be unique to humans, and provide what (Tomasello 2008) calls the social cognitive infrastructure necessary for the evolution of modern human language. Other species than humans have a capacity for imitative learning of complex vocalizations. This has not (yet) been sufficient to lead to the evolution of human-like language among these species. Nonhuman primates have the ability to plan actions, and to recognize regularities in behavior of other creatures, enough to manipulate their behavior. These abilities are preconditions for executing complex actions such as joint actions, but they are not sufficient for doing so.

Research on primate behavior in natural and experimental settings suggest that some primates even have the ability to recognize conspecifics as beings with intentional states like themselves in some circumstances (Tomasello 2008; this ability develops in humans only at around nine months of age). This is a prerequisite for the emergence of common ground, the recognition by individuals that knowledge and beliefs can be shared, which in turn is a precondition for the achievement of shared cooperative activity, joint attention, and convention. Nevertheless, nonhuman primates have not been demonstrated to have the ability to engage in shared cooperative activity as defined above. (Tomasello 2008) suggests that in particular, helpfulness, Bratman's condition (d) for shared cooperative activity described above, may be critical to the evolution of the ability to carry out joint actions.

These social cognitive abilities constitute a complex and interconnected suite, all of which are necessary for modern human language and the role it plays in human social interaction. Open questions that are difficult to answer in the absence of empirical evidence are whether the evolution of these social cognitive abilities was gradual, and hence whether the evolution of modern human language was also gradual, involving stages that are one or more steps beyond the social interactional and communicative skills observed among nonhuman species today, but less than the full suite of social interactional and linguistic capacities of modern humans.

## 5.3 Sources of Variation in Language Use

The model of language use described in the preceding section might suggest that human beings have evolved the ability to successfully carry out joint actions using language. If so, then one might not expect languages to change, or for that matter for languages to maintain the vast differences that exist. Of course, that flies in the face of the facts: there is a high degree of diversity, even if it is being tragically reduced by the massive loss of languages around the world; and languages continue to change. It also flies in the face of facts of communication and social interaction: communication often fails, sometimes subtly, sometimes spectacularly; and joint actions also fail despite the best intentions of the participants (I leave aside uncooperative human social behavior, which also exists on a large scale). These facts all demonstrate that the story presented in the preceding section is not so neat, and the cracks in that story are also the sources of variation, the starting point of language change and the resulting linguistic diversity that we observe.

### *5.3.1 Mechanisms Generating Phonetic Variation*

The lowest level, the phonetic realization of language and its perception by the listener, gives us the best documentation of variation and why it might arise and persist. Instrumental phonetics has demonstrated that the phonetic realization of phonemes (the categories of sounds in a language) is highly variable. Not only do different speakers not produce the same phonetic values for a single phoneme, but the same speaker produces different phonetic values for a single phoneme on different occasions. The variation comes from two sources: the production process and the perception process. On the production side, the realization of a specific phonetic target requires highly fine-grained coordination and timing of a large number of articulatory gestures. Hitting the precise target every time is simply impossible. Also, sounds are not produced in isolation. Part of the variation in production is due to coarticulation effects in the transition of gestures from one phoneme to the next (in some cases, the effects spread across many phonemes in an utterance; see e.g. Ohala 1983). Finally, sounds are part of words, and the frequency of use of those words will lead to changes in the phonetic realization of the sounds (e.g. high frequency of a word will lead to the shortening or attenuation of the articulatory gestures making up the phonemes in that word; Bybee 2006).

The listener has the mirror-image problem. The listener must recognize the phonemes in the sound stream he perceives, despite the variation in the phonetic realization in those phonemes, both in the current utterance and in prior utterances with the same phonemes that the listener has heard. The listener must also correctly disentangle those parts of the phonetic signal that represent the phoneme in question and those parts that are the result of coarticulation effects from any number of neighboring phonemes. The listener must also factor out properties of individual voice quality, the physical condition of the speaker (e.g. if she has a cold), and ambient sounds. It could very well be that the listener does not recognize the same phoneme that the speaker intended to produce (or thought she had produced).

This description of the fact of phonetic variability and the reasons for it in production and perception might lead one to think that successful communication is so difficult, even at the phonetic level, that it is impossible. As a matter of fact, automatic speech recognition is a very difficult task. Yet most of the time, human beings are successful in linguistic interactions. This is partly because native speakers and listeners have had a lot of practice. Native speakers may not hit some ideal phonetic target, but they are close enough most of the time. Native listeners are accustomed to the variability in the input and are skilled in factoring out coarticulations, voice quality and ambient sounds in the phonetic signal. Very recently, there have been proposals to redefine phonological systems—the speaker's knowledge about their language—in terms of clusters of heard tokens or exemplars of the phoneme in phonetic space, rather than a fixed ideal point representing the ideal phonetic realization of that phoneme (Pierrehumbert 2003; Bybee 2001).

However, success in perceiving phonemes is also due to the role that language plays in social interaction, as described above. The role of language in social interaction is essentially a supporting role. All a speaker must do is produce sounds that allow the listener to identify the words and grammatical forms that are intended at the next level up in Clark's action ladder for language use—and so on up the action ladder to the joint action. So variation is tolerable as long as the phonetic signal succeeds in helping the listener ultimately achieve the joint action with the speaker. Moreover, there is a feedback effect, in that if I identify the word you intended to produce, then I can conclude which phonemes you were trying to utter, and add the new phonetic realization to my exemplar-based category for that phoneme. Thus for instance communication can succeed even with a speaker with a non-native accent (and non-native grammatical constructions, etc.), because language is essentially a tool to another end. As long as the end is successfully achieved, the linguistic behavior toward that end was effective enough, and that is all that matters.

One consequence of the tolerance of phonetic variation is that physiological biases in the production or auditory perception of sounds may tend to "push" sounds to change in a certain direction; this results in what have been identified as phonetically motivated tendencies or directions in sound change (see Box 10). However, sounds sometimes change in the opposite direction, which suggests that the story is more complex than what might be expected based on the physiology of sound production and perception. At the very least, the directionality observed in sound change is a statistical result of a probabilistic process.

**Box 10: Phonetically motivated sound change (Ohala 1983: 204)**

In 1806, Benjamin Franklin phonetically recorded his pronunciation of the English word *natural*: it was like 'natyural' (phonetically [nætju□el]). Today it is generally pronounced with a 'ch' sound (phonetically [næt□□el]). Some dialects of English pronounce *Tuesday* with a 'ch' sound as well. This sound change—'t' changing to 'ch' before a 'y' sound—is phonetically motivated. A semivowel like 'y' is pronounced with a narrow vocal opening. This results in a relatively higher velocity of airflow in the oral cavity, leading to higher frication, i.e. an 'sh' pronunciation that, combined with the preceding 't', results in a tendency towards a 'ch' pronunciation, realized in *natural* and dialectal *Tuesday*.

### *5.3.2 Mechanisms Generating Grammatical Variation*

The same phenomena appear to operate at the next level up, the level of the lexical and grammatical structure of utterances. As was indicated above, the structures at this level are linguistic conventions, which allow a hearer to identify the conceptualization of the experience being communicated by the speaker. However, this process is, if anything, even more complex than what goes on at the phonetic level.

Linguistic convention can function as a coordination device for communication because there are recurrent coordination problems in communication: people have repeatedly wished to convey similar intentions formulated in similar concepts. The conventions of human languages are particularly powerful because they can be combined and recombined in novel ways that allow a speaker to communicate an experience that, as a whole, is in fact not recurrent but unique and new. Convention, linguistic or otherwise, is a regularity of behavior that emerges in a community or society.

But convention must emerge from previous successful communication events where a convention did not previously exist (see also Garrod and Pickering, this volume). In other words, there must be a precedent: you and I use a coordination device because we used it before (or observed it used before), and it worked. Following a precedent is a coordination device, but it is not (yet) convention: it is based not on regular behavior that is mutually known in the community, but only on previous successful uses that the participants directly engaged in, or at least witnessed (Lewis 1969).

Following a precedent cannot be the ultimate root of convention either: it always requires a successfully coordinated communicative act as a precedent. The ultimate coordination device is joint salience: each participant can assume that in a particular situation, certain features are salient to both participants (Lewis 1969). As noted above, joint salience is possible because humans have the social cognitive capacity for joint attention to entities in their environment (Tomasello 1999).

Linguistic convention, however, is not perfect, even after it has become established beyond its origins in joint salience and precedent. This is partly because of the kind of conventions found in language, and partly because of the nature of convention itself.

Linguistic conventions are incomplete largely because of the phenomena of indexicality and ambiguity (Clark 1996). A linguistic convention such as *sweep* or *bedroom* represents a type, but on a particular occasion of use, we usually intend to convey a particular token of the category. Thus, *I swept the bedroom* communicates a particular sweeping event involving a specific bedroom. In order to identify which sweeping event and which bedroom, the interlocutors must rely on joint salience in the context, to coordinate successfully on the right sweeping event and the right bedroom. This effort is facilitated in part by the past tense of *sweep* and the article *the* combined with *bedroom*, as part of the verbalization process described above. But even these words only narrow the possibilities; they do not uniquely specify a particular bedroom or a particular sweeping event. Linguistic shifters such as the pronoun *I* more explicitly require joint salience, namely identifying who is the speaker in the context. Proper names appear to denote tokens, but even a proper name such as *Jack Straw* may be (and is) used for more than one individual.

Most words are also highly ambiguous: that is, the same regularity of behavior is used as a coordination device to solve different recurrent coordination problems. For example, *patient* is ambiguous between the linguistic semantic role (*The patient in sentence 25 is Roland*) and a role in the domain of medicine (*The patient in room 25 is Roland*). Linguistic convention alone cannot tell which meaning is intended by the speaker. Only joint salience, provided in the example sentences by the meanings of the other words and the broader context of conversation, will successfully solve the coordination problem of what is meant by *patient*.

Indexicality and ambiguity are so pervasive in language that no utterance can be successfully conveyed without recourse to nonconventional coordination devices. But convention itself is also incomplete. This is because every situation being communicated is unique, and can be construed as the recurrence of different coordination problems. For example, a comparison of speakers verbalizing the same scenes observed in a short film, the “Pear film” (Chafe 1980) demonstrates a wide variety of words and constructions used to describe the same scene (Croft 2010). Variation was found even in individual words selected to describe the parts of the scene that were otherwise decomposed in the same way by the different speakers. One might argue that this is due to alternative conceptualizations of the same scene by different speakers. While this may be true up to a point, there are reasons to believe that grammar and lexicon are not that precise. In addition to the fact that every situation is unique, every scene is subject to alternative conceptualizations. The hearer's conceptualization of the scene, derived from the words chosen by the speaker, will be based on the hearer's prior experiences in the use of those words, which is not the same as the speaker's prior experiences of those same words. And finally, the listener cannot read the speaker's mind, so the listener must fall back on his perspective on the current context, and on his past experiences with situations he construes as similar. Thus there is plenty of room for variation in formulation and interpretation as well as in vocalization.

Again, it may appear that successful communication is impossible. And it is indeed difficult; although the assumption of most people is that understanding is the default case, it is probably wiser to consider misunderstanding the default, and successful understanding as the accomplishment. Indeed, in some domains such as scholarly discourse on abstract theoretical concepts there are often alternative construals of what is intended by particular scholars: what do we take Saussure to have meant? But again, the formulation of a proposition is merely playing a supporting role to the recognition of the speaker's intention by the hearer with respect to the joint action to be considered (and success in the join action can be facilitated by processes such as alignment; Garrod and Pickering, this volume). A certain amount of variation in verbalization is thus tolerable. And again, cognitive "biases" in conceptualization may push constructional variants in certain directions, which are then described after the fact as unidirectional grammatical and lexical changes. One example of this is the variation in the insertion of an overt marker of negation in a context with implicit negation: *I miss having him around* and *I miss not having him around* are intended to mean the same thing in most circumstances: he is not around, and the speaker regrets that. These two assertions, contradictory on the face of it, are tolerated because we understand what is being communicated.

### *5.3.3 Intentional and Nonintentional Mechanisms of Variation*

Thus there are fundamental characteristics of the nature of language use that give rise to variation at both the phonological level and the grammatical level. This sort of variation arises even though the interlocutors are trying to conform to linguistic convention. There is also the possibility that speakers are intentionally violating linguistic convention. This is plausible in the model of language use described in this chapter, because linguistic convention is a means to an end; other means may be employed to achieve the same end, if they are successful. These other means are in fact mechanisms for the generation of variation that have been appealed to in one form or another for many decades: to be more expressive; to avoid misunderstanding;

to save energy (by producing shorter or terser utterances), and so on. In the cases, a speaker is intentionally violating convention in order to achieve a social interactional goal. While I do not intend to downplay these "intentional" mechanisms to generate variation (see Keller 1994 for a particularly good explication of these mechanisms), I do wish to emphasize that the model of language use described in this chapter actually predicts that variation is continually generated even in the process of "normal", conventional language use, at both the phonological and grammatical levels. It is not the case that there is a fixed linguistic system that must be violated for variation to take place; variation is part of the linguistic system, because it is part of the nature of language use.

### 5.3.4 Speech Communities and the Selection/Propagation of Variants in Language Use

The generation of variation occurs in conversational interaction, which is in fact just one facet of the social interaction that is facilitated by language. Thus, the generation of variation takes place at the very small time scale, the time scale of lingueme replication, described in the first section of this chapter. The propagation of variation also occurs in language use: it depends on the selection of one variant as opposed to another. But its effect is defined over the entire population at a larger time scale, namely the time scale encompassing the survival, fixation or extinction of the linguistic variant over very many generations of replication. Hence, to understand the selection or propagation process, we must consider the population as a whole. The population of linguemes (the replicators)—the variants that are propagated or go extinct—is defined by the population of speakers (the interactors), who replicate the linguemes in language use. Thus we must examine the speech community and its role in the linguistic selection process.

The definition of speech community is tied closely to the definition of a language. A speech community is said to be defined by the language it speaks. But it is difficult to identify what counts as the same language. Are dialects parts of the same language, linguistically speaking? In some cases, such as Bulgarian and Macedonian, the differences among Bulgarian dialects is as great as the difference between Bulgarian and Macedonian, based on linguistic structures.

A more satisfying definition of a language, which follows evolutionary theory, is a population approach. We also start with the speakers rather than the language itself. Just as a biological species is defined as a reproductively isolated population, a speech community, and hence the language it speaks, is defined as a communicatively isolated population. Hence a speech community is defined by its social interactions involving language: members of the speech community communicate with each other, and the community is defined by its communicative isolation from other communities. Communicative isolation is relative, of course, as is reproductive isolation among biological populations. In fact, the structure of human speech communities appears to be far more complex than the structure of biological populations. A language is therefore the population of utterances produced by the members of the speech community as they communicate with one another. One can observe the results of the selection process in the lingueme pool, parallel to the gene pool of a biological population.

### 5.3.5 Defining Speech Communities: Common Ground and Shared Practice

Two different properties have been proposed to define what members of a community, including a speech community, have in common. The first property, common ground, links the speech community to the characteristics of language use, namely joint action and convention, which presuppose common ground. The second property, shared practice, links the speech community to the population definition just given: the shared practice for a speech community is communicative interaction, and the joint actions that it serves. These two definitions are interconnected and should not be thought of as contradicting each other.

Common ground plays an essential role in defining joint action and convention, both central to understanding the nature of language. Common ground consists of knowledge, beliefs and attitudes presumed by two or more individuals to be shared between them. The shared knowledge that constitutes common ground emerges from a shared basis. A shared basis for common ground has the following properties: the basis provides information to the persons involved that it holds; the basis indicates to each person that it provides information to every person that it holds; and the basis indicates to every person the proposition in the common ground (Clark 1996: 94). A basis for common ground varies in how well it is justified (in other words, Clark's definition must be recast in more gradient terms); hence we may not always be certain of what is or is not common ground with our fellow community members.

Common ground is divided into two types: personal common ground and communal common ground (Clark 1996, Chap. 4). Personal common ground is shared directly in face-to-face interaction by the persons who share it. The networks of individuals who are defined by their personal common ground correspond to social networks, which are instrumental in language maintenance and change (Milroy 1987).

Personal common ground has two bases. The first is called the perceptual basis by Clark: we share knowledge of what is in our shared perceptual field. The perceptual basis is provided by virtue of joint attention and joint salience, which also plays a central role in any sort of social interaction, including language use. The second basis for perceptual common ground is the discourse basis. When I communicate situations I have experienced to you in conversation, and vice versa, these become part of our personal common ground. Although we did not experience them perceptually together, we did experience the reporting of them linguistically together. The discourse basis thus involves joint attention (to the linguistic signal), as well as the common ground of a shared language.

The other type of common ground is communal common ground. Communal common ground is shared by virtue of common community membership. A person can establish common ground with a stranger if they belong to a common community (e.g. Americans, linguists, wine enthusiasts, etc.). Some communities are quite specialized, while other communities are very broad and even all-encompassing, such as the community of human beings in this world. The fact that there is some common ground defining all human beings in this world gives rise to the possibility of communication in the first place.

Clark argues that the basis of communal common ground is shared expertise. Wenger (1998), on the other hand, defines communities of practice in terms of shared practice: individuals engage in joint actions together, and this gives them common ground and creates a community. Wenger's definition of a community of practice therefore requires face-to-face interaction. Hence communities of practice resemble social networks. Clark defines social networks in terms of personal common ground. Wenger's and Clark's definitions of a social network are not incompatible, as personal common ground emerges from shared practice.

Shared practice can be passed on as new members enter the community and share practice with remaining current members. This is cultural transmission, and can lead to individuals being members of the same community through a history of shared practice even if they do not interact directly with every other member of the community. Hence shared practice is more like shared expertise, albeit with an emphasis on the joint actions that give rise to knowledge rather than the knowledge itself. Hence shared practice can be the basis for communal common ground as well as personal common ground.

Since communities are defined by shared practice, and human beings engage in a great variety of joint actions with different groups of people, the community structure of human society is extremely complex. Every society is made up of multiple communities. Each person in the society is a member of multiple communities, depending on the range of shared activities he or she engages in. The different communities have only partially overlapping memberships. Hence the structure of a human society consists of multiple overlapping communities. Communities also differ in their degree of inclusiveness. The community of historical linguists is less inclusive than the community of linguists, which in turn is less inclusive than the community of academic scholars.

As a consequence, a language of a population is equally complex. A linguistic structure—a pronunciation, a word, a construction—is associated with a particular community, or set of communities, in a society. A pronunciation is recognized as an accent characteristic of a particular community. Words will have different meanings in different communities (e.g. *drift* is parallel directed change for linguists but random change for biologists). The same concept will have different forms in different communities (e.g. *Cabernet Sauvignon* for the general layperson, *Cab* to a wine aficionado). Thus, a linguistic convention is not just a symbol—a pairing of form and meaning—but has a third part, the community in which it is a convention.

This is in fact part (iii) of the definition of convention given above. In other words, this "semiotic triangle" of form-meaning-community is implicit in the definition of convention, including linguistic conventions.

Hence a language as a population is neither homogeneous nor sharply delineated. It is a complex system of multiple, partially overlapping populations of linguistic structures of different degrees of community inclusiveness. Each individual has a linguistic *repertoire* which reflects her knowledge and exposure to the communities in the society in which she engages in joint actions (shared practice). One consequence of this complex population structure is that an individual is familiar with linguistic variants for (roughly) the same meanings from different communities in which she interacts. Hence a speaker has multiple variants available to her, which are associated to greater or lesser extent with specific communities in her mind. Those variants may belong to what linguists would call different languages, in a multilingual society, even though they may be combined into a single utterance (what sociolinguists call code-switching).

### *5.3.6 Mechanisms of Propagation (Selection) in Language Use*

Against this background of the population structure of a speech community, we can now consider how linguistic selection takes place. As in the case of the generation of variation, the selection or propagation of variants may be attributable to intentional mechanisms or nonintentional mechanisms. And again, there has been a recent shift of focus from intentional to nonintentional mechanisms.

Many analyses of intentional selection processes involve an act of identity: the choice of a linguistic form on the part of a speaker is an act of identification with the community that uses it (LePage and Tabouret-Keller 1985). In some cases, this might be due to prestige: a speaker chooses a variant that is associated with a community in a higher socioeconomic class. Or it may be that the variant is chosen with a different community that is not in a higher socioeconomic class, but has other social traits that lead certain speakers to identify with it. More recent work in sociolinguistics has argued that linguistic acts of social identity are not always passive: individuals institute linguistic conventions to construct an identity as well as to adopt one (Eckert 2000). Whatever the social factor might be for these theories of propagation, they involve some differential social valuation assigned to different variants, such that one is more likely to be chosen in a particular occasion of use than another. In an evolutionary framework, this model corresponds to associating linguistic variants (linguemes) with different fitness values: the classical selection model in evolutionary theory. Baxter et al. (2009) describe this as *replicator selection*, since the fitness value is associated with the replicator (the lingueme or linguistic variant), not the interactor.

Other linguists have argued that one need not invoke any social valuation of linguistic variants to account for the propagation of language change, at least for certain situations of language change. All that is necessary is differential exposure to the different variants: relative frequency that a speaker hears the variants will determine the likelihood that a variant is propagated. This corresponds to an evolutionary *drift* model of propagation.

Another mechanism is one based solely on social network structure: a member of the speech community interacts with certain speakers more often than other speakers (for any reason—by choice or by circumstance), and therefore is exposed to the variant preferred by the former group more than the variant used by the latter group. Baxter et al. (2009) describe this mechanism as *neutral interactor selection*: the differential weighting of variants is due to the different likelihoods of interaction with interactors, but the variants produced by any interactor are weighted equally.

A final mechanism is that a member of the speech community weights variants differentially based on who produces the variant. For this mechanism, what matters is not how often one interacts with a particular speaker, but what value is placed on that speaker's speech compared to the speech of others. In other words, one weights the utterances on one speaker more than another speaker by virtue of some social valuation, regardless of how often one interacts with either of those speakers. Baxter et al. (2009) describe this mechanism as *weighted interactor selection*. In weighted interactor selection, unlike replicator selection, the value is associated with the speaker, not the variant: if another speaker happens to produce the same variant, its weighting by the hearer is determined by the hearer's valuation of the speaker, not the particular variant produced. Although this mechanism has not been explicitly described as such, it appears to be the best model for theories in which the linguistic variants produced by a subset of speakers (the leaders of a change) are valued more highly than the variants produced by the rest of the community (e.g. Milroy and Milroy 1985; Labov 2001, Chap. 10). The mechanisms of propagation (selection) are summarized in Box 11.

**Box 11: mechanisms of propagation (selection) (Baxter et al. 2009; Blythe and Croft 2012)**

(1) Drift: random fluctuations in replicator (e.g. a linguistic variant) frequencies lead to propagation (ideally, to fixation); this implies that the more frequent replicator is more likely to be propagated, other things being equal.
(2) Neutral interactor selection: the likelihood that a replicator is propagated depends on the likelihood that an interactor (e.g. a speaker) producing that replicator interacts with other interactors (e.g. the hearers), other things being equal.
(3) Weighted interactor selection: the likelihood that a replicator is propagated depends on the value associated with the interactor (speaker) by the interactor being interacted with (hearer), which assigns a weight to the value of the variant.

(4) Replicator selection: the likelihood that a replicator is propagated depends on the value associated with the replicator, independent of which interactor produces that replicator.

At this point, it is too early to evaluate the operation of these mechanisms in actual cases of language change. Trudgill (2004) argues that drift and neutral interactor selection are sufficient to account for propagation of language change in new-dialect formation in an isolated speech community, such as the emergence of New Zealand English in the latter half of the nineteenth century. Baxter et al. (2009) argue based on a statistical mechanical model that these two mechanisms are not sufficient, although they do play a role in the process. Blythe and Croft (2012) argue that replicator selection is virtually necessary in order to obtain the S-curve found in the propagation of novel linguistic variants. In this author's view, it is likely that all four mechanisms play a role in the propagation of language changes in most if not all social situations.

## 5.4 Conclusion

This chapter has described an evolutionary model of cultural transmission, specifically the transmission or replication of linguistic elements (linguemes) by interactors (speakers). An evolutionary model is a model of change by replication. An evolutionary model is also a two-step model: the generation of variation and the selection of variants. Thus the processes described in this chapter operate on two different timescales: the very short timescale of conversational interaction for the generation of variation (i.e. innovation of a language change), and the much longer timescale of decades or centuries or even millennia of replicator generations—not human biological generations—for the selection of linguistic variants (i.e. the propagation of a change).

This chapter has presented a theory of language use, based largely on Clark (1996) and his antecedents in philosophical pragmatics, that provides plausible mechanisms for both the generation of variation and the selection/propagation of variants. Generation of variation occurs all of the time, due to basic indeterminacies in the processes of language use in both production and comprehension, at both the phonological and grammatical levels. Propagation can and apparently does occur as the result of a variety of mechanisms operating over an extremely complex population structure based on common ground and ultimately shared practice among members of a society.

The model for analyzing processes of language change described in this chapter represents research that has been developed and/or integrated only in the past decade and a half, and represents the first steps at empirically testing and elaborating the hypotheses embodied in the model. It is hoped that future research will allow us to increase our understanding of the role that language use plays in language change, and its relation to other evolutionary processes related to language behavior.

## References

Aunger, R.: The Electric Meme: A New Theory of How We Think. Free Press, New York (2002)

Austin, J.L.: How to do Things with Words. Harvard University Press, Cambridge (1962)

Baxter, G.J., Blythe, R.A., Croft, W., McKane, A.J.: Modeling language change: an evaluation of Trudgill's theory of the emergence of New Zealand English. Lang. Var. Change **21**, 157–196 (2009)

Blackmore, S.: The Meme Machine. Oxford University Press, Oxford (1999)

Blythe, R.A., William, C.: S-curves and the mechanisms of propagation in language change. Language **88**, 269–304 (2012)

Bratman, M.: Shared cooperative activity. Philos. Rev. **101**, 327–341 (1992)

Bybee, J.L.: Phonology and Language Use. Cambridge University Press, Cambridge (2001)

Bybee, J.L.: Frequency of Use and the Organization of Language. Oxford University Press, Oxford (2006)

Bybee, J.L.: Language, Usage and Cognition. Cambridge University Press, Cambridge (2010)

Chafe, W.: The recall and verbalization of past experience. In: Cole, P. (ed.) Current Issues in Linguistic Theory, pp. 215–246. Indiana University Press, Bloomington (1977)

Chafe, W. (ed.): The Pear Stories. Ablex, New York (1980)

Clark, H.H.: Arenas of Language Use. University of Chicago Press and the Center for the Study of Language and Information, Chicago & Stanford (1992)

Clark, H.H.: Communities, commonality and communication. In: Gumperz, J.J., Levinson, S.C. (eds.) Rethinking Linguistics Relativity, pp. 324–355. Cambridge University Press, Cambridge (1996)

Clark, H.H.: On the origins of conversation. Verbum **21**, 147–161 (1999)

Croft, W.: Explaining Language Change: An Evolutionary Approach. Longman, Harlow (2000)

Croft, W.: The origins of grammar in the verbalization of experience. Cogn. Linguist. **18**, 339–382 (2007)

Croft, W.: The origins of grammaticalization in the verbalization of experience. Linguistics **48**, 1–48 (2010)

Croft, W.: Explaining Language Change: An Evolutionary Approach, 2nd ed. Oxford University Press, Oxford (In preparation)

Dawkins, R.: The Selfish Gene. Oxford University Press, New York (1976)

Dawkins, R.: The Extended Phenotype. Freeman, San Francisco (1982)

Eckert, P.: Language Variation as Social Practice: The Linguistic Construction of Identity in Belten High. Blackwell, Oxford (2000)

Grice, H.P.: Meaning. In: Grice, H.P. (ed.) Studies in the Way of Words, pp. 213–223. Harvard University Press, Cambridge (1948/1989)

Griesemer, J.R.: Development, culture, and the units of inheritance. Philos. Sci. **67**(Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers), S348–S368 (2000)

Griffiths, P.E., Gray, R.D.: Developmental systems and evolutionary explanation. J. Philos. **90**, 277–304 (1994)

Griffiths, P.E., Russell D.G.: Darwinism and developmental systems. In: Oyama, S., Griffiths, P.E., Gray, R.D. (eds.) Cycles of Contingency: Developmental Systems and Evolution (pp. 195–218). MIT Press, Cambridge (2001)

Hull, D.L.: Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science. University of Chicago Press, Chicago (1988)

Keller, R.: On Language Change: The Invisible Hand in Language. Routledge, London (1994) (Translation and expansion of Sprachwandel, 1990)

Labov, W.: Principles of Linguistic Change. Internal Factors, vol. 1, Basil Blackwell, Oxford (1994)

Labov, W.: Principles of Linguistic Change: Social Factors, vol. 2, Blackwell, Oxford (2001)

Lakoff, G.: Women, Fire and Dangerous Things: What Categories Reveal About the Mind. University of Chicago Press, Chicago (1987)

Langacker, R.W.: Foundations of Cognitive Grammar, Theoretical Prerequisites, vol. I, Stanford University Press, Stanford (1987)

LePage, R.B., Tabouret-Keller, A.: Acts of Identity. Cambridge University Press, Cambridge (1985)

Lewis, D.: Convention. MIT Press, Cambridge (1969)

Mayr, E.: The Growth of Biological Thought: Diversity, Evolution, Inheritance. Belknap Press, Cambridge (1982)

Milroy, J., Milroy, L.: Linguistic change, social network and speaker innovation. J. Linguist. **21**, 339–384 (1985)

Milroy, L.: Language and Social Networks, 2nd edn. Basil Blackwell, Oxford (1987)

Ohala, J.: The origin of sound patterns in vocal tract constraints. In: MacNeilage, P.F. (ed.) The Production of Speech, pp. 189–216. Springer, New York (1983)

Pierrehumbert, J.: Phonetic diversity, statistical learning, and acquisition of phonology. Lang. Speech **46**, 115–154 (2003)

Searle, J.R.: Speech Acts: An Essay in the Philosophy of Language. Cambridge University Press, Cambridge (1969)

Tomasello, M.: The Cultural Origins of Human Cognition. Harvard University Press, Cambridge (1999)

Tomasello, M.: Origins of Human Communication. MIT Press, Cambridge (2008)

Trudgill, P.: New-Dialect Formation: The Inevitability of Colonial Englishes. Edinburgh University Press, Edinburgh (2004)

Weinreich, U., Labov, W., Herzog, M.I.: Empirical foundations for a theory of language change. In: Lehmann, W.P., Malkiel, Y. (eds.) Directions for Historical Linguistics, pp. 95–195. University of Texas Press, Austin (1968)

Wenger, É.: Communities of Practice: Learning, Meaning and Identity. Cambridge University Press, Cambridge (1998)

# Chapter 6
# Transitions: The Evolution of Linguistic Replicators

**Simon Kirby**

**Abstract** Maynard Smith and Szathmáry (1995) propose a series of major transitions in the evolutionary history of life. Their work provides a rich framework for thinking about replication. They identified the importance of language in this light, but language is a new system of replication in more than one sense: it is both an enabler of cultural replicators with unlimited heredity, and also a new kind of evolutionary system itself. Iterated learning is the process of linguistic transmission, and it drives both language change and the transitions to qualitatively new kinds of linguistic system. By seeing language as an evolutionary system, the biggest payoff we get may be the ability to take biologists' insights into the evolution of life and apply them to the evolution of language.

## 6.1 Introduction: The Major Transitions

In their thought-provoking survey, Maynard Smith and Szathmáry (1995, hereafter MS&S) propose a series of "major transitions" in the evolutionary history of life. These take us from the emergence of populations of replicating molecules out of solitary replicating molecules as the earliest transition, all the way to the most recent, the emergence of human language-using societies out of primate a-lingual ones. In the introduction to their work, they list eight transitions in particular:

---

S. Kirby (✉)
School of Philosophy, Psychology and Language Sciences,
University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh, UK
e-mail: simon@ling.ed.ac.uk

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_6,

Replicating molecules → Populations of molecules
Independent replicators → Chromosomes
RNA → DNA
Prokaryotes → Eukaryotes
Asexual clones → Sexual populations
Protists → Animals, plants, fungi
Solitary individuals → Colonies
Primate societies → Human societies, **Language**

This list is not merely a rag-bag of interesting evolutionary developments, of course. The insight that MS&S had was that there were recurring themes and commonalities among these crucial points in evolutionary history. Despite their apparent diversity, by considering these transitions together we can reveal underlying similarities such that advances in understanding any one can deepen our understanding of the others. In this chapter, I will suggest that we can extend the same logic to the evolution of language and find transitions which are remarkably similar in kind to the ones that MS&S discuss.

But first, I will quickly survey some of the common themes that recur throughout the major transitions:

1. **Loss of independent replication**: Examples given by MS&S include the origins of social groups such as in the social insect species in which individuals can only survive as part of a group. Another example is the origin of the eukaryotes. Ancestors of mitochondria and chloroplasts, which now replicate only as part of a cell, were once independently viable.
2. **Division of labour**: An example of this feature of transitions is given by MS&S's characterisation of a transition from a hypothetical "RNA world" where RNA worked as both carrier of genetic information and as a catalyst, to the DNA world in which DNA carries the code, but other functions are carried out by proteins.
3. **New ways of transmitting information**: There are many examples of this in MS&S's book and in some sense it is what underpins the whole idea of evolutionary transitions. With the origins of RNA, the origins of the genetic code, and so on we see the emergence of important new systems of inheritance.

It is striking that MS&S put language on their list as part of the last transition they discuss. Why have they done this? The answer is that they feel language is a significantly new system of information transmission. Language, for MS&S, is part of the last transition because it supports human cultural transmission. Cultural information that humans possess can be *replicated* in other humans through the medium of language. Szathmáry (2000) refers to the units of information passed on in this way as human memes, and argues that these are a new type of replicator that could only be possible with language.

Memes are an example of a *phenotypic replicator* in Szathmáry's (2000) terminology. They are transmitted without direct copying of a code. Other examples of phenotypic replication in nature include prions, which are proteins with have an infective phenotype that replicates by re-shaping proteins it comes into contact with directly—again, without use of a separate copied code. What makes human memes supported by language special for MS&S, and qualitatively different from what has gone before, is that they have *unlimited heredity*:

> A crucial distinction is between systems of 'limited heredity', in which only a few distinct states can be transmitted, and systems of 'unlimited heredity', capable of transmitting an indefinitely large number of messages (Maynard Smith and Szathmáry 1995, p. 13).

So, in summary, the emergence of language saw the appearance of a qualitatively new kind of transmission of cultural information in the primate lineage. Prior to this, cultural information was transmitted via phenotypic replicators with limited heredity, afterwards we made the transition to unlimited phenotypic replication.

## 6.2 Linguistic Replication

It is perhaps striking that Maynard Smith and Szathmáry's (1995) work has not made a more significant impact on the evolutionary linguistics community, let alone linguistics itself. Part of the problem may be that it is unclear what placing language among the major evolutionary transitions actually tells us about the evolution of language. One approach might be to pursue an adaptationist account whereby our capacity for language was selected for on the basis of its meme-carrying capacity. But it is unclear how such an account would be spelled out, beyond the simplistic but common notion that language must have evolved "for communication" (see, e.g., discussion in Pinker and Bloom 1990, and its associated commentaries). The most obvious difficulty is in determining how a capacity for sharing information not only among others in a group, but also to future generations could be promoted by selection operating on individuals.

Here I want to propose another way of looking at the relationship between the evolution of language and the major transitions. If you consider an utterance, what information does that utterance convey? Most obviously, it conveys *semantic* information, and ultimately this is the information that MS&S are talking about in their discussion of unlimited cultural heredity. In other words, utterances carry *content*. But there is another kind of information that can be conveyed by any linguistic production, and that is information about the linguistic system itself. Along with content comes *form*. When I produce the sentence "these berries are good" I may be propagating cultural information about the edibility of items in the environment via the content of the sentence. At the same time I may also be propagating information about the construction of sentences in my language: its phonology, its lexical entries, its syntactic structure and so on.[1] Obviously, the relevance of these two kinds of information transmission will vary depending on the listener. Children in the process of language acquisition will be making the most significant use of the latter kind, although the process of language change is driven to a large part by ongoing adaptation of adults' linguistic representations throughout life (see, e.g., Croft 2000, for discussion).

So, language is not only a mechanism for cultural replication, it is *itself* culturally[2] transmitted. By transmitting information about their own construction, utterances are part of a system of linguistic replication. There has been a great deal of debate in the literature about the best way to characterise language change in broadly Darwinian terms, and a thorough review of this debate is beyond the scope of this chapter (see, e.g., Croft 2000; Ritt 2004; Kirby 1999; McMahon 1994). Suffice it to say that there are a number of different ways of applying ideas such as replication and selection to the linguistic system, each with their own plausibility and merits. Here, I want to relate linguistic transmission to Szathmáry's (2000) characterisation of replicators.

One way of thinking about the process of language acquisition is in terms of *reverse engineering*. This is a process whereby an engineering team will attempt to copy a competitor's product by inferring the (hidden) inner workings of that product through careful observation of its external behaviour. The goal of reverse engineering is not so much to reproduce just the observed behaviours, but to generalise these to all possible behaviours of the product being copied. This is likely to result in a copy of the inner working of the product, but this is not guaranteed. Reverse engineering is appropriate when the underlying mechanism is hidden and only external behaviour is accessible. This is exactly the case for language acquisition: children cannot directly observe the state of the (neural) mechanisms that generate adult linguistic behaviour but instead must attempt to create their own internal state that will produce similar behaviour.

Replication involving reverse engineering in this way is *non-Weissmannian* (Brighton et al. 2005) in that it does not involve direct copying of a code (genetic replication, on the other hand, is the classic example of a Weissmannian process because DNA provides a coding scheme for phenotypes that is copied directly). This makes linguistic structure another example of a phenotypic replicator, like memes; one that is transmitted culturally rather than biologically.

---

[1] Croft (2000, this volume) elaborates a specific proposal along these lines, motivated by a theoretical framework for linguistic representation known as Construction Grammar. Croft treats constructions as replicators—linguemes in his terminology.

[2] Note that there are potentially confusing difference in the interpretation of "cultural". I will use "cultural" to mean any behaviour that is inherited socially rather than genetically. "Culture" in the sense of the specific identifying practices of a society is transmitted culturally, but so too are behaviours like language.

## 6.3 Transitions in Linguistic Evolution

It would seem, then, that language qualifies for inclusion on Maynard and Szathmáry's (1995) list of major transitions for two reasons. It enables a new system of transmission of cultural content with unlimited heredity (the reason the authors themselves gave), but also its appearance heralds a completely new evolutionary system of linguistic structure itself through phenotypic replication by repeated reverse engineering.

In this chapter I propose that with this insight we can take MS&S's ideas much further and look *within* this new system of replication for evolutionary transitions that bear striking similarities to those seen in the broader evolutionary history of life. These major transitions in linguistic evolution are primarily non-biological, involving an evolution of the systems of linguistic phenotypic replication, but they are likely to interact with ongoing biological evolution of humans in interesting ways that we are only beginning to understand (see also, Zuidema 2005, for discussion of transitions in linguistic evolution that have both cultural and biological implications).

I will discuss here three candidates for major transitions:

Vocal trajectories → Combinatorial phonotactics
Holophrastic utterances → Compositionality
Monolithic lexicon → Contentive/functional split

It is important to note that these are essentially preliminary suggestions for significant shifts in the cultural evolution of language that mirror the other evolutionarily earlier major transitions. It is very likely that there will be other ones that also fit the general pattern. Equally, it is important to stress that each transition does not deliver up the linguistic system ready for the next[3] (this is also true of MS&S's list). Nevertheless, these are plausible transitions that derive from a change in the system of replication similar to that which recurs again and again in the history of life. In particular, we will see that these linguistic transitions are characterised by a loss of independent replication, and new divisions of labour—key features of the major evolutionary transitions listed in the introduction.

Before I cover each of these proposed transitions in turn, it is worth expanding a little on the process of linguistic transmission by repeated reverse engineering. In the growing literature on the cultural evolution of language, this process has been termed *iterated learning* (e.g., Kirby 2000; Kirby and Hurford 2002; Kirby et al. 2004; Brighton et al. 2005; Kirby et al. 2007; Kalish et al. 2007):

> **Iterated Learning:** a process whereby some behaviour is acquired by an individual by observing an equivalent behaviour in another individual who acquired it in the same way (Kirby et al. 2008, p. 10681).

---

[3] See Jackendoff (1999) for a more detailed proposal of a sequence of trajectories language could have gone through. Jackendoff's trajectories may be broadly compatible with the view presented here, but note that the driving mechanism I propose is cultural rather than biological evolution.

Language persists through iterated learning via two states: an internal state private to the individual (i.e. a grammar), and a public external state in the form of utterances (see Hurford 1990, for extensive discussion of the implications of this). Language is transformed from its internal to external state through a process of language production, and back again by perception and learning—the reverse engineering process discussed above. Note that there are many ways we could view "learning" here. At one end of the scale it could involve the acquisition of completely new structures, lexical items and so on, whereas at the other it could simply involve the adjustment of frequencies of variants driven by something like priming (see, e.g., Croft 2000; Rosenbach and Jäger 2008, Garrod and Pickering, this volume).

We normally think of this cycle of repeated production and perception/learning as driving language *change* (see, e.g., Croft 2000; Briscoe 2000; Kirby 1999; Andersen 1973), but typically iterated learning is used in models of language *evolution*. What, if any, is the difference between language change and the cultural evolution of language, except that they are discussed in different conferences? Some authors have suggested that the term "language evolution" should be reserved for the biological evolution of the language faculty (Bickerton 2007), whereas others have used the term "evolutionary" to refer to the ongoing process of historical language change (e.g., Blevins 2004). This is basically a matter of terminology, so we should not get too hung up on it, but I think there is a useful contrast that can be made in the light of the present argument: language evolution (in the cultural, rather than biological, sense) involves language change *but also major transitions*.

It is appropriate to talk about language evolution when we are discussing the emergence of language as we recognise it today out of qualitatively different non-linguistic precursors. For some, such qualitative changes in language *require* biological changes, and hence it makes no sense to talk about cultural evolution of language. But if iterated learning leads to changes in the process of replication as I will suggest, then talking about this in terms of "mere" change fails I think to capture what is being claimed, even though the basic mechanism underlying both change and evolution are the same. To put it another way, one of the central dogmas of linguistics has been the uniformitarianist assumption: that all languages are in some sense qualitatively equivalent and have always been so (although see Newmeyer 2002, for critical review). This assumption of uniformity applies in the main to theories of diachronic linguistics, but it can hardly make sense when considering the evolutionary origins of language! Instead, it makes sense to maintain as far as possible an assumption of uniformity of *process* rather than *state*. In other words, the same process of phenotypic replication through iterated learning may drive both evolutionary transitions of language from one qualitatively different state to the next and at the same time the normal process of ongoing change within each of those evolutionary stages.

## 6.4 Three Example Transitions

With these preliminaries out of the way, we will now turn to the three candidates for major transitions in the evolution of language. None of these are new proposals, and the first two in particular have been described in substantial detail elsewhere as being driven by a process of iterated learning. My purpose in highlighting them here is to show that they can be viewed in the light of Maynard Smith and Szathmáry's (1995) work as instances of a more general pattern of evolutionary transitions.

### *6.4.1 The Emergence of Compositionality*

A number of authors have proposed that an evolutionarily early form of language may have once existed with some of the features of language as we know it, but not all (Bickerton 1990; Wray 1998; Jackendoff 1999). There are various suggestions about what such a *protolanguage* may have looked like, but one particularly prominent theory is that it consisted initially of solely *holophrastic* utterances (Wray 1998; Kirby 2000; Arbib 2005; Mithen 2005). In a holophrastic utterance, the mapping between meanings and signals is non-compositional, with the whole string of phonemes in a sentence corresponding to the whole meaning. In this respect, protolanguage utterances would have been like non-compositional forms in language today such as one-word utterances (*fire!*), idioms (*bought the farm*), and holistic formulae (*how's it going?*).[4]

An obvious issue is how language evolved from this holophrastic stage into a compositional one that we are now familiar with where the meaning of most sentences is a function of the meanings of parts of the sentence. Wray (1998) proposes a process of "fractionation" by analysis where chance correspondences between sets of holistic expressions lead to learners breaking the previously unanalysed strings into sub-parts that can then go on to be combined. Although the plausibility of this process has been questioned by some (Tallerman 2007), it gains strong support from a large range of computational simulations of iterated learning (see, e.g., Brighton et al. 2005, for review), and, more recently, experimental studies of the cultural transmission of artificial languages in human subjects (Kirby et al. 2008). These simulations and experiments were not set up to explicitly test Wray's (1998) hypothesis, and were developed

---

[4] Idioms and holistic formulae are multi-word, so may appear to be compositional rather than holophrastic. However, in an idiom like *bought the farm* the meaning (*died*) has no systematic relationship to the meaning of the parts of the idiom so it is non-compositional. In the case of holistic formulae such as *how's it going?* although the meaning is arguably compositional, it is likely that in actual use this utterance is not processed by either speaker or hearer compositionally, but rather as a holistic chunk. See Wray (1998) for more discussion. These apparently compositional holophrases exist in fully modern human language because it is indeed largely compositional. In the hypothesised protolanguage stage, the holophrases would have been entirely idiosyncratic in form, rather like the mono-morphemic lexicon of any modern language.

largely independently, but nevertheless the observed process of gradual emerging fractionation by analysis is strikingly similar to that predicted by Wray (1998).

A typical simulation uses a population of agents (i.e. simulated individuals) that each learn a language mapping strings of arbitrary symbols onto a structured "world" of meanings. Agents produce a finite subset of the possible utterances in their language by being prompted with a randomly picked set of meanings drawn from the world. The set of pairs of meanings and strings thus produced are what provide the training input for the next generation of agents. Obviously, within this broad characterisation there is a lot of individual variation in different studies, for example with respect to the learning algorithm the agents employ, the space of meanings in the world, the size and structure of the population, the existence of population turnover and so on. The sentences below are examples taken from a fairly typical simulation (Kirby et al. 2004): the first set are produced by an agent early in the simulation, whereas the second are from an agent thousands of generations later, after the language has evolved (purely culturally) through repeated cycles of learning and production (glosses of the meaning of the sentences are given in English):

(1) a. ldg
"Mary admires John"
b. xkq
"Mary loves John"
c. gj
"Mary admires Gavin"
d. axk
"John admires Gavin"
e. gb
"John knows that Mary knows that John admires Gavin"

(2) a. gj h f tej m
John Mary admires
"Mary admires John"
b. gj h f tej wp
John Mary loves
"Mary loves John"
c. gj qp f tej m
Gavin Mary admires
"Mary admires Gavin"
d. gj qp f h m
Gavin John admires
"John admires Gavin"
e. i h u i tej u gj qp f h m
John knows Mary knows Gavin John admires
"John knows that Mary knows that John admires Gavin".

In this example, we see clearly the emergence of compositional encoding of meanings from an initially randomly constructed holistic protolanguage. There are obvious advantages to a compositional scheme—most notably, individuals can generalise from a sub-sample of utterances in their language to unseen examples and thus communicate reliably to other individuals about novel situations. This is impossible with a holistic protolanguage since the signal for each meaning is essentially completely independent of the others and must be learned separately by rote.

Note, however, that although the compositional language is clearly superior from a communicative point of view and would therefore appear to be the sort of thing that would arise by a process of selective adaptation for communicative function, this is *not* what is driving the process in the simulation. There is no biological evolution in these simulation (nor, of course, in the parallel experimental models using human participants by Kirby et al. 2008!). So what is going on here? Why do compositional languages evolve?

Rather than think about this evolution in terms of the benefits to the communicating individuals, it actually makes more sense to consider the benefits of the emergence of compositionality for the system that is being transmitted—in other words, language itself. In these models, as in reality, language persists despite having to be passed through a transmission *bottleneck*. Learners never get perfect training data because of the finite time available to them. In the simulations this is normally modelled by deliberately restricting the number of training examples so that it is lower than the total number of meanings. In the case of real language, this will always be the case because the set of meanings we may wish to convey is unbounded. Holistic languages *cannot* be reliably transmitted in the presence of a bottleneck like this, since generalisation to unseen examples cannot be reliable. A compositional language, on the other hand, enables completely reliable transmission in the presence of a bottleneck (assuming that enough training examples are available to cover the basic vocabulary of the language).[5]

So, the language appears to have adapted simply through the process of iterated learning in such a way as to become more learnable. In the presence of a bottleneck, the structure of the strings in the language has changed in order to leverage pre-existing structure in meanings.[6] In the simulations we can observe exactly how this is happening. It turns out to be precisely the mechanism Wray (1998) suggested: similarities between strings that by chance correspond to similarities between their associated meanings are being picked up by the learning algorithms that are sensitive to such substructure. Even if the occurrences of such correspondences are rare, they are amplified by the iterated learning process. A holistic mapping between a single

---

[5] Another way of thinking about this is that the bottleneck sets the level of granularity of dividing up the meanings at which a language can afford to be idiosyncratic and non-compositional. Simulation results demonstrate that frequently expressed meanings can be described holistically, whereas infrequent ones must be compositional (Kirby 2001). Similarly, languages typically exhibit irregular morphology among frequent items (Francis and Kucera 1982). Consider the past tenses of frequent verbs in English such as *be/was*, *go/went* etc. for example. In this view, the word or morpheme is simply the level at which the language can afford to be holistic and still be transmitted faithfully.

[6] Although, see Kirby (2007) for a model where the structure of meanings themselves can similarly be seen to adapt to improve learnability.

meaning and a single string will only be transmitted if that particular meaning is observed by a learner. A mapping between a sub-part of a meaning and a sub-string on the other hand will be provided with an opportunity for transmission every time *any meaning is observed that shares that sub-part*. Because of this differential in the chance of successful transmission, these compositional correspondences tend to snowball until the entire language consists of an interlocking system of regularities.

If we now return to Szathmáry's (2000) characterisation of replicators, it becomes clear that there are striking parallels between this transition from protolinguistic holophrasis to syntactic compositionality and the other evolutionary transitions discussed. In the early phase, there is a population of individual replicators corresponding to meaning-signal pairs. The survival of each of these replicators depends on it being heard by a learner. In Szathmáry's terms, these are *solitary* phenotypic replicators. To put it simply, each stands or falls on its own, they do not interact in any significant way (see Fig. 6.1, left).

After the transition, sub-parts of meanings are associated with sub-parts of strings (Fig. 6.1, right). The replicators now are no longer single utterances, but rather individual words/morphemes and systems of combination (i.e. grammatical rules).[7] As we have noted, words and rules are more prolific than single holophrases, since each occurs over many meanings, but on the other hand they now rely on more examples before they can be acquired. More importantly, there is a loss of independent replicability—one of the identifying features of MS&S's major transitions. They have become *ensemble* replicators. Along with this development comes another feature typical of an evolutionary transition: division of labour. Whereas in the holistic language, there is no distinction between a lexicon of words and the set of rules for grammatical utterances, after the transition there is a clear division between the words of the language and the systems for combining those words in grammatical or ungrammatical ways.

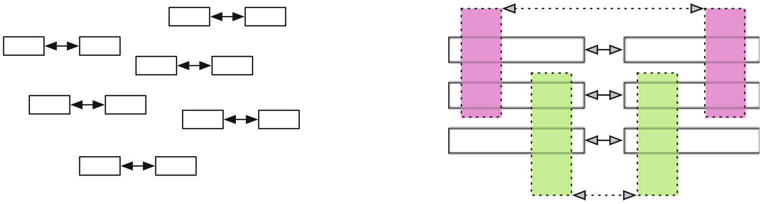


**Fig. 6.1** Before (*left*) and after (*right*) the transition from a holophrastic to compositional language. Prior to the transition, replicators are individual meaning, signal pairs. After the transition, these pairs are analysable into recurring sub-pairings (e.g. words, morphemes etc.), which become the new replicators

---

[7] There is no particular need here to take a position on exactly how best to represent these replicators formally. For example, in the simulation model discussed in this section, vocabulary items and rules were represented formally using a traditional grammar formalism from computational linguistics (a DCG). Other modellers have formalised construction grammar (e.g. Steels 2009), which lends itself well to an analysis in terms of replicators.

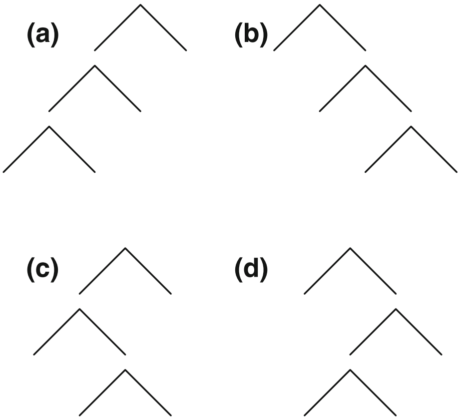


**Fig. 6.2** Branching consistency. Languages vary in the order of categories in a sentence. For example, verbs may precede or follow objects, adpositions may precede or follow noun-phrases and so on. In general this will lead to different patterns of branching when a sentences surface syntax is shown diagrammatically as a tree. It turns out that languages in which trees branch consistently on the left (**a**) or consistently on the right (**b**) are more common than those that branch inconsistently across the syntactic categories of a language (**c** and **d**). In other words, generalisations about the structure of a language can be made above the level of the syntactic category

Although I have only discussed compositionality here, it is possible that the same process can deliver further linguistically relevant generalisations at higher levels. For example, we might consider that the major syntactic categories such as noun, verb and so on are actually generalisations over words and are thus replicators that have even greater chance of being copied once the whole ensemble of other parts of the system is in place. Once major categories are in place, then we can imagine generalisations over these categories will produce the next level of replicator. The fact that, cross-linguistically, the word-order of languages tends to place non-branching categories consistently to the left or to the right across categories (Dryer 1992) is just the kind of higher-level generalisation we might expect (see Fig. 6.2). Finally, simulations like Batali's (2002) suggest that even basic features of language like recursion may be a side-effect of the evolution of linguistic replicators, although much more work on this is needed.

### 6.4.2 *The Emergence of Combinatorial Phonotactics*

Despite the exciting results from models of the emergence of compositionality, there are some remaining concerns. For example, one question is how plausible is the reanalysis of signals? For this to happen, the protolinguistic utterances must have had analysable internal structure, otherwise no fractionation would have been possible. On the other hand, signals must not have been too complex otherwise chance alignments that drive the process would not have occurred. In other words, protolanguage would need to have had segmentable, but non-random signals.

The next candidate for a major transition in the evolution of language delivers just such a set of signals that could feed into the one we discussed in the last section. Oudeyer (2005a,b, 2006, this volume) proposes a mechanism for the emergence of combinatorial phonotactics using a computational model of interacting agents. In this section, I will summarise his model very briefly—for more details please consult the original references, particularly Oudeyer (2006) which presents many interesting simulations and has insightful discussion into their significance.

In Oudeyer's (2006) initial stage, vocalisations of the agents are arbitrary trajectories through acoustic/articulartory space. The replicators in this initial stage are these vocalisations. In the simplest version of the model, vocalisations are simple linear trajectories between a start point and an end point in a one dimensional space. One way of thinking about these trajectories is as a continuous signal moving from one part of the vocal tract to another, for example, but it is important to realise that this is really an idealisation for the purposes of understanding the general process, and Oudeyer (2005b) demonstrates that this simple model can be scaled up to more complex and realistic representations of acoustic/articulatory trajectories.

Agents store representations of these vocalisations in a neural network, which is set up in such a way that start and end points of trajectories can gradually adapt to vocalisations that an agent perceives (by shifting the receptive fields of neurons towards heard vocalisations and pruning neurons that bridge between start and end points if they are rarely activated). In the simulations, a population of agents located randomly in a spatial environment are initialised with different random neural networks. At random times, an agent will produce one of the vocalisations in its repertoire at random, and then both it and the nearest other agent will adjust their neural networks in response to hearing that vocalisation.

In this way, vocalisations are culturally transmitted from one agent to another (and back again) through gradual modifications of the population's neural networks responding to vocal behaviour. This then, is another instance of iterated learning. In terms of the initial population of replicators, the survival of a vocalisation will depend on hearers tuning their neural networks to that specific vocalisation, and then going on to produce that vocalisation in turn.

So, what happens over time in this system? The left panel of Fig. 6.3 shows the repertoire of vocalisations of a typical agent at the beginning of the simulation. Vocalisations here have arbitrary start and end points. The space of vocalisations is essentially continuous, and there are no relationships between vocalisations. Rather like the case of the holistic language in the last section, these are solitary phenotypic replicators.

After repeated interactions between agents in the population, the situation looks very different however. In the right panel of Fig. 6.3, the space of vocalisations has been discretised, with trajectories only starting or ending in particular places. Furthermore, only certain *combinations* of start and end are possible. So, for example, a trajectory ending around 0.2 must start at either 0.5 or around 0.9, whereas a trajectory ending at 0.5 may start from a range of different places.

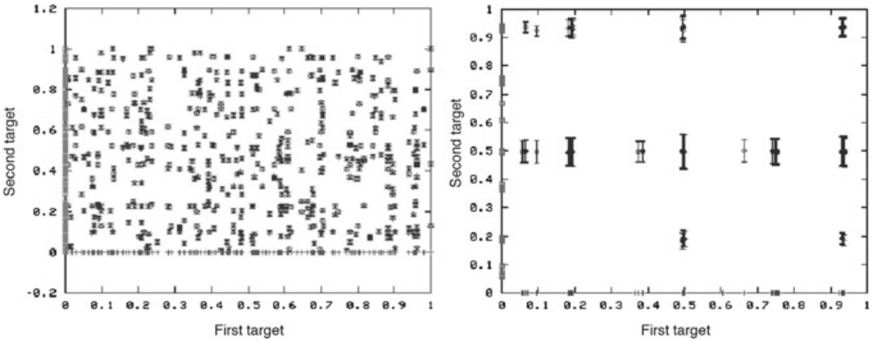


**Fig. 6.3** Before (*left*) and after (*right*) the transition from a set of vocal trajectories to a repertoire exhibiting combinatorial phonotactics. The graphs show results early and late in a simulation. The two axes are start and end points of a vocalisation on an arbitrary continuum. Points plotted within the graph show possible combinations represented in a neural network architecture. See Oudeyer (2005b), from where these graphs are taken, for more information

Arguably, whereas the replicators previously were individual whole continuous trajectories, now the replicators are *phonemes* and *phonotactic constraints*. Just like the words and rules of the previous section, these can no longer replicate independently. The discretised set of possible start and end points for vocalisations forms a mutually-reinforcing system of ensemble replicators. The similarity across vocalisations stabilises the neural representations of the population of agents because it is far easier for a replicator to survive that specifies a particular point as being a valid start or end of a vocalisation than one that specifies a particular trajectory through the space. The former is reinforced by a large set of vocalisations produced by an agent that possesses it, whereas the latter is only reinforced by the particular vocalisation it encodes.[8]

Once again, we see a typical transition involving loss of independent replicability and the emergence of division of labour by the evolution of a new form of replicator that introduces a higher level of linguistic structure. Interestingly, this transition also shows another change in replicator type that Szathmáry (2000) points out is typical in evolution: the change from analog to digital replicators.[9]

---

[8] More recently, similar results have been obtained in a laboratory experiment by Verhoef and colleagues (Verhoef et al. 2011; Verhoef and de Boer 2011). Here, participants learn a system of whistled signals that are transmitted culturally from one participant to the next. Over these artificial generations, discrete sub-signals emerge that are reused and recombined systematically.

[9] Szathmáry (2000) uses the terms *holistic* or *processive* for the former, and *modular* for the latter, but this terminology may be confusing given the different use of the term "holistic" in the discussion of Wray's (1998) protolanguage.

### *6.4.3 The Emergence of a Contentive/Functional Split*

Given the right conditions (e.g. learning machinery, social population and so on), the basic dynamics of iterated learning seem to lead inevitably from arbitrary, continuous vocalisations to segmentable non-random signals. Once paired with structured meanings, these provide the right kind of input to a further process of evolution by iterated learning of compositional syntax out of a holophrastic protolanguage, and then potentially on to major syntactic categories, word order universals and recursion.

It would be a mistake, however, to think the output of the second transition—to compositionality—delivers up the complexity of syntax that we see in human language. If anything, what this gives us is another kind of protolanguage: one more like fully-modern language than the holistic protolanguage of Wray (1998), but still missing much of what characterises the syntax of human language.

One of the striking differences between the kinds of languages that arise in the computational models of the emergence of compositionality, and real human language lies in the nature of the lexicon. Although we see major categories arising in the simulation models these are largely semantically motivated. In many ways the lexicon that emerges is homogeneous in its role in the communication system: each word maps onto part of the content of an utterance and combines with other words to compose a whole message.

Obviously there are a large number of lexical items in real languages that behave in this way. These *contentive* elements, such as nouns, verbs, adjectives and so on form the *open-class* subset of any language's lexicon. However, in language there is another class of lexical items —the *functional* elements, such as determiners, prepositions, markers of case, aspect, tense, number and so on, which form the *closed-class* subset of the lexicon.

This divide between contentive and functional is perhaps the most fundamental in language. Contentive elements are open-ended, phonologically diverse, and meaningful, whereas functional elements are closed-class, have atypical phonology, and serve grammatical roles. There is evidence that these two sets of lexical items are processed differently in the brain, and the distinction between them has an important role in the early acquisition of language (see, Cann 2000, for review of the various aspects of this division).

Functional elements in particular are a crucial part of the syntax of fully-modern human language. They underpin much of what makes the structure of language so interesting and unusual, and in some theories of grammar they have a basic role to play in defining the particular structure of a language and parameterising cross-linguistic variation (Cann 2000).

Arguably, the change from a monolithic lexicon to one with a basic split between contentive and functional is the last major transition that leads us to language as we know it today (see Fig. 6.4). Once again, this fundamental split in the lexicon reflects a very common feature of the evolutionary transitions: division of labour. Replicators in an ensemble will over time become increasingly co-dependent and tend to diversify in function. But what drives this transition in language?

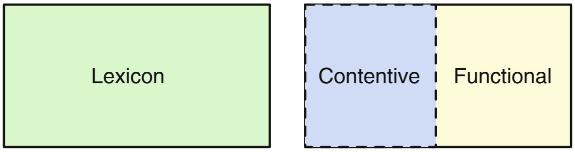


**Fig. 6.4** Before (*left*) and after (*right*) the transition from a monolithic lexicon to a contentive/functional split. Prior to the transition, the lexicon contained contentive elements only distinguished by broadly semantically defined syntactic categories. Afterwards, the lexicon specialises into an open class of elements carrying content, and a closed class of those with a grammatical function

At this stage, we do not have a model that is as simple as the ones described above for the other transitions, but some computational modellers are now turning to the problem of the features like case-coding (which, when marked overtly, involves closed-class morphemes) and how it may emerge from the dynamics of socially interacting communicative agents (e.g., Steels 2009).

Furthermore, it is likely we can learn much from studies of the ongoing emergence in languages of functional elements out of previously contentive items through the process of *grammaticalisation* (e.g., Hopper and Traugott 1993). Ultimately, grammaticalisation is driven by iterated learning in that it is a product of the cultural transmission of language. So, we can say with some confidence that the phenotypic replication of language can lead to the emergence of new replicators with a specifically syntactic function because we can actually observe this process in languages today. What remains an open question at this stage is if we can demonstrate that the same process lead to the emergence of the very first functional elements in the longer term emergence of language. In other words, did the process of grammaticalisation lead to the original split in the lexicon in the first place? The answer to this question must await further modelling.

## 6.5 Summary and Conclusions

In this paper, I have tried to develop parallels between ongoing research into the cultural evolution of language and the generalisations made by Maynard Smith and Szathmáry (1995) about transitions in evolutionary systems more broadly. I have argued that ongoing change in language is driven by phenotypic replication through iterated learning, and that this same process also delivers major transitions in the emergence of language which justify us calling this non-biological process "evolution" rather than merely "change".

The transitions I have surveyed all involve the emergence of key aspects of linguistic structure out of a previous state in which that structure was not present. They have suggestive similarities with each other, and with evolutionary transitions in general. In particular, they involve a shift (at various levels) from solitary replicators with independent replicability, to ensemble replicators that lose independence and lead to diversification and division of labour as a result.

Although these transitions are purely cultural, in the sense that they can arise out of the process of repeated transmission of language through production, observation, and learning even among biologically identical agents, this does not mean that there is no role for biological evolution. For example, we need to consider what the conditions must be in order for iterated learning to happen at all. Species-specific social and cognitive adaptations must be involved to make us such avid transmitters of linguistic replicators. Secondly, the emergence and ongoing evolution of a structured linguistic system is likely to have a knock-on effect on the biological evolution of its carriers (us). It is entirely reasonable to expect co-evolution of the culturally transmitted linguistic system and the biologically determined apparatus for supporting it (see e.g. Smith and Kirby 2008; Dediu, this volume; Zuidema, this volume, for discussion).

Maynard Smith and Szathmáry's (1995) work provides a rich framework for thinking about replication. They themselves identified the importance of language in this light, but language is a new system of replication in more than one sense: it is both an enabler of cultural replicators with unlimited heredity, and also a new kind of evolutionary system itself. Iterated learning is the process of linguistic transmission, and it drives both language change and the transitions to qualitatively new kinds of linguistic system. By seeing language as an evolutionary system, the biggest payoff we get may be the ability to take biologists' insights into the evolution of life and apply them to the evolution of language.

## References

Andersen, H.: Abductive and deductive change. Language **40**, 765–793 (1973)

Arbib, M.A.: From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. Behav. Brain Sci. **28**, 105–124 (2005)

Batali, J.: The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In: Briscoe, E. (ed.) Linguistic Evolution Through Language Acquisition: Formal and Computational Models, pp. 111–172. Cambridge University Press, Cambridge (2002)

Bickerton, D.: Language and Species. University of Chicago Press, Chicago (1990)

Bickerton, D.: Language evolution: A brief guide for linguists. Lingua **117**, 510–526 (2007)

Blevins, J.: Evolutionary Phonology. The Emergence of Sound Patterns. Cambridge University Press, Cambridge (2004)

Brighton, H., Smith, K., Kirby, S.: Language as an evolutionary system. Phys. Life Rev. **2**, 177–226 (2005)

Briscoe, E.: Grammatical acquisition: inductive bias and coevolution of language and the language acquisition device. Language **76**, 245–296 (2000)

Cann, R.: Functional versus lexical: a cognitive dichotomy. In: Borsley, R.D. (ed.) The Nature and Function of Syntactic Categories, pp. 37–78. Academic Press, New York (2000)

Croft, W.: Explaining Language Change: An Evolutionary Approach. Longman, London (2000)

Dryer, M.: The Greenbergian word order correlations. Language **68**, 81–138 (1992)

Francis, N., Kucera, H.: Frequency Analysis of English Usage: Lexicon and Grammar. Houghton Mifflin, Boston (1982)

Hopper, P.J., Traugott, E.C.: Grammaticalization. Cambridge University Press, Cambridge (1993)

Hurford, J.R.: Nativist and functional explanations in language acquisition. In: Roca, I.M. (ed.) Logical Issues in Language Acquisition, pp. 85–136. Foris, Dordrecht (1990)

Jackendoff, R.: Possible stages in the evolution of the language capacity. Trends Cogn. Sci. **3**, 272–279 (1999)

Kalish, M.L., Griffiths, T.L., Lewandowsky, S.: Iterated learning: intergenerational knowledge transmission reveals inductive biases. Psychon. Bull. Rev. **14**, 288–294 (2007)

Kirby, S.: Function, selection and innateness: the emergence of language universals. Oxford University Press, Oxford (1999)

Kirby, S.: Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In: Knight, C., Studdert-Kennedy, M., Hurford, J. (eds.) The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form, pp. 303–323. Cambridge University Press, Cambridge (2000)

Kirby, S.: Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Trans. Evol. Comput. **5**, 102–110 (2001)

Kirby, S.: The evolution of meaning-space structure through iterated learning. In: Lyon, C., Nehaniv, C., Cangelosi, A. (eds.) Emergence of Communication and Language, pp. 253–268. Springer, London (2007)

Kirby, S., Cornish, H., Smith, K.: Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. Proc. Nat. Acad. Sci. USA **105**, 10681–10686 (2008)

Kirby, S., Dowman, M., Griffiths, T.L.: Innateness and culture in the evolution of language. Proc. Nat. Acad. Sci. USA **104**, 5241–5245 (2007)

Kirby, S., Hurford, J.: The emergence of linguistic structure: an overview of the iterated learning model. In: Cangelosi, A., Parisi, D. (eds.) Simulating the Evolution of Language, pp. 121–148. Springer, London (2002)

Kirby, S., Smith, K., Brighton, H.: From UG to universals: linguistic adaptation through iterated learning. Stud. Lang. **28**, 587–607 (2004)

Maynard Smith, J., Szathmáry, E.: The major transitions in evolution. Oxford University Press, Oxford (1995)

McMahon, A.: Understanding Language Change. Cambridge University Press, Cambridge (1994)

Mithen, S.: The Singing Neanderthals: The Origins of Music, Language, Mind and Body. Weidenfeld and Nicolson, London (2005)

Newmeyer, F.J.: Uniformitarian assumptions and language evolution research. In: Wray, A. (ed.) The Transition to Language, pp. 359–375. Oxford University Press, Oxford (2002)

Oudeyer, P.-Y.: From holistic to discrete speech sounds: The blind snow-flake maker hypothesis. In: Tallerman, M. (ed.) Language Origins: Perspectives on Evolution, pp. 68–99. Oxford University Press, Oxford (2005a)

Oudeyer, P.-Y.: The self-organization of combinatoriality and phonotactics in vocalization systems. Connection Sci. **17**, 325–341 (2005b)

Oudeyer, P.-Y.: Self-organization in the Evolution of Speech. Oxford University Press, Oxford (2006)

Pinker, S., Bloom, P.: Natural language and natural selection. Behav. Brain Sci. **13**, 707–784 (1990)

Ritt, N.: Selfish sounds and linguistic evolution: A Darwinian approach to language change. Cambridge University Press, Cambridge (2004)

Rosenbach, A., Jäger, G.: Priming and unidirectional language change. Theor. Linguist. **34**, 85–113 (2008)

Smith, K., Kirby, S.: Cultural evolution: implications for understanding the human language faculty and its evolution. Philos. Trans. Roy. Soc. B **363**, 3591–3603 (2008)

Steels, L.: Cognition and social dynamics play a major role in the formation of grammar. In: Bickerton, D., Szathmáry, E. (eds.) Biological Foundations and Origin Syntax, pp. 345–368. MIT Press, Cambridge (2009)

Szathmáry, E.: The evolution of replicators. Philos. Trans. Roy. Soc. B **355**, 1669–1676 (2000)

Tallerman, M.: Did our ancestors speak a holistic protolanguage? Lingua **117**, 579–604 (2007)

Verhoef, T., de Boer, B.: Cultural emergence of feature economy in an artificial whistled language. In: Zee, E., Lee, W. (eds.) Proceedings of the 17th International Congress of Phonetic Sciences, pp. 2066–2069. City University of Hong Kong, Hong Kong (2011)

Verhoef, T., Kirby, S., Padden, C.: Cultural emergence of combinatorial structure in an artificial whistled language. In: Carlson, L., Hölscher, C., Shipley, T. (eds.) Proceedings of the 33rd Annual Conference of the Cognitive Science Society, pp. 483–488. Cognitive Science Society, Austin (2011)

Wray, A.: Protolanguage as a holistic system for social interaction. Lang. Commun. **18**, 47–67 (1998)

Zuidema, W.: The Major Transitions in the Evolution of Language. Ph.D. thesis, University of Edinburgh (2005)

# Chapter 7
# Genes: Interactions with Language on Three Levels—Inter-Individual Variation, Historical Correlations and Genetic Biasing

**Dan Dediu**

**Abstract** The complex inter-relationships between genetics and linguistics encompass all four scales highlighted by the contributions to this book and, together with cultural transmission, the genetics of language holds the promise to offer a unitary understanding of this fascinating phenomenon. There are inter-individual differences in genetic makeup which contribute to the obvious fact that we are not identical in the way we understand and use language and, by studying them, we will be able to both better treat and enhance ourselves. There are correlations between the genetic configuration of human groups and their languages, reflecting the historical processes shaping them, and there also seem to exist genes which can influence some characteristics of language, biasing it towards or against certain states by altering the way language is transmitted across generations. Besides the joys of pure knowledge, the understanding of these three aspects of genetics relevant to language will potentially trigger advances in medicine, linguistics, psychology or the understanding of our own past and, last but not least, a profound change in the way we regard one of the emblems of being human: our capacity for language.

## 7.1 Introduction

The phenomenon of language can be explored at many scales, as so eloquently illustrated by the present volume, ranging, in time, from the human life extending over tens of years up to evolutionary timescales of (tens of) millions of years and, in social complexity, from the single individual up to interconnected societies at the continental scale. However, these levels cannot and must not be treated separately

---

D. Dediu (✉)
Language and Genetics Department, Max Planck Institute for Psycholinguistics,
Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands
e-mail: Dan.Dediu@mpi.nl

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_7,

except for very clearly set pedagogical goals, as they interact in complex ways, each shaping all the others and being shaped by them.

As clearly shown by the previous contributions, the two main factors conferring unity to language across all these diverse scales are represented by the *cultural transmission of language across generations* (see also Croft, Kirby and Oudeyer, this volume) and the *genetic bases* of the capacity for language (Zuidema, this volume). As expected, these two factors interact at all these levels to produce the bewildering complexity and diversity of language and languages.

This chapter will explore the main aspects of these interactions and will introduce some fundamental concepts, methods and findings by focusing both on well-known and accepted cases as well as recent and provocative hypotheses.

## 7.2 Being Different

It is obvious that we are different in many respects, including, for example, height, hair color, memory, mathematical abilities or language and speech. As some of us are shorter and some taller, some of us, despite normal or above-average intelligence, struggle with words while others have an almost artistic way of speaking and writing. Some of us speak clearly, articulating in a distinctive manner while others mumble at the limit of intelligibility. But why are we different in the first place?

One well-known reason is the experience of *different environments* by different people: eating well will allow you to grow taller than starving during childhood, being encouraged to read early by your parents will increase your chances of succeeding at school and exercising regularly will make your muscles stronger and bigger.

But it is equally well-known that some differences are down to innate factors, to *genetic differences* between individuals: those unfortunate enough to have a supplementary chromosome 21 will have the many problems associated with Down's syndrome (Plomin et al. 2001) while being born to taller parents will probably make you taller (Weedon et al. 2007).

However, the distinction between these two causes is not clear-cut and they are not somehow opposed, as the old question of "nature versus nurture" would imply. In fact, most inter-individual variability results from a combination of genetic and environmental differences, usually interacting in complex ways. A fascinating illustration of this is given by the recent study of the relationship between IQ and breastfeeding (Caspi et al. 2007), where it was found that breastfeeding tended to increase children's IQ but only if they had a certain allele of the *FADS2* gene. Alleles are possible variants of a given gene (Jobling et al. 2004) and this particular one allows its carriers to be able to process the human-specific fatty acids in mothers' milk, which are seemingly involved in brain development (Caspi et al. 2007).

What this shows is that if there happens to be the right nurture (breast feeding) but not the right nature (*FADS2* allele), or the other way around, there is no increase in the IQ. *Both* nature *and* nurture must be there and work together to bring about the inter-individual phenotypic variation (Ladd et al. 2008).

Understanding these inter-individual differences will allow us not only to better conceptualize human nature but also to both treat those characteristics perceived as pathological and enhance those seen as desirable.

### *7.2.1 Heritability*

A first step is trying to somehow disentangle the relative contributions of genetic and environmental differences to inter-individual phenotypic variation in a given population. A popular way of doing this is represented by *heritability*, defined as the proportion of the phenotypic variation accounted for by genetic variation between individuals in a certain population (Plomin et al. 2001; Halliburton 2004) and reflecting the expected degree of similarity between parents and offspring on genetic grounds.

Heritability can be estimated in many ways through, for example, adoption, family aggregation and pedigree studies, and most commonly, by comparing *monozygotic* (MZ) and *dizygotic* (DZ) twins (Stromswold 2001; Plomin et al. 2001). While DZ (or fraternal) twins form from two separate ova fertilized by two different spermatozoa and are genetically as related as any two regular siblings, MZ (or identical) twins form from a single ovum fertilized by a single spermatozoon which then divides into two embryos, providing them with identical genomes (barring new mutations). By comparing the similarities between MZ ($r_{MZ}$) and DZ ($r_{DZ}$) twins, it is possible to estimate the heritability of the considered phenotypic trait in the target population as $h^2 = 2(r_{MZ} - r_{DZ})$, because presumably any increase in similarity between MZ relative to DZ twins is due to their greater (on average, two times bigger) genetic similarity (Stromswold 2001; Plomin et al. 2001).

However, it must be noted that there are a number of caveats concerning heritability estimates (see, for example, Charney 2012) including the fact that they are meaningful only for phenotypes which show variation in the target population, they are higher in homogeneous compared to variable environments, they differ between populations and can vary with age (Bishop 2003; Plomin et al. 2001). Therefore, heritability is not an absolute measure of some kind of intrinsic genetic contribution to the trait but a *relative* estimate fundamentally reflecting the specific population and environmental conditions.

Heritability estimates for aspects of speech and language generally point to the involvement of genes in explaining the inter-individual variation.[1] For language and speech disorders—by far the best studied to date—the heritabilities of their various aspects seems to be high (typically $h^2 > 0.50$) with differing heritabilities for different aspects of language (Stromswold 2001; Bonneau et al. 2004; Fisher et al. 2003; Felsenfeld 2002; Plomin et al. 2002; Plomin and Kovas 2005). For example, liability to stuttering is highly heritable ($h^2 \approx 0.70$; Felsenfeld 2002) and the heritability of

---

[1] Probably the best review of the heritability of aspects of speech and language remains Stromswold (2001).

SLI (*Specific Language Impairment*, a complex and controversial language pathology; Bishop 2003, OMIM[2] 602081) is also high (Bishop 2003, cites $h^2 \approx 0.76$, 0.48 and 0.54, while Bonneau et al. (2004) give $h^2 \approx 0.70$ and 0.73, depending on the study).

Likewise, normal inter-individual variation in aspects of speech and language seems to also have a genetic component, ranging from a very low $h^2 \approx 0.02$ for expressive vocabulary at 14 months to $h^2 \approx 0.38$ at 24 months and $h^2 \approx 0.72$ for WISC-R vocabulary (Stromswold 2001). As a comparison (Halliburton, 2004, p. 540), the heritability of height in humans is $h^2 \approx 0.65$, of schizophrenia $h^2 \approx 0.70$ and of fingerprint ridge count $h^2 \approx 0.92$.

### *7.2.2 Linkage Studies and the FOXP2 Gene*

Having established that a phenotypic trait is heritable is not, of course, the end of the story and the next interesting question concerns the nature of the gene(s) involved and the specific mechanisms bridging the causal gap between genetic and phenotypic variation. A first step in this direction is knowing where the gene(s) might be, on which of the 23 pairs of chromosomes (Jobling et al. 2004) and in which specific position.

The main idea is to use the fact that genes on the same chromosome are not transmitted independently, and that this *linkage* between them generally increases the closer together they are, offering a way to build *genetic maps*. A *linkage study* exploits the association between the trait under consideration and various *genetic markers* across generations in a family tree, on the assumption that the genes(s) responsible and the linked markers will tend to be inherited together (Jobling et al. 2004; Halliburton 2004; Gibson and Gruen 2008; for more details see Box 1). This design was instrumental to the discovery of *FOXP2*, probably the most discussed gene when it comes to speech and language.

**Box 1: Genetic linkage and linkage studies**

Two *loci* or "positions" on different chromosomes or on the same chromosome but far apart are transmitted independently to the offspring, but two loci near to one another on the same chromosome are not. Chromosomes are of two types: *autosomes* come in identical pairs while the two *sex chromosomes* come in an identical pair in women (XX) but are different in men (XY) (Jobling et al. 2004; Halliburton 2004). Let's consider two loci on an autosome, *1* and *2*; therefore, each individual will have two alleles for each locus. Let's consider an example individual *heterozygous* for both loci, carrying two different alleles for each of them (as opposed to *homozygous* loci for which (s)he would carry the same

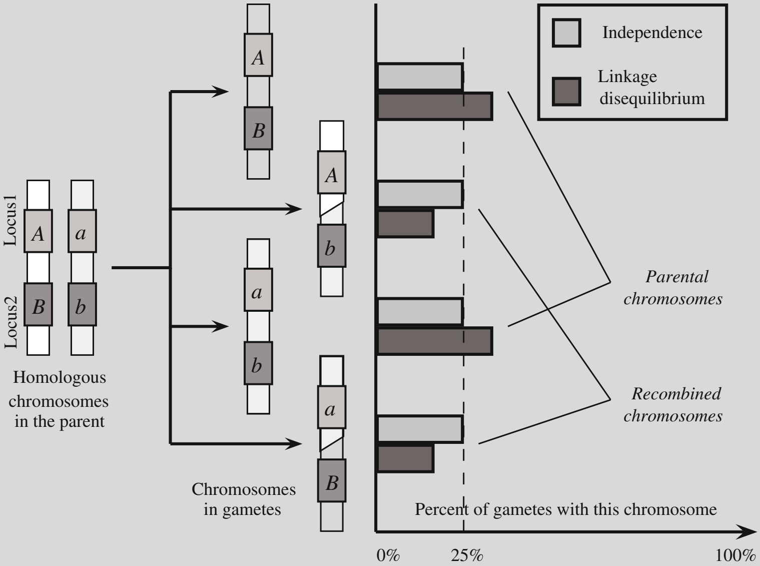
---

[2] *Online Mendelian Inheritance in Man*, an online comprehensive database of genetic disorders freely browsable at http://www.ncbi.nlm.nih.gov/omim.

allele): *A* at locus *1* and *B* at locus *2* on one chromosome and *a* at locus *1* and *b* at locus *2* on the other chromosome of the pair.

If loci *1* and *2* are transmitted independently, then this individual's offspring would have the same probability of inheriting allele *A* (locus *1*) and allele *B* (locus *2*)—*AB* for short—, *Ab*, *aB* or *ab* from this parent and something else (and irrelevant for now) from the other parent. However, if loci *1* and *2* are on the same chromosome and linked, then the probability of the offspring inheriting *AB* or *ab* will be greater than that of inheriting the *recombinants Ab* or *aB*. The strength of the linkage is given by this difference in probabilities and is measured in *centiMorgans*, cM, representing a 1% probability of recombination between two loci in a generation (Jobling et al. 2004), with independence for separations bigger than 50 cM. Very roughly, there is a linear relationship between strength of linkage and physical distance between the loci, with 1 cM ≈ 1 Mb at the scale of the entire human genome, but, as always, things are really not that simple (Jobling et al. 2004; Speed and Zhao 2007). Physical distances are measured in units of the fundamental building blocks of the DNA molecule, the *base pair*, bp, and multiples, like millions of base pairs, Mb (Jobling et al. 2004).

The figure below illustrates the concept of linkage disequilibrium: the two *homologous* chromosomes contain alleles *A* and *a* at locus *1* and alleles *B* and *b* at locus *2* in the parental generation. If these two loci are independent (light gray bars), then it is expected that all four possible combinations of alleles to be present in the same proportion (25% each) in the gametes (sex cells, ova and spermatozoa) producing the next generation. However, if there is linkage between them (dark gray bars), the *parental* combinations (*AB* and *ab*) are expected more frequently than the *recombined* ones (*Ab* and *aB*).



The British "KE" family has a very interesting three-generations pedigree, with half the members (15 out of 31) affected by a complex pathology, involving speech and language (Hurst et al. 1990; Gopnik and Crago 1991), classified as *developmental verbal dyspraxia* (OMIM 602081) and included in the SLI category. The exact phenotype is very complex (Fisher et al. 2003; Lai et al. 2001; Vargha-Khadem et al. 1998; Marcus and Fisher 2003) and includes *articulatory problems*, due to troubles with coordinating complex oro-facial movements, *cognitive impairments* and *language impairments*, affecting spoken (expressive and receptive) and written language, as well as the comprehension and production components of grammar (understanding complex sentences, inflectional and derivational morphology).

This pathology follows a *dominant autosomal* mode of inheritance and is *fully penetrant*, meaning that the gene (named *SPCH1*, OMIM 602081) is located on an autosome (chromosome 7; Fisher et al. 1998) and an individual carrying even a single mutation of the gene will manifest the pathology (Hurst et al. 1990; Bishop 2003; Lai et al. 2001). The gene was finally identified as *FOXP2* (Lai et al. 2000, 2001), a member of the numerous *Forkhead box* (Fox) family of genes which act as *transcription regulators*, controlling the expression of other genes (Scharff and White 2004). Recently, Vernes et al. (2008) have shown that *FOXP2* downregulates the gene *CNTNAP2* which was previously involved in language delay in autism Alarcón et al. (2008) and is strongly associated with non-word repetition (Vernes et al. 2008). The specific mutation affecting the "KE" family[3] disrupts the regulatory function of this gene, leading to a cascade of events resulting in the pathological phenotype.

It is not entirely clear what *FOXP2* does in humans, but it is certainly *not* "the language gene", as sometimes boasted in the media. As always, the story is much more fascinating, as it turns out that *FOXP2* has complex roles in neural (and not only) development, with many targets in the developing brain (Spiteri et al. 2007) and affecting brain structure and functioning (Vargha-Khadem et al. 1998; Watkins et al. 2002; Liégeois et al. 2003). This picture is complicated by studies involving birds with learned and non-learned song, echolocating bats and vocal-learning non-human mammals (Li et al. 2007; Webb and Zhang 2005; Teramitsu et al. 2004; Scharff and Haesler 2005; Haesler et al. 2004; Shu et al. 2005), including the recent engineering of mice having the human-specific gene (Enard et al. 2009), suggesting a complex role for *FoxP2*.

Evolutionary studies seem to imply that, overall, *FoxP2* is very conserved across taxa (Enard et al. 2002) but that there seem to exist two "human-specific" mutations, fixed in the human population and under strong selection (Enard et al. 2002; Zhang et al. 2002). It is, however, unclear what exactly these mutations have been selected for, nor when this has happened, with earlier estimates (Enard et al. 2002; Zhang et al. 2002) around 100–200 thousand years ago, apparently coinciding with the emergence of anatomically modern *Homo sapiens*. Recently, Krause et al. (2007) found that these "human-specific" mutations are shared with Neandertals and conclude that the human *FOXP2* originated and was selected before the split between us and them,

---

[3] This is not the only *deleterious* mutation affecting *FOXP2*: for example, two different mutations in two unrelated individuals produce similar pathologies (Lai et al. 2001; MacDermot et al. 2005).

more than 300 thousand years ago, but Coop et al. (2008) read the data as suggesting low rates of gene flow between modern humans and Neadertals. Nevertheless, the publication of the Neandertal draft genome (Green et al. 2010) confirmed that we share this variant of *FOXP2*, most probably present in the last common ancestor.

In conclusion, *FOXP2*, the first gene specifically identified for its role in language and speech, vividly shows how complex the questions concerning the genetic influences on this human-specific phenotype are (Fisher and Scharff 2009). It is to be expected that there is no "language gene" which would suddenly have allowed a lucky mutant somewhere down the human line to speak, nor even a small set of such genes. However, the unravelling of these complex interactions between genes and environmental factors will undoubtedly help us better understand what language really is.

### *7.2.3 Association Studies and Future Directions*

In *association studies*, a group of unrelated individuals displaying the trait of interest and a group of individuals without it are compared, looking for those genetic markers which correlate with the presence of the trait (Jobling et al. 2004; Plomin et al. 2001). If, based on previous research, one or more *candidate genes* are hypothesized to be involved in the target phenotype, the search will be constrained to variants of these genes. In contrast, the "hypothesis-free" approach of *genome-wide association studies* (GWAS) consists of looking for correlations between the phenotype of interest and genetic markers (such as SNPs[4]) covering the entire genome in the hope that very strong correlations are due to real genetic influences (Pearson and Manolio 2008; McCarthy et al. 2008).

The GWAS are very promising and currently intensively used in topics ranging from the genetic bases of cancer or diabetes (The Wellcome Trust Case Control Consortium 2007) to schizophrenia and bipolar disorder (Maher et al. 2008; Lewis et al. 2003; Segurado et al. 2003). However, there are a number of issues mainly concerning the fact that when hundreds of thousands of genetic markers are used the chance of "discovering" non-existent correlations with the trait (*false positives*) is huge and techniques for *multiple testing correction* (which lower the threshold of statistical significance from the "standard" 0.05 to much smaller values, like $10^{-8}$) must be employed (Speed and Zhao 2007) together with caution in interpreting the results of the study (Pearson and Manolio 2008; McCarthy et al. 2008). These and other difficulties have resulted in a low degree of replication of the early findings and the recent publication of stringent guidelines (Chanock et al. 2007). Concerning specifically language and speech, it is to be expected that large GWAS comparing both pathological to normal phenotypes as well as the range of normal variation will

---

[4] Single Nucleotide Polymorphisms, one of the simplest types of genetic variation where a single DNA "letter" (nucleotide) varies between individuals and is present at relatively high frequencies.

be areas of intense research in the near future, but they will require better definition and measurement of interesting phenotypes.

Another very promising avenue is represented by *copy-number variation* (CNVs), which represent inter-individual variations in the number of times a certain region of the genome is present (McCarroll and Altshuler 2007; Redon et al. 2006). Due to the fact that, as opposed to SNPs, which have only two alleles, CNVs can have a number of possible repeats, they can alter the *dosage* of gene products, a plausible mechanism behind complex phenotypic variation (McCarroll and Altshuler 2007) and very promising for the study of language and speech.

*Epigenetic* processes (Jiang et al. 2008), whereby non-genetic marks (chemical changes of the nucleotides or the proteins associated with the DNA) alter the expression of genes, seem to play an important role not only in development and disease (Hirst and Marra 2009) but also in normal neuronal plasticity (Fagiolini et al. 2009). These mechanisms also seem to have evolutionary relevance because they are transmitted across multiple generations (Youngson and Whitelaw 2008; Jablonka and Raz 2009). They subtend *parent of origin effects* where the same allele has different effects depending on the parent it has been inherited from, as has been suggested for the gene *LRRTM1* involved in handedness and schizophrenia (Francks et al. 2007). In the future, epigenetics will probably play an important role in understanding language development and evolution.

## 7.3 Changing Together

As briefly reviewed above, it is clear that human individuals differ from each other at the genetic level (with the possible exception of identical twins) by carrying different alleles of the same gene. But do populations also differ in what concerns their genes and, if so, in what ways?

It is generally agreed that humans are genetically very uniform when compared with other mammals (Jobling et al. 2004; Relethford 2001; Templeton 1998), and that the existing genetic variation is apportioned mostly (approximately 85%) *within* populations (between the individuals from the same group), but there is some variation (approximately 15%) due to *inter-population* genetic differences (*between* populations) (Jobling et al. 2004; Lewontin 1972). These inter-group differences are in great majority due to different *allele frequencies* in one group as compared to another and not to group-specific genetic variants.

When considering together many genetic loci, the variation in the frequencies of their alleles between populations combine and allows the study of the genetic structure of humans at the population level, the reconstruction of demographic history and even the inference of the population of origin for a given individual (Rosenberg et al. 2002; Bamshad et al. 2003; Falush et al. 2003). However, it is still debated if this variation between populations is gradual or if there are some continental-level boundaries, where the changes are abrupt (Barbujani and Belle 2006). Nevertheless, it must be highlighted that the existence of such genetic structure offers no support to racist theories (Jobling et al. 2004; Banton 1998).

Having established that there are (small but detectable) differences between human populations, the next question concerns their origin and dynamics. Some are caused by *genes under selection*, whereby variants of a gene are (dis-)favored in certain environments, leading to differences between populations experiencing these different environmental pressures (Halliburton 2004; Jobling et al. 2004; Skelton 1993). Well-known examples are represented by *Sickle cell disease*, where a recessive autosomal mutation affecting haemoglobin confers resistance to heterozygotes against malaria but inflicts sickle cell anemia to homozygotes (*heterozygote advantage*; see Box 2), and *lactase persistence* where the ability to digest milk is retained into adulthood in dairying populations (Jobling et al. 2004; see Box 3).

### Box 2: Sickle cell anemia: natural selection and heterozygote advantage

The red blood cells' function is to transport oxygen from the lungs throughout the body and, to this end, they contain vast quantities of *haemoglobin*, a complex iron-containing protein. In some people, a *point mutation* in the *HBB* gene on chromosome 11 (this mutation is known as *HbS*), results in the substitution of one *amino acid* for another and the production of a malfunctioning haemoglobin molecule. When an individual is homozygous for this mutation, it will produce only malfunctioning haemoglobin, manifested as *sickle cell anemia*, a complex disease which usually kills before puberty (Guégan et al. 2007, see OMIM 603903 for details).

However, a heterozygous *carrier* will have a more or less normal phenotype, as the normal (*wild-type*, denoted *HbA*) copy of the gene will produce enough functioning red blood cells to compensate for the mutated ones. Interestingly, these heterozygous individuals actually suffer less from *malaria*, a deadly infectious disease caused by a parasite of the red blood cells (genus *Plasmodium*, the most dangerous being the *falciparum* species), transmitted through mosquito (genus *Anopheles*) bites.

Thus, the *HbS* mutation, on one hand, is very deleterious in homozygous form (*HbS-HbS*) by causing sickle cell anemia, but is better than the normal homozygous (*HbA-HbA*) when in heterozygous form (*HbS-HbA*) *in environments* where malaria is *endemic* (Guégan et al. 2007; Cavalli-Sforza et al. 1994). This phenomenon is called *heterozygote advantage* and explains why some alleles deleterious in homozygous form are, nevertheless, maintained at relatively high frequencies, in the population (Cavalli-Sforza et al. 1994; Skelton 1993): in a sense, this is a cost some have to pay for the benefit of many.

**Box 3: Lactose tolerance: culture impacting on genes**

One of the defining characteristics of mammals is represented by the production of milk to feed their young. *Lactose* is one of the most important sugars in milk and digesting it is ensured by the enzyme *lactase*. After weaning, lactase production gradually decreases, resulting in a total inability or greatly reduced capacity to digest lactose in adults. However, in some human populations, the proportion of individuals *retaining* lactose tolerance is very high and the molecular mechanism seems to involve changes in a non-coding region of the *MCM6* gene, which upregulates the *LCT* gene responsible for the production of the lactase enzyme (Ingram and Swallow 2008, OMIM 223100).

Interestingly, the capacity to digest lactose is retained in populations with a history of milking, reaching high frequencies in North–West Europe and some parts of Africa (see map in Ingram and Swallow 2008, p. 1128) and it seems that there were several independent origins of this capacity. Therefore, the mutations allowing adults to digest milk were selected in those groups with a culture favoring milking, clearly showing that culture can lead to new selective pressures on genes (Ingram and Swallow 2008; Cavalli-Sforza et al. 1994).

However, this dramatic example is just the simplest possible type of feedback from culture onto genes, with more complex cases being discussed in the *niche construction* (Odling-Smee et al. 2003) or *cultural evolution* (Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981) literatures.

But most are due to *selectively neutral alleles*, whose frequencies change through *random drift*, going up or down in a given population as chance dictates (Jobling et al. 2004; Halliburton 2004). When this is the case, the genetic differences between populations reflect their history of *splits* and *admixture*, as genetically similar populations share a more recent common ancestor and/or have recently been in intense contact, while more dissimilar populations have diverged further in the past (Cavalli-Sforza et al. 1994; Jobling et al. 2004).

When populations split, they carry with them not only their genes, but also their languages, so that the two will diverge in parallel ways, albeit at very different rates. Also, when populations meet, they not only exchange genes through inter-marriages but their languages converge through borrowing. Therefore, it looks as if it would be possible to find *correlations between genetic and linguistic diversities* across populations due to common processes making them change together (Cavalli-Sforza et al. 1994; Jobling et al. 2004). However, such correlations are far from perfect as languages and genes can become decoupled through *language shift* (or language replacement, where one population replaces its original language with a new one) or *elite dominance* (a small elite, military, religious, etc., manages to impose its language and culture but not its genes), for example (Cavalli-Sforza et al. 1994).

The seminal work of Cavalli-Sforza et al. (1994) introduced the methodology used to find and investigate such correlations, including the building of population trees—

based on the *genetic distances* between such populations—and their comparison with *historical linguistic* trees in the attempt to capture the historical processes of descent with modification at the demographic and linguistic levels. However, this approach encounters a number of issues, like the usage of contentious linguistic classifications, the problematic concept of "population" and population trees, and even the method of comparing linguistic and population trees.

This early literature tended to relay on contentious linguistic classifications, like Ruhlen's (Ruhlen 1991, 1994) or Greenberg's (Greenberg 1971, 1987, 1998), which transcend the accepted *historical linguistic comparative method* (McMahon and McMahon 2005; Matisoff 1990; Sims-Williams 1998), producing such debated constructs like "Indo-Pacific" (Dixon 1997), "Australian" (Dixon 2001; Dench 2001) and especially "Amerind" (Bateman et al. 1990; Bolnick et al. 2004; Matisoff 1990; Sims-Williams 1998). Moreover, this work uses even more contentious "macrofamilies" like "Nostratic" and "Eurasiatic", which purport to represent a level beyond linguistic families, but which seem to lack any linguistic validity (Campbell 1999; McMahon and McMahon 2005; Trask 1999; Appleyard 1999). The main drive behind this systematic and voluntary error, which still persists despite vehement critique by linguistics (see, for a recent example, Belle and Barbujani 2007), is represented by the perceived need to have a fully hierarchical classification of languages to feed into the comparison, no matter how complex the linguistic reality actually is (see, for example, the clear statement in Cavalli-Sforza 2000, p. 139).

Another critique is represented by the "populations" used, which are generally sampling units based on language criteria and cultural significance (McMahon 2004) and which are not equivalent in the sense that, for example, the Hadza of Tanzania (population of $\approx$1000), South Chinese (population $\approx$500 million) and French (population $\approx$ 60 million) cannot be treated as comparable (MacEachern 2000). Moreover, it is assumed that there is a one-to-one relationship between tribal and language names, leading to a the concept of "language as a powerful ethnic guidebook" (Cavalli-Sforza 1994, p. 23), but this is totally misleading (Sims-Williams 1998; McMahon 2004; MacEachern 2000). However, this sampling procedure is here to stay and the best we can do is try to minimize the impact it has on studies of correlations between languages and genes.

The third, and probably the most criticized aspect, concerns the methods for comparing linguistic and population classifications. Tree classifications of populations were produced from genetic distances, resulting in *phenetic* (Skelton 1993) populations trees, which cluster populations based on their similarity, but assumed (contentiously; Bateman et al. 1990; MacEachern 2000) to represent the actual genealogical history of those populations. These trees were then compared to the linguistic classifications, apparently supporting a high concordance between the two (Cavalli-Sforza et al. 1994), but this appearance seems totally due to the method used (Bateman et al. 1990), as it exploited the ability of branches in a tree to freely rotate in order to artificially increase the visual similarity between the two compared trees.

In recent work, however, there is a tendency to move away from tree-based methods and controversial linguistic classifications, using, for example, *Mantel correlations* (correlations adapted for the case of distance matrices where the "observations" are not independent) between genetic and linguistic distances (e.g., Poloni et al. 1997; Rosser et al. 2000), *boundary detection* (e.g., Manni et al. 2004) or *AMOVA* (Analysis of Molecular Variance; Excoffier et al. 1992).

Probably the best-known example of language-genes correlation hypothesis is represented by the *language/farming co-dispersal* theory, which tries to explain the modern distribution of a major component of genetic and linguistic diversity as being the result of the spread of farming from its primary origins, carrying together the genes and languages of the original farmers (Jobling et al. 2004; Cavalli-Sforza et al. 1994; Mithen 2003; Diamond 1998). To focus on the *Indo-European* family of languages (Fortson 2004), it is postulated that its spread started around 10,000 years ago when agriculturalists from somewhere around Anatolia expanded westward and eastward (Diamond 1997, 1998; Diamond and Bellwood 2003; Cavalli-Sforza et al. 1994; Renfrew 2002). This theory is highly contentious on linguistic (Mallory 1991) as well as genetic and archaeological (Jobling et al. 2004; Sims-Williams 1998) grounds, and the SE to NW genetic gradients discovered across Europe (Cavalli-Sforza et al. 1994), initially taken to support such a demic diffusion, seem to reflect a far more complex history (Jobling et al. 2004; Sims-Williams 1998). However, recent (and no less controversial) applications of *phylogenetics* to Indo-European *cognate sets*[5] seem to support an origin of this family around the time of agricultural expansion from Anatolia (Gray and Atkinson 2003; Atkinson and Gray 2006).

However, in general, it seems that the bulk of the correlations between linguistic historical classifications and genetics is due to geography, with great distances and major obstacles being the best predictors of large differences in languages and genes (Belle and Barbujani 2007; Rosser et al. 2000), but that there is also a residual correlation between languages and genes after controlling for geography (Cavalli-Sforza et al. 1994; Belle and Barbujani 2007), due to shared history and *linguistic assortative mating*, whereby people tend to marry within their own linguistic community (Cavalli-Sforza et al. 1994).

## 7.4 Pushing and Pulling at Language

If the previous two sections tried to review in a more or less objective manner well-established aspects of the interaction between genes and language, the present one will concern a very new and controversial development in which the author is directly involved, so that some subjectivity is bound to remain despite all efforts to the contrary. The differences between individuals (Sect. 7.2) are limited in both time

---

[5] Roughly, two words in two different languages are cognate if they descend from the same original proto-word.

and scale to the ontogenetic and personal, while the inter-population correlations between genes and languages (Sect. 7.3) concern historical (cultural) and continental (macroregional) scales.

The correlations between linguistic classifications and genetic differences between populations discussed above consider explicitly the *phylogenetic diversity*, characterized by the number of different *language lineages* (language families, subfamilies, etc.) and reflecting historical processes (Nettle 1999). There are around 100 language families more or less accepted to date (Gordon 2005), vastly unequal with respect to their geographical distribution and size (measured either as number of speakers or number of languages; Nettle 1999; Gordon 2005), but, given the difficulties involved, these data are very approximate (Trask 1996; Campbell 2004; Lass 1997).

But another dimension along which languages differ is represented by their *structure*, allowing the definition of *structural diversity* (Nettle 1999). Languages can be described (to a good approximation) using a set of abstract categories (variously named *features*, *variables* or *parameters*) which can have different values for different languages and covering every aspect of language, from phonetics and phonology, through morpho-syntax to semantics and pragmatics (studied by *linguistic typology*; Croft 1990; Comrie 1981; Haspelmath et al. 2005). One example of such a feature is “number of vowels” (Maddieson 2008): some languages have a lot (like English or German) while some very few (like Berber or many Australian aboriginal languages). Another well-known example is represented by the order of Subject, Object and Verb in declarative clauses (Dryer 2008): many languages (like Japanese) prefer SOV, others (like English) prefer SVO, while the others go for VSO, VOS, OVS, OSV or show no particular preference.

### *7.4.1 Linguistic Tone*

Another linguistic feature is represented by *tone*. *Non-tone* languages (like English or Arabic) use *voice pitch* to conveys certain meanings at the sentence level (*intonation*) like marking an utterance as a question or an exclamation. In addition, *tone languages*, like Chinese or Yoruba, use voice pitch to distinguish words or grammatical structures in very much the same way all languages use vowels and consonants (Yip 2002; Dediu and Ladd 2007): for example, “nian” in Mandarin Chinese means “year” when pronounced with rising pitch and “read” when pronounced with falling pitch. The number of tones varies between tone languages, but usually is between two and seven (Yip 2002).

The classification of languages as tonal or not is not straightforward, with many borderline cases (like Swedish or Norwegian), but a bit more than half the world’s 6000 or so languages (Gordon 2005) seem to be tonal (Haspelmath et al. 2005). Their geographic distribution is not random, with clusters predominating in sub-Saharan Africa, continental and insular South–East Asia, and Central America and

Amazonia[6] (Haspelmath et al. 2005; Dediu and Ladd 2007) and a very interesting question concerns the causes of this distribution: is it purely “random”, following the laws of cultural transmission and the vagaries of history and geography, or are there some other causes, as well?

It is known that languages sharing a common ancestor tend to inherit its value for tone (in a way, tonal begets tonal and non-tonal, non-tonal, but this is highly simplifying) and it is also known that tone (or non-tone, for that matter) tends to spread to neighboring languages through regular language contact (Yip 2002; Dediu and Ladd 2007). Also, we know that *tonogenesis* can happen through regular historical linguistic processes (Hyman 1978; Yip 2002) and that the reverse process, of tone loss, appears, for example, in situations of usage as *lingua franca* (like Swahili).

But the hypothesis D. Robert Ladd and I have recently put forward (Dediu and Ladd 2007) suggests that there might be a slight *genetic biasing* (Dediu 2011) at work, contributing to the dynamics and distribution of tone across the world’s languages.

### *7.4.2 Two Brain-Related Genes*

This genetic biasing is proposed to be due to the so-called *derived* or *adaptive*[7] *haplogroups* of two human genes, *ASPM* and *Microcephalin* (in this chapter, we will denote these two haplogroups as *ASPM-D* and *MCPH-D*, respectively). Both *ASPM* and *Microcephalin* are clearly involved in brain growth and development because people carrying deleterious mutations of any of these two genes develop *microcephaly*, having heads much smaller than the average (but these are not the only genes causing microcephaly; Gilbert et al. 2005; Cox et al. 2006; Woods 2004). The exact mechanism by which the deleterious mutations of both *ASPM* and *Microcephalin* induce this pathology are not fully understood yet, but they very probably affect the formation of neural cells (Dediu and Ladd 2007; Caviness et al. 1995).

However, the derived haplogroups *ASPM-D* and *MCPH-D* do not cause microcephaly, or, for that matter, any other obvious phenotypic effects. Previous research failed to find any correlations with, for example, variation in intelligence or head size in the normal population[8] (Mekel-Bobrov et al. 2007; Woods et al. 2006) or any association with schizophrenia (Rivero et al. 2006). But they are fascinating because they seem to be under strong natural selection in humans,[9] have a skewed geographical

---

[6] For a map of tone, see the World Atlas of Language Structures (Haspelmath et al. 2005) Online, the chapter “Tone” by Ian Maddieson, at http://wals.info/chapter/13.

[7] The names “derived” and “adaptive” come from the original work by Bruce Lahn’s group (Mekel-Bobrov et al. 2005; Evans et al. 2005) which, in fact, identified them. They are called “derived” as opposed to the original (ancient) form of the genes and “adaptive” because of the selection pressures acting on them, that the authors claim to have detected. For our purposes here, these are just labels.

[8] But see the recent claim that another SNP of *Microcephalin* is associated with cranial volume in normal Chinese males (Wang et al. 2008).

[9] However, the methodology used to infer this is probably not adequate and the claimed recent selection has not been replicated.

distribution and have relatively recent origins (Mekel-Bobrov et al. 2005; Evans et al. 2005; Dediu and Ladd 2007).

It was found (Dediu and Ladd 2007; Ladd et al. 2008) that the geographic distribution of *ASPM-D* and *MCPH-D* correlates very well with the distribution of tone languages, even after controlling for the main two sources of such *spurious* correlations (see Sect. 7.3): the historical relatedness between languages and language contact mediated by spatial closeness (this was done by computing the *partial Mantel correlations* between linguistic, genetic, historical linguistic and geographical distances between populations—see Dediu and Ladd 2007 for technical details). Moreover, when comparing the correlations between tone, *ASPM-D* and *MCPH-D* with all the possible correlations between 25 or so other linguistic features and 1000 genes, it was found that it was well in the upper tail of the empirical distribution of this type of correlation, suggesting that this association is very "special", in a statistical sense (Dediu and Ladd 2007). Therefore, it was concluded that the relationship between *ASPM-D* and *MCPH-D* and tone is stronger than expected by chance and not fully explained by classical genes-languages co-dispersal processes (Dediu and Ladd 2007; Ladd et al. 2008).

### *7.4.3 Biasing Language Transmission*

Populations having *low* frequencies of both derived haplogroups tend to speak *tone* languages, while populations with *high* frequencies tend to speak *non-tone* languages, with populations with *low ASPM-D* and *high MCPH-D* showing no preference[10] (see also the figure in Dediu and Ladd 2007). Therefore, we suggested that the derived haplogroups of these two genes might be able to bias language *towards non-tonality* or *against tonality* (Dediu and Ladd 2007; Ladd et al. 2008).

It is obvious that such a bias, if it indeed exists, does not have any major effects at the individual level, as any normal child can acquire the language(s) of his/her community irrespective of their genetic background (Ladd et al. 2008). What such a bias does is to have an extremely small, almost invisible impact at the individual level, but it will be amplified and made manifest by the cultural transmission of language across generations, by "pushing" or "pulling" language towards its preferred state (see also various other contributions in this volume). Again, such a bias will emphatically *not* determine the fate of language, as many other factors play a major role in language change and evolution (contact, history, sheer accident), but it will statistically bias the distribution of structural diversity (see Ladd et al. 2008 for a thorough discussion and relevant examples). Moreover, such a bias probably does not represent the reason why *ASPM-D* and *MCPH-D* are under natural selection, but is simply a byproduct of their effects on brain development (Dediu and Ladd 2007; Ladd et al. 2008).

---

[10] There are no *high ASPM-D* and *low MCPH-D* in the sample (Dediu and Ladd 2007; Ladd et al. 2008).

Confirmation or falsification of this hypothesis rests with experimental approaches of the type discussed above (Sect. 7.2), by trying to define the effects of this bias at the individual level, design appropriate measures and test if the variation in such measures correlates with the presence or absence of the derived haplogroups. Interestingly, it was reported very recently (Wong et al. 2012) that *ASPM-D* seems to be involved in pitch processing at the individual level, but more work is needed in order to draw any clear conclusions. However, the methodology introduced by Dediu and Ladd (2007) holds the potential to detect more candidate genetic biases and their effects in shaping linguistic diversity, being fundamentally a *hypothesis-generating* mechanism (Nettle 2007).

## 7.5 Conclusions

It should come as no surprise that our genes impact on our language and speech but what might be unexpected is the complexity of the mechanisms involved. It is clear that the old question of "nature vs nurture" is profoundly misguided, as genes and environment interact in subtle ways and require each other in order to produce their phenotypic effects (see, for example, Ladd et al. 2008 for a detailed discussion of this issue in the context of language).

It is also such inter-individual differences which, if structured in appropriate ways, can influence the trajectory of language change. People with the derived haplogroups of *ASPM* and *Microcephalin* might indeed turn out to be different from those without them in ways relevant to learning, processing or producing linguistic tone, but if only one such individual appears each generation in a population, it will not be enough to bias language change towards non-tone languages. However, intuition and computer models (Dediu 2008b) seem to suggest that when enough such individuals exist in a population for long enough time, language will indeed be influenced by their combined biases.

On the other hand, inter-individual differences and genetic biases are embedded in the larger context of language dispersal, birth, change and death (Nettle 1999) and these three levels interact in complex ways to produce the patterns of language and genetic diversity. And, finally, such small genetic biases could form the basis of language evolution, viewed as the co-evolution between our languages and our brains and bodies (Dediu 2008a; Christiansen and Chater 2008).

But understanding this interaction—and the multifaceted *phenomenon of language*—requires understanding all the intervening levels: brain, hearing, speech and gesturing, discourse now and here, and language change on the historical and evolutionary timescales. It requires, thus, a truly integrative approach.

## References

Alarcón, M., Abrahams, B.S., Stone, J.L., Duvall, J.A., Perederiy, J.V., Bomar, J.M., Sebat, J., Wigler, M., Martin, C.L., Ledbetter, D.H., Nelson, S.F., Cantor, R.M., Geschwind, D.H.: Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am. J. Human Genet. **82**(1), 150–159 (2008). doi:10.1016/j.ajhg.2007.09.005

Appleyard, D.: Afroasiatic and the Nostratic hypothesis. In: Renfrew, C., Nettle, D. (eds.) Nostratic: Examining a Linguistic Macrofamily, pp. 289–325. The McDonald Institute for Archaeological Research, Oxbow Books, Oxford (1999)

Atkinson, Q.D., Gray, R.D.: How old is the Indo-European language family? Illumination or more moths to the flame? In: Clackson, J., Forster, P., Renfrew, C. (eds.), Phylogenetic Methods and the Prehistory of Languages, pp. 91–109. MacDonald Institute, Cambridge (2006)

Bamshad, M., Wooding, S., Watkins, W., Ostler, C., Batzer, M., Jorde, L.: Human population genetic structure and inference of group membership. Am. J. Hum. Genet. **72**, 578–589 (2003)

Banton, M.: Racial Theories, 2nd edn. Cambridge University Press, Cambridge (1998)

Barbujani, G., Belle, E.: Genomic boundaries between human populations. Hum. Hered. **61**(1), 15–21 (2006)

Bateman, R., Goddard, I., O'Grady, R., Funk, V., Mooi, R., Kress, W., Cannell, P.: Speaking of forked tongues: the feasibility of reconciling human phylogeny and the history of language. Curr. Anthropol. **31**, 1–13 (1990)

Belle, E., Barbujani, G.: Worldwide analysis of multiple microsatellites: language diversity has a detectable influence on DNA diversity. Am. J. Phys. Anthropol. **133**(4), 1137–1146 (2007)

Bishop, D.V.M.: Genetic and environmental risks for specific language impairment in children. Int. J. Pediatr. Otorhinolaryngol. **6751**, S143–S157 (2003)

Bolnick, D., Shook, B., Campbell, L., Goddard, I.: Problematic use of Greenberg's linguistic classification of the Americas in studies of native American genetic variation. Am. J. Hum. Genet. **75**, 519–523 (2004)

Bonneau, D., Verny, C., Uzé, J.: Les facteurs génétiques dans les troubles spécifiques du langage oral. Archives de pédiatrie **10**, 1213–1216 (2004)

Boyd, R., Richerson, P.J.: Culture and the Evolutionary Process. University of Chicago Press, Chicago (1985)

Campbell, L.: Nostratic and linguistic palaeontology in methodological perspective. In: Renfrew, C., Nettle, D. (eds.) Nostratic: Examining a Linguistic Macrofamily, pp. 179–230. The McDonald Institute for Archaeological Research, Oxbow Books, Oxford (1999)

Campbell, L.: Historical Linguistics: An Introduction. Edinburgh University Press, Edinburgh (2004)

Caspi, A., Williams, B., Kim-Cohen, J., Craig, I.W., Milne, B.J., Poulton, R., Schalkwyk, L.C., Taylor, A., Werts, H., Moffitt, T.E.: Moderation of breast feeding effects on the IQ by genetic variation in fatty acid metabolism. Proc. Natl. Acad. Sci. USA **104**(47), 18860–18865 (2007)

Cavalli-Sforza, L.: Genes, Peoples, and Languages. Allen Lane: The Penguin Press, London (2000)

Cavalli-Sforza, L., Menozzi, P., Piazza, A.: The History and Geography of Human Genes. Princeton University Press, Princeton (1994). (abridged paperback edition)

Cavalli-Sforza, L.L., Feldman, M.W.: Cultural Transmission and Evolution: A Quantitative Approach. Princeton University Press, Princeton (1981)

Caviness, V.J., Takahashi, T., Nowakowski, R.: Numbers, time and neocortical neuronogenesis: a general developmental and evolutionary model. Trends Neurosci. **18**(9), 379–383 (1995)

Chanock, S., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D., Thomas, G., Hirschhorn, J., Abecasis, G., Altshuler, D., Bailey-Wilson, J., Brooks, L., Cardon, L., Daly, M., Donnelly, P., Fraumeni, J., Freimer, N., Gerhard, D., Gunter, C., Guttmacher, A., Guyer, M., Harris, E., Hoh, J., Hoover, R., Kong, C., Merikangas, K., Morton, C., Palmer, L., Phimister, E., Rice, J., Roberts, J., Rotimi, C., Tucker, M., Vogan, K., Wacholder, S., Wijsman, E., Winn, D., Collins, F.: Replicating genotype-phenotype associations. Nature **447**(7145), 655–660 (2007). doi:10.1038/447655

Charney, E.: Behavior genetics and post genomics. Behav. Brain Sci. **35**(6), 1–80 (2012)

Christiansen, M., Chater, N.: Language as shaped by the brain. Behav. Brain Sci. **31**(5), 489–508 (2008). (discussion 509–58)

Comrie, B.: Language Universals and Linguistic Typology. Basil Blackwell Publisher Limited, Oxford (1981)

Coop, G., Bullaughey, K., Luca, F., Przeworski, M.: The timing of selection at the human FOXP2 gene. Mol. Biol. Evol. **25**(7), 1257–1259 (2008). doi:10.1093/molbev/msn091

Cox, J., Jackson, A.P., Bond, J., Woods, C.G.: What primary microcephaly can tell us about brain growth. Trends Mol. Med. **12**(8), 358–366 (2006)

Croft, W.: Typology and Universals. Cambridge University Press, Cambridge (1990)

Dediu, D.: Causal correlations between genes and linguistic features - the mechanism of gradual language evolution. In: Smith, A.D.M., Smith, K., Ferrer, R., Cancho, I. (eds.) The Evolution of Language: Proceedings of the 7th International Conference (EVOLANG7), pp. 83–90. World Scientific Press, Singapore (2008a)

Dediu, D.: The role of genetic biases in shaping language-genes correlations. J. Theor. Biol. 254, 400–407 (2008b)

Dediu, D.: Are languages really independent from genes? If not, what would a genetic bias affecting language diversity look like? Hum. Biol. **83**(2), 279–296 (2011). doi:10.3378/027.083.0208

Dediu, D., Ladd, D.: Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and microcephalin. Proc. Natl. Acad. Sci. USA **104**(26), 10944–10949 (2007)

Dench, A.: Descent and diffusion: the complexity of the pilbara situation. In: Aikhenvald, A., Dixon, R. (eds.) Areal Diffusion and Genetic Inheritance, pp. 105–133. Oxford University Press, Oxford (2001)

Diamond, J.: The language steamrollers. Nature **389**, 544–546 (1997)

Diamond, J.: Guns, Germs and Steel: A Short History of Everybody for the Last 13,000 Years. Vintage, London (1998)

Diamond, J., Bellwood, P.: Farmers and their languages: the first expansions. Science **300**, 597–603 (2003)

Dixon, R.: The Rise and Fall of Languages. Cambridge University Press, Cambridge (1997)

Dixon, R.: The Australian linguistic area. In: Aikhenvald, A., Dixon, R. (eds.) Areal Diffusion and Genetic Inheritance, pp. 64–104. Oxford University Press, Oxford (2001)

Dryer, M.S.: Order of subject, object and verb. In: Haspelmath, M., Dryer, M.S., Gil, D., Comrie, B. (eds.) The World Atlas of Language Structures Online. Max Planck Digital Library, Munich (2008)

Enard, W., Gehre, S., Hammerschmidt, K., Hölter, S.M., Blass, T., Somel, M., Brückner, M.K., Schreiweis, C., Winter, C., Sohr, R., Becker, L., Wiebe, V., Nickel, B., Giger, T., Müller, U., Groszer, M., Adler, T., Aguilar, A., Bolle, I., Calzada-Wack, J., Dalke, C., Ehrhardt, N., Favor, J., Fuchs, H., Gailus-Durner, V., Hans, W., Hölzlwimmer, G., Javaheri, A., Kalaydjiev, S., Kallnik, M., Kling, E., Kunder, S., Mossbrugger, I., Naton, B., Racz, I., Rathkolb, B., Rozman, J., Schrewe, A., Busch, D.H., Graw, J., Ivandic, B., Klingenspor, M., Klopstock, T., Ollert, M., Quintanilla-Martinez, L., Schulz, H., Wolf, E., Wurst, W., Zimmer, A., Fisher, S. E., Morgenstern, R., Arendt, T., de Angelis, M.H., Fischer, J., Schwarz, J., Pääbo, S.: A humanized version of Foxp2 affects cortico-basal ganglia circuits in mice. Cell **137**(5), 961–971 (2009). doi:10.1016/j.cell.2009.03.041

Enard, W., Przeworski, M., Fisher, S., Lai, C., Wiebe, V., Kitano, T., Monaco, A., Pääbo, S.: Molecular evolution of FOXP2, a gene involved in speech and language. Nature **418**, 869–872 (2002)

Evans, P., Gilbert, S., Mekel-Bobrov, N., Vallender, E., Anderson, J., Vaez-Azizi, L., Tishkoff, S., Hudson, R., Lahn, B.: Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science **309**(5741), 1717–1720 (2005)

Excoffier, L., Smouse, P., Quattro, J.: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**, 479–491 (1992)

Fagiolini, M., Jensen, C.L., Champagne, F.A.: Epigenetic influences on brain development and plasticity. Curr. Opin. Neurobiol. **19**(2), 207–212 (2009). doi:10.1016/j.conb.2009.05.009

Falush, D., Stephens, M., Pritchard, J.K.: Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**(4), 1567–1587 (2003)

Felsenfeld, S.: Finding susceptibility genes for developmental disorders of speech: the long and winding road. J. Commun. Disord. **35**, 329–345 (2002)

Fisher, S., Lai, C., Monaco, A.: Deciphering the genetic basis of speech and language disorders. Annu. Rev. Neurosci. **26**, 57–80 (2003)

Fisher, S., Vargha-Khadem, K., Monaco, A., Pembery, M.: Localization of a gene implicated in a severe speech and language disorder. Nat. Genet. **18**, 168–170 (1998)

Fisher, S.E., Scharff, C.: FOXP2 as a molecular window into speech and language. Trends Genet. **25**(4), 166–177 (2009). doi:10.1016/j.tig.2009.03.002

Fortson, B.: Indo-European Language and Culture: An Introduction. Blackwell Publishing, Oxford (2004)

Francks, C., Maegawa, S., Laurén, J., Abrahams, B.S., Velayos-Baeza, A., Medland, S.E., Colella, S., Groszer, M., McAuley, E.Z., Caffrey, T.M., Timmusk, T., Pruunsild, P., Koppel, I., Lind, P.A., Matsumoto-Itaba, N., Nicod, J., Xiong, L., Joober, R., Enard, W., Krinsky, B., Nanba, E., Richardson, A.J., Riley, B.P., Martin, N.G., Strittmatter, S.M., Möller, H.J., Rujescu, D., Muglia, P., Roos, J.L., Fisher, S.E., Wade-Martins, R., Rouleau, G.A., Stein, J.F., Karayiorgou, M., Geschwind, D.H., Ragoussis, J., Kendler, K.S., Airaksinen, M.S., Oshimura, M., DeLisi, L.E., Monaco, A.P.: LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. Mol. Psychiatry **12**, 1129–1139 (2007)

Gibson, C.J., Gruen, J.R.: The human lexinome: genes of language and reading. J. Commun. Disord. **41**(5), 409–420 (2008)

Gilbert, S.L., Dobyns, W.B., Lahn, B.T.: Genetic links between brain development and brain evolution. Nat. Rev. Genet. **6**(7), 581–590 (2005)

Gopnik, M., Crago, M.: Familial aggregation of a developmental language disorder. Cognition **39**, 1–50 (1991)

Gordon, R.G. (ed.): Ethnologue: Languages of the World, 15th edn. SIL International, Dallas (2005)

Gray, R.D., Atkinson, Q.D.: Language-tree divergence times support the anatolian theory of Indo-European origin. Nature **426**, 435–439 (2003)

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.-Y., Hansen, N.F., Durand, E.Y., Malaspinas, A.-S., Jensen, J.D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H.A., Good, J.M., Schultz, R., Aximu-Petri, A., Butthof, A.,Höber, B., Höffner, B., Siegemund, M., Weihmann, A., Nusbaum, C., Lander, E.S., Russ, C., Novod, N., Affourtit, J., Egholm, M., Verna, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Doronichev, V.B., Golovanova, L.V., Lalueza-Fox, C., de la Rasilla, M., Fortea, J., Rosas, A., Schmitz, R.W., Johnson, P.L.F., Eichler, E.E., Falush, D., Birney, E., Mullikin, J.C., Slatkin, M., Nielsen, R., Kelso, J., Lachmann, M., Reich, D., Pääbo, S.: A draft sequence of the Neandertal genome. Science **328**(5979), 710–722 (2010). doi:10.1126/science.1188021

Greenberg, J.: The Indo-Pacific hypothesis. In: Sebeok, T. (ed.) Current Trends in Linguistics, Linguistics in Oceania, vol. 8, pp. 807–871. Mouton, The Hague (1971)

Greenberg, J.: Language in the Americas. Stanford University Press, Stanford (1987)

Greenberg, J.: The convergence of Eurasiatic and Nostratic. In: Salmons, J., Joseph, B. (eds.) Nostratic: Sifting the Evidence, pp. 51–60. John Benjamins Publishing Co., Amsterdam (1998)

Guégan, J.-F., Prugnolle, F., Thomas, F.: Global spatial patterns of infectious diseases and human evolution. In: Stearns, S.C., Koella, J.C. (eds.) Evolution in Health and Disease, pp. 19–29. Oxford University Press, Oxford (2007)

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E., Lints, T., Jarvis, E., Scharff, C.: FoxP2 expression in avian vocal learners and non-learners. J. Neurosci. **24**, 3164–3175 (2004)

Halliburton, R.: Introduction to Population Genetics. Pearson Eduction Inc., Upper Saddle River (2004)

Haspelmath, M., Dryer, M. S., Gil, D., Comrie, B. (eds.): The World Atlas of Language Structures. Oxford University Press, Oxford (2005)

Hirst, M., Marra, M.A.: Epigenetics and human disease. Int. J. Biochem. Cell Biol. **41**(1), 136–146 (2009). doi:10.1016/j.biocel.2008.09.011

Hurst, J., Baraitser, M., Auger, E., Graham, F., Norell, S.: An extended family with a dominantly inherited speech disorder. Dev. Med. Child Neurol. **32**, 352–355 (1990)

Hyman, L.: Historical tonology. In: Fromkin, V. (ed.) Tone: A Linguistic Survey, pp. 257–269. Academic, London (1978)

Ingram, C.J.E., Swallow, D.M.: Population genetics of lactase persistence and lactose intolerance. In: Cooper, D.N., Kehrer-Sawatzki, H. (eds.) Handbook of Human Molecular Evolution, vol. 2, pp. 1127–1132. John Wiley and Sons Ltd, Chichester (2008)

Jablonka, E., Raz, G.: Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q. Rev. Biol. **84**(2), 131–176 (2009)

Jiang, Y., Langley, B., Lubin, F.D., Renthal, W., Wood, M.A., Yasui, D.H., Kumar, A., Nestler, E.J., Akbarian, S., Beckel-Mitchener, A.C.: Epigenetics in the nervous system. J. Neurosci. **28**(46), 11753–11759 (2008). doi:10.1523/JNEUROSCI.3797-08.2008

Jobling, M., Tyler-Smith, C., Hurles, M.: Human Evolutionary Genetics: Origins, Peoples and Disease. Garland Science, New York (2004)

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R., Burbano, H., Hublin, J.-J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S.: The derived FOXP2 variant of modern humans was shared with neandertals. Curr. Biol. **17**(21), 1908–1912 (2007)

Ladd, D., Dediu, D., Kinsella, A.: Languages and genes: reflections on biolinguistics and the nature-nurture question. Biolinguistics **2**(1), 114–126 (2008)

Lai, C., Fisher, S., Hurst, J., Levy, E., Hodgson, S., Fox, M., Jeremiah, S., Povey, S., Jamison, D., Green, E., Vargha-Khadem, F., Monaco, A.: The SPCH1 region on human 7q31: genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. Am. J. Hum. Genet. **67**, 357–368 (2000)

Lai, C., Fisher, S., Hurst, J., Vargha-Khadem, F., Monaco, P.: A forkhead-domain gene is mutated in a severe speech and language disorder. Nature **413**, 519–523 (2001)

Lass, R.: Historical Linguistics and Language Change. Cambridge University Press, Cambridge (1997)

Lewis, C., Levinson, D., Wise, L., DeLisi, L., Straub, R., Hovatta, I., Williams, N., Schwab, S., Pulver, A., Faraone, S., Brzustowicz, L., Kaufmann, C., Garver, D., Gurling, H., Lindholm, E., Coon, H., Moises, H., Byerley, W., Shaw, S., Mesen, A., Sherrington, R., O'Neill, F., Walsh, D., Kendler, K., Ekelund, J., Paunio, T., Lönnqvist, J., Peltonen, L., O'Donovan, M., Owen, M., Wildenauer, D., Maier, W., Nestadt, G., Blouin, J.-L., Antonarakis, S., Mowry, B., Silverman, J., Crowe, R., Cloninger, C., Tsuang, M., Malaspina, D., Harkavy-Friedman, J., Svrakic, D., Bassett, A., Holcomb, J., Kalsi, G., McQuillin, A., Brynjolfson, J., Sigmundsson, T., Petursson, H., Jazin, E., Zoëga, T., Helgason, T.: Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: schizophrenia. Am. J. Hum. Genet. **73**(1), 34–48 (2003)

Lewontin, R.: The apportionment of human diversity. In: Dobzhansky, T., Hecht, M.K., Steere, W.C. (eds.) Evolutionary Biology, vol. 6, pp. 381–398. Appleton-Century-Crofts, New York (1972)

Li, G., Wang, J., Rossiter, S., Jones, G., Zhang, S.: Accelerated FoxP2 evolution in echolocating bats. PLoS ONE **2**(9), e900 (2007)

Liégeois, F., Baldeweg, T., Connelly, A., Gadian, D., Mishkin, M., Vargha-Khadem, F.: Language fMRI abnormalities associated with FOXP2 gene mutation. Nat. Neurosci. **6**, 1230–1237 (2003)

MacDermot, K., Bonora, E., Sykes, N., Coupe, A., Lai, C., Vernes, S., Vargha-Khadem, F., McKenzie, F., Smith, R., Monaco, A., Fisher, S.: Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am. J. Hum. Genet. **76**, 1074–1080 (2005)

MacEachern, S.: Genes, tribes, and african history. Curr. Anthropol. **41**, 357–384 (2000)

Maddieson, I.: Vowel quality inventories. In: Haspelmath, M., Dryer, M.S., Gil, D., Comrie, B. (eds.) The World Atlas of Language Structures Online. Max Planck Digital Library, Munich (2008)

Maher, B., Riley, B., Kendler, K.: Psychiatric genetics gets a boost. Nat. Genet. **40**, 1042–1044 (2008). doi:10.1038/ng0908-1042

Mallory, J.: In Search of the Indo-Europeans: Language, Archaeology and Myth. Thames and Hudson, London (1991)

Manni, F., Guérard, E., Heyer, E.: Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using monmonier's algorithm. Hum. Biol. **76**, 173–190 (2004)

Marcus, G., Fisher, S.: FOXP2 in focus: what can genes tell us about speech and language? Trends Cogn. Sci. **7**, 257–262 (2003)

Matisoff, J.: On megalocomparison. Language **66**, 106–120 (1990)

McCarroll, S.A., Altshuler, D.M.: Copy-number variation and association studies of human disease. Nat. Gen. **39**(7 Suppl), S37–S42 (2007)

McCarthy, M., Abecasis, G., Cardon, L., Goldstein, D., Little, J., Ioannidis, J., Hirschhorn, J.: Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat. Rev. Genet., **9**, 356–369 (2008). doi:10.1038/nrg2344

McMahon, A., McMahon, R.: Language Classification by Numbers. Oxford University Press, Oxford (2005)

McMahon, R.: Genes and languages. Commun. Genet. **7**, 2–13 (2004)

Mekel-Bobrov, N., Gilbert, S., Evans, P., Vallender, E., Anderson, J., Hudson, R., Tishkoff, S., Lahn, B.: Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. Science **309**(5741), 1720–1722 (2005)

Mekel-Bobrov, N., Posthuma, D., Gilbert, S., Lind, P., Gosso, M., Luciano, M., Harris, S., Bates, T., Polderman, T., Whalley, L., Fox, H., Starr, J., Evans, P., Montgomery, G., Fernandes, C., Heutink, P., Martin, N., Boomsma, D., Deary, I., Wright, M., de Geus, E., Lahn, B.: The ongoing adaptive evolution of ASPM and Microcephalin is not explained by increased intelligence. Hum. Mol. Genet. **16**(6), 600–608 (2007)

Mithen, S.: After the ice: a global human history. Phoenix: Orion Books Ltd, London (2003)

Nettle, D.: Linguistic Diversity. Oxford University Press, Oxford (1999)

Nettle, D.: Language and genes: a new perspective on the origins of human cultural diversity. Proc. Natl. Acad. Sci. USA **104**(26), 10755–10756 (2007)

Odling-Smee, F.J., Laland, K.N., Feldman, M.W.: Niche Construction: The Neglected Process in Evolution. Princeton University Press, Princeton (2003)

Pearson, T., Manolio, T.: How to interpret a genome-wide association study. J. Am. Med. Assoc. **299**(11), 1335–1344 (2008)

Plomin, R., Colledge, E., Dale, P.: Genetics and the development of language disabilities and abilities. Curr. Paediatr. **12**, 419–424 (2002)

Plomin, R., Defries, J.C., McClearn, G.E., McGuffin, P.: Behav. Genet., 4th edn. Worth Publishers, New York (2001)

Plomin, R., Kovas, Y.: Generalist genes and learning disabilities. Psychol. Bull. **131**, 592–617 (2005)

Poloni, E., Semino, O., Passarino, G., Santachiara-Benerecetti, A., Dupanloup, I., Langaney, A., Excoffier, L.: Human genetic affinities for Y-chromosome P49a, f/TaqI haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. **61**, 1015–1035 (1997)

Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., González, J.R., Gratacòs, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., Hurles, M.E.: Global variation in copy number in the human genome. Nature **444**(7118), 444–454 (2006)

Relethford, J.: Genetics and the Search for Modern Human Origins. Wiley-Liss, New York (2001)

Renfrew, C.: 'The Emerging Synthesis': the archaeogenetics of farming/language dispersals and other spread zones. In: Bellwood, P., Renfrew, C. (eds.) Examining the Farming/Language Dispersal Hypothesis, pp. 3–16. The McDonald Institute for Archaeological Research, Cambridge (2002)

Rivero, O., Sanjuán, J., Moltó, M.-D., Aguilar, E.-J., Gonzalez, J.-C., de Frutos, R., Nájera, C.: The microcephaly ASPM gene and schizophrenia: a preliminary study. Schizophr. Res. **84**(2–3), 427–429 (2006)

Rosenberg, N., Pritchard, J., Weber, J., Cann, H., Kidd, K., Zhivotovsky, L., Feldman, M.W.: Genetic structure of human populations. Science **298**, 2381–2385 (2002)

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G., Brede, G., Cooper, G., Côrte-Real, H.B., de Knijff, P., Decorte, R., Dubrova, Y.E., Evgrafov, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E.W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S.A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L.A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T.A., Mikelsaar, A.V., Mitchell, R.J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P.C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M.J., Previderé, C., Roewer, L., Rootsi, S., Rubinsztein, D.C., Saillard, J., Santos, F.R., Stefanescu, G., Sykes, B.C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M.A.: Y-chromosomal diversity in europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. **67**, 1526–1543 (2000)

Ruhlen, M.: A Guide to the World's Languages, vol. I. Classification (with a Postscript on Recent Developments). Arnold, London (1991)

Ruhlen, M.: On the Origin of Languages: Studies in Linguistic Taxonomy. Stanford University Press, Stanford (1994)

Scharff, C., Haesler, S.: An evolutionary perspective on FoxP2: strictly for the birds? Curr. Opin. Neurobiol. **15**, 694–703 (2005)

Scharff, C., White, S.: Genetic components of vocal learning. Ann. N. Y. Acad. Sci. **1016**, 325–347 (2004)

Segurado, R., Detera-Wadleigh, S., Levinson, D., Lewis, C., Gill, M., Nurnberger, J., Craddock, N., DePaulo, J., Baron, M., Gershon, E., Ekholm, J., Cichon, S., Turecki, G., Claes, S., Kelsoe, J., Schofield, P., Badenhop, R., Morissette, J., Coon, H., Blackwood, D., McInnes, L., Foroud, T., Edenberg, H., Reich, T., Rice, J., Goate, A., McInnis, M., McMahon, F., Badner, J., Goldin, L., Bennett, P., Willour, V., Zandi, P., Liu, J., Gilliam, C., Juo, S.-H., Berrettini, W., Yoshikawa, T., Peltonen, L., Lönnqvist, J., Nöthen, M., Schumacher, J., Windemuth, C., Rietschel, M., Propping, P., Maier, W., Alda, M., Grof, P., Rouleau, G., Del-Favero, J., Broeckhoven, C., Mendlewicz, J., Adolfsson, R., Spence, M., Luebbert, H., Adams, L., Donald, J., Mitchell, P., Barden, N., Shink, E., Byerley, W., Muir, W., Visscher, P., Macgregor, S., Gurling, H., Kalsi, G., McQuillin, A., Escamilla, M., Reus, V., Leon, P., Freimer, N., Ewald, H., Kruse, T. A., Mors, O., Radhakrishna, U., Blouin, J.-L., Antonarakis, S., Akarsu, N.: Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: bipolar disorder. Am. J. Hum. Genet. **73**(1), 49–62 (2003)

Shu, W., Cho, J., Jiang, Y., Zhang, M., Weisz, D., Elder, G., Schmeidler, J., De Gasperi, R., Sosa, M., Rabidou, D., Santucci, A., Perl, D., Morrisey, E., Buxbaum, J.: Altered ultrasonic vocalization in mice with a disruption in the foxp2 gene. Proc. Natl. Acad. Sci. USA **102**, 9643–9648 (2005)

Sims-Williams, P.: Genetics, linguistics, and prehistory: thinking big and thinking straight. Antiquity **72**, 505–527 (1998)

Skelton, P.: Evolution: a biological and palaeoantological approach. The Open University, Milton Keynes (1993)

Speed, T., Zhao, H.: Chromosome maps. In: Balding, D., Bishop, M., Cannings, C. (eds.) Handbook of Statistical Genetics, vol. 1, 3rd edn. Chapter 1, pp. 3–39. John Wiley and Sons., Ltd., Chichester (2007)

Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., Vernes, S., Fisher, S., Ren, B., Geschwind, D.: Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. Am. J. Hum. Genet. **81**(6), 1144–1157 (2007)

Stromswold, K.: The heritability of language: a review and metaanalysis of twin, adoption, and linkage studies. Language **77**, 647–723 (2001)

Templeton, A.: Human races: a genetic and evolutionary perspective. Am. Anthropol. **100**, 632–650 (1998)

Teramitsu, I., Kudo, L., London, S., Geschwind, D., White, S.: Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J. Neurosci. **24**, 3152–3163 (2004)

The Wellcome Trust Case Control Consortium.: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature **447**, 661 (2007). doi:10.1038/nature05911

Trask, R.: Why should a language have any relatives? In: Renfrew, C., Nettle, D. (eds.) Nostratic: Examining a Linguistic Macrofamily, pp. 157–178. The McDonald Institute for Archaeological Research, Oxbow Books, Oxford (1999)

Trask, R.L.: Historical Linguistics. Arnold, London (1996)

Vargha-Khadem, F., Watkins, K., Price, C., Ashburner, J., Alcock, K., Connelly, A., Frackowiak, R., Friston, K., Pembrey, M., Mishkin, M., Passingham, R.: Neural basis of an inherited speech and language disorder. Proc. Natl. Acad. Sci. USA **95**, 12695–12700 (1998)

Vernes, S.C., Newbury, D.F., Abrahams, B.S., Winchester, L., Nicod, J., Groszer, M., Alarcón, M., Oliver, P.L., Davies, K.E., Geschwind, D.H., Monaco, A.P., Fisher, S.E.: A functional genetic link between distinct developmental language disorders. New Eng. J. Med. **359**(22), 2337–2345 (2008)

Wang, J.-K., Li, Y., Su, B.: A common SNP of MCPH1 is associated with cranial volume variation in Chinese population. Hum. Mol. Genet. **17**(9), 1329–1335 (2008)

Watkins, K., Vargha-Khadem, F., Ashburner, J., Passingham, R., Connelly, A., Friston, K., Frackowiak, R., Mishkin, M., Gadian, D.: MRI analysis of an inherited speech and language disorder: structural brain abnormalities. Brain **125**, 465–478 (2002)

Webb, D., Zhang, J.: FoxP2 in song-learning birds and vocal-learning mammals. J. Hered. **96**, 212–216 (2005)

Weedon, M.N., Lettre, G., Freathy, R.M., Lindgren, C.M., Voight, B.F., Perry, J.R.B., Elliott, K.S., Hackett, R., Guiducci, C., Shields, B., Zeggini, E., Lango, H., Lyssenko, V., Timpson, N.J., Burtt, N.P., Rayner, N.W., Saxena, R., Ardlie, K., Tobias, J.H., Ness, A.R., Ring, S.M., Palmer, C.N.A., Morris, A.D., Peltonen, L., Salomaa, V., Initiative, D.G., Consortium, W.T.C.C., Smith, G.D., Groop, L.C., Hattersley, A.T., McCarthy, M.I., Hirschhorn, J.N., Frayling, T.M.: A common variant of HMGA2 is associated with adult and childhood height in the general population. Nat. Genet. **39**(10), 1245–1250 (2007). doi:10.1038/ng2121

Wong, P.C.M., Chandrasekaran, B., Zheng, J.: The derived allele of aspm is associated with lexical tone perception. PLoS One **7**(4), e34243 (2012). doi:10.1371/journal.pone.0034243

Woods, C.G.: Human microcephaly. Curr. Opin. Neurobiol. **14**(1), 112–117 (2004)

Woods, R., Freimer, N., Young, J.D., Fears, S., Sicotte, N., Service, S., Valentino, D., Toga, A., Mazziotta, J.: Normal variants of Microcephalin and ASPM do not account for brain size variability. Hum. Mol. Genet. **15**(12), 2025–2029 (2006)

Yip, M.: Tone. Cambridge University Press, Cambridge (2002)

Youngson, N.A. Whitelaw, E.: Transgenerational epigenetic effects. Annu. Rev. Genomics Hum. Genet. **9**, 233–257 (2008). doi:10.1146/annurev.genom.9.081307.164445

Zhang, J., Webb, D., Podlaha, O.: Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics **162**, 1825–1835 (2002)

# Chapter 8
# Language in Nature: On the Evolutionary Roots of a Cultural Phenomenon

**Willem Zuidema**

**Abstract** What could an evolutionary explanation for language look like? Here I review relevant evidence from linguistics, comparative biology, evolutionary theory and the fossil record, which suggest vocal imitation and hierarchical compositionality as the essential and uniquely human biological foundations of language. I also outline a plausible scenario for how human language evolved, and propose that language preceded, and facilitated the development of, other cognitive domains such as reasoning, the ability to plan, and consciousness.

## 8.1 Introduction

What distinguishes Man from beast? For all of human history we have been wondering about that question, and over the centuries we have learned to dismiss some of the popular answers. Humans might walk upright more than any other ape, have less hair, be better at long distance running, use tools more readily, have more advanced reasoning skills, be more aware of the thoughts of others or behave more cooperatively. But all of these features, it has turned out, are differences of degree and not of kind. One answer, however, has survived all serious scrutiny: humans have language. In other animals we find elaborate communication systems, sometimes with one or two properties vaguely reminiscent of language, but always differing radically in many other properties.

Although it is difficult to list the defining properties of language, there simply is no other animal that comes close to having something like human language, and, inversely, there is no human population that does not have it. Moreover, we use language typically for many hours per day, and language is involved in all parts

---

Willem Zuidema (✉)
Institute for Logic, Language and Computation, University of Amsterdam,
Amsterdam, The Netherlands
e-mail: zuidema@uva.nl

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_8,

of human life: in gossiping, shopping, education, politics, fighting, courtship, and everything else. And, although much remains ill-understood, many scientists suspect that language somehow facilitates other cognitive skills that are dear to us: music, reasoning, consciousness, planning, mathematics and more. Hence, it is no overstatement to say that, from an evolutionary point of view, language is the most striking aspect of the human phenotype and cries out for an evolutionary explanation.

What could an evolutionary explanation for language look like? Libraries are filled with books on this issue, but many of the proposals are very speculative and, in fact, inconsistent with available evidence. It's worthwhile, therefore, to step back a bit and first consider some of the sources of information that could constrain the scenarios we might want to propose. The relevant evidence for evaluating evolutionary scenarios—consisting of particular starting and end points, and a mechanism that drives the steps in between—comes from many different fields. The end point, in our case, is the human capacity for language, and the obvious field to provide data is linguistics (although this field can offer less clear answers than we would perhaps wish). The starting point is the set of abilities of the last common ancestor that humans share with chimpanzees, our closest relatives. Our best guesses on these abilities come from a comparison of the abilities of other living great apes, i.e., from behavioural biology. The steps in between are largely unknown, but we find some hints in the fossil record. The mechanisms driving the evolution of language are also largely unknown, but evolutionary theory offers at least some constraints on the form of evolutionary scenarios. Finally, evidence on the abilities of more distantly related animals, such as songbirds, helps assess the plausibility of these scenarios (by reasoning about 'convergent evolution' as explained below).

In this chapter I will survey some of these sources of information to get an idea what form an evolutionary explanation for the human-specific, and possibly language-specific, linguistic abilities should take. But before we embark on a discussion of the anatomy and abilities (Sect. 8.4 and 8.5) of humans and other animals, we must first consider how we can apply the standard approach from evolutionary biology—the comparative method—to a culturally evolved system like language (Sect. 8.2) and why we don't take one of the elaborate theories from linguistics as our starting point (Sect. 8.3).

## 8.2 The Comparative Method in the Light of Cultural Evolution

In investigating the evolution of language we will of course pay special attention to those traits that are unique to humans among the apes—the human-specific traits—which are likely to have evolved since that common ancestor. Moreover, we might want to distinguish, as well as we can, between traits that emerged in human evolution independently from their function in language and those that are in fact language-specific. However, it would be a mistake, for three reasons, to limit our attention to such uniquely human or uniquely linguistic abilities alone. First, one of the most successful approaches in biology for understanding the evolution of particular traits is in fact based on trying to identify commonalities between different species: by comparing many different species and considering the evolutionary relationships and similarities and differences in their ecology, biologists can try to reconstruct the evolutionary history of a trait, and attribute commonalities between two species to homology (the two species inherited the trait from a common ancestor) or analogy (the trait evolved independently in both species due to similar selection pressures, a process known as ‘convergent evolution’). Applying such a comparative method to language turns out to yield a more powerful approach than many armchair theorists stressing the uniqueness of language realized (Fitch 2005).

Second, as explored in other chapters of this book, language is a rather unique system in nature, because it is transmitted culturally from generation to generation and can undergo cultural evolution. For research on the biological evolution of language abilities this is a very relevant fact, because it radically changes what counts as evidence for one theory or another. In particular, it is important to realize that not every difference between humans and other apes is equally interesting, not even if we limit ourselves to traits that are demonstrably relevant to language. To see why, consider that when we compare the vocalizations or learning abilities of any two species, we will necessarily find many differences that are accidental in some sense. In the case of language, we know that the cultural evolution process, where languages adapt to language learners, will result in languages that reflect such accidental properties. The very fact that the peculiarities of languages and those of humans ‘match’ is thus expected even in the absence of biological adaptation.

Adaptations are traits that evolved because they conferred a fitness advantage, that is, because individuals with the traits on average obtained more offspring than individuals without them (‘fitness’ of an individual in evolutionary biology is defined as the expected number of offspring of that individual). When looking for biological adaptations for systems like language that can undergo cultural evolution, we need to look for differences in traits that still have effects on fitness after the process of cultural evolution has unrolled. It would be a mistake to classify as an adaptation every uniquely human trait that is more useful for learning and using human language than an ancestral trait, because the ancestral trait might in fact have been equally good for learning an ancestral language and the good match between humans and modern language a result of cultural rather than biological evolution. Unfortunately, many discussions of language in a comparative perspective make that mistake. For example, Pinker and Jackendoff (2005) list many properties of speech perception that they take to be unique to humans and adaptations, including differences in preferred category boundaries for humans and nonhuman animals and the fact that human neonates have a preference for speech sounds. These features might be unique for humans, however they are more likely accidental features that language adapted to than biological adaptations for language.

Third, another consequence of the cultural evolution of language is that there is no one-to-one correspondence between the ‘human capacity for language’ and the features of individual languages. Human children can learn any natural language, but languages can be very different and not all features of the human language capacity are necessarily exploited by any particular language. Similarly, any particular communication system found among a population of non-human apes might not reflect their capacities to the full. As human languages have evolved culturally to adapt to features of the human brain, the possibility remains that human languages reveal previously hidden talents of the ape brain: features shared with other apes even if they have left no observable effects on ape communication.

## 8.3 Linguistics and Language Evolution

Investigations of the evolution of language naturally start with the question: what is language? The good news is that, at a very general level, linguists all agree: languages are complex, acquired systems of conventions about relations between forms (e.g., spoken or signed utterances) and meanings. The forms are built up by combining elementary units from a basic inventory (phonemes, syllables, hand shapes), and utterances are built up by combining meaningful units (morphemes, words, gestures) into phrases, sentences and discourse, following rule-like patterns. Every human population has language, and in practice, linguists have no difficulty determining which behaviors in an unknown culture count as language, and which as nonlinguistic sounds (e.g., music) or gestures (e.g., dance).

However, the bad news is that the consensus ends at this very general level. The moment we want to make more precise what language in modern humans exactly is, controversies pop up everwhere. For instance, what are those elementary units of form? Even when describing a single language, like English, disagreements abound. Some theories assume the elementary units are phonemes, others that the atomic level is that of 'distinctinctive features' (e.g., Chomsky and Halle 1968). More recently, a popular position is to take larger units—syllables or exemplars—as atomic (Levelt and Wheeldon 1994). And this is only the beginning; much more controversy surrounds more complex units, further removed from direct observation, such as morphemes or grammar rules.

The lack of consensus is even more apparent when considering the full diversity of languages in the world. Languages differ beyond imagination (Evans and Levinson 2009). Some languages build up incredibly long words that convey the meaning of a complete sentence in English; some languages have an almost completely free word order, but mark with a complex system of inflections the roles that various words in a sentence play. Other languages obey strict word order rules, but lack any kind of word morphology, including even plural markers like -s in English. Some language use only a handful of phonemes, others have well over a hundred distinctive atomic sounds. The usefulness of even the most basic concepts of linguistics—'word', 'phoneme', 'subject', 'rule', 'category'—is regularly questioned in the description of one language or another.

Nevertheless, comparison with other animals does quickly make clear that human language is qualitatively very different from any other communication system in nature, even if a convincing, integrated theory of how language works remains elusive. There are interesting questions to be asked about why linguistics is in this state, and why descriptive and theoretical linguistics seem to have so little to offer to solving questions about the evolution of language. I suspect that cultural evolution, and the fact that languages have adapted to the messy idiosyncracies of the human brain, has much to do with it. For the purposes of this chapter, however, the best way forward is to take a pragmatic approach and focus on those aspects of language and speech where empirical research comparing humans and other animals has revealed important qualitative differences—these differences are candidates for the adaptations for language (and speech) that we are after.

## 8.4 Anatomy and Language

### *8.4.1 Speech Production*

When we look at the anatomy of the human speech production and perception apparatus, we see a strong continuity with the other great apes and even the broad class of mammals. To produce sounds, many mammals, like humans, let air flow from the lungs through the larynx, the throat and the nose and mouth. The larynx contains special membranes, the vocal folds, which vibrate in the air flow and can be tightened or loosened to produce higher or lower pitched sounds. The cavities between larynx and the lips together form the vocal tract, which effectively filters the mesh of sounds created by the larynx, by reinforcing some frequencies (resonances) and attenuating others. Three features of the human anatomy used in speech production stand out (see Fig. 8.1): the fact that the larynx is very low in the throat, that humans, unlike all other apes, have no air sacs, and that humans have detailed and rapid control over the shape of the vocal tract (see Crystal 1997 for an accessible review of the human anatomy involved in speech production and perception).

The human larynx is high in the throat in babies (allowing them to breathe and drink at the same time), but descends to the lower position as they get older. In males, there is a second descent of the larynx during puberty. The position of the larynx is very relevant for speech as it determines the length of the vocal tract, and the size and shape of the vocal tract in turn determine the quality of the sound that comes out. Naturally, human vocal tracts are ideal for producing human speech sounds, but is the system as a whole ‘better’ in some way? Philip Lieberman (1984) has argued that the descended larynx allowed a much richer repertoire of speech sounds, and could thus confer a fitness advantage that offsets the disadvantage of an increased probability of choking. Lieberman went as far as claiming that this was the crucial innovation in the evolution of language. Although recent modelling work (de Boer 2010) upholds his claim that the human vocal tract is optimal for producing a range of distinctive sounds, the effects are small and unlikely to be the crucial factor in the evolution of speech.

Moreover, the permanently descended larynx turns out to be not uniquely human but is also found in red deer and other species without language (Reby et al. 2005),

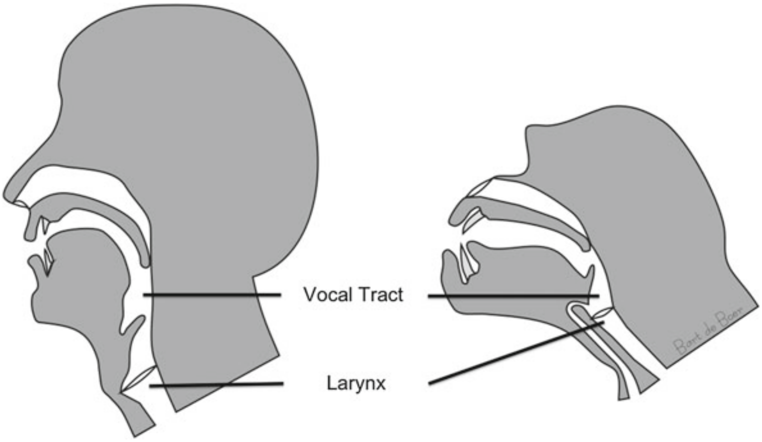


**Fig. 8.1** Human (*left*) and chimpanzee (*right*) vocal anatomy differ in three important respects: in humans, the larynx is positioned lower in the throat and the tongue is rounder, yielding a vocal tract with equivalent controllable back and front cavities; humans have no air sacs attached to the larynx (chimpanzee air sacs are attached to the larynx through the narrow tube that can be observed in front of the larynx), further improving the range of sounds that can be produced; humans have voluntary control over the movements of the vocal folds. Diagram courtesy of Bart de Boer; based on FMRI data in Fitch (2000)

strongly suggesting that there is at least one other biological function for a descended larynx. Fitch (2000) suggests this other function might be size exaggeration: with a larynx low in the throat one can make sounds that otherwise only much large animals could make. Finally, there is at least one mammal without a permanently descended larynx that is, under exceptional circumstances, very good at imitating human speech: recordings from the harbor seal Hoover, raised in a fisherman's bath tub, contain a few intelligible sentences (Ralls et al. 1985). Hence, the position of the larynx might very well have been a target of natural selection for speech once rich languages had emerged among hominids, but it is unlikely to be a crucial factor in the emergence of a rich language in the first place.

Much less attention has been given to the absence of air sacs. All other apes have such sacs: cavities attached to the larynx that can range from modest in size (chimpanzees) to clearly visible balloons in the neck (gorillas). It's clear that air sacs have an acoustic effect on the vocalization produced, and various researchers in the last century have formulated the hypothesis that humans lost air sacs because of a detrimental effect on speech comprehensibility. In recent modelling and experimental work, de Boer (2009) confirms the suspicion that air sacs have such a detrimental effect. However, as with the descended larynx, the effects are not enormous. The loss of air sacs is likely to have been affected by evolutionary pressures for speech, but it is unlikely to be the key event that set all the rest in motion.

One fascinating aspect of air sacs is that they have left traces in the archeological record: the shape of the 'hyoid bone' in the throat correlates with the presence or absence of airsacs, and this bone is occasionally but very rarely fosselized in ape and hominin fossils. Based on the few findings reported, we can make a rough estimate of the disappearance of air sacs: Australopithecus afarensis, an human ancestor that lived about 3.3 million years ago still had air sacs (Alemseged et al. 2006), while Homo heidelbergensis, who lived some 600,000 years ago did not (Martínez et al. 2008; for a review of homin fossils and approximate timelines, see Jones et al. 1992).

A third anatomical oddity of human speech is the extremely rapid control over vocalizations, with precise, millisecond-level synchronisation of movements at distant places in the vocal tract, from the larynx to the lips. On high-speed x-ray films of the human vocal apparatus one can see complex, extremely fast and accurate movements that tongue, lips and other articulators make when producing a string of words. Although it is difficult to quantify, nothing comparable has been reported in the vocalizations of other primates. In song birds, however, we do see extremely fast and complex vocalizations as well, with precisely timed simultanous movements in syrinx (the bird equivalent of the larynx) and beak. An open question is whether birds, like humans, deliberately manipulate the resonance frequencies of the vocal tract (e.g., by moving the tongue or by opening air sacs), but preliminary evidence (Ohms et al. 2010) seems to point in this direction. In both humans and song birds, but not other primates as far as we know, forebrain regions seem to be involved in the control over vocalizations (Deacon 2000). Combined with the fact that in other primates we find only limited vocal repertoires and relatively simple and slow vocalizations, the findings on the extremely versatile articulatory control in humans suggests that evolutionary innovations could have been essential for the high rate of information transfer through speech that modern humans are capable of.

In short, there are some likely biological adaptations to the anatomy of the vocal tract that would have improved communication through speech, but none, it seems, that would have been necessary for language to emerge in the first place. There are some further likely biological adaptations in vocal control; these, in contrast, might have been essential for language, in the vocal-auditory channel at least, to confer a fitness advantage.

### *8.4.2 Speech Perception*

In human speech perception, the relevant anatomical structures seem even more similar to what is common among mammals. Although the shape of the outer ears vary widely, the middle and inner ears of other (land) mammals—exquisitely complex organs—are very comparable (although the shape of the outer and inner ears might be responsible for increased sensitivity of humans in the frequency range needed for speech perception, Martínez et al. 2004). Behind the ear drum, we find the same hammer, anvil and stirrup that conduct the vibrations to the cochlea, where they are translated into neural activation patterns in the Organ of Corti. The auditory nerve then transports these signals to higher processing levels in the brain.

Behaviorally, however, human speech perception does seem special. In the 1960s, pioneering research by Alvin Liberman and colleagues (Liberman et al. 1967) revealed that subjects could still perceive differences between two different speech sounds at high levels of background noise where differences between other sounds are lost. On the other hand, subjects did not perceive much difference between two versions of the same phoneme, even if the physical difference was of similar magnitude. This phenomenon, that differences between categories are perceived much more clearly than differences within a category, is called "categorical perception". Liberman and colleagues further found that subjects especially perceive physically different sounds as similar if they are produced with similar movements. This led the authors to propose a "motor theory of speech perception" which states that speech is perceived in terms of the movements necessary to produce it.

Together, these results led to the "speech is special" hypothesis, which states that humans are biologically specialized for speech perception. Much interesting research has been published since, comparing human and non-human perception of linguistic and non-linguistic stimuli. Where human speech perception was found to differ from non-speech or non-human perception, such findings were often claimed to be 'adaptations' (e.g., Pinker and Jackendoff 2005). However, as we saw before, the close fit between language (in this case, speech sounds) and human abilities (in this case, speech perception) is not in itself conclusive evidence for biological adaptation.

One recent finding by Smith and Lewicki (2006) is telling in this respect. They considered the different ways in which neural firing patterns can encode auditory information; such neural encoding is ultimately what the inner ear does with auditory input, to send information for further processing to the brain. Some neural codes are very efficient for one type of acoustic input, others for other types of input. Smith and Lewicki discovered, to their surprise, that the code used in the inner ears of cats (as derived from neurophysiological studies) appears optimised for human speech. As it is very unlikely that cat's hearing has indeed been adapted to human speech since their (in evolutionary terms) recent domestication, the only sensible explanation is that the causality is the other way around: human speech exploits those sounds that the mammalian auditory system can most efficiently process. And, indeed, in the same study, Smith and Lewicki find that the same encoding is also optimal for the sample of sounds they created from a mixture of ambient sounds (water flowing, cracking twigs) and animal vocalizations.

Hence, although the empirical discoveries by Liberman et al. still stand in broad outline, Smith and Lewicki (2006) and many other studies since the 1960s have put the original interpretation into perspective. On the perception side, it looks like humans make use of a biological apparatus that hasn't fundamentally changed from our prelinguistic ancestors, although there are many human peculiarities that languages would have adapted to culturally.

## 8.5 Design Features of Language

### *8.5.1 Cultural Transmission and Vocal Learning*

When we turn from aspects of human anatomy to less tangible, structural aspects of language, a list of 'design features' of human language by Hockett (1960) is a useful starting point. This list has since been a focus for research comparing language with natural vocalizations in other animals, in particular the four main design features: cultural transmission, symbolism, duality of patterning and hierarchical phrase-structure. I will first focus on cultural transmission, which refers to the fact that languages, within the constraints set by our biology, are conventional systems that persist through time by repeated learning. This is true for the elementary sounds of spoken languages, for the elementary shapes and movements in sign language, and for all the grammatical rules and constructions. Cultural transmission is not unique to language or humans—we also observe it in, e.g., music and bird song—but rare among primates and a key qualitative feature of language.

Focusing on sounds, cultural transmission is possible thanks to the ability for vocal imitation: the ability to relate perceived sounds back to the articulatory movements that can produce it. Vocal imitation is, as far as we know now, absent or very limited among other primates (Janik and Slater 2000), with the possible exception of gibbons. Early language-training studies with apes famously failed to get the apes to produce any speech-like sound, and natural vocalizations in monkeys and apes appear to be innate (i.e., develop independently from exposure to those sounds, cf. Seyfarth and Cheney 1997). Among mammals, the only groups other than humans with vocal imitation are seals, bats, elephants, dolphins and whales. For instance, humpback whales sing long songs, that are shared among members of one population of multiple generations, but differ from population to population and gradually change over time (Payne and McVay 1971). Among birds, finally, there are very many vocal learning species, but they are limited to three groups: songbirds, humming birds and parrots.

Although it is not completely clear what the criteria for true vocal learning are (evidence for vocal learning in the mentioned species sometimes comes from experiments with controlled training stimuli, sometimes from field observations of imitation or cultural transmission; Janik and Slater 2000), it does seem clear that advanced vocal learning is rare but found in multiple species scattered over the evolutionary tree of life. This presents a wonderful opportunity to investigate the possibility of convergent evolution: in various branches of the evolutionary tree similar solutions evolved for similar problems. Why is vocal learning rare? What are the difficulties or disadvantages preventing most species from having the ability, and what are the advantages that drove its evolution in the species, including humans, that do have it?

The question about difficulties is all the more pertinent, because vocal learning—from a computational point of view—is not something particularly complex. Some existing computer models of vocal learning might help to find an answer. For instance, Westermann and Miranda (2004, see also Oudeyer, this volume) present an elegant model of neural structures that can learn mappings from articulatory movements to sounds, and vice versa, and thus implement vocal learning. The model consists of two neural maps, one representing motor activity and the other perceptual input. It assumes a babbling phase, where the learner initially produces sounds at random, and ‘articulatory feedback’, meaning that he can hear himself. Given those assumptions, the model shows how (Hebbian) connections between neurons in both maps can come to encode the relation between movements and sounds. When learning is complete, the model can be prompted with just a sound and then produce, in the motor map, the pattern of activity needed to generate that sound. Hence, it shows the potential for vocal imitation. A straightforward extension of the model with a visual map also makes correct predictions about the influence of seeing lip movements on the sounds perceived (the so-called McGurk effect).

Comparing the model to real brains, there are two clues to what might make vocal learning difficult in reality. First, the connections between maps in the model are bidirectional: the same connections are used for predicting sounds given motor parameters as for vocal imitation. In real brains, the ability to predict sounds given motor activity is likely to be common among animal species, but because neurons are not bidirectional, a dedicated pathway might be necessary to also learn the inverse mapping. Second, motor and sound map activity are static in the model; in real brains, the motor maps will already be involved in planning for the next vocalization by the time the articulatory feedback arrives (Dave and Margoliash 2000). This thus necessitates a memory-motor map in addition to the motor and perceptual map in the model, and a dedicated pathway to transfer information from the memory-motor map back to the actual motor map for the production of vocal imitations.

Brain research on song birds has indeed found evidence for such dedicated pathways; intriguingly, the solutions found in the independent evolution of vocal learning in song birds, humming birds, parrots and humans appear to be very similar (Jarvis 2004). Jarvis (2006) argues that this is indeed a case of convergent evolution, and observes that there is one thing the species with vocal learning have in common that distinguishes them from many non-vocal learning sister species: they are at the top of the food chain and often have few, if any, predators. This observations needs more systematic research, but it could be a key factor, because vocal learning of a complex repertoire of sounds requires practice. Practice sessions, where infants spend time and energy (that could otherwise be used for more direct ways to increase survival) and make noise that attracts predators, might simply not be a viable option for species that are under predation pressure.

Given that vocal imitation is possible, but not trivial and likely to be costly, the question arises what the evolutionary advantages could be in song birds and, ultimately, in humans. Jarvis surveys a number of popular ideas in the literature, including the idea that learned vocalizations allow for individual identification and for cultural adaptation to diverse habitats where different types of sounds might transmit better. Coming from bird song research, he favors the hypothesis that the variability allowed by vocal learning played a key role in mate attraction; given the attractiveness of human singers to members of the opposite sex this is indeed a serious candidate selection pressure in human evolution as well.

However, the increase in the number of distinct sounds that can be produced could also have played a different role in humans than in other vocal learners, in particular in ‘semantic communication’. Jarvis is skeptical of a role for semantic communication in the evolution of vocal learning, and correctly points out that in song birds signals that carry meaning (like food or alarm calls) tend, in fact, not to involve vocal learning, while vocally learned song has no referential role. However, modern human language is very different in this respect, and Jarvis might stretch the songbird analogy a little too far here. In any case, I see no reason why the selection regime that allowed large, meaningful vocabularies to emerge in humans could not have played a role in the emergence of vocal learning in the first place.

In conclusion, vocal learning is a rare trait in nature but crucial for spoken language. The comparative record provides some clues about the questions as to why we have this trait and many other animals do not. Our position as a top predator might have removed selection pressures against it, while the need for a great variety of sounds in communication—useful in mate attraction or semantic communication—might have provided selection pressures for it. Firm conclusions, however, cannot be drawn at the current state of knowledge and alternative (but not necessarily contradictory) theories exists. For example, Lachlan and Slater (1999) propose, based on a mathematical model, a “cultural trap hypothesis” which states that once vocal learning has emerged and a variable repertoire is used in a species, for whatever reason, vocal learning is favored over innate vocalizations. Oliphant (1999) proposes that the difficulty of identifying the intended referents in learning a lexicon was a crucial obstacle in the evolution of cultural transmitted semantic communication in other species.

Perhaps genetic evidence, as is now starting to emerge (see Dediu, this volume) will play a role in the future in understanding the evolution of vocal learning. FoxP2 is a gene involved in speech and language, as discovered in the study of a family with a heritable disorder affecting several speech, language and motoric abilities. Through the careful work of Vargha-Kadem and colleagues (Vargha-Khadem et al. 1995) it has become clear that the gene is not specific for language (as proponents of an extremely modular view of the mind were perhaps hoping), although it does indeed seem to affect linguistic abilities over and above the indirect effects one can explain from effects on general intelligence and motoric abilities. Interestingly, the same gene also plays a role in vocal learning in birds (Haesler et al. 2004). Studies of variants of the gene in other species, including the extinct Neandertals (Krause et al. 2007), are starting to provide a fascinating look on the evolution of the gene, but given the many unknowns about the exact function of the gene it is too early to directly relate it to scenarios of the evolution of vocal learning and language.

### *8.5.2 Symbolism and Arbitrariness*

A second feature of natural language that is often said to be unique is its ‘symbolism’, but this term can mean various things. One aspect of symbolism, featuring in most definitions, is that the relationship between the words or morphemes in a language and what they refer to is arbitrary. Thus, there is nothing in the sound of words like 'sleep', 'green' or 'democracy' that is in any way similar to what these words denote. Even onomatopoeia—words that do mimick the sound they describe, such as 'cock-a-doodle-do' in English—are to a large part conventionalized, as can be seen from the fact that the same rooster's calls are referred to as 'kukeleku' or 'cocorico' in other languages (Dutch, Italian). Thus, there is no doubt that humans have the ability to assign arbitrary meanings to arbitrary sounds, and they do so all the time: adult native language users typically know many tens of thousands of words (Bloom 2000). In that massive vocabulary, some words sound or look somewhat like what they denote, but the vast majority of word-meaning mappings are arbitrary (e.g., Tamariz 2005).

This 'arbitrariness of the sign' is a feature of natural language that is cherished by many linguists, but arbitrariness per se might be less relevant from the comparative and evolutionary point of view than has often been assumed. There are many alarm call systems—in birds, primates, rodents—where particular sounds denote particular predators (or better, perhaps: denote the appropriate response to the presence of a given predator) and where there seems to be no relation between the sound and its meaning. Learning these associations is common too: although the production of calls is typically thought to be innate, the interpretation of calls is somewhat flexible, and different species of monkeys are known to be able to learn to interpret each others alarm calls (e.g., Zuberbühler 2002). Moreover, arbitrariness is also not the all-or-nothing phenomenon that it has often been taken to be. In human sign languages (known to often be much more iconic than spoken languages; Frishberg 1975) and ape gesturing (e.g., Tomasello et al. 1997), it has often been observed that gestures that start out as iconic, can gradually become more and more arbitrary (i.e., for an external observer the original iconic relationship between gesture and meaning is less obvious or even unobservable in later stages).

Other features of human vocabularies might be truly unique to humans, but also don't necessarily point to language-specific adaptations. Quantitatively, the readiness with which humans acquire a vocabulary is remarkable. Children start understanding and using their first words around their first birthday, and after a slow start, are estimated to learn 10 words a day between age two and six, reaching a vocabulary of about 14,000 words by age six, which further increases to perhaps as much as 60,000 words at high school age (O'Grady 2005). These numbers dwarf any vocabulary size found in non-human animals, where chimpanzee gesture repertoires are estimated to contain at most a few hundred signs. The record-holder in vocabulary size, as established in a controlled experiment, is held by a dog: border collie Rico can recognize about 200 names for objects (Kaminski et al. 2004). Thus, there is a huge quantitative gap between humans and other animals, but this does not prove that humans have language-specific mechanisms for word learning and usage. The gap could also be an indirect consequence of differences in reasoning abilities, in particular in abilities to reason about the intentions of others (Oliphant 1999; Bloom 2000). Moreover, it could be an indirect effect of evolutionary innovations in grammar (as discussed in Sects. 8.5.3 and 8.5.4): most words are learned concurrently with the grammar of a language, and the grammatical context provides additional clues to word meaning (Cruse 1986).

Returning to the qualitative differences, another well-known observation about word learning is a "mutual exclusivity bias" in children: a preference for 1-to-1 mappings from words to objects, without synonymy (several words with the same meaning) and homonymy (identical words with multiple meanings). Interestingly, the mentioned border collie Rico seems to share this bias with humans: when confronted with a novel word, he was more likely to associate that with an object that he did not already know a word for. This suggests we do not need to assume a language-specific adaptation to explain this bias in humans, but can rely on general cognitive and communicative processes.

Finally, in most definitions symbolism is more than just the arbitrariness of word meanings. Harnad (2003) defines a symbol as an object that not only has an arbitrary meaning, but is also part of a symbol system. A symbol system, in turn, is a system of symbols and rules, where rules apply to symbols regardless of their meaning. Thus, the word 'cat' is a symbol, not just because its sound is in no way similar to the animal it denotes, but also because it participates in a system of rules and many other symbols: the English language. The rules of English (such as those that put the determiner 'the' in front of it, or the plural marker '-s' behind it) apply to it because of its syntactic category ('noun') and not because of what it means. The part-of-a-system requirement in Harnad's definition means that it makes no sense to speak of symbols in isolation (as Harnad himself point out); therefore, it seems to me that symbolism is an inseparable consequence of compositionality, which I will discuss in the next section.

Using similar arguments as Harnad, Paul Bloom (2004) has warned against over-interpreting the analogies between human word learning and the 'words' learned by dogs and other animals. He points out that the research with animals has not demonstrated that they can combine words for objects with all kinds of action words. This, however, is again the part-of-a-system requirement, and, I would argue, inseparable from compositionality.

For many scholars, there is even more to the symbolic nature of language than the part-of-a-system aspect. They feel there is something special about the relation between human words and the concepts they symbolize, but typically do not give a precise definition of that special, symbolic relationship. This doesn't stop many of them from speculating about the relevance of the appearance of art, from about 100,000 years ago, in the archeological record for scenarios of language evolution. This, they argue, is a strong clue for language, as art, like language, requires symbolic thought. I am skeptical about the confidence with which they argue for this point: it is hard to imagine any hominin species with only part of the human suite of cognitive abilities, but that is exactly our job. I don't see a priori reasons why art without language would be harder than language without art, and the comparative record has unfortunately little to say on this issue (see Botha 2008 for a detailed critique of several such proposals).

### *8.5.3 Duality of Patterning*

To a first approximation, sentences in languages are built up from meaningful words (or rather: morphemes), and words are built up from meaningless phonemes. Although the situation is more complicated than that, it seems fair to say that human languages employ at least two combinatorial systems: a combinatorial, phonological system that regulates which basic sounds can be combined into possible words, and a compositional, semantic system that regulates how words and their associated meanings can be combined to give sentences and compound meanings. Both combinatorial systems generalize to unseen sequences: we can interpret sentences we have never seen before, and distinguish impossible from non-existing but possible words in our native language (e.g., the French word "pluie" is an impossible word in English, because of the onset /plj/, while the Dutch word "vonk" is a nonexisting but perfectly possible word in English). Hockett (1960) used the term 'duality of patterning' for the marriage of a combinatorial phonology and a compositional semantics (somewhat earlier, Martinet 1949 had already made the same observation using the term 'double articulation').

Most animal vocalizations, in contrast, are holistic: a single vocalization has a particular function, but there is no sense in which we want to analyze the vocalization as built-up from components that are reused in other vocalizations. It is interesting to look at the exceptions to this general statement, where there are three aspects to pay attention to: are vocalizations built up from several elementary units? Do these vocalizations have a referential meaning? And is the meaning of a combination somehow a function of the meaning of the parts? As always in biology, we find an enormous variety in nature and we do observe that combinatorial phonology and compositional semantics have their echos in other animals' communication systems. However, there are important qualitative differences and the presence of both in one species has, as far as I know, only been attested in humans.

When we only focus on the combination of vocal elements, there are in fact still quite a lot of examples in primates and cetaceans, and especially among songbirds. In many song birds we find distinct repertoires of basic elements that can be combined in various ways but according to quite strict rules. A good example is chaffinch song: Riebel and Slater (2003) describe the repertoire of a population of chaffinches, and the rules that govern the structure of the songs. Each (male) bird sings two or three different songs, and each song follows a stereotypical AxByF or AxByCzF pattern. Elements in the A, B and F part are repeated a varying number of times, but the elements that make the x and y part, called transitional elements, are never repeated although they can be omitted. Extremely similar A elements can be found combined with different Bs and Cs. From such findings, it has become clear that chaffinches have a combinatorial system in place. It differs, however, qualitatively from what we find in language. Most importantly, the songs do not convey a referential meaning: songs have a function in attracting females and defending the male's territory against rivals, but the message to them is always the same: come here! or go away! Variations in, for instance, the number of repetitions do not change that message.

That means we cannot speak of duality of patterning, as there is only one system of combination, and we can perhaps not even speak of combinatorial phonology, as the term ‘phonology’ is usually reserved for meaning-carrying vocalizations/gestures. It also potentially limits the usefulness of looking at bird song for understanding the evolution of combinatorial phonology, because a crucial constraint that presumably operated in the evolution of phonology—the system must remain useful for encoding and decoding information—was missing in bird song evolution.

Combination of vocalizations that do carry meaning is, in contrast, very rare. Arnold and Zuberbühler (2006) describe a communication system in putty-nosed monkeys that fits the bill. The monkeys use distinct, loud alarm calls to warn each other of predators: they emit the so-called pyow call when a leopard is detected in the vicinity, and a hack call for eagles. Additionally, the monkeys sometimes produced pyow-hack sequences, consisting of 1–3 pyows followed by 1–4 hacks. These sequences are produced in response to both eagles and leopards, and are typically followed by the whole group of monkeys moving to a different area of the forest. In their study, Arnold and Zuberbühler demonstrated experimentally that the pyow-hack sequences indeed mean something different than the individual pyows or hacks. They played leopard growls to 17 groups of monkeys, each consisting of a single adult male, with several females and their offspring; in about half of the groups, the adult male responded with a pyow-hack sequence, and those groups were found to have moved significantly further away 20 min later than the groups that only responded with pyows—the leopard alarm call. Putty-nosed monkeys thus have a rudimentary form of combinatorial phonology: elementary sounds, used to denote the two predators, are reused to form a third signal which roughly means: “let’s go”. But the putty-nosed monkeys do not exhibit compositional semantics, as the meaning of the combined signal is not somehow derived from the meanings of its component parts.

In another study, however, Zuberbühler (2002) did find rudimentary compositionality: Campbell’s monkeys also have a system of alarm calls for various predators, and aside from the usual unitary calls, also sometimes produce sequences of two calls. The first call is a so-called ‘boom’, and modifies (weakens) the meaning of the second call, a leopard or eagle alarm. Zuberbühler experimentally demonstrated that Diana monkeys—another monkey species living in the same habitat and eavesdropping on the Campbell’s calls—withhold the usual response to the alarm calls if they are preceded by the boom. The meaning of the whole, it seems, is thus a function of the meanings of the parts. That places the Campbell’s monkeys in an odd class of species for which some form of compositional semantics has been attested. The only other members of that class are humans and some species of bees, who convey the location of a food source through a dance where two components of the form (direction, length) map onto two components of the meaning (direction with respect to the sun, distance from the beehive; von Frisch 1974).

The results from Zuberbühler et al. are important, but they do only demonstrate very rudimentary forms of either side of the duality of patterning. They add to the evidence that combining two signals to mean something new, and combining meanings to create a compound meaning, are feats that do not necessarily require a language-adapted brain. Although it's novel to see such evidence in natural communication systems, we already knew from trained apes, dogs and other domesticated animals that a rough combination of the meaning of two sounds is within reach of those animals (e.g., Truswell, in prep). Compositional semantics in natural language, however, is quite a bit more sophisticated. For instance, in many languages word order is a crucial variable, such that 'dog bites man' means something different from the sentence in the reverse. Such a phenomenon has never been attested in any non-human animal.

We shouldn't be surprised if a monkey communication system is soon discovered that combines the tricks of the Campbell's and putty-nosed monkeys, and thus provides a rudimentary duality of patterning. But exciting as such a finding would be, such a rudimentary form would not tell us much about the evolution of duality in human language. What we would really like to know is whether the ability for an extensive duality of patterning is already lurking inside the primate brain, but species other than humans lack the motivation to use it, or whether we need dedicated brain structures to be able to process it. To answer these questions, we need to know much more about the neural and cognitive mechanisms that underlie duality of patterning in humans. From a purely computational point of view, it is hard to see why compositional semantics would be particularily difficult for a monkey, ape or bird brain that can already readily process combinatorial conceptual structure (such as needed for planning, vision and social cognition) and combinatorial signals. The only obvious difficulty derives from the fact that compositional semantics requires combinatory operations to apply to representations of meaning and representations of form in synchrony, in such a way that the system become bidirectional: language users must be able to compute the meanings of a given form, or the appropriate form for an intended meaning using the same rules of language. There might be difficulties, though, in integrating pieces of information that are processed in different parts of the brain, similar to what we saw in the case of vocal learning. More research is necessary on the particularities of how primate brains handle conceptual structure and communicative signals.

Meanwhile, we can ask what the evolutionary costs and benefits of duality of patterning could be, assuming that brains can implement it. Martin Nowak and others studied a number of interesting mathematical models that bear on this question (e.g., Nowak and Krakauer 1999; Zuidema and de Boer 2009). The basic insight underlying this work is simple: combinatorial systems can convey many more messages than holistic ones with the same number of elementary units; e.g., a system with 10 nouns and 10 verbs can handle $10 \times 10 = 100$ distinct noun-verb combinations, whereas an holistic system of the same size can only convey $10 + 10 = 20$ distinct messages. How many messages do we need to convey? Nowak and Krakauer reason that species will differ in how many distinct signals they will want to communicate with; we will call that number N. Now, it is reasonable to assume that the number of distinct elements that can be learned, remembered or distinguished from each other is limited to some number M. Moreover, we must assume a cost to using a combinatorial strategy (for instance, because learning 20 nouns and verbs is more difficult than learning 20 holistic signals, or because combinatorial strategies use up more memory and energy), but for a given M and cost there is always some number N at which combinatorial strategies will outperform holistic strategies. Nowak and colleagues therefore suggest that a possible explanation for why humans have combinatorial phonology and compositional semantics is that they were more cooperative and wanted to communicate more distinct messages than other primate species. In other words, the human N is above the threshold for combinatoriality, while the chimpanzee N is not.

A more difficult question is how natural selection could have driven the transition from holistic systems, to communication codes with either combinatorial phonology or compositional semantics, or both. In Nowak and Krakauer (1999) and related papers, Nowak et al. show that in species that have both a holistic and combinatorial system in parallel, the evolutionary dynamics will, under reasonable assumptions, always lead to using the combinatorial system more and more. However, Zuidema (2003) and Zuidema and de Boer (2009) argue that assuming two systems in parallel makes for a rather unrealistic scenario, and show that in a single system optimization for noise robustness can yield systems with both rudimentary compositional semantics and combinatorial phonology.

In sum, the extensive duality of patterning of human language—with its combination of meaningless phonemes into words, and of meaningful words into meaningful sentences (compositional semantics)—is unique in Nature. From a computational point of view, the most likely obstacle in the evolution of compositional semantics has been the necessity to perform operations on phonetic form and semantic structure in synchrony, perhaps requiring dedicated neural pathways. The most likely driving force for its evolution has been a selection pressure for an expressive, robust and learnable communication system under circumstances for learning and communicating with noise and time pressure.

### *8.5.4 Hierarchical Structure, Syntax and Recursion*

Even simple utterances in natural languages go far beyond the rudimentary compositionality of the Campbell's monkeys. First of all, they are not limited to combining two elements to create a third; the result of one combinatory operation in languages is usually again the input for the next combinatory operation. Thus, in a sentence like 'happy people sing', we first combine the meanings of happy and people, and then combine the resulting compound with sing. Human languages thus show hierarchical compositionality.

Moreover, words and phrases come in different categories. 'Happy' is an adjective that can modify the noun 'people'; combined they form a noun phrase that can be the argument of the verb 'sing'. Importantly, the syntactic categories of words and phrases, that determine what can be combined with what, are not always predictable from their semantics. Not only can we assess the grammaticality or ungrammaticality of nonsensical sentences (such as Chomky's famous pair 'colorless green ideas sleep furiously' vs. 'furiously sleep ideas green colorless', Chomsky 1957), but syntactic constraints can also make sentences impossible that would semantically work perfectly well ('*the asleep child', '*John sang the Marseillaise his heart out', Culicover et al. 2004). Thus, natural languages employ a system of syntactic constraints that functions, at least in part, independently from semantics. The parts of a sentence over which syntactic constraints are defined are called phrases or constituents. Also for such syntactic phrases it is true that they do not always correspond one-to-one to semantic units; the hierarchical structure in the syntactic domain is called hierarchical phrase-structure.

Finally, in natural language sentences we can observe that a phrase of one particular syntactic category can be embedded in a phrase of the same syntactic category. Thus, a phrase like 'the man on the moon' is a noun phrase, but embedded in it we find another noun phrase: 'the moon'. A sentence like 'Luggage people leave behind is destroyed', contains 'people leave behind x' that linguists analyse as being of category sentence. This property of language is called recursion. In the debates on language and evolution a further distinction has played a key role: if the embedded phrase always ends up on the right or left edge of the larger phrase (as in the first example) this is called tail recursion; if phrases get nested in the middle of the larger phrase it is called center-embedding (as in the second example).

There is much disagreement in linguistics about the exact nature of hierarchical compositionality, phrase-structure, syntax and recursion, but there is no doubt that human languages show patterns that invite descriptions in these terms. In animal communication systems, in contrast, there is very little that comes close. In some song bird species, the song repertoires invite a description in terms of so-called finite-state machines (or hidden markov models): many songs here share a similar overall structure, but, for particular parts of a song, they differ in the number of repetitions of one or more elements or the choice for one variant or another (e.g., Okanoya 2004). Although birdsong researchers describe this as 'song syntax', it's clear that it's very different from language: there is no semantics that the syntax can be independent of, and there is no real sense in which the system is recursive (let alone exhibiting center-embedding).

In humpback whale song researchers have also discovered relatively complex structure. Researchers describe the songs as being built up from themes, consisting of phrases, consisting of units, in turn built up from subunits. Hence, whale song might rightly be characterized as hierarchical (and a similar case can be made, though less pronounced, for many bird song species). However, there is no reason to assume a compositional semantics for these songs or a recursive structure. Also Suzuki et al.'s 2006 sophisticated analysis of humpback song does not establish the need for a descriptions in terms of center-embedding, even if it does reinforce the conclusion that the songs are hierarchical.

Finally, a strong animal contender for the ability to process center-embedded, hierarchical structure is the bonobo Kanzi, who was exposed to human language from birth. Kanzi has been at the center of a long standing controversy about the language-abilities of apes. Unfortunately, the facts about what Kanzi could and could not do are hard to obtain. One side, represented by lead researcher Sue Savage-Rumbaugh, has tried to make the case for very advanced abililities ( e.g., Savage-Rumbaugh and Lewin 1994), but much of the presented evidence consists of video footage (which lacks crucial statistical information) or experimental data from designs that aren't up to today's standards in the behavioral sciences. The other side has often been dismissive without access to the relevant data and seems to have been driven in part by preconceptions about an innate language faculty (e.g., Pinker 1994).

An interesting exception to this state of affairs is a recent paper by Rob Truswell (in prep), who reanalyzed a database composed by Savage-Rumbaugh et al. of spoken instructions to Kanzi and his responses. Truswell finds that Kanzi's performance is impressive in general: the ape seems very well capable of combining the meanings of several words. However, in most of the sentences used in the database, a correct interpretation of the instruction is not dependent on sensitivity to the hierarchical structure of the sentence. Truswell identifies a class of sentences where this sensitivity is crucial (sentences with NP-coordination) and finds that Kanzi's performance on those sentences is at chance level. These are sentences like 'Kanzi—put the coke and the milk in the fridge'. Assuming that Kanzi knows the meanings of all content words and knows that fridges can't be put into coke or milk (which is indeed an impressive achievement already), there are four possibilities for what goes into the fridge: nothing, coke, milk or both. Averaging over all 18 such cases in the database, Truswell finds that Kanzi is only correct 22% of the time.

A quite different approach to comparing syntactic abilities between humans and other species is pioneered by Fitch and Hauser (2004). They tested Tamarin monkeys on their ability to detect particular patterns in sequences of syllables. When one group of Tamarins had heard sounds conforming to the pattern ABAB or ABABAB, they reacted with surprise when confronted with the sound patterns AABB or AAABBB. However, another group of Tamarins first heard AABB/AAABBB and then failed to notice the change to the other patterns. Because the second pattern is typically used in mathematical work on center-embedded recursion, Fitch and Hauser interpreted these results as showing that the monkeys were unable to process such center-embedding. Subsequent work has shown that starlings (Gentner et al. 2006) and zebrafinches (van Heijningen et al. 2009) can, like humans, learn to distinguish between the two types of patterns. However, although the earlier papers generated much debate about whether or not animals can process recursive structures, van Heijningen et al. argue that the experimental set-up used in the experiments is problematic and that none of the results so far has really demonstrated the ability or inability to process center-embedding in any species. Rather, the results of van Heijningen et al. show that each of the zebrafinches in their experiment exploit one of many possible non-recursive strategies to successfully distinguish grammatical from nongrammatical stimuli, and that the statistical analysis from earlier papers, where results from multiple subjects were averaged, fail to correctly control for these alternative explanations. Hence, although this type of study might become important in the future to answer comparative questions about grammar, current results are inconclusive and more research is needed on this issue.

In conclusion, the way humans combine meaningful words to form complex sentences, guided by a system of semantic and syntactic categories and rules (collectively labeled grammar), is unique in nature. The computational complexity of this behavior, the absence of anything similar in animal communication, the failure of extremely intelligent apes to master it, and the fact that it makes language an extremely powerful system, together make a strong case that there is a true adaptation at play here.

Interestingly, for many of the claimed unique design features of language, the uniqueness seems to depend on the presence of grammar: words are symbols because of grammar, words might be learned efficiently because of grammar, talking about things remote in space and time (Hockett's displacement) is possible because of grammar, and human compositional semantics differs fundamentally from that of bees and Campbell's monkeys because of grammar. Hierarchical phrase structure and the possibility of recursion and center-embedding, follow, it seems to me, from the way grammar allows us to combine words. The core component of grammar is hierarchical compositionality (other components are the syntactic constraints that are independent from semantics, but these are less crucial for communication and plausibly the result of preexisting idiosyncracies of the human brain); hence, hierarchical compositionality is at the top of the list of candidate features that make human language unique.

## 8.6 Towards an Evolutionary Scenario

### *8.6.1 Evolutionary Scenarios: Why and How?*

So far, I have reviewed a number of traits of humans that seem directly involved in speech and language, and enquired to what extent they are shared with other animals. This exercise has led me to identify a number of candidate adaptations, some of which seem essential for a spoken, complex language to have emerged at all (vocal learning, vocal control, grammar), whereas others are more likely to be consequences of the new selection pressures that the use of a spoken language brought (optimized vocal tract shape, loss of air sacs). How do those fit into an evolutionary scenario that explains why humans and not other species have language?

By formulating a specific scenario I risk being accused of entering the realm of speculation, as so many theories on language evolution did before. However, as long as we emphasize the hypothetical nature of any favoured scenario, I don't think much harm is done. Moreover, complete scenarios are in fact necessary if we want to investigate the relation between various proposed adaptations and evaluate the plausibility of each step in the context of the other steps. Knowing the place in a particular scenario further helps to focus our attention on the relevant evolutionary innovations, and evaluate their likelihood using modelling and the detailed analysis of data where available. Scenario building is thus actually necessary to move beyond speculation.

For evaluating the plausibility of various evolutionary scenarios that account for the comparative data discussed in this chapter, we can turn to various other fields, including comparative psychology for data on non-linguistic behavioral differences between humans and other primates and paleoanthropology for data on the evolutionary history of the human species. Additionally, evolutionary theory sets constraints on such scenarios, in particular by clarifying which components a scenario must involve. This is not the place to review the many findings from these various fields, but a few observations are useful to decide on what shape our scenario should have.

First, there is a whole suite of behavioral or cognitive abilities, other than language, that make humans stand out among animals, including advanced reasoning, consciousness, music, social cognition and theory of mind (knowing about the thoughts of others), the ability to imitate movements and sounds and our willingness to cooperate and share resources and knowledge. Together with uniquely human features of our life history (long helpless period in infancy, delayed sexual maturity, long post-reproductive life) and anatomy (reduced hair cover, sweat glands and upright posture), some researchers speak of a complete package of 'humanness' (e.g., Jones et al. 1992). It is clear that, a priori, an evolutionary scenario that assumes a common cause, or several common causes, for all of these different aspects of humanness is more plausible than an evolutionary scenario that assumes a distinct selection regime and evolutionary adaptation for each of them separably.

Second, from genetic and archeological data we know that the last common ancestor of chimpanzees and humans lived about 7 million years ago. A whole range of hominin species has been identified from fossil findings, ranging from the more ape-like Australopithecus afarensis closer to that common ancestor, to the much more recent Homo heidelbergensis occurring just before the appearance of anatomically modern humans about 200,000 years ago. One thing that is striking about those 7 million years is that most of that period involved only very slow changes. For instance, from about 2.6 million years ago hominins used simple stone tools, which remained virtually the same for a million years until hand axe technology first appeared in Homo ergaster. Then, in the last 100,000 years developments start to pick up speed. Art appears 80–100,000 years ago, modern humans spread around the globe (including the Americas about 12,000 years ago), agriculture is invented about 10,000 years ago, writing about 7,000 years ago and human history took off from there. A key factor in judging the plausibility of an evolutionary scenario of humanness, including language, is whether it can account for such a sudden speed-up in the evolutionary development.

Third, an evolutionary scenario describes a sequence of innovations, and evolutionary theory tells us to consider, at each step, whether the variation required for selection to operate would have been present and whether selection would have favored the proposed innovations among the many other possibilities. Focusing on the role of selection, there are two major obstacles in scenarios of language evolution. The first is that selection for linguistic traits is typically frequency dependent: the advantages of a trait usually depend on how many other people in a population already have it. For instance, knowing a particular word or grammatical construction is of little use if no-one else is able to understand it. As novel traits are initially always rare (because innovations in biological evolution are generated by rare mutations), this creates a kind of catch-22 situation: each innovation, even if it represents a true improvements when adopted, is initially selected against and therefore never becomes abundant enough to start conveying its advantage and thus be selected for.

The second major obstacle can be called the ‘problem of cooperation’: linguistic innovations that improve the efficiency of information transfer are often not in the interest of the speaker, but only in that of the hearer. Hence, although not impossible, it is difficult to see why evolution would lead to speakers to adopting it.

Both obstacles thus have to do with the fact that language is a social phenomenon. Both can be overcome in various ways, for instance through the mechanism called ‘kin selection’: if an individual interacts preferentially with other individuals that are genetically closely related (e.g., one’s brothers or sisters), natural selection can under some particular conditions favor the evolution of altruistic traits. For the various steps in any proposed scenario, we need to check, as well as we can, whether these conditions are met.

For the plausibility of any proposed scenarios, this means that those that involve very many genetically specified linguistic innovations under social selection pressures, are a priori less likely. This is the case, for instance, of the scenarios proposed by Jackendoff and Pinker: to overcome the discussed obstacles these scenarios need language-external circumstances for millions of years to be continuously unusually favourable. Moreover, during those same millions of years none of those human- and language specific tricks were selected for in other great apes. More probable scenarios, in contrast, involve a positive feedback mechanism: a mechanism where the emergence of a rudimentary form of language fundamentally changes the evolutionary dynamics and makes selection for further linguistic traits more probable. In such a scenario, favorable circumstances during a shorter stretch of human evolution could have provided the seed for a self-enforcing process leading to full-blown language.

To be sure, these arguments do not establish to correctness or falsehood of any scenario, but only establish that, before we have considered any data on the biology of language, scenarios are a priori more likely if they involve common causes, explain the speed-up and do not involve too many population-dependent genetic innovations.

### *8.6.2 A Scenario of the Evolution of the Cultural Phenomenon ‘Language’*

Combining these desiderata with the comparative evidence from Sect. 8.4 and 8.5, I arrive at the following scenario—hypothetical, but more plausible in my assessment than the alternatives and worthy as a working hypothesis.

The scenario starts out with the traits of the last common ancestor (LCA) with chimpanzees. Given the comparative evidence, I assume that the LCA, like modern apes, had an ability to handle hierarchical, conceptual structure in reasoning about the physical world, in reasoning about the behaviour of conspecifics and other animals (prey, predators, competitors) and in making plans. I assume the LCA had, like modern apes, a relatively rich communication system, with tens of vocal and gestural signals, that involved some learning (especially on the receiving side) but no true vocal learning and no compositionality. I assume the LCA, like modern apes, lived in groups with a limited form of cooperativity and at least a minimal degree of social cognition. Finally, I assume it had a complex brain, with quite advanced cognitive abilities compared to other mammals, well adapted for survival in its contemporary environment but also with a ‘hidden potential’ to develop even more complex cognitive skills under the right circumstances.

The first step is a process of biological evolution after the split of the chimpanzee and human lineage, adapting the hominin species to function in larger social groups, probably as a result of moving from the forest to savannah environments (a change of niche also thought to be involved in the evolution of bipedalism, sweating and running skills). Selection pressures for surviving in a group and as a group (with the typical mix of selectional mechanisms studied in social evolution theory, including kin selection, altruistic punishment and knowledge-for-status; see, e.g., West et al. 2007) then led to increases in social intelligence, in cooperativity, in the willingness to share information and in the size of signal repertoires. The need for larger signal repertoires, in turn, led to the increased reliance on learned vocalizations (vocal learning, gestural imitation), with learned, conventional meanings and combinatorial phonology (reuse of articulatory programs, but no compositional semantics).

With the appearance of a learned signal system, the circumstances were ready for the second step: cultural evolution kicked in, and the signals adapted culturally to pre-existing biases of the hominin brain, ears, hands and mouth. Because of the cultural adaptation, the communication system could become more complex than it could have become otherwise. I take the highest achievements of any non-human primate today as an estimate of what these hominins could achieve: large repertoires of conventional, arbitrary signals (vocal and gestural) and a rudimentary form of compositional semantics.

Step three is that once this communication system, due to cultural evolution, had started to form such an important aspect of life, it also started to change the course of biological evolution. The rudimentary language served as a medium to transmit knowledge from generation to generation, for instance by learning about food sources or relatively rare but grave dangers. Hence, language made those individuals that mastered it well more knowledgeable than their less talkative competitors, and thus more likely to survive and more attractive to mate with. More and more complex language thus led to more complex cognition and to increased biological selection pressure for both general cognitive and specific linguistic abilities, including those subserving speech such as vocal learning, vocal control and acoustic range. Moreover, proficient language users were likely to seek each other’s company, and thus profit from their advanced abilities even in populations where those were rare. This provides a positive feedback mechanism through which the presence of language makes overcoming the obstacles from social selection pressures more likely.

In that run-away evolutionary process, I assume a kind of arms race between language users emerged in which at some point, step four, the ability for hierarchical compositional structure emerged. It is difficult to say how much the primate brain had to change to allow for hierarchical compositionality without even the beginnings of an understanding of how it is implemented in modern human brains. However, its striking absence in animal communication and in the achievements of language-trained apes and birds, combined with the fact that many uniquely human linguistic traits seem linked to it, strongly suggest a biological basis. Also from a computational point of view, hierarchical compositionality is special as it requires dedicated computational mechanisms to perform operations in the signal domain and meaning domain in synchrony. I would speculate that a neural pathway for synchronizing preexisting combinatorial operations in the conceptual and motor domain (i.e., combinatorial phonology) was the crucial innovation.

Once hierarchical compositionality (HC) emerged, cultural evolution could take the languages spoken (or signed) to unprecedented levels of complexity in step five. Given the enormous diversity in languages spoken and signed today, and the fact that any human child can learn any of them, I suspect there is little biological specialization for language beyond HC. Symbolism, duality of patterning, phrase-structure, recursion are all potentially indirect consequences of HC, and I see the vast variety of intricate patterns in phonology, morphosyntax and pragmasemantics as likely to be the result of cultural evolution adapting to the pre-existing features of the human brain and body under communicative pressures.

The final and sixth component of this scenario concerns the impact that the discovery of complex language could have had on other aspects of cognition. It has often been proposed that language facilitates reasoning, planning, music, consciousness, social cognition and other cognitive domains, but equally often scientists have made proposals where the direction of influence is the other way around. Although with the current state of knowledge it would be unwise to claim much certainty on any position, I favour a language-first scenario. From an evolutionary point of view, the main argument in favour of language as the foundation for the rest is that it is the only of the uniquely human cognitive functions that plausibly plays a role in all the other functions and can plausibly have facilitated its own evolution through the positive feedback mechanism discussed above. Without positive feedback, it would remain a mystery why there are no non-human animal species that share at least some of those functions. Example of scenarios in which general intelligence is the driving force need to postulate millions of years of selection for intelligence until the various thresholds for language, music, consciousness etc. are met, while no other animal species was apparently under selection long enough for intelligence to reach even one of those thresholds.

## 8.7 Conclusions

A solid, scientific understanding of the evolutionary origins of language will remain elusive for some time. This means that the field of language evolution will continue to be an attractive domain for speculation and fantasizing. However, we need not (and, indeed, should not) accept this as a final verdict of the field. Solid comparative research and formal modelling, often inspired by more speculative theories, have led to many new findings on how aspects of natural language relate to animal abilities and under which circumstances biological and cultural evolution will favor particular changes in those abilities. Taken together, this evidence points to a central role for vocal and gestural imitation as the basis for cultural evolution, and to hierarchical compositionality, as the essential and uniquely human feature of language needed in definitions of symbolism, duality of patterning, phrase-structure and recursion.

The evidence also allows us to evaluate the relative plausibility of various scenarios. Although different researchers might reach different conclusions, this exercise leads me to conclude that a central focus for research in this field ought to be on those steps in the scenario for which there still is embarrassingly little empirical and modelling evidence: the neural basis of hierarchical compositionality, the feedback mechanism of language fostering its own evolution and the possible roles of language in not just influencing but facilitating consciousness, reasoning, planning and music.

## References

Alemseged, Z., Spoor, F., Kimbel, W.H., Bobe, R., Geraads, D., Reed, D., Wynn, J.G.: A juvenile early hominin skeleton from Dikika, Ethiopia. Nature **443**, 296–301 (2006)

Arnold, K., Zuberbühler, K.: Language evolution: semantic combinations in primate calls. Nature **441**, 303 (2006)

Bloom, P.: How Children Learn the Meanings of Words. MIT Press, Cambridge (2000)

Bloom, P.: Can a dog learn a word? Science **304**, 1605–1606 (2004)

Botha, R.: Prehistoric shell beads as a window on language evolution. Lang. Commun. **28**, 197–212 (2008)

Chomsky, N.: Syntactic Structures. Mouton, The Hague (1957)

Chomsky, N., Halle, M.: The Sound Pattern of English. Harper and Row, New York (1968)

Cruse, D.A.: Lexical Semantics. Cambridge University Press, Cambridge (1986)

Crystal, D.: The Cambridge Encyclopedia of Language. 2nd edn. Cambridge University Press, Cambridge (1997)

Culicover, P., Nowak, A., Borkowski, W.: Learning constructions and the theory of grammar. In: Proceedings of the 2004 Stanford Child Language Research Forum (2004)

Dave, A., Margoliash, D.: Song replay during sleep and computational rules for sensorimotor vocal learning. Science **290**, 812–816 (2000)

de Boer, B.: Acoustic analysis of primate air sacs and their effect on vocalization. J. Acoust. Soc. Am. **126**, 3329–3343 (2009)

de Boer, B.: Modelling vocal anatomy's significant effect on speech. J. Evol. Psychol. **8**, 351–366 (2010)

Deacon, T.W.: Evolutionary perspectives on language and brain plasticity. J. Commun. Disord. **33**, 273–290 (2000)

Evans, N., Levinson, S.C.: The myth of language universals: language diversity and its importance for cognitive science. Behav. Brain Sci. **32**, 429–492 (2009)

Fitch, W.T.: The evolution of speech: a comparative review. Trends Cogn. Sci. **4**, 258–267 (2000)

Fitch, W.T.: The evolution of language: a comparative review. Biol. Philos. **20**, 193–230 (2005)

Fitch, W.T., Hauser, M.D.: Computational constraints on syntactic processing in a nonhuman primate. Science **303**, 377–380 (2004)

Frishberg, N.: Arbitrariness and iconicity: historical change in American sign language. Language **51**, 696–719 (1975)

Gentner, T.Q., Fenn, K.M., Margoliash, D., Nusbaum, H.C.: Recursive syntactic pattern learning by songbirds. Nature **440**, 1204–1207 (2006)

Haesler, S., Wada, K., Nshdejan, A., Morrisey, E.E., Lints, T., Jarvis, E.D., Scharff, C.: FoxP2 expression in avian vocal learners and non-learners. J. Neurosci. **24**, 3164–3175 (2004)

Harnad, S.: Symbol-grounding problem. In: Encylopedia of Cognitive Science. Macmillan, London (2003)

Hockett, C.F.: The origin of speech. Sci. Am. **203**, 88–96 (1960)

Janik, V.M., Slater, P.J.B.: The different roles of social learning in vocal communication. Anim. Behav. **60**, 1–11 (2000)

Jarvis, E.D.: Learned birdsong and the neurobiology of human language. Annals of the New York Academy of Sciences **1016**, 749–777 (2004)

Jarvis, E.D.: Selection for and against vocal learning in birds and mammals. Ornithol. Sci. **5**, 5–14 (2006)

Jones, S., Martin, M., Pilbeam, D. (eds.): The Cambridge Encyclopedia of Human Evolution. Cambridge University Press, Cambridge (1992)

Kaminski, J., Call, J., Fischer, J.: Word learning in a domestic dog: evidence for "fast mapping". Science **304**, 1682–1683 (2004)

Krause, J., Lalueza-Fox, C., Orlando, L., Enard, W., Green, R.E., Burbano, H.A., Hublin, J.J., Hänni, C., Fortea, J., de la Rasilla, M., Bertranpetit, J., Rosas, A., Pääbo, S.: The derived FOXP2 variant of modern humans was shared with neandertals. Curr. Biol. **17**, 1908–1912 (2007)

Lachlan, R.F., Slater, P.J.: The maintenance of vocal learning by gene-culture interaction: the cultural trap hypothesis. Proc. R. Soc. Lond B Biol. Sci. **266**, 701–706 (1999)

Levelt, W.J., Wheeldon, L.: Do speakers have access to a mental syllabary? Cognition **50**, 239–269 (1994)

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M.: Perception of the speech code. Psychol. Rev. **74**, 431–461 (1967)

Lieberman, P.: The Biology and Evolution of Language. Harvard University Press, Cambridge (1984)

Martinet, A.: La double articulation linguistique. Travaux du Cercle Linguistique de Copenhague **5**, 30–37 (1949)

Martínez, I., Arsuaga, J.-L., Quam, R., Carretero, J.-M., Gracia, A., Rodríguez, L.: Human hyoid bones from the Middle Pleistocene site of the Sima de los Huesos. J. Human Evol. **54**, 118–124 (2008). (Sierra de Atapuerca, Spain)

Martínez, I., Rosa, M., Arsuaga, J.-L., Jarabo, P., Quam, R., Lorenzo, C., Gracia, A., Carretero, J.-M., Carbonell, E.: Auditory capacities in Middle Pleistocene humans from the Sierra de Atapuerca in Spain. Proc. Natl. Acad. Sci. USA **101**, 9976–9981 (2004)

Nowak, M.A., Krakauer, D.C.: The evolution of language. Proc. Natl. Acad. Sci. USA **96**, 8028–8033 (1999)

O'Grady, W.D.: How children learn language. Cambridge University Press, Cambridge (2005)

Ohms, V.R., Snelderwaard, P.C., ten Cate, C., Beckers, G.J.L.: Vocal tract articulation in zebra finches. PLoS ONE **5**, e11923 (2010)

Okanoya, K.: Song syntax in bengalese finches: proximate and ultimate analyses. Adv. Study Behav. **34**, 297–346 (2004)

Oliphant, M.: The learning barrier: moving from innate to learned systems of communication. Adapt. Behav. **7**, 371–384 (1999)

Payne, R.S., McVay, S.: Songs of humpback whales. Science **173**, 585–597 (1971)

Pinker, S.: The Language Instinct: How the Mind Creates Language. Harper Perennial, New York (1994)

Pinker, S., Jackendoff, R.: The faculty of language: what's special about it? Cognition **95**, 201–236 (2005)

Ralls, K., Fiorelli, P., Gish, S.: Vocalizations and vocal mimicry in captive harbor seals, *Phoca vitulina*. Can. J. Zool. **63**, 1050–1056 (1985)

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W.T., Clutton-Brock, T.: Red deer stags use formants as assessment cues during intrasexual agonistic interactions. Proc. R. Soc. Lond. B **272**, 941–947 (2005)

Riebel, K., Slater, P.J.B.: Temporal variation in male chaffinch song depends on the singer and the song type. Behaviour **140**, 269–288 (2003)

Savage-Rumbaugh, S., Lewin, R.: Kanzi: The Ape at the Brink of the Human Mind. Wiley, New York (1994)

Seyfarth, R.M., Cheney, D.L.: Some general features of vocal development in nonhuman primates. In: Snowdon, C.T., Hausberger, M. (eds.) Social Influences on Vocal Development, pp. 249–273. Cambridge University Press, Cambridge (1997)

Smith, E.C., Lewicki, M.S.: Efficient auditory coding. Nature **439**, 978–982 (2006)

Suzuki, R., Buck, J.R., Tyack, P.L.: Information entropy of humpback whale songs. J. Acoust. Soc. Am. **119**, 1849–1866 (2006)

Tamariz, M.: Configuring the phonological organization of the mental lexicon using syntactic and semantic information. In: Proceedings of the 27nd Annual Conference of the Cognitive Science Society, pp. 2145–2150 (2005)

Tomasello, M., Call, J., Warren, J., Frost, G., Carpenter, M., Nagell, K.: The ontogeny of chimpanzee gestural signals: a comparison across groups and generations. Evol. Commun. **1**, 223–259 (1997)

van Heijningen, C.A.A., de Visser, J., Zuidema, W., ten Cate, C.: Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. Proc. Natl. Acad. Sci. USA **106**, 20538–20543 (2009)

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., Passingham, R.: Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proc. Natl. Acad. Sci. USA **92**, 930–933 (1995)

von Frisch, K.: Decoding the language of the bee. Science **185**, 663–668 (1974)

West, S.A., Griffin, A.S., Gardner, A.: Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. **20**, 415–432 (2007)

Westermann, G., Miranda, E.: A new model of sensorimotor coupling in the development of speech. Brain Lang. **89**, 393–400 (2004)

Zuberbühler, K.: A syntactic rule in forest monkey communication. Anim. Behav. **63**, 293–299 (2002)

Zuidema, W.: Optimal communication in a noisy and heterogeneous environment. In: Banzhaf, W., Christaller, T., Dittrich, P., Kim, J.T., Ziegler, J. (eds.) Advances in Artificial Life (Proceedings of the 7th European Conference on Artificial Life), pp. 553–563. Springer, Berlin (2003)

Zuidema, W., de Boer, B.G.: The evolution of combinatorial phonology. J. Phonetics **37**, 125–144 (2009)

# Chapter 9
# Self-Organization: Complex Dynamical Systems in the Evolution of Speech

**Pierre-Yves Oudeyer**

**Abstract** Human vocalization systems are characterized by complex structural properties. They are combinatorial, based on the systematic reuse of phonemes, and the set of repertoires in human languages is characterized by both strong statistical regularities—universals—and a great diversity. Besides, they are conventional codes culturally shared in each community of speakers. What are the origins of the forms of speech? What are the mechanisms that permitted their evolution in the course of phylogenesis and cultural evolution? How can a shared speech code be formed in a community of individuals? This chapter focuses on the way the concept of self-organization, and its interaction with natural selection, can throw light on these three questions. In particular, a computational model is presented which shows that a basic neural equipment for adaptive holistic vocal imitation, coupling directly motor and perceptual representations in the brain, can generate spontaneously shared combinatorial systems of vocalizations in a society of babbling individuals. Furthermore, we show how morphological and physiological innate constraints can interact with these self-organized mechanisms to account for both the formation of statistical regularities and diversity in vocalization systems.

**Keywords** Self-organization · Natural selection · Evolution of speech · Combinatoriality · Computational model

Human vocalization systems are characterized by complex structural properties. They are combinatorial, based on the systematic reuse of phonemes, and the set of repertoires in human languages is characterized by both strong statistical regularities—universals—and a great diversity. Besides, they are conventional codes culturally shared in each community of speakers. What are the origins of the forms

---

P.-Y. Oudeyer (✉)
INRIA Bordeaux Sud-Ouest, 13, C1 route du Brassan,
33640 Ayguemorte-les-Graves, France
http://www.pyoudeyer.com
e-mail: pierre-yves.oudeyer@inria.fr

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_9,

of speech? What are the mechanisms that permitted their evolution in the course of phylogenesis and cultural evolution? How can a shared speech code be formed in a community of individuals? This chapter focuses on the way the concept of self-organization, and its interaction with natural selection, can throw light on these three questions.

The tendency of many complex physical systems to generate spontaneously new and organized forms, such as ice crystals or galactic spirals, is indeed present as much in the inorganic world as in the living world. Thus, the explanation of the origins of forms and structures in the living can not only rely on the principle of natural selection, but should be complemented by the understanding of physical mechanisms of form generation in which self-organization plays a central role. This applies to the social and cultural forms of the living, in particular to the forms of speech and language. As a consequence, I will begin by articulating in a general manner the relationships between self-organization, natural selection and Neo-Darwinism in explanations of the genesis of forms in the living. Then, I will instantiate these relations in the context of the three questions formulated above. After that, I will explain why the use of computer models and simulations is fundamental for progress in our understanding of these issues. Finally, I will present the example of an experiment based on a computer model which shows that certain simple mechanisms coupling perception and production of sounds can generate combinatorial systems of vocalizations, characterized by the universal/diversity duality, and shared culturally by the members of a speech community. I will conclude with the presentation of evolutionary scenarios that this computer experiment complements or renews.

## 9.1 Self-Organization and the Evolution of Forms in the Living

### *9.1.1 Physics, the Caldron of Self-Organized Forms*

Nature, especially inorganic nature, is full of fascinatingly organized forms and patterns. The silhouettes of mountains are the same, whether one views them at the scale of a rock, a summit, or a whole mountain range. Sand dunes often arrange themselves in long parallel stripes. Water crystallizes into symmetrical serrated flakes when the temperature is right. And when water flows in rivers and hurtles over cataracts, trumpet-shaped vortices appear and the bubbles collect together in structures which are sometimes polyhedral. Lightning flashes draw plant-like branches in the sky. Alternating freezing and thawing of the rocky ground of the tundra leaves polygonal impressions in the earth. The list of these forms rivals many human artefacts in complexity, as can be seen in Fig. 9.1. And yet they are not designed or conceived by anyone or anything, not even natural selection, Dawkins' 'blind watchmaker' (Dawkins 1986). What, then, are the mysterious factors that explain their existence?

In fact, all these organized structures have a feature in common: they are the macroscopic outcomes of local interactions between the many components of the system

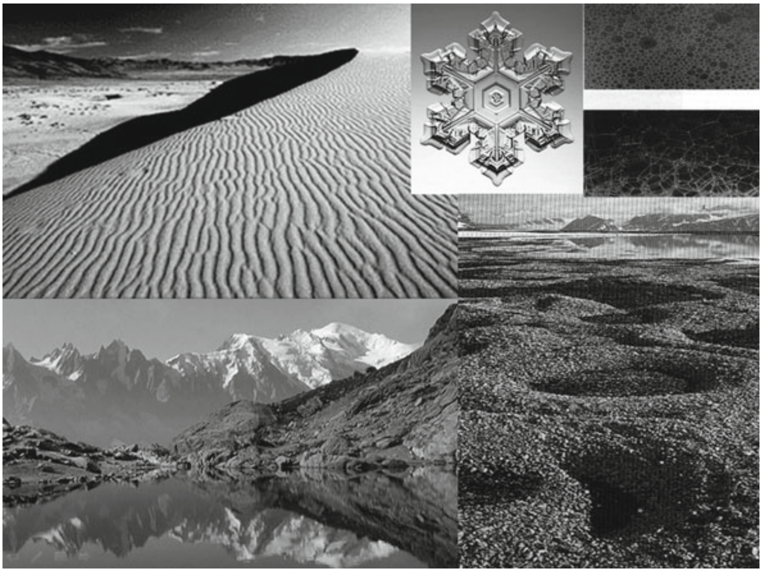


**Fig. 9.1** Nature is full of organized forms and patterns without there being anywhere any plans which might have served to build them; they are said to be self-organized. Here, parallel stripes running through sand dunes, water bubbles on the surface of liquid which has been stirred up and the polyhedral structures which are left when they dry out, an ice crystal, mountains whose shapes are the same whether one views them on the scale of a rock or a whole peak (Photos: Nick Lancaster, Desert Research Institute, Nevada)

from which they emerge. Their global organizational properties are not to be found at the local level. Indeed the properties of the shape of a water molecule, as well as its individual physico-chemical properties, are qualitatively different from the properties of ice crystals, whirlpools, or polyhedral bubbles. The polygonal impressions in the tundra do not correspond to the shape of the stones composing them, and have a spatial organization quite different from the temporal organization of freezing and thawing. This is the hallmark of a newly discovered phenomenon—self-organization.

In nature self-organization characterizes very diverse physical systems, but several typical properties can be identified: non-linearity, symmetry breaking, presence of dynamical systems, "attractors", and historicity. For example, when one heats from below a thin layer of oil spread out on a flat surface, convection currents with peculiar geometric shapes (lines or polygons) self-organize and these shapes change dramatically when the temperature goes over given thresholds (see Fig. 9.2). On the contrary, between these thresholds the shapes remain globally stable even if they are perturbed, constituting attractors. Another property of many self-organized dynamical systems is historicity, often associated with the sensitivity to initial conditions in chaotic systems: the attractor in which the systems falls, i.e. the shapes/forms that are produced by the complex system, can be very different depending on slight variations in the initial conditions. For example, this is the case of ferromagnetization: each of the

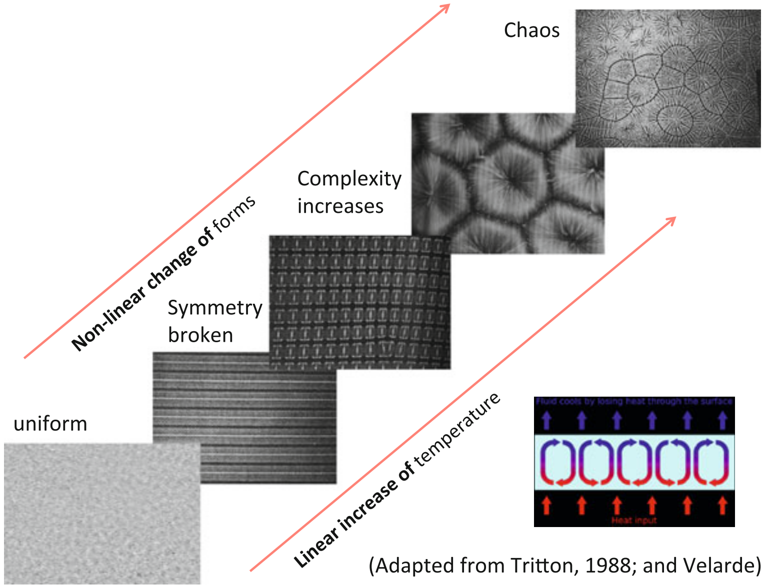


**Fig. 9.2** Rayleigh-Bénard cells: when one heats from below a thin layer of oil spread out on a plane surface, convection currents with peculiar geometric shapes (lines or polygons) self-organize and these shapes change brutally when the temperature goes over given thresholds. This kind of non-linearity characterized many self-organized systems both in the inorganic and in the living world (Photos adapted from Tritton (1988), and Manuel Velarde, Universidad Complutense, Madrid)

atoms of an iron plate can be viewed as a sort of magnet that can have several possible orientations, and whose orientation is random if the temperature is high enough. Yet, if the temperature goes below a certain threshold, a self-organized phenomenon happens: all atoms spontaneously adopt the same magnetic orientation. This shared orientation is quasi-impossible to predict beforehand and tiny random variations of initial orientations of atoms can lead the plate to be magnetized in a quite different orientation. These variations in initial conditions are typically linked to contingent events that interacted with the iron plate: this is why the final state of the plate depends both on its history and on its intrinsic physical mechanisms, whence the term “historicity”.

This fundamental concept is the touchstone of the paradigm shift undergone by the sciences of complexity in the twentieth century (Ashby 1956; Nicolis and Prigogine 1977; Kauffman 1996; Ball 2001). Ever since Newton, good science was supposed to be reductionist, and consisted in decomposing natural systems into simpler subsystems. For example, to understand the functioning of the human body, it was appropriate to study the respective parts, such as the heart, the nervous system, or the limbic system. Moreover, things did not stop there, and the study of the nervous system, for example, was subdivided into study of the cortex, of the thalamus, or of the peripheral motor innervations, and each of these sub-parts was studied by hyper-specialists in separated dedicated university departments. This method has obviously enabled us to accumulate an impressive bank of knowledge. But the prophets of complexity have broken up this paradigm. Their credo is "the sum of the parts is greater than the parts taken independently".

### *9.1.2 The Impact of Self-Organization on the Origins of Forms in the Living*

Complex systems, i.e. systems composed of many interacting sub-systems, abound in nature and have a strong tendency to self-organize. The examples of the previous section were chosen deliberately from inorganic systems to show that the property of self-organization can be found in systems subject to laws which have nothing to do with natural selection. However, self-organization applies similarly to living systems. It is a concept widely used in several branches of biology. It is particularly central to theories which explain the capacity of insect societies to build nests or hives, to hunt in groups or to explore in a decentralized and effective way the food resources of their environment (Camazine et al. 2002). In developmental biology, it is used, for example, to explain the formation of coloured patterns on the skins of animals like butterflies, zebras, jaguars or ladybirds (Ball 2001).

Thus, it seems possible that there are shape- and pattern-forming mechanisms in biological systems which are orthogonal to natural selection, through their property of self-organization. Now natural selection is at the heart of almost all the arguments of biologists when it comes to explaining the presence of a shape, a pattern or a structure in an organism. What, then, is the relationship between the theory of natural selection and self-organization?

Some researchers have suggested that self-organization casts doubt on the centrality of natural selection in explaining the evolution of living organisms. Waldrop explains:

> Complex dynamical systems can sometimes go spontaneously from randomness to order; is this a driving force in evolution? Have we missed something about evolution—some key principle that has shaped the development of life in ways quite different from natural selection, genetic drift, and all the other mechanisms biologists have evoked over the years? . . . Yes! And the missing element. . . is spontaneous self-organization: the tendency of complex dynamical systems to fall into an ordered state without any selection pressure whatsoever (Waldrop 1990).

However, this is not the position I take in this article. Rather than seeing self-organization as a concept which minimizes the role of natural selection by suggesting competing form-creating mechanisms, it is more accurate to see instead as belonging to a somewhat different level of explanation and more importantly as describing mechanisms which actually increase the power of natural selection by orders of magnitude. Systems with the self-organizing property are completely compatible with the mechanism of natural selection in explaining the evolution of forms and structures in biology.

### *9.1.3 Classic Neo-Darwinism*

To see the matter clearly, it is first necessary to recall what the mechanism of natural selection, or Neo-Darwinism, comprises. It is a mechanism characterizing a system composed of individuals each having particular traits, shapes or structures. In addition, the individuals in this system are capable of replication. This replication must occasionally produce individuals which are not exact copies of their ancestors, but are slight variants. These variations are the source of diversity among individuals. Finally, each individual has a greater or lesser capacity for replication, according to its surrounding environment. This generates differential replication of individuals and gives rise to "selection" of those who are most capable of replicating themselves. The combination of the processes of replication with variation and selection means that, over the generations, the structures or traits of individuals which help them to reproduce themselves are preserved and improved upon.

Now there is one crucial point on which the theory of natural selection is neutral: it is the way in which variation is generated, and more generally the ways in which the individuals with their shapes, traits and structures are produced. A number of Neo-Darwinist arguments consider the mechanisms of variation of forms as secondary in comparison with the reproductive advantages of these forms when it comes to explain their evolution. This states implicitly that the relation between the level of genes, considered as the main space in which variations operate through mutations and cross-overs, and the level of phenotypes, considered as an isomorphic image of the space of genes, is simple and linear. According to this vision, exploration of the space of phenotypes (which determines, along with the environment, the relative effectiveness of the genes at replicating) can simply be carried out by studying the way things change in the space of genotypes. Now the mechanisms of mutation which actually bring about these changes are of small amplitude (most mutations only affect a minimal fraction of the genome when replication succeeds), and thus random variations in genes lead to uniform exploration of the space of genotypes. What this means is that under the hypothesis that phenotypic and genotypic space have the same structure and can be mapped approximately linearly, the space of possible biological forms can be searched quasi-continuously, by successive little modifications of pre-existing forms. Fortunately for the appearance of complex life-forms, this is not the case. In fact, although this mechanism of small successive variations in form is notably effective in the delicate regulation of the structures of organisms, it would make the search for forms as complex as those of human organisms equivalent to the search for a needle in a haystack because genomes are much too high-dimensional (Keefe and Szostak 2001).

### *9.1.4 Self-Organization Constrains the Space of Forms to be Explored: Not All Forms can Emerge Equally Easily*

It is here that the concept of self-organization comes to the rescue of this naive search mechanism in the space of phenotypic forms in the Neo-Darwinian theoretical framework. In fact the relation between genes and the forms of organisms is characterized by its complexity and its non-linearity that are expressed through the ontogenetic and epigenetic development of each organism. Organisms are constructed starting from a stem cell containing a whole genome. This stem cell can be seen as a dynamic system parameterized by its genome and under the influence of perturbations imposed by the environment. This dynamic system is above all a self-organizing system with the same sorts of properties as the self-organizing systems described in the previous section. The genome is a set of parameters analogous to temperature and the viscosity of liquids in Bénard systems, and the environment is analogous to noise (but evidently highly structured noise!). Thus the development of an organism from a stem cell shares many properties with physical systems: shapes, structures and patterns appear at the global level, and are qualitatively different from those implementing functioning at the local level, that is, different from the patterns characterizing the structure of the stem cell and its genome. The hexagonal pattern which can appear as a result of a simple difference in temperature in a homogeneous liquid gives an idea of the way in which a simple sequence of nucleotides enclosed in a system of molecules which transforms them automatically into proteins can generate a bipedal organism endowed with two eyes and ears and an immensely complex brain.

As with Bénard systems or ferromagnetic plates, dynamic systems defined by the cells and their genomes are characterized by a landscape of attractors: there are large regions in the parameter space within which the dynamic system systematically adopts behaviour which is more or less the same. For Bénard systems, there is a range of temperatures giving rise to parallel stripes which is wide enough to locate easily. For ferromagnetic plates the range of temperature in which the system settles to global magnetic coherence is also very wide. Thus for living organisms it is not only possible to generate self-organizing structures with complex global properties, but in addition these structures are generated by genomes belonging to broad subspaces of genome space called basins of attraction. The structuring of genome space into basins of attraction by this kind of dynamic system facilitates the evolutionary search of the space of forms so that it does not resemble a search for a needle in a haystack.

As in ferromagnetic systems, structured noise imposed by the environment on the development of the dynamic system can lead it to take different developmental pathways. For pieces of iron at low temperatures, this corresponds to magnetization in one direction or another. For a living organism, this corresponds to its possible shapes; this is how it happens that even monozygotic twins can show quite important morphological differences. This is also the reason why the relationship between genes and the forms of organisms is not only complex and nonlinear, but also nondeterministic. Moreover, and as in Bénard systems where search of the parameter space of temperature can sometimes lead to fast and qualitative changes in the behaviour of the system (for example the change from parallel stripes to square cells), which have been called phase-transitions, the search within genome space can also lead to fast qualitative changes. This possibly corresponds to many observations of rapid form-changes in evolution, as witnessed by the fossils studied by anthropologists, and which are the basis of the theory of punctuated equilibrium proposed by Elredge and Gould (1972).

To summarize, the self-organizing properties of the dynamic system composed by the cells and their DNA brings essential structuring to the phenotypic space by constraining it, making the discovery of complex robust forms by natural selection much easier. On the one hand, these properties enable a genome to generate complex, highly organized forms without the need for precise specification of each detail in the genome (in the same way as Bénard's polygonal shapes are not specified precisely, or encoded in a plan, in the properties of the liquid's molecules). On the other hand, the self-organizing properties structure the landscape of these possible forms into basins of attraction within which they resemble each other greatly (here is where gradual evolution happens, involving fine tuning of existing structures), and between which there can be substantial differences among forms (transitions from one basin to another are what provide abrupt and powerful innovations in evolution). To give a simple picture, self-organization provides a catalogue of complex forms distributed over a landscape of valleys in which and between which natural selection moves and makes its choices: self-organization proposes, and natural selection disposes. Obviously this is only an image to facilitate understanding, because with its movements natural selection actually enables new mechanisms, themselves self-organized, to appear, and these in turn structure the space of forms within which it moves; thus natural selection participates in the formation of these mechanisms which help it to move effectively in the space of forms; vice versa, the mechanism of natural selection certainly appeared in the history of life due to the self-organized behaviour of systems which were as yet completely unconnected to natural selection; natural selection and self-organizing mechanisms thus help each other reciprocally in a sort of spiral which enables complexity to increase during the course of evolution.

The consequence of this interplay between natural selection and self-organization is that any explanation of the origins and evolution of forms and structures in living organisms requires at least two kinds of argumentation. The first one, classic, is the neo-Darwinian functionalist argumentation: it consists in identifying the ecological context in which a given trait may have appeared and in articulating the balance between the associated costs and reproductive advantages. The second kind of argumentation is more rarely used but is equally essential: it consists in identifying the developmental/epigenetic mechanisms, and the associated constraints, which may have permitted, or made difficult, the genesis of these novel traits. And the concept of self-organization is central to the way developmental mechanisms impact the genesis of forms.

## 9.2 Self-Organization and the Evolution of Forms and Structures of Language and Languages

The question of how speech and language arose in humans, and the question of how new languages form and evolve, are among the most difficult that science has to address. After being put aside from scientific enquiries during most of the twentieth century, partly because of the ban pronounced by the Société Linguistique de Paris, they are now again the focus of attention of a whole scientific community. There is an emerging consensus among researchers who are today getting down to questions of the origin of the human language faculty and the evolution of languages: this research must be interdisciplinary. It in fact poses a puzzle with immense ramifications which go beyond the competence of each individual discipline on its own. Firstly, it is because the two big questions, that of the origins of language and that of the origins of languages, must be decomposed into subquestions which are themselves already quite complex: What, in fact, is the language faculty? What is a language? How are sounds, words, sentences and representations of meaning related to each other? How does the brain represent and process these sounds and sentences and the concepts which they convey? How do we learn to speak? What are the respective roles of nature and nurture? What is language for? What is its role in a community? How does a language form and change in the course of successive generations of speakers (Croft, this volume; Kirby, this volume)? What do we know about the history of each particular language? Why are languages and the language faculty the way they are? Why do we see universal tendencies and at the same time great diversity in languages? How does language influence the way we perceive and understand the world? What do we know about the history of the human capacity for speech? Is it mainly the result of genetic evolution, like the evolution of the eyes, or a cultural invention, like writing? Is language an adaptation to a changing environment? An internal change in an individual which increased its chances of reproduction? Is it an exaptation, a side effect of changes which were not at first tied to communicative behaviour? What are the evolutionary prerequisites which paved the way for the capacity of speech? And how did these prerequisites appear? Independently? Genetically? Culturally? (Dediu, this volume; Zuidema, this volume).

Placed against the diversity of these questions is an even greater diversity of research disciplines and methods. Linguists, even though they continue to provide critical data on the history of languages, are no longer the unique actors. Developmental and cognitive psychologists and neuropsychologists carry out behavioural studies of language acquisition and language pathology, and these often reveal cognitive mechanisms involved in language processing (Wonnacott, this volume). Neuroscience (Müller, this volume), especially with equipment for brain imaging allowing us to see which brain regions are active for given tasks, attempts to find neural correlates of verbal behaviour, to discover its organization in the brain. Some researchers also study the physiology of the vocal tract, to try to understand how we produce speech sounds. The physiology of the ear, the essential receptor in the speech-decoding chain (or vision, in the case of signed languages), is also a focus of research. Archeologists examine fossils and artefacts left by the first hominids, and try on the one hand to deduce our anatomical evolution (especially of the larynx) and on the other hand to get an idea of what activities they were engaged in (What tools did they make? How did they use them? What can these tools tell us about the degree of cognitive development?). Anthropologists do fieldwork on isolated peoples, and report on cultural differences, especially those related to languages and the meanings they convey. Primatologists try to report on the communicative capacities of animals that may have some ancestors in common with human and to compare them with our own communicative capacities. Geneticists on the one hand sequence the human genome and that of potential ancestral species when it is possible to specify their phylogenetic relatedness, and on the other hand use genetic information from different people across the planet to help in reconstructing the history of languages, which is often correlated with the genetic history of their speakers (Dediu, this volume).

Thus language involves a multitude of components interacting in complex ways in parallel on several timescales: the ontogenetic timescale, characterizing the growth of an individual person, the glossogenetic or cultural timescale which characterizes the evolution of cultures, and the phylogenetic timescale, which characterizes the evolution of species (see Fig. 9.3). In particular, language is characterized by complex physical and functional interactions among multiple cerebral circuits, several organs, the individuals who are equipped with them, and the environment in which they live. Now, as we have seen in previous paragraphs, not only is it essential to study each

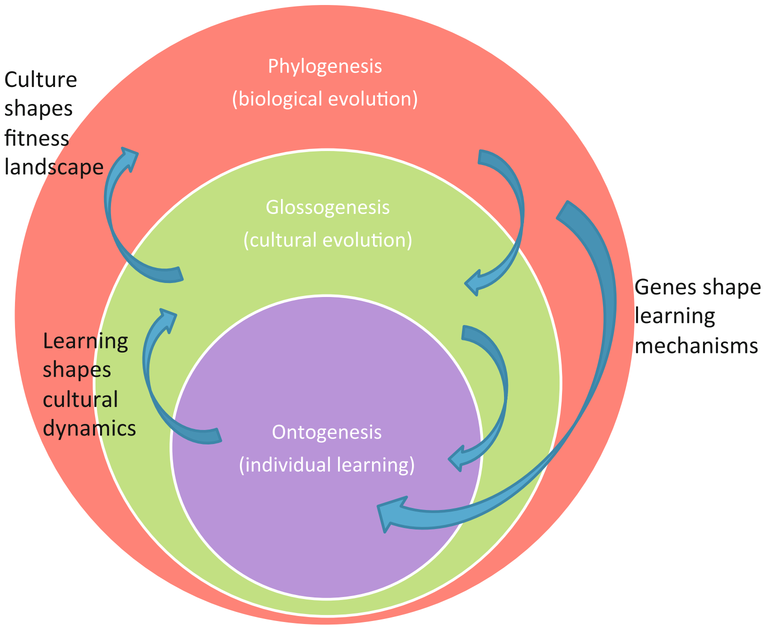


**Fig. 9.3** Multiple interaction scales involved in the origins of language

of these components independently to reduce the complexity of the problem; it is also necessary to study their interactions. Thus, a growing number of researchers have proposed the idea that many properties of language and languages may not be encoded in any particular component involved, i.e. in certain specific cerebral structures or in properties of the auditory and vocal apparatus, or even in an individual considered independently of others, but rather may be the self-organized outcomes of the complex dynamic interactions among the components and individuals. Yet, these self-organizational phenomena are often complicated to understand or to foresee intuitively, and to formulate in words, whence the necessary use of mathematical and computer modelling that I will now present.

## 9.3 Computer Models and Simulations of the Evolution of Language

### *9.3.1 Experimenting with Complex Dynamical Systems*

Nowadays, one of the most efficient ways to improve our understanding of the dynamics of self-organized systems is the use of computers or robots. Indeed, they allow us to implement operational models of which we know all the assumptions, to run them, and to observe the resulting behaviour as a function of the values of parameters set in the context of these models. This is why, in addition to linguists, psychologists, anthropologists, neuroscientists, geneticists, and physiologists, mathematicians and computer scientists/roboticists have now a critical role in this research.

An operational model is one which defines the set of its assumptions explicitly and above all shows how to calculate their consequences, that is to prove that it leads to a certain set of conclusions. There are two main types of operational model. The first, used by mathematicians and some theoretical biologists, consists in abstracting from the phenomenon of language a certain number of variables along with the rules of their evolution in the form of mathematical equations. Most often this resembles systems of coupled differential equations, and benefits from the framework of dynamic systems theory. The second type, which allows for modelling of more complex phenomena than the first, is that used by researchers in artificial intelligence: it consists in the construction of artificial systems implemented in computers or in robots. These artificial systems are made of programs which most often take the form of artificial software or robotic agents, endowed with artificial brains and bodies. These are then allowed to interact with an artificial environment (or a real environment in the case of robots), and their dynamics can be studied. This is what one calls the “method of the artificial” (Steels 2001) or the “synthetic methodology” (Pfeifer and Scheier 1999). The use of computational machines to simulate and study natural phenomena is not new: Early computers were used by Lorenz to study the behaviour of climatological models, by Fermi to simulate non-linear interactions among magnetized particles, by Turing to imagine how morphogenetic processes could self-organize, and by Von Neumann to study self-replication.

More recently, this method has allowed ethology to progress significantly in the understanding of behaviour performances of social insects (Bonabeau et al. 1997). Computer simulations of social insects were built based on the concept of software or robotic agents modelling each insect individually, hence the term agent-based modelling. This has permitted to establish sufficient characteristics of behaviour and insect capabilities that lead to the formation of collective structures, such as the construction of termite nests, or the formation of organized groups for hunting or foraging in ants, or the formation of fish shoals, thermoregulation in beehives or the formation of social structures in wasps. In general, these computer simulations have shown that it was often not necessary that insect be equipped with complex cognitive structures in order to produce the collective formation of complex structures.

Physicists have also a tradition of using computers to make simulations of complex systems that allow them to elaborate their intuitions. For example, through experimentation with cellular automata—sorts of grids which cells can be in an "on" or "off" state and evolution depends on the state of their neighbours according to simple rules—they discovered how, starting either from initially random structures or completely uniform structures, complex patterns with non-trivial symmetries could be formed, resembling those observed in ice crystals, in the distribution of avalanches in sand piles or in mountains, dunes in the desert, the shape of fluvial deltas, galaxies or polyhedral bubbles in water cascades. For physicists, cellular automata are not what could be called physical models of ice crystals or avalanches, but they have played the role of metaphors and analogies which triggered a renewal in the way their community perceived and understood these phenomena (Vichniac et al. 1989; Weisbuch 1991; Bak 1996; Ball 2001).

### 9.3.2 *Computer Science and the Origins of Language and Languages*

It is also possible to use computers and agent-based simulations not only to help us understand the phenomena that characterize self-organization of matter, simple biological structures, or insect societies, but also to help us understand phenomena that characterize humans and their societies. The time has come to use computers and robots as scientific tools in human sciences. Thus, building artificial systems in the context of research into language origins and the evolution of languages is enjoying a growing popularity in the scientific community, exactly because it is a useful tool for studying the phenomena of language in relation to the complex interactions of its components (Steels 1997; Oudeyer and Kaplan 2007; Kaplan et al. 2008). These systems are put to two main types of use: (1) they serve to evaluate the internal coherence of verbally expressed theories already proposed, by clarifying all their hypotheses and verifying that they do indeed lead to the proposed conclusions (and quite often one discovers errors in the assumptions as well as in the conclusions, which need to be revised); (2) they serve to explore and generate new theories, which themselves often appear when one simply tries to build an artificial system reproducing the verbal behaviour of humans.

A number of notable results have already been obtained and have opened the way for resolution of previously unanswered questions: the decentralized generation of lexical and semantic conventions in populations of agents (e.g. Kaplan 2001), the formation of shared inventories of vowels or syllables in groups of agents (e.g. Berrah et al. 1996; de Boer 2001; Oudeyer 2001, 2005a,b, 2006), with features of structural regularities greatly resembling those of human languages (e.g. Pierrehumbert 2001; Wedel 2006), the formation of conventionalized syntactic (e.g. Batali 1998) and grammatical structures (e.g. Steels 2005), the conditions under which combinatoriality, the property of systematic re-use, can be selected (Kirby 2001).

It is important to note that in the context of research on the origins of language, this methodology of the artificial is a *methodology for exploration*. It fits within an abductive scientific logic, i.e. a logic in which one searches for the premises that can lead to a given conclusion (instead of a deductive logic in which one searches for the conclusions that can be reached from a given set of premises).

The word ***model*** has here a different meaning than in its traditional use. Indeed, traditionally, modelling consists in observing a natural phenomenon and then abstracting fundamental mechanisms and variables on which a formalism is constructed that permits to predict reality as precisely as possible. Here, we are rather interested in qualitative investigations of broad types of mechanisms that may have been implemented in nature to solve given problems. Language is such a complex phenomenon that observations alone cannot allow researchers to *deduce* explanatory mechanisms. On the contrary, it is necessary to have beforehand a structured and rich conceptualisation of the space of hypotheses and mechanisms that might explain the complex properties of language evolution. This is where artificial computational systems, for which we use the term ***model***, come into play: they are used to refine our intuitions on the dynamics of language and languages formation, as well as to sculpt the space of hypotheses.

As a consequence, the aim of these computational models is not to establish the list of mechanisms responsible for the origins of given aspects of language. Rather, the objective is more modestly to try to build a list of potential candidates, and then to constrain the space of hypotheses, in particular by showing examples of mechanisms that are sufficient and examples of mechanisms that are not necessary.

## 9.4 The Speech Code

I will now illustrate this computer modelling work about the evolution of language and languages with the description of an experiment which focuses on the problem of the origins of speech, i.e. sound systems as physical vehicles of language (as gestures can also be in signed languages). The goal of this experimentation is to foster the reconceptualisation of this scientific issue, through the evaluation of both existing and novel scientific hypotheses.

**Discreteness and combinatoriality**. Humans have a complex system of vocalizations. They are discrete and combinatorial, i.e. they are built from elementary units, "sculpted" in the auditory and phonatory continuum, which are systematically recombined and reused. These units exist at several levels (motor primitives to obstruct the air flow in the vocal tract, called *gestures*; gestures coordinations, called phonemes and which define vowels and consonants; syllables; etc. . .). Whereas the articulatory space is continuous and potentially permits an infinity of gestures and phonemes, each language discretizes this space in its own way, carving a repertoire of gestures and phonemes both small and finite (Studdert-Kennedy and Goldstein 2003). This is why it is sometimes referred as *phonemic coding*.

**Universals and diversity**. In spite of the great diversity of these elementary units in world languages, one can also find strong statistical regularities. For example, certain vowel systems are much more frequent than some others, such as the five-vowel system composed of vowels [e], [i], [o], [a] and [u]. The same fact can be stated for consonants. The way units are combined is also peculiar: one the one hand, not all sequences of phonemes are allowed to form syllables in each particular language, and on the other hand the associated sets of possible phoneme combinations can be grouped into generic types. This organization into generic types means that for example, one can summarize the possible phoneme combination to form syllables in Japanese ("moras") with types "CV/CVC/VC", where "VC" denotes syllables composed of two slots, with any Japanese vowel in the first slot and any Japanese consonant in the second slot.

**Cultural sharing**. Speech is a conventional code. Whereas statistical regularities can be observed across human languages, each linguistic community possesses its own way to perceive and categorize sounds as well as its own repertoire of rules for combinations. For example, native speakers of Japanese do not hear the difference between the *r* in *read* and the *l* in *lead*. How can a linguistic community come to form a code shared by all individuals and without a central coordinated control of the code?

Since the work of de Boer (2001) and Kaplan (2001), we have convincing hypotheses about how a new sound or a new word can propagate and be accepted in a given population. But these negotiation mechanisms, also called "consensus dynamics", assume the pre-existence of linguistic interaction conventions (Croft, this volume). Thus, the associated models concern mainly the formation and evolution of languages, but do not address directly the question of the origins of language. Indeed, when there was not already a conventional linguistic communication system, how could the first conventions have bootstrapped?

The model I will present focuses on this later question. It is obviously linked to the question of the origins of languages, because it is about understanding how a speech code may have formed to be used as a basis for the first languages. The main difference among the two questions lies in the properties that shall characterize the mechanisms we are searching for. If one is interested in the origins of speech, one must search for an explanatory mechanism which assumes neither the existence of linguistic conventions, nor the existence of cognitive structures that are specific to language. Indeed, this would imply that we would have models of individuals that can already speak, and thus for which language would already have appeared.

## 9.5 Self-Organization and the Evolution of Speech

How did the first human speech codes form in a society with no language? As argued above, two kinds of answers must be given. First, a functional answer: they establish the function of vocalization systems, and shows that human systems, with the properties we described, are efficient to achieve this function. Liljencrantz and Lindblom (1972) proposed such an answer, and showed that the statistical regularities of human repertoires of phonemes were the most efficient in terms of the perceptual distinctiveness/articulatory cost compromise. This kind of answer is necessary, but not sufficient: it does not allow us to explain how evolution (genetic or cultural) could have found those quasi-optimal structures, and does not allow us either to explain how a given linguistic community can “choose” one solution among several quasi-optimal ones. In particular, it is possible that “naïve” Darwinian search with random mutations is not efficient enough to find complex structures such as those of the speech code: the search space is too large.

This is why a second kind of answer is necessary: we need to investigate how biological evolution might have generated and selected these structures. One possibility is to study how self-organization may have constrained the search space to help natural selection. This would consist in showing that a system much simpler than the structure we want to explain spontaneously *self-organizes into this structure*.

I will now present such a system and show how relatively simple premises—from an evolutionary point of view—can lead to the self-organized formation of speech codes.

### *9.5.1 A Computer Investigation of the Formation of Fundamental Structures of Speech*

This computer model is agent-based: it consists in setting up virtual robots equipped with models of the auditory and phonatory apparatus coupled with a network of artificial neurons that connect perceptual and motor modalities. These artificial neurons determine the robots’ behaviour, mainly consisting in vocal babbling. The babbling activity, coupled with the properties of plasticity characterizing neural networks, allow the robots to learn the correspondences between the space of auditory perceptions and the space of vocal tract gestures. Finally, these robots are placed together in a shared environment where they can hear the vocalizations of their neighbours, which influence their own vocalizations, and wander around. I will show that a number of properties characterizing the vocalizations produced by robots in a given population emerge spontaneously.

More technically,[1] agents possess an artificial ear (whose properties can be modified to study their specific role: see below) capable of transforming an acoustic signal into neural impulses which stimulate neurons in an artificial perceptual neural map. They also possess a motor neural map whose neurons activation produce movements of a vocal tract model, which itself produces an acoustic wave (and which degree of realism can equally be modified). Both perceptual and motor maps are totally interconnected (see Fig. 9.4). Initially, internal parameters of all neurons, as well as the parameters of their connections, are random. To produce a vocalization, a robot randomly activates several motor neurons, whose internal parameters encode articulatory configurations which shall be reached in sequence, which in turn produces an articulatory trajectory and, through the vocal tract model, an acoustic signal that can

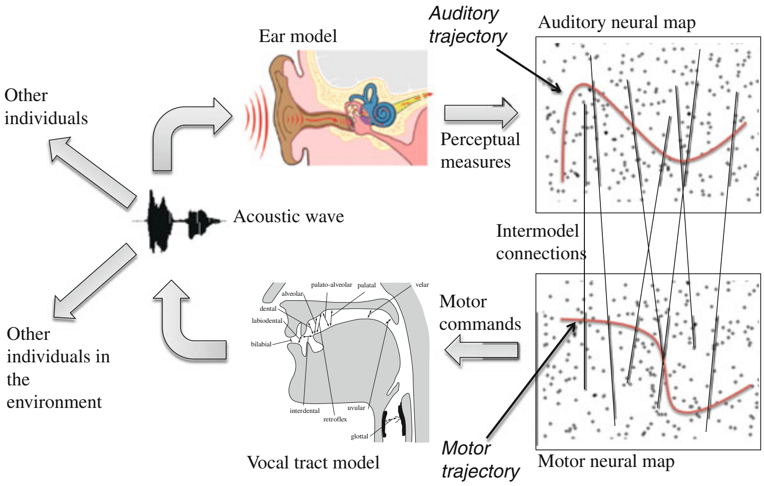


**Fig. 9.4** The robot architecture in the artificial system. Each robot is equipped with an artificial ear, an artificial vocal tract, and two neural maps (perceptual and motor) that couple the ear and the vocal tract. These neural maps are initially random but characterized by two forms of plasticity: (1) intermodal connections evolve in such a way that the robot can learn the correspondences between auditory trajectories and associated motor trajectories when the robot is babbling; (2) neurons in each map evolve in such a way that they track the distribution of sounds heard by the robots. Thus, if one exposes a robot with a continuous flow of speech of a given language, its babbling will tune/align itself to the distribution of sounds in this language.

---

[1] We only give here a general description of the system: a detailed mathematical description is available in (Oudeyer 2006).

be perceived by the ear model. This is the basis of babbling, and explains why initially, robots produce vocalizations randomly spread in the vocal continuum. These neural networks are characterized by two forms of plasticity: (1) intermodal connections evolve in such a way that the robot learns the correspondences between auditory and motor trajectories perceived and produced when it is babbling[2]; (2) neurons in each map evolve in such a way that they tend to model the distribution of sounds heard by the robot[3]; (3) The connections between both neural maps are such that the distribution of sounds encoded in the motor map follows roughly the distribution of sounds encoded in the perceptual map. This implies that the neural architecture is such that robots have the tendency to produce the same distribution of sounds as the distribution of sounds that they hear around them. Thus, if one exposes a robot with a continuous flow of speech of a given language, its babbling will tune/align itself on the distribution of sounds in this language. For example, is this language contains the vowels [a,e,i] but not [o], the robot's babbling vocalizations will quickly contain [a,e,i] much more often than [o]. This behaviour corresponds to what is observed in young infants, and referred as "phonological attunement" (Vihman 1996).

### 9.5.2 A Unified Mechanism for the Self-Organization of Combinatoriality, of the Universals/Diversity Duality, and of Cultural Sharing

This type of architecture has frequently been used in the literature to model speech acquisition in children (Kohonen 1988; Sanguineti et al. 1998), in experiments in which the system learned to pronounce sounds/syllables of a language it was exposed to. Yet, the experiment I present here is different: one does not assume that a constituted speech code exists initially. On the contrary, one places a population of babbling robots together in a shared environment, such that they can both perceive their own babblings and those of their neighbours (see Fig. 9.5). Given that the properties of plasticity of their brains make them align their babbling vocalizations on those they hear around them, and as initially they all produce random vocalizations uniformly spread in the vocal space, the initial state is an equilibrium.

Yet, if one runs the simulation, one observes that this equilibrium is not stable. Indeed, noise—stochasticity—makes that from time to time, certain types of vocal-

---

[2] Connections between the two maps evolve according to Hebb's law: those that link neurons that are often activated in a correlated manner are reinforced, whereas those that link neurons with uncorrelated activation become weaker. These connections are initially random, and through babbling and Hebb's law, they self-organize and finally allow the robot to find motor commands that correspond to a given sound that he perceives.

[3] Neurons adapt to stimuli through sensitization: their dynamics is such that if a stimulus S is perceived, then they are modified such that if the same stimulus S would be presented just afterwards they would be more activated than the first time, and the amount of modification depends exponentially on their activation (strongly activated neurons are modified most).

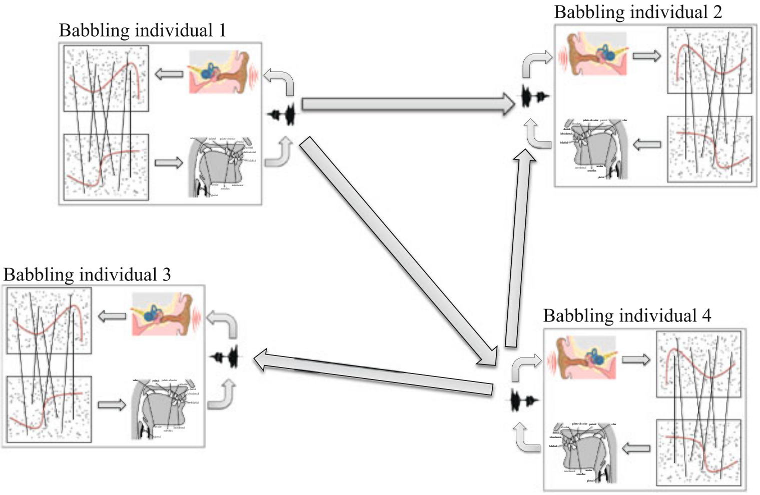


**Fig. 9.5** In the experiment, babbling robots are placed in a shared environment and can hear each other's vocalizations in addition to their own. Their neural networks are initially random, thus their vocalizations are initially random and unorganized. Given that the properties of plasticity of their brains push them to align their vocalizations with the one they hear, and as they statistically all produce vocalizations randomly uniformly spread in the articulatory space, the initial state is an equilibrium. Yet, this equilibrium is unstable and the noise inherent to their interactions and to the neural dynamics provokes a symmetry breaking: vocalizations crystallize and become discrete, combinatorial and shared by all the members of the population.

izations are pronounced a little more often than others. Now, the mechanism that couples perception and production creates a positive feedback loop: these deviations from the mean get amplified when they are big enough, and the system's symmetry breaks. Neural maps self-organize into clusters of neurons encoding particular acoustic and articulatory configurations in the space of vocalizations (see Fig. 9.6). In brief, the continuous space of vocalization has been discretized. The vocalizations produced by agents are not holistic anymore, but discrete and combinatorial: they are systematically built through the sequencing of key configurations, that we can call phonemes. One sees the formation of phonemic coding, i.e. discrete combinatorial speech codes as described earlier. Besides, the system of phonemes that self-organize is shared by all robots of a given simulation, and is different in different simulations. Thus, one observes the formation of a "cultural convention" which can be diverse across groups.

As a matter of fact, several variants of this experiment can be set up and permit to refine the conclusions. First, it is possible to experiment what happens when there is only one single robot which is listening to itself babbling. In this case, one also observes a crystallization of vocalizations: it quickly ends up producing vocal trajectories that systematically reuse few key articulatory configurations. One can deduce that the formation of phonemic coding, i.e. discreteness and combinatoriality, is not

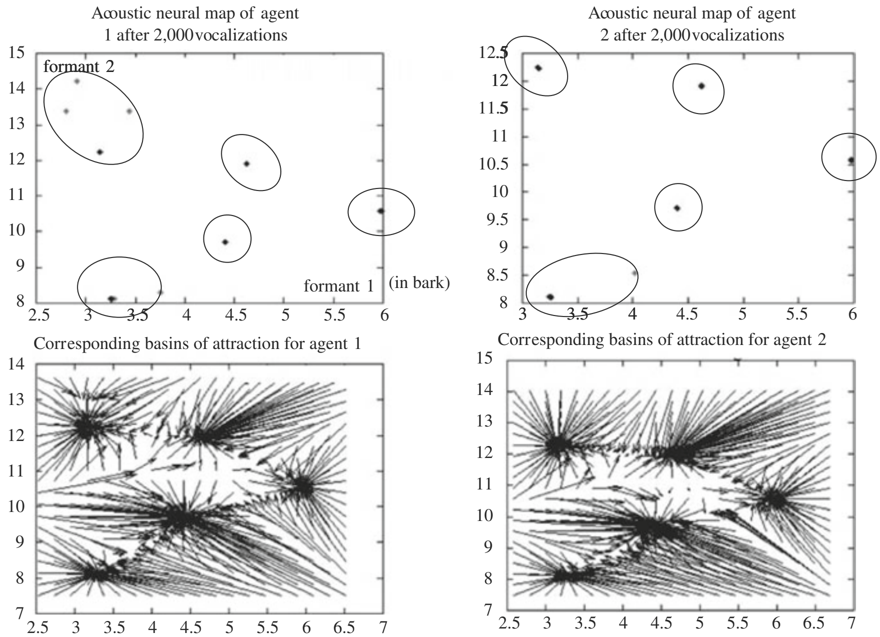


**Fig. 9.6** Very quickly, initial symmetry in the system is broken, and neurons which initially encoded random vocal configurations now encode a small number of configurations which are systematically reused by agents when they babble: the vocal space has been discretized. Besides, these self-organized elementary configurations are the same in all robots of the same population, but different across populations. One can see it on this figure which represents the perceptual neural maps of two agents after 2000 vocalizations, where the configurations encoded by neurons are clustered (*top*), as well as their corresponding distributions (*bottom*). The auditory space is here projected on the first and the second effective formant, expressed in barks, which allows us to visualize the self-organized vowel systems

the result of social interactions but rather of the internal coupling between speech perception and production. Yet, whereas the vocalizations of isolated babbling agents will crystallize on different vocalization systems, these systems will spontaneously synchronize when they share the same environment and are capable of hearing each other: in this case, the self-organized systems are approximately the same in all agents on the same population.

A second significant variant of this experiment consists in varying the morpho-physiological properties of the auditory and phonatory systems so as to determine the impact of these properties on the systems that form (or do not form). In particular, an essential property of the speech organs is the non-linearity of the function that maps articulatory configurations to acoustic waves to auditory perceptions. The human vocal tract is indeed such that for certain articulatory configurations, small variations produce small variations in the perceived sound, while for other articulatory configurations, small variations produce large variations in the perceived sound.

Now, this property is central in several theories that propose to explain why speech is phonemically coded, such as in Stevens' quantal theory (Stevens 1989) or in the DRM model (Mrayati et al. 1988). It is possible to use models of the ear and the vocal tract that are realistic and include this type of non-linearity, but it is also possible to construct on purpose a non-realistic model to evaluate the specific impact of the non-linearities. These experiments were run, and with a linear model one observes that in a population of babbling robots the crystallisation we presented above still happens: vocalizations self-organize into a combinatorial system in which particular articulatory configurations are systematically reused in vocal trajectories. Thus, we can draw a first conclusion: these simulations show that phonemic coding can appear spontaneously without non-linearities in the auditory-phonatory system. This does not imply that non-linearities do not accelerate the formation of phonemic coding, but that they are not necessary, as proposed in the quantal theory or in the DRM model.

Yet, if one looks at the distribution of self-organized key vocal configurations when one uses the linear auditory-phonatory system (which one can see as kinds of phonemes), one observes that these configurations are globally positioned randomly uniformly in the space of possible vocal configurations. But when one uses a realistic ear and vocal tract model, reproducing in particular the properties of vowel production and perception[4], one observes an additional phenomenon. Besides the crystallization which is the qualitatively same as with the linear model, the vocalization systems that form are characterized by statistical regularities that share many similarities with human vowel systems. For example, one can collect statistics on the vowel systems that appear as key reused configurations in the self-organized systems when one runs many simulations. The results, illustrated in Fig. 9.7, show that on the one hand a diversity of systems appear, and on the other hand that certain vowel systems appear much more often than others. As a conclusion, one observes the same duality of universals and diversity that characterized human languages, and the simulation suggests a unified explanation:

(1) The dynamical system composed by the set of babbling robots and the internal sensorimotor couplings possesses a number of attractors which are culturally shared combinatorial vocalization systems;
(2) Under the influence of noise and small variations of initial conditions, the dynamical system falls in a particular attractor, which allows us to explain the "decentralized collective choice" made by the population to adopt a system rather than another;
(3) Non-linearities in the auditory and phonatory systems introduce asymmetries between attractors: some of them have a larger basin of attraction, in particular those for which phonemes are in zones where small articulatory variations provoke small perceptual variations, which increases the probability that the system falls in such attractors.

---

[4] See (Oudeyer 2006) for a precise description of the model based on the work of (de Boer 2001).

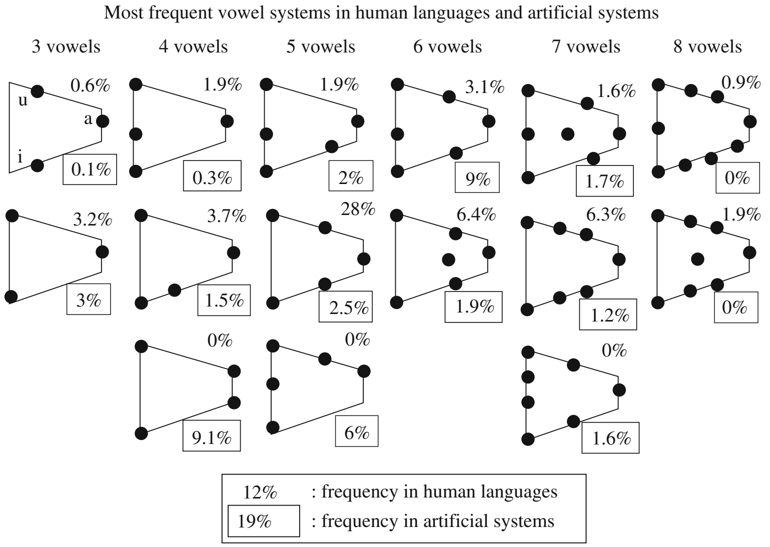


**Fig. 9.7** Comparison between the distribution of vowel systems formed by populations of robots and in human languages of the UPSID database (Maddieson 1984). Vowel systems are represented on the vocalic triangle, which horizontal dimension corresponds to the first formant, and vertical dimension corresponds to the second effective formant. One observes that the most frequent systems in the artificial system are the same as in human languages (in particular the symmetric 5 vowel system /a,e,i,o,u/ with 25 % in artificial systems and 28 % in human languages)

Furthermore, there is not only a structural correspondence between simulations and reality, but the vowel systems that appear most frequently in robot populations are approximately the same, and in the same proportion, as those that appear most frequently in human languages. Thus, there is a quantitative relationship. One can conclude that the non-linearities of the auditory and phonatory systems are decisive to explain why certain systems of phonemes are statistically more frequent than others. Yet, the very existence of these phonemes, i.e. the existence of a vocalization system in which invariant articulatory and auditory configurations are systematically reused, is not necessitated by these non-linearities.

It is important to note that in these simulations, neural architectures are characterized by several parameters, and that all values of these parameters do not lead to the above mentioned results. Nevertheless, only one parameter has a critical influence on the results (Oudeyer 2006): neurons are characterized by a selectivity to stimuli ($\sigma$) that can be focused or wide. If this selectivity is too focused, no crystallization happens, but the robots remain capable to learn the relationships between the auditory space and the phonatory/motor space. If the selectivity is too wide, the system crystallizes in a degenerate state in which all vocalization are exactly the same and are stationary: there is only one phoneme. Yet, there is a large range of intermediary values between these two extremes that allows us to observe a crystallization in which a combinatorial system with multiple phonemes self-organizes in a population of babbling robots.

### *9.5.3 Towards a Novel Vision of Evolutionary Scenarios of the Origins of Coded Speech*

The preceding discussion has important consequences if one uses the model to imagine evolutionary scenarios that may have led to the formation of the first vocalization systems featuring the fundamental properties of contemporary human speech. Indeed, they imply that many parameter variations in the neural architecture permit to generate combinatorial speech systems shared by all members of a community. Also, they imply that with such a neural architecture, a combinatorial phonemically coded vocalization system can appear without assuming special properties of the auditory and phonatory system: the only assumption is that a certain variety of sounds can be produced, but non-linearities are not required. Finally, the architecture is in itself relatively primitive: it includes neural units which intrinsic properties (dynamics and plasticity) are classical and are functionally similar to most of neural units in mammal brains (Oudeyer 2006). The specificity of this architecture relies in the systematic and direct plastic connections between the auditory and motor maps. Yet, this specificity characterizes the basic building blocks of the capacity to learn to imitate holistic sounds, which we call here ***holistic adaptive vocal imitation*** and is sometimes also called adaptive vocal mimicry or vocal learning in the animal literature, which is a capacity functionally more restricted than the capacity to learn and share combinatorial phonemically coded speech sounds. This leads us to the following evolutionary scenario to conceptualize the origins of combinatorial vocalization systems culturally shared by all individuals of a community:

(1) Adaptive vocal imitation is present in many animals (Snowdown and Hausberger 1997; Hauser 1997) which possess shared and learnt vocalization systems, but which do not possess language. Ethologists have identified many potential reproductive advantages characterizing the capacity of adaptive vocal imitation in a community of individuals (i.e. this allows individuals to mark their group membership). Thus, it is reasonable to think that before being capable of speech and language, humans may have evolved the capacity to imitate vocally;

(2) Being capable of adaptive vocal imitation, as well as the related reproductive advantages identified in non-human animals, does not imply and does not necessitate a combinatorial phonemically coded vocalization system. As an example, the range of parameter values for which the selectivity $\sigma$ is focused allows the robots to learn efficiently the vocal perceptuo-motor correspondences without generating a phonemic system;

(3) Now, if one imagines an ecological context in which the presence of a combinatorial vocalization system would provide a reproductive advantage to those that possess it, then the experiments that we described permit to state that a simple

> change in the value of the $\sigma$ parameter in motor and perceptual neural map would lead to the spontaneous formation of vocalization systems which feature several fundamental properties of contemporary human speech systems. This allows us to understand that what may have been a great leap forward for language, i.e. the formation of shared combinatorial vocalizations, may be the consequence of a small biological change together with the self-organizing properties of neural matter and multi-modal sensorimotor coupling.

Moreover, this scenario in which phonemically coded vocalization systems would have been selected thanks to the reproductive advantage that they may have provided is not the only one that the computational model can support and refine. Indeed, I explained above that in the range of $\sigma$ values that allows combinatorial systems to be formed, the capacity of adaptive vocal imitation is intact and equally efficient. Besides, performance being equal, the transition of $\sigma$ among this range and the range of more focused values for selectivity does not a priori imply a metabolic cost. This implies that in an ecological context in which those neural structures appeared under a selective pressure for adaptive vocal imitation, neutral mutation/variation and neutral drift may have happened and generated spontaneously shared phonemically coded vocalization systems without a selective pressure for language. One observation makes this scenario particularly stimulating: among animal species capable of adaptive vocal imitation in which culturally shared sound systems exist, but which do not possess language, several of them produce vocalizations or songs structured around the systematic reuse of basic units. For example, this can be observed in zebra finches (Brenowitz and Beecher 2005) or in humpback whales (Tyack 1981). The function of this quasi-phonemic structuring has been only little understood so far in ethology. Besides, because the model I presented is neutral with respect to many properties of the auditory and phonatory systems, and because the neural architecture that it assumes corresponds to the minimal equipment for adaptive vocal imitation, it can be applied to the formation of songs in these animals. In this case, it provides a hypothesis reinforced by the current scientific uncertainty about the function of combinatorial coding in these songs: combinatorial and systematically reused units may have been generated spontaneously as a collateral effect of the biological equipment for adaptive vocal imitation. Thus, it is also reasonable to imagine that this may have been the case in humans: combinatorial speech systems may have been recruited only later on to achieve their current linguistic function. This implies that several fundamental properties of contemporary human speech systems may be exaptations.[5]

---

[5] This term was introduced in (Gould and Vrba 1982). It refers to the use of a biological feature/structure for a function A which is different than the function B for which it was initially evolutionary selected.

## 9.6 Conclusion

Through the implementation and use of a computer model, I have shown how a relatively simple neural architecture coupling auditory and phonatory modalities permitted, via a self-organizing dynamics, the spontaneous formation of combinatorial phonemically coded vocalization systems shared by all members of a community, and characterized by the duality of universals and diversity. The first contribution of this work is that for the first time it provides a unified explanation of these three phenomena.

Besides, this multimodal coupling architecture corresponds to the minimal neural kit required by adaptive vocal imitation, and does not include biological elements that are specific to human speech. Given that crystallization happens in a large range of the parameter space, this shows that the transition from inarticulated vocalization systems to human-like speech codes may have been largely due to a modest biological innovation. Indeed, the model indicates that neuronal structures that encode a priori and specifically phonemic organization, as well as typical regularities of speech, do not need to be innately generated to allow the formation of such speech code. This is the second contribution of this work: it allows us to understand how the self-organizing properties of simple neural structures may have constrained the space of biological vocalization structures and how speech codes may have been generated and selected during phylogenesis.

These new hypotheses may not have been identified without the use of computer models and simulations, because the underlying dynamics are complex and difficult to anticipate just through verbal reasoning. This illustrates the potential importance of these new methodological tools in human and biological sciences. Yet, such computer models abstract many biological and behavioural mechanisms, and provide primarily a theoretical investigation of the space of hypotheses: once this space is reconceptualised, and the internal coherence of hypotheses evaluated through computer simulations, validation work and grounding of these hypotheses in biological field observations remains to be done. Thus, the third contribution of this work is, more than just the elaboration of specific hypotheses, also the construction of a framework and of tools that allow us to develop new intuitions and new concepts for our understanding of the origins and evolution of speech.

**Acknowledgments** *This work was in major part achieved in the Sony Computer Science Laboratory, Paris, and benefited from the support of Luc Steels.*

## References

Ashby, W.R.: An Introduction to Cybernetics. Chapman & Hall, London (1956)
Bak, P.: How Nature Works: The Science of Self-Organized Criticality. Copernicus, New York (1996)
Ball, P.: The Self-Made Tapestry : Pattern Formation in Nature. Oxford University Press, Oxford (2001)

Batali, J.: Computational simulations of the emergence of grammar. In: Hurford, J.R., Studdert-Kennedy, M., Knight, C. (eds.) Approaches to the Evolution of Language: Social and Cognitive Bases. Cambridge University Press, Cambridge (1998)

Berrah, A., Glotin, H., Laboissière, R., Bessière, P., Boë, L.J.: From form to formation of phonetic structures: an evolutionary computing perspective. In: 13th International Conference on Evolutionary Computing and Machine Learning (ICML), pp 23–29. MIT Press, Bari (1996)

Bonabeau, E., Theraulaz, G., Deneubourg, J.L., Aron, S., Camazine, S.: Self-organization in social insects. Trends in Ecology and Evolution **12**, 188–193 (1997)

Brenowitz, E.A., Beecher, M.D.: Song learning in birds : diversity and plasticity, opportunities and challenges. Trends in Neuroscience **28**(3), 127–132 (2005)

Camazine, S., Deneubourg, J.-L., Franks, N.R., Sneyd, J., Theraulaz, G., Bonabeau, E.: Self-Organization in Biological Systems. Princeton University Press, Princeton (2002)

Dawkins, R.: The Blind Watchmaker. W. W. Norton and Company, New York (1986)

de Boer, B.: The Origins of Vowel Systems. Oxford University Press, Oxford Linguistics, Oxford (2001)

Elredge, N., Gould, S.J.: Punctuated equilibria : an alternative to phylogenetic gradualism. In: Schopf, T.J.M. (ed.) Models in Paleobiology, pp. 82–115. Freeman, San Francisco (1972)

Gould, S.J., Vrba, E.S.: Exaptation: a missing term in the science of form. Paleobiology **8**(1), 4–15 (1982)

Hauser, M.D.: The Evolution of Communication. MIT Press, BradfordBooks, Cambridge (1997)

Kauffman, S.: At Home in the Universe : The Search for Laws of Self-Organization and Complexity. Oxford University Press, Oxford (1996)

Keefe, A., Szostak, J.: Functional proteins from a random sequence library. Nature **410**, 715–718 (2001)

Kaplan, F.: La naissance d' une langue chez les robots. Hermès (2001)

Kaplan, F., Oudeyer, P.-Y., Bergen, B.: Computational models in the debate over language learnability. Infant Child Dev. **17**(1), 55–80 (2008)

Kirby, S.: Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Trans. Evol. Comput. **5**(2), 102–110 (2001)

Kohonen, T.: The neural phonetic typewriter. Computer **21**(3), 11–22 (1988)

Liljencrantz, J., Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrast. Language **48**, 839–862 (1972)

Maddieson, I.: Patterns of Sounds. Cambridge University Press, Cambridge (1984)

Mrayati, M., Carre, R., Guerin, B.: Distinctive regions and modes: a new theory of speech production. Speech Commun. **7**, 257–286 (1988)

Nicolis, G., Prigogine, I.: Self-Organization in Non-equilibrium Systems: From Dissipative Structures to Order Through Fluctuations. Wiley, New York (1977)

Oudeyer, P.-Y.: Origins and learnability of syllable systems, a cultural evolutionary model. In: Collet, P., Fonlupt, C., Hao, J., Lutton, E., Schoenauer, M. (eds.) Artificial Evolution; LNCS 2310, pp. 143–155 (2001)

Oudeyer, P.-Y.: The self-organization of speech sounds. J. Theor.l Biol. **233**(3), 435–449 (2005a)

Oudeyer, P.-Y.: The self-organization of combinatoriality and phonotactics in vocalization systems. Connection Sci. **17**(3–4), 325–341 (2005b)

Oudeyer, P-Y.: Self-Organization in the Evolution of Speech. Oxford University Press, Oxford (2006)

Oudeyer, P.-Y., Kaplan, F.: Language evolution as a darwinian process: computational studies. Cogn. Process. **8**(1), 21–35 (2007)

Pierrehumbert, J.: Exemplar dynamics: word frequency, lenition, and contrast. In: Bybee, J., Hopper, P. (eds.) Frequency Effects and the Emergence of Linguistic Structure. John Benjamins, Amsterdam (2001)

Pfeifer, R., Scheier, C.: Understanding Intelligence. MIT Press, Cambridge (1999)

Sanguineti, V., Laboissière, R., Ostry, D.J.: A dynamic biomechanical model for neural control of speech production. J. Acoust. Soc. Am. **103**(3), 1615–1627 (1998)

Snowdown, C.T., Hausberger, M.: Social Influences on Vocal Development. Cambridge University Press, Cambridge (1997)
Steels, L.: The synthetic modeling of language origins. Evolution of communication, **1**(1), 1-34 (1997).
Steels, L.: The methodology of the artificial. Behav. Brain Sci. **24**(6), 1077–1078 (2001)
Steels, L.: The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. Connection Sci. **17**(3–4), 213–230 (2005)
Stevens, K.N.: On the quantal nature of speech. J. Phonetics **17**, 3–45 (1989)
Studdert-Kennedy, M., Goldstein, L.: Launching language: the gestural origin of discrete infinity. In: Christiansen, M., Kirby, S. (eds.) Language Evolution: The State of the Art, pp. 235–254. Oxford University Press, Oxford (2003)
Tritton, D.J.: Physical Fluid Dynamics. Oxford University Press, Oxford (1988)
Tyack, P.: Interactions between singing hawaian humpback whales and conspecifics nearby. Behav. Ecol. sociobiol. **8**(2), 105–116 (1981)
Vichniac, G., Manneville, P., Boccara, N., Bidaux, R. (eds.): Cellular automata and modeling of complex systems. In: Workshop Les Houches. Springer, Heidelberg (1989)
Vihman, M.: Phonological Development: The Origins of Language in the Child. Blackwell, Cambridge (1996)
Waldrop, M.: Spontaneous order, evolution, and life. Science **247**, 1543–1545 (1990)
Wedel, A.: Exemplar models, evolution and language change. Linguist. Rev. **23**, 247–274 (2006)
Weisbuch, G.: Complex systems dynamics. Addison-Wesley, Redwood City, CA (1991)

# Chapter 10
# Environment: Language Ecology and Language Death

**Suzanne Romaine**

**Abstract** Global linguistic diversity is rapidly declining. As the world becomes less linguistically diverse, it is becoming culturally less diverse as well as the world's tribes and languages are dying out or being assimilated into modern civilization because their habitats are being destroyed. At the same time the world is experiencing a substantial decline in biodiversity. The extinction of languages is part of the larger picture of near total collapse of the worldwide ecosystem, and languages are vital parts of complex local ecologies that must be supported if global biodiversity is to be maintained.

## 10.1 Introduction

As a species humans display remarkable cultural and linguistic diversity despite a high degree of genetic uniformity (Pagel and Mace 2004). This diversity is at risk when traditional cultures and languages disappear along with the complex ecosystems that sustained them. According to some estimates, 50–90 % of the world's 6,900 or more languages may soon become extinct (Nettle and Romaine 2000). Global biodiversity also faces an extinction crisis, with annual losses of plant and animal species estimated at 1000 times or more greater than historic background rates (Millennium Ecosystem Assessment 2005). Much of the world is now being covered by a few species of Eurasian origin such as wheat, barley, cattle, rice. These monocultures are replacing a profusion of endemic diversity. The situation for languages is strikingly similar, but the spreading varieties are large languages like English, Spanish and Chinese, etc.

The idea of language ecology has become a prominent theme in recent discussions of language endangerment (Nettle and Romaine 2000; Mufwene 2001; Mühlhäusler 2003).

---

S. Romaine (✉)
Merton College, Oxford OX1 4JD, UK
e-mail: suzanne.romaine@ling-phil.ox.ac.uk

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_10,

The ecological approach stresses the whole rather than the parts by attending to interrelations and interdependencies among the entities inside the ecosystem. The basis of the ecological perspective rests on the premise that objects are principally networks of relationships embedded in larger networks. Evolution is therefore always coevolution because all organisms evolve interdependently with the others in their environment (Capra 1997, p. 37). Ultimately all extinctions have as their cause environmental change. As Wilson (1998, p. 328) acknowledges in connection with the disappearance of species, "the outright elimination of habitat … is the leading cause of extinction. But the introduction of aggressive exotic species … come[s] close behind in destructiveness".

Although languages have no tangible existence like people or trees, they are, nevertheless, like species, highly adapted to their environments. To understand why languages die entails looking not just at the languages themselves but at all aspects of the lives of the people who speak them. Some time ago Haugen (1972a, p. 325) called on linguists to take account of what he referred to as "the ecology of language". Inspired by what he believed to be the first use of the term by Voegelin and Voegelin (1964, p. 2), who suggested that "in linguistic ecology, one begins not with a particular language but with a particular area", Haugen defined this domain of inquiry as the "study of interactions of any given language and its environment". Language as a system and as a set of verbal practices and behaviors evolves by means of a process of competition and selection. This chapter examines the utility of the ecological framework in understanding the dynamics of language endangerment and death by looking at ecological niche theory, social networks and language competition. In addition to sketching out some of the theory underlying these ideas, this chapter will also examine some of the computational modeling techniques used in understanding populations and ecosystems and show some of the ways in which they are helpful in explaining the evolution of linguistic diversity, and in particular language death.

## 10.2 The Ecology of Language: Metaphors and Models

In many respects the ecology of language framework and the notion of 'ecology' (from Greek *oikos* 'home') are apt in connection with the focus of this book on describing linguistic phenomena at various scales, and examining how these affect or are affected by phenomena at other scales. Language is a complex object that can be variously regarded as system, process, activity, behavior and product depending on the locus of our observations and the nature of the phenomena we choose to focus on. Although Haugen does not actually use the term 'complexity' or refer to language as a complex object, the idea is implicit in his pioneering effort to clarify the requirements for a proper analysis of the ecology of language. In (Haugen's 1972a, p. 334) view this entailed not only a description of the "sociological and psychological situation of each language, but also the effect of this situation on the language itself".

> The true environment of a language is the society that uses it as one of its codes. Language exists only in the mind of its users, and it only functions in relating these users to one another and to nature, i.e. their social and natural environment. Part of its ecology is therefore psychological: its interaction with other languages in the minds of bi- and multilingual speakers. Another part of its ecology is sociological: its interaction with the society in which it functions as a medium of communication. The ecology of a language is determined primarily by the people who learn it, use it, and transmit it to others (Haugen 1972a, p. 325).

In drawing attention to the struggle between dominated and dominant groups for the right to survive, Haugen (1972a) also emphasized that the preservation of language is part of human ecology. The use of terms like ‘death’, ‘endangerment’, ‘extinction’, ‘preservation’ and ‘revitalization’ in relation to languages may at first glance sound strange or inappropriate. In the nineteenth century when linguistics was influenced by biological theories of evolution and similarities between the process of language divergence and biological speciation, it was common to talk about the ‘life of language’ and to speak of languages as if they were living organisms. Indeed, Haugen (1972a, p. 326) dismissed such parlance as “a metaphor only, which brought out certain analogues between languages and biological organisms, but could not be pushed too far”. While acknowledging with Haugen (1972a, p. 326) that language “has no life of its own apart from those who use it”, Nettle and Romaine (2000) contend that more is at stake than analogies and metaphors in their elaboration of the links between language survival and the environment. Language shift and death occur as responses to pressures of various types (social, cultural, economic and even military) on a community. Communities can thrive only where their members have a decent environment to live in and a sustainable economic system. Languages can only survive if there is a viable community speaking them and passing them on from parent to child at home. When languages lose their speakers, they die.

Consider the fate of the Ugong language in western Thailand. In the late 1970s the Electricity Generating Authority of Thailand built two hydroelectric dams on the two branches of the River Kwai. These dams flooded the locations of two Ugong villages and the inhabitants were relocated elsewhere. With the unity of the villages destroyed and their speakers scattered, the older speakers who still preserve the language have few, if any, people to speak to in Ugong. Ugong has literally been swamped and the speakers immersed in Thai villages. Children speak Thai as first language and fewer than a hundred adults use the language at home (Bradley 1989).

The fate of Ugong demonstrates clearly how languages, like cultures, are complex adaptive systems that evolve in specific environments to meet the communicative requirements of their speakers. A language is enmeshed in a local social and geographical matrix just as a rare species is enmeshed in an ecosystem. Even a small amount of environmental change can cause a cascade of extinctions as the dependent species become stressed. Take, for example, the rare Kirtland’s warbler, whose numbers plummeted during the latter half of the nineteenth century when its breeding range was restricted by destruction of its highly distinctive habitat. It lives in just a few counties in Michigan in dense stands of young 5–20 year old Jack pines. The trees providing the warblers’ nesting area must also be just the right height (ca. 5–16 feet tall) with just enough space between them to let sunlight through to the ground.

The birds nest on the ground under the bushy lower branches. If the trees grow too tall, the upper branches block the sunlight and the lower branches die. The birds then look for new nesting grounds. The pines themselves have a restricted climatic and topological range, growing best in sandy soil in a cool temperate climate. Historically, they were maintained by naturally occurring wildfires that swept through the region and removed the older trees. However, once human settlement, logging and the spread of farming altered the distinctive habitat on which the warbler depended, the cowbird was able to expand its range into the warbler's territory. The cowbirds do not build their own nests, but instead invade the warblers' nests to lay their eggs.

The good news for the warbler, however, is that several decades of forest management by the U.S. Department of Agriculture Forest Service and the Michigan Department of Natural Resources in the form of logging, burning, seeding, and replanting along with measures to control cowbirds have facilitated the recovery of birds' nesting habitat and with it an increase in the Kirtland warbler population. Nevertheless, this brief sketch underlines the vulnerability of the Kirtland's warbler; if for any reason the Jack pines were to be threatened or destroyed, through, say, climate change or logging, the warbler would disappear.

Unfortunately, the prognosis for Ugong and many of the world's other languages spoken by small groups is much less certain. In the linguistic domain too the problem lies in the changing ecology of the speakers of a language. In discussions of language maintenance, revitalization, etc. there is a tendency to reify languages as systems and structures, when it is communities, communicative practices and language ecologies we should be talking about. The analogies between the Kirtland's warbler and Ugong are clear. If we want to save the warbler, we must preserve its habitat. Similarly, to revitalize and maintain languages, we must try to preserve language ecologies. To understand what is happening to languages, and what it means for the future of linguistic diversity, however, we have to examine the broader and more fundamental social pressures active in the world today that are altering language ecologies, such as the huge differences in numbers and economic power between the peoples of the world. Only by understanding the various historical forces that have shaped the evolution of languages and led to the expansion of some languages and the contraction of others will we be able to do anything about the loss of linguistic diversity.

## 10.3 Survival of the Fittest or Survival of the Most Powerful?

For most of the many millennia of human history people lived in small communities quite distinct from their neighbors. The world was close to linguistic equilibrium with the number of languages lost roughly equaling the new ones created. No one knows exactly how many languages or cultures there are in the world either historically or currently, but as recently as 500 years ago there may have been some 12,000–14,000 cultural groups. More than half of this diversity has possibly been lost (Sponsel 2000). Perhaps half the known languages in the world have disappeared over this same time period (Nettle and Romaine 2000).

Until recently in human history, there were no massive, enduring differences between the expansionary potential of different peoples, of the kind that might cause the sustained expansion of a single, dominant language. Now all but a handful of hunter-gatherer societies live outside their local ecosystems. Over the past 10,000 years, however, a variety of events have punctured this equilibrium forever. Firstly, the invention and spread of agriculture, colonialism, later the Industrial Revolution, and today the globalization of economies and mass media (particularly the internet) have created the global village phenomenon. These forces have enabled a few languages- all Eurasian in origin- to spread over the last few centuries.

The rise to dominance of a few Eurasian languages and the global spread of western technology and culture do not however represent a case of survival of the fittest, or a triumph of some innately superior civilization. Instead, they are the result of deeper, complex, structural conditions occurring in Eurasian societies at just the right time. Eurasia had by far the world's most productive farming and livestock complex. Although this was no more than a fluke of biogeography, it allowed Eurasian populations to boom, and eventually expand beyond their shores. However, it also made them hosts to great killer diseases, which, paradoxically, gave them an edge over other peoples when the continents collided. Finally, the dense population and the high agricultural productivity, in Europe at least, unleashed a process of diversification and specialization that set those economies on the path to industrialization.

As large language communities expanded, others contracted. The spread of large languages in modern times has resulted in marked demographic disparities in the size of populations speaking the world's languages. If languages were equal in size, each would have around 917,000 speakers. However, the median number of speakers for the languages of the world is only 5,000–6,000. The world's top nine languages each with more than 100 million speakers are spoken by just over 41 % of the world's population. Nearly 80 % of the world's population speaks a total of only 83 languages. Each of these has 10 million or more speakers. Leaving aside the world's largest languages, however, almost 82 % of the languages of the world have fewer than 100,000 speakers; almost 56 % are used by only ten thousand or fewer speakers. There are around 548 languages with fewer than 99 speakers, comprising nearly 1/10 of the world's languages (Gordon 2005).

Over the last five hundred years small languages nearly everywhere have come under acute threat. Most, if not all, of these may be at risk because small languages can disappear much faster than larger ones due to the vulnerability of small groups to external pressures in a rapidly changing world. The fate of Ugong is a case in point. The developed nations of the world are now rapidly destroying the habitats sustaining much of the world's biological and linguistic diversity. Resource extraction, the spread of mechanized agriculture and development projects damage the environment at the same time as they displace and marginalize people from places they traditionally relied on for their food, shelter, cultural practices and spiritual well-being. Since 1900, 90 of Brazil's 270 Indian tribes have completely disappeared. More than 2/3 of the remaining ones have fewer than 1,000 members. The fate of most of the world's linguistic diversity, and by implication its cultural diversity, lies in the hands of a small number of people who are the most vulnerable to pressures of globalization.

The changing face of linguistic diversity in the modern world is thus not really a case of survival of the fittest, but rather the story of how a few metropolitan languages expanded very rapidly at the expense of the rest, as smaller communities have been pulled into the orbit of more powerful ones. These power disparities have allowed a few metropolitan groups a virtual stranglehold upon global resources and global power. This power takes many forms, among them controlling the flow of information, through radio, television and the internet. The scenario I have sketched of the changing dynamics of the evolution of linguistic diversity over the course of human history would argue strongly against Friedman's (2005) flat world theory. Friedman contends that globalization has created a potentially level playing field for all competitors who supposedly have equal opportunity in the new interlinked global world of commerce. In discussing the advantages and disadvantages of the flattening of the world, he mentions that the greatest downside affects individuals, organizations and societies that are unable or choose not to participate.

From the vantage point of the ecological view of society developed in this chapter, people are actors in a complex field whose boundaries are set by physical geography and natural resources, by their own knowledge and opportunities, and by the behavior of others around them. Within this constantly changing dynamic people make choices in everyday conversational interaction that affect their linguistic repertoires by deciding, among other things, which languages and language varieties to use, and which of these to pass on to their children. Although these choices are not always conscious, over time they become cumulative, leading to language shift and death. The creole-based approach advocated by Mufwene (2001) recognizes that the ecological setting in which language contact takes place accounts for variation in the way competition and selection among features is resolved in the birth of new languages such as pidgins and creoles. Similar selective pressures likewise operate in favor of one or more competing variant forms in one language or variety. Thus, the same forces drive language change and the evolution of language more generally (see, for example, Croft, this volume). In the next two sections I will draw on ecological niche theory and social network theory to illustrate how these evolutionary pressures work in the case of language extinction.

## 10.4 Ecological Niche Theory and Social Networks

The concept of 'ecological niche' has played a central role in understanding the evolution and survival of species by describing how organisms or populations respond to the distribution of resources and competitors. Due to pressure from, and interactions with other organisms, especially superior competitors, species are usually forced to occupy a narrow niche to which they are mostly highly adapted. The Kirtland's warbler is again a case in point. An examination of the global distribution of species and languages reveals some significant parallels and overlaps. Linguistic diversity, like biodiversity, is heavily concentrated through the tropics and tails off towards the poles. Tropical ecosystems are typically rich in number of species, but poor in number of organisms; this is the opposite of northern latitudes. Thus, the population of any one species may be relatively small; it is variety which is great- a characteristic of a stable ecosystem. If we apply this biological analogy to languages, then we would expect to find great numbers of languages spoken by relatively small numbers of speakers in the tropics. That is precisely what we tend to find. The greatest linguistic diversity lies in some of the ecosystems richest in biodiversity that are home to indigenous peoples, who represent around 4 % of the world's population, but speak at least 60 % of its languages (Nettle and Romaine 2000). Like the Kirtland's warbler, the majority of the world's languages are narrowniched, confined to a single small group, village, or territory, while relatively few languages, like English, Arabic, Chinese, Spanish etc., which have spread far beyond their original communities of users, are broadniched. For example, Hórom, spoken by 500 people in northern Nigeria, is restricted to a single village.

Because regions containing exceptionally high degrees of biological diversity also contain very high linguistic diversity, the areas identified as biodiversity hotspots by conservationists such as Myers et al. (2000) and most recently Mittermeier et al. (2004) are also hotbeds of linguistic diversity, e.g. New Guinea. Nettle and Romaine (2000) use the term 'biolinguistic diversity' to refer to the rich spectrum of life encompassing all the earth's species of plants and animals along with human cultures and their languages (see also Maffi 2001, 2005; Oviedo et al. 2000; Loh and Harmon 2005). Not only do biodiversity and linguistic diversity share the same geographic locations, they face common threats. The same amount of habitat destruction in the tropics, home to 2/3 of Earth's biodiversity, would lead to many more extinctions than would occur in the higher latitudes. Although the precise details of the direct and indirect causes of these parallel declines in linguistic diversity and biodiversity are extremely complex, both species and languages are in danger of becoming extinct in the same places, with languages at even greater risk than species (Sutherland 2003).

Nettle and Romaine (2000) argue that the explanation for the linkages between biological and cultural-linguistic diversity will need to be sought in a sophisticated ecological theory that takes account of peoples' interactions with their environment. Ecological risk has a significant influence on the formation and persistence of linguistic groups. This factor refers to the amount of variation people face in their food supply over time, which in turn is related to other variables such as climate, diversification of productive and income-generating activities, food storage, mobility, and patterns of social exchange. In areas where rainfall is continuous throughout the year and communities are able to produce their own food supply, they are not so dependent on their neighbors for subsistence. Nettle's (1999a) study of the distribution of West African languages revealed a correlation between the length of the rainy season and the number of languages in a region. Only 20 languages are found in Niger, a vast, arid region, while farther south 430 languages are found in equally large but wetter Nigeria. These findings add weight to the conclusion that distinct languages may be more likely to evolve and be maintained in small, self-sufficient communities. The greater the ecological risk, the more people must develop larger social networks to ensure a reliable food supply. Because language norms spread through social networks, the average size of a language group increases in proportion to ecological risk. A wetter climate also enhances the evolution and coexistence of more plant species. Areas containing high biodiversity provide more niches for human populations, allowing higher cultural diversification.

## 10.5 Language Competition, Social Networks and Language Death

The previous section has shown how ecological niche theory can be useful for understanding processes of diversification by providing a general framework for examining the interlocking matrix of political, geographical and economic factors that support the maintenance of biological, linguistic and cultural diversity. Such a theory can also help us to analyze some of the forces threatening to disrupt language ecologies. Although all languages may be potentially capable of fulfilling all communicative functions, in practice they are unequal because they evolve differently through processes of adaptation, variation and selection. Where two or more language compete for the same communicative functions, speakers have to make choices. Every time a language stops performing a particular function, it will lose some ground to another that takes its place. Death occurs when one language replaces another over its entire functional range, and parents no longer transmit the language to their children.

Bourdieu (1977, 1991) developed the metaphor of a linguistic market in referring to language as a form of capital, which, like all other forms of capital, is unequally distributed in society. Those who control particular linguistic resources are in a position of power over others. Their power is economic, cultural and social as well as symbolic. Some languages have primarily symbolic and cultural capital that enables speakers to accrue local prestige within particular social networks. Indeed, such forms of capital may be the main forces sustaining the maintenance of non-standard and minority languages. Otherwise, speakers would not maintain varieties that deny or hamper their access to power (Haugen 1981, p. 100). Other languages, however, allow access to domains linked with economic power. Thus, a key question regulating the forces of the linguistic market is whether symbolic and cultural forms of capital can be converted into economic capital. The higher the profit to be achieved through knowledge of a particular language, the more it will be viewed as worthy of acquisition. Here too Haugen (1973, p. 34) was perhaps conceptually ahead of his time when he wrote "That *learning* is the key to every language problem is so obvious as to be almost a truism". Haugen (1981, p. 114), who also used the marketplace metaphor, put it this way:

> Wherever languages are in contact, they are in competition for users. They may be seen as commodities on a *language market*, and they will live only as long as they find customers who will buy them. Language competence is a skill with market value that determines who will acquire it. The price of a language is the effort required to learn it and its value is the benefits its use will bring to the learner.

Languages of colonial conquest and dominant languages of nation-states penetrate into, transform and undermine a minority community's ability to maintain its language, culture and identity in various ways. Usually the dominant language prevails in all areas of official life (e.g. in government, school and media), necessitating bilingualism on the part of the minority. Eventually, the dominant language tends to invade the inner spheres of the subordinate language, so that its domains of use become even more restricted. The prestige of the dominant language and its predominance in public institutions leads the community to devalue their own language, culture, and identity as part of a process of symbolic domination. Hindley (1990, p. 179), for instance, observed how the association between Irish and poverty was reflected in common folk sayings such as "Irish will butter no bread". Writing of Scottish Gaelic speakers who emigrated to Cape Breton, Nova Scotia, Mertz (1989, p. 12) remarked that young people's denials of any knowledge of Gaelic represent attempts to deny an image of themselves as poor or lower-class. As knowledge of English was required for assimilation to and social mobility within mainstream Canadian English speaking society, the symbolic linkage between Gaelic, rural 'backwardness' and economic hardship propelled language shift.

At the micro-level, specific patterns of sociolinguistic interactions among the individuals comprising the population, in particular those of key agents who may exert greater influence over others, is what drives speakers' choices in favor of some varieties over others (and likewise, within those varieties, of some variant forms over others). Networks are a critical vector for change. Milroy (2002, p. 558), for example, has argued that "networks constituted chiefly of strong ties function as a mechanism to support minority languages, resisting institutional pressures to language shift, but when these networks weaken, language shift is likely to take place." Relying on the notion of network power, Grewal (2008) contends that globalization is best understood as the rise to dominance of shared standards involving media, trade, as well as language and culture. Although in theory we can choose among the standards, in practice our choices tend to narrow over time because all networks have standards embedded in them. The price of gaining access to a global network is adopting a dominant standard at the expense of alternative ones. A global standard that has become dominant threatens to eliminate local and less powerful ones. From Grewal's perspective then, the evidence that Friedman (2005) offers for the flattening of the world might better be understood as arguments about how the world has become more networked.

Changes in global networks of communication have altered the linguistic ecology and transformed the world's linguistic landscape. This means that the evolution of linguistic diversity needs to be studied within the context of a language ecology that takes into account the social networks mediating communication and the distribution of resources that determine the utility and value of particular languages and varieties.

## 10.6 Dynamic Modeling of Language Competition and Language Death

The ecological view of languages as evolving, competing sets of communicative practices and resources in time and space rather than as abstract structures has been the subject of computational modeling in a series of studies simulating dynamic and complex interactions between populations speaking different languages. Although no model approaches the level of algorithmic complexity that actually exists in the real world in terms of number of variables affecting speakers' behavior, computational models are nevertheless useful tools despite some of the simplifying assumptions they rely on. It is a truism that reality is complex and messy and that all models are simplifications in some respects, but as Schulze et al. (2008, p. 289) point out, "reductionism is at the heart of any scientific enterprise". Beginning with simplified models allows researchers to focus on a few key variables in order to understand the fundamental dynamic of competition and then to add other parameters to see if they increase explanatory power. Models vary according to the mathematical computations performed on the data, number of speakers, their organization in a social network, and the behavior of speakers. Typically the state of the model is described as the proportion of speakers who have adopted a certain language at a particular time.

Abrams and Strogatz (2003), for instance, presented a simple competition model involving two languages, which they tested for accuracy against historical demographic data for three endangered languages: Scottish Gaelic, Welsh and Quechua. Their model considered two languages competing with one another for speakers. Speakers are attracted to choose one language over another based on the status and size of the language. For simplicity's sake, it was assumed that the population was monolingual, highly interconnected, had a uniform social structure and age distribution, and that individuals interacted with each other at the same rate. The model predicts that two languages cannot coexist stably; the more attractive one will eventually drive the less attractive one to extinction. Thus, attractiveness is essentially a zero-sum game; it is impossible to make one language more attractive without making the other less so.

This model relied on a number of simplifications in order to test the predictive power of status as a variable. In reality, most people in the world are bi- or multilingual rather than monolingual and language shift normally involves transitional bilingualism rather than a shift from monolingualism in one language to monolingualism in another (Romaine 1995). Indeed, stable bilingualism may persist for centuries under certain circumstances without leading to shift or death of one of the languages. By focusing on the expected aggregate behavior of the population as a whole, the model also ignored the possible differential social impact of the behavior of individual speakers, especially the role of bilingual agents in different kinds of social networks. Language maintenance and language shift are the long-term, collective consequences of consistent patterns of language choice, typically unconscious, on the part of these agents. Real speakers are of course complex interacting individuals with a variety of social characteristics (e.g. age, sex, ethnicity, etc.) whose language use patterns are influenced by many factors other than status. Moreover, status is itself a multidimensional variable, having to do not just with the prestige, power and wealth of the speakers of a language, but also with the domains in which a language is used, e.g. government, education, media, and speakers' attitudes. In addition, the model also did not distinguish the vertical (i.e. across generations) and horizontal transmission (i.e. across individuals) of language. Decisions made by bilingual parents about which language(s) to pass on to their children are critical to the survival of endangered languages, but their choices are not made in a social vacuum (Romaine 1995).

A number of subsequent studies extended the Abrams-Strogatz model in various ways by taking into account other variables such as bilingualism, and/or by altering the mathematical computations. Patriarca and Leppänen (2004), for example, incorporated population dynamics and reaction-diffusion equations. Their model predicted that two competing languages can coexist if the main concentrations of their speakers are in two separate geographical areas, with diffusion in the border zones and higher status in one geographical area. Although this model correctly predicts the survival of a minority language occupying a particular region if that minority language is spoken as the majority language in some other region (e.g. Swedish in Finland and Finnish in Sweden), it does not take account of cases where two languages co-exist and compete for speakers in the same zone. Wickstrom (2005) introduced an element of decision-making into a model incorporating the competing advantages of the communicative function of language as opposed to the emotional attachment speakers had to it as a carrier of cultural identity. The trade-off between the communicative value of a language and its importance as an identity marker was assumed to provide the basis for parents' choices in raising their children. In this way transmission was not purely mechanically determined by the type of parent, but partially due to decisions made by parents. The model predicted that various combinations of status and emotional attachment can result in stable co-existence of two languages.

Pinasco and Romanelli (2006) applied a Lotka-Volterra model of a type used to model competition between a predator and its prey or the propagation of an epidemic or virus to examine competition between languages in the same area. They found that stable co-existence was possible without spatial segregation if speakers of the minority population reproduced at a high enough rate to counterbalance losses due to language shift. They concluded that the death of a language seems to be caused by a hostile environment rather than by the influence of a more attractive language. In addition, Pinasco and Romanelli draw an analogy between language competition and the spread of infectious diseases to suggest possible mechanisms for stabilizing the coexistence of two competing languages. One way of preventing infection is to isolate a population from contact with the infected agents, but they recognize that one cannot segregate populations speaking different languages with one another.

The really big challenge for maintaining linguistic diversity lies not in preventing language contact, increasingly impossible in today's highly interconnected world, but in managing it to ensure fair conditions under which speakers of different languages can co-exist without one language threatening the survival of another. The kinds of policy and planning required to accomplish this goal are the subject of an extensive literature, but lie beyond the scope of this chapter (see, for example, Romaine 2006, 2007). Attempts at producing computation models for planners are only just beginning. This is a welcome move because computational models are not easily understood by non-mathematicians, thus rendering them of limited use not only to linguists but also to the social planners, politicians, educational reformers, applied linguists, and market analysts to whom their predictions are of most interest. As Wyburn and Hayward (2009, p. 626) point out, most computational models produced thus far have been concerned with attainment of equilibria over the long-term, perhaps over centuries, during which time their parameters are unlikely to remain constant (see also Kandler and Steele 2008).

Many instances of language shift and death occur under duress and stressful social circumstances. Shift can sometimes be quite rapid, especially in a small group under threat. It makes intuitive sense that change is more easily diffused among a small group of 500 people than among a group of thousands or millions and Nettle (1999b) has shown that small languages tend to change faster than large ones. A language that has been demographically stable for centuries may suddenly reach a critical tipping point as a result of a series of factors, and then dramatically experience what (Dorian 1989, p. 9) calls an abrupt transmission failure. Some of the local varieties of German brought to the United States by Anabaptist immigrants such as the Amish, Hutterites, Mennonites, and others have survived alongside English for nearly 400 years. This persistence is quite remarkable in view of the fact that for most immigrant minorities in the US, bilingualism is largely a temporary and transitional stage in intergenerational language shift occurring over three and sometimes even two generations. Among the most conservative groups like the Old Order Mennonites a strict and stable situation of diglossia with bilingualism exists with no mixing of English and German. Functional compartmentalization seems to work in a way similar to territorial segregation, but among other less conservative groups and particularly among nonsectarians, as soon as English intrudes into what were German domains, shift to English is swift and complete.

In practice, measures to increase and promote social or economic opportunities for using minority languages and revitalizing endangered languages offer no guarantee that people will use them. Language vitality depends not just on capacity to use the language but also opportunities and desire to do so. Minett and Wang (2008) extended the Abrams-Strogatz model to examine the role of bilingualism and social structure in order to asses the impact of two strategies for language maintenance, i.e. the status of the endangered language and the availability of monolingual and bilingual educational resources. They found that by increasing the status of an endangered language or the availability of educational resources in each of the competing languages the dynamics could be altered so that both languages are maintained with non-negligible probability. Intervention to enhance the viability of the endangered language should, however, be undertaken within a certain time window, before it becomes moribund. Although most scholars of language shift have lamented the fact that most efforts aimed at reversing language shift are generally undertaken too late (Fishman 1991; Romaine 2007), Minett and Wang (2008, p. 42) model produced the somewhat counter-intuitive finding that greater success was obtained by implementing the maintenance strategy as late rather than as soon as possible. This requires more testing.

Wyburn and Hayward (2008) also applied a predator-prey model to predict the viability of Welsh, Breton, Irish, and Scottish Gaelic. Their predictions, however, are subject to the substantial caveat that available resources such as census and other data on the size of speaker populations do not provide unique estimates of the linguistic parameters. Moreover, census information is not comprehensive and multiple resources must be used. Despite these limitations and some reservations about some aspects of the predator-prey interaction paradigm, the model made apparently successful predictions except in the case of Irish. It predicted death for Breton and Scottish Gaelic and maintenance for Welsh. The prognosis of maintenance for Welsh is based on the most recent 2001 census statistics, which showed an increase of 2 % in the number of Welsh speakers over the 1991 census. This is the first ever reported increase; ca. 20–23 % of the population reports being able to speak Welsh. The largest number of speakers (40.8 %) is concentrated in the 5–15 year old group, testament to the role of strong education policies favoring Welsh.

The case of Irish is particularly complicated. The 2006 census numbers actually represent slight declines over the 2002 figures of 42.8 % reporting knowledge of Irish for the country as a whole, and 72.6 % for the Gaeltacht (i.e. Irish-speaking areas). More telling, however, is the 2002 census finding that nearly half a million of those who said they *could* speak Irish reported that they never *used* it. Another half-million said they used it less than weekly. The 2002 figures do not distinguish frequency of use inside and outside education and are therefore not directly comparable to the 2006 figures. A refinement of the question in 2006 reveals that only 3.2 % reported using Irish daily outside the education context; in the Gaeltacht, the figure is 27.5 % of the 64,265 speakers. Comparison of more recent figures from 2002 and 2006 reveals a continuing fundamental weakness in intergenerational transmission, within and without the Gaeltacht. Irish declines over the life cycle, beginning with 15–19 year olds and continuing through the childbearing years of adulthood. Moreover, the loss of speakers is twice as great in the Gaeltacht (1.8 %) as in the country as a whole (0.9), indicating that the gains made in terms of number of people reporting themselves as Irish-speakers in the census are largely the result of school-based reproduction. The proportion of pre-school age children who are returned as Irish-speakers has hardly budged since state intervention in 1921.

These statistics reveal a complex picture of competing forces leading to continuing loss of the remaining heartland and to a degree of renewal through school (Romaine 2008a). The maintenance of more or less stable rates of bilingualism over recent decades in Ireland has been due more to the capacity of the schools to produce competent bilinguals than to the capacity of the bilingual community to reproduce itself (Ó Riagáin 2001, p. 204). Meanwhile, in the Gaeltacht the historical process of language shift has progressed to the point where Irish is ceasing to be a community language and is becoming instead the language of particular social networks (Ó Riagáin 1997, p. 107).

A study done by Fernando et al (2010) relied on a social evolution model based on differential equations and was also designed to be useful to language planners by showing the effects of public and private intervention strategies on endangered languages. The model took into account three types of speakers (i.e. monolingual speakers of a high-status language, monolingual speakers of a low-status language, and bilinguals) as well as intergenerational transmission (i.e. parents teaching their language(s) to their children), and horizontal transmission (i.e. language acquired outside the home and learned formally in school). The model tested the efficacy of three strategies for intervention, namely, promoting learning of the minority language by raising its perceived status, increasing the amount of the language in society, and formal teaching to children. Either increasing the amount of the minority language in the community or providing formal language instruction proved effective, but interventions directed at increasing the amount of the minority language in the community were particularly effective, about twice as effective as teaching. Without intervention the minority language becomes extinct over a relatively short period of time. Moreover, stable co-existence obtained only with continuous intervention. When support measures are abandoned, the low status language dies out, unless its number of speakers is dramatically increased so that they dominate the higher status language. The lower status language can also persist if it has a high initial proportion of speakers, as in Quebec, for example, where 80 % of the population already spoke French when concerted policy efforts were initiated in support of its use.

The need for continuous and early intervention is more in line with linguists' recommendations than Minett and Wang's (2008, p. 42) finding which suggested that greater success was obtained by implementing intervention as late rather than as soon as possible. Indeed, French in Quebec has emerged as a symbolic case "for language minorities of the world, for it shows that sustained language planning can reverse language shift even relative to the most powerful language of this millennium:English" (Bourhis 2001, p. 101). Yet most small language communities can only wish for such an enviable position of strength from which to launch policies supportive of maintenance. Moreover, (Bourhis 2001, pp. 105, 111) remarks that language planners in Quebec were well placed to intervene in the 1970s in favor of French with strong intergenerational transmission on their side, even though a sociolinguistic analysis would have led to the conclusion that such planning was unnecessary. Fernando et al (2010) also showed that continuous intervention by increasing the use in public domains is approximately twice as effective as interventions using formal teaching. Only rarely will formal teaching lead to intergenerational-transmission, a prediction supported in theory by Fishman (1991) views on reversing language shift as well as in practice by the Irish case and many others.

Despite the increasing number of mathematical models being applied to language competition and death, our knowledge of the underlying dynamics of language extinction, i.e. why language death is occurring, and more importantly, how the dynamics could be changed, is still limited. The predictive power is constrained by the lack of accurate statistical data needed for validating computational models. In addition, the precise quantitative relationships among the various maintenance mechanisms and the parameters of models are still only poorly understood. These deficiencies leave a substantial gap between the sociolinguistic processes being modeled and the theoretical mechanisms underpinning them.

## 10.7 Conclusion

This chapter has examined the complexity of human-environment relationships on earth, by outlining a complex web of relationships and fundamental links between human languages and cultures, nonhuman species, and the earth's ecosystems. The ecological approach to understanding the dynamics of the evolution of linguistic diversity regards languages as sets of communicative practices embedded in networks of relationships linked into larger networks. This chapter has also shown that languages, like human cultures and species, adapt to specific ecological niches, and as cultural artifacts they have historicity. (Haugen 1972b), for instance, referred to the Scandinavian languages as 'cultural artifacts'. The problem of language endangerment raises critical issues about the survival of knowledge that may be of strategic use in the conservation of the world's ecosystems. Through being used for particular purposes by a particular group of people living in the same place over generations, each of the world's languages has come to express the history, values and beliefs important to that culture in a distinctive and efficient way. Elaborate culturally-specific terminologies and complex grammatical distinctions are often among the first linguistic casualties when languages become endangered. The disappearance of hundreds of species along with their names and related knowledge of their habitat and behavior represents a huge loss to science at precisely the time when we need most urgently to manage local ecosystems more effectively.

Many people now recognize that resources must be managed if we are to survive, but few people think of languages in the same way as they do of other natural resources such as air, water, oil, etc., which need careful planning. Much of the professional linguistic literature on language maintenance and preservation has been concerned with preserving the structures of individual languages in grammars and dictionaries, or has directed its attention to education programs in the threatened language. There is a distinction to be made between documenting language data and sustaining a language. What is being saved or preserved? Nora Marks Dauenhauer and Richard Dauenhauer put it well when they write that "Preservation [...] is what we do to berries in jam jars and salmon in cans [...]. Books and recordings can preserve languages, but only people and communities can keep them alive" (Lord 1996, p. 68). While salvage operations aimed at documenting endangered languages for preservation are worthwhile endeavors, and may be all that can be accomplished for some severely eroded languages, they do not address the root causes of language decline and without further action they do not contribute substantially to language maintenance efforts in the long term (Romaine 2008b). Grammars and dictionaries are artificial environments for languages. They reflect only a fraction of the diversity of a language in its everyday use and cannot capture the ever changing nature of language. Documentation does not directly ensure survival of a language, just as building a museum where we can display stuffed owls does nothing to preserve the bird in its natural habitat or guarantee that it can reproduce itself. Likewise, there is an important distinction to be made between learning a language in the artificial environment of the classroom and transmitting it in the natural environment of homes and communities.

Extinctions in general, whether of languages or species, are part of a pattern of human activities contributing to radical alterations in our ecosystem. If there is any ‘good’ news here, then it is that we are dealing with one large inter-connected problem rather than several independent extinction crises. Given the need to conserve the fundamental balance enabling the maintenance of ecosystems and all their components, the following questions are critical in addressing language vitality. What environmental conditions do linguistic varieties require to achieve their continuity? What is the minimal ecological niche that a particular language needs in order to ensure its permanence and reproduction? Because regions containing high biological diversity also represent areas of considerable importance to maintaining Earth’s linguistic and cultural diversity, conservation strategies that promote the persistence of economic and cultural systems likely will help sustain biological diversity in many settings as well. In view of the strong relationship between areas of great risk for species and for languages and their associated cultures, integrated strategies need to be developed to ensure the survival of both human diversity and biological diversity on our rapidly globalizing planet. Because the historical causes of the threats facing the earth’s languages, cultures, and biodiversity are the same, the solutions are also likely to come from the same place: empowering local people and communities. The maintenance of the world’s many languages has a vital role to play in preserving biodiversity because the measures most likely to preserve small languages are the very ones that will help increase their speakers’ standard of living in a longterm, sustainable way.

## References

Abrams, D.M., Strogatz, S.H.: Modelling the dynamics of language death. Nature **424**, 900 (2003)

Bourdieu, P.: The economics of linguistic exchanges. Soc. Sci. Inf. **16**(6), 645–668 (1977)

Bourdieu, P.: Language and Symbolic Power. Trans. Raymond, G., Adamson, M. Cambridge, Harvard University Press (1991)

Bourhis, R.Y.: Reversing language shift in Quebec. In: Fishman, J. (ed.) Can Threatened Languages Be Saved? Reversing Language Shift, Revisted: A 21st Century Perspective, pp. 101–141. Multilingual Matters, Clevedon (2001)

Bradley, D.: The disappearance of the Ugong in Thailand. In: Dorian, N.C. (ed.) Investigating Obsolescence: Studies in Language Contraction and Death, pp. 33–40. Cambridge University Press, Cambridge (1989)

Capra, F.: The Web of Life. Flamingo, Hammersmith (1997)

Dorian, N.C.: Introduction. In: Dorian, N.C. (ed.) Investigating Obsolescence: Studies in Language Contraction and Death, pp. 1–10. Cambridge University Press, Cambridge (1989)

Fernando, C., Valijärvi, R.-L. and Goldstein, R.: A model of the mechanisms of language extinction and revitalization strategies to save endangered languages. Human Biology 82(1):47–75 (2010).

Fishman, J.A.: Reversing Language Shift. Theoretical and Empirical Foundations of Assistance to Threatened Languages. Multilingual Matters, Clevedon (1991)

Friedman, T.: The world is Flat: A Brief History of the Twenty-First Century. Farrar, Straus and Giroux, New York (2005)

Gordon, R.G. (ed.): Languages of the World. 15th edn. SIL International, Dallas (2005)

Grewal, D.S.: Network Power. The Social Dynamics of Globalization. Yale University Press, New Haven (2008)

Haugen, E.: The ecology of language. In: Dil, A.S. (ed.) The Ecology of Language. Essays by Einar Haugen, pp. 340–344. Stanford University Press, Stanford (1972a.).

Haugen, E.: The Scandinavian languages as cultural artifacts. In: Dil, A. S. (ed.) The Ecology of Language. Essays by Einar Haugen, pp. 265–286. Stanford University Press, Stanford (1972b)

Haugen, E.:. The curse of babel. In: Haugen, E., Bloomfield, M. (eds.) Language as a human problem, pp. 33–45. W.W. Norton & Co., Inc., New York (1973)

Haugen, E.: Language fragmentation in Scandinavia: revolt of the minorities. In: Haugen, E., McClure, J.D., Thomson, D. (eds.) Minority Languages Today, pp. 100–119. Edinburgh University Press, Edinburgh (1981)

Hindley, R.: The Death of the Irish Language: A Qualified Obituary. Routledge, London (1990)

Kandler, A., Steele, J.: Ecological models of language competition. Bio. Theor. Integr. Dev. Evol. Cogn. **3**(2), 164–173 (2008)

Loh, J., Harmon, D.: A global index of biocultural diversity. Ecol. Indic. **5**, 231–241 (2005)

Lord, Nancy: Native tongues. Sierra Mag. **81**(6), 46–69 (1996)

Maffi, L. (ed.): On Biocultural Diversity. Linking Language, Knowledge, and The Environment. Smithsonian Institution Press, Washington, DC (2001)

Maffi, L.: Linguistic, cultural and biological diversity. Annu. Rev. Anthropol. **29**, 599–617 (2005)

Mertz, E.: Sociolinguistic creativity: Cape Breton Gaelic's 'linguistic tip'.In: Dorian, N.C. (ed.) Investigating Obsolescence: Studies in Language Contraction and Death, pp. 103–116. Cambridge University Press, Cambridge (1989)

Millennium Ecosystem Assessment.: Ecosystems and Human Well-Being. Island Press, Washington, DC (2005)

Milroy, L.: Social networks. In: Chambers, J.K., Trudgill, P., Schilling-Estes, N. (eds.) The Handbook of Language Variation and Change. Blackwell, Oxford (2002)

Minett, J.W., Wang, W.S.-Y.: Modelling endangered languages: the effects of bilingualism and social structure. Lingua **118**, 19–45 (2008)

Mittermeier, R.A., Gil, P.R., Hoffmann, M., Pilgrim, J., Brooks, T., Mittermeier, C.G., Lamoreux, J., de Fonseca, G.A.B.: Hotspots Revisited: Earth's Biologically Richest and Most Threatened Terrestrial Ecoregions. CEMEX, Mexico City (2004)

Mufwene, S.S.: The Ecology of Language Evolution. Cambridge University Press, Cambridge (2001)

Mühlhäusler, P.: Language of Environment: Environment of Language. A Course in Ecolinguistics. Battlebridge, London (2003)

Myers, N., Mittermeier, R.A., Mittermeier, C.G., de Fonseca, G.A.B., Kent, J.: Biodiversity hotspots for conservation priorities. Nature **403**, 853–858 (2000)

Nettle, D.: Linguistic Diversity. Oxford University Press, Oxford (1999a)

Nettle, D.: Using social impact theory to simulate language change. Lingua **108**, 95–117 (1999b)

Nettle, D., Romaine, S.: Vanishing Voices. The Extinction of the World's Languages. Oxford University Press, New York (2000)

Ó Riagáin, P.: Language Policy and Social Reproduction. Ireland 1893–1993. Oxford University Press, Oxford (1997)

ÓRiagáin, P.: Irish language production and reproduction 1981–1996. In: Fishman, J.A. (ed.) Can Threatened Languages be Saved? Reversing Language Shift, Revisited: A 21st Century Perspective, pp. 195–215. Multilingual Matters, Clevedon (2001)

Oviedo, G., Maffi, L., Larsen, P.B.: Indigenous and Traditional Peoples of the World and Ecoregion Conservation: An Integrated Approach to Conserving the World's Biological and Cultural Diversity. World Wildlife Foundation International and Terralingua, Gland (2000)

Pagel, M., Mace, R.: The cultural wealth of nations. Nature **428**, 275–278 (2004)

Patriarca, M., Leppänen, T.: Modelling the dynamics of language death. Phys. A Stat. Mech. Appl. **338**, 296–299 (2004)

Pinasco, J.P., Romanelli, L.: Coexistence of languages is possible. Phys. A Stat. Mech. Appl. **361**(1), 355–360 (2006)

Romaine, S.: Bilingualism. 2nd edn. Blackwell, Oxford (1995)

Romaine, S.: Planning for the survival of linguistic diversity. Lang. Policy **5**(2), 443–475 (2006)

Romaine, S.: The impact of language policy on endangered languages. In: Koenig, M., De Guchteneire, P. (eds.) Democracy and Human Rights in Multicultural Societies, Chapter 10, pp. 217–236. Ashgate/UNESCO, Aldershot (2007)

Romaine, S.: The Irish language in a global context. In: Nic Pháidín, C., Ó Cearnaigh, S. (eds.) A New View of the Irish Language, pp. 11–25. Cois Life, Dublin (2008a)

Romaine, S.: Linguistic diversity, sustainability, and the future of the past. In: King, K.,, Schilling-Estes, N., Fogle, L., Lou, J., and Soukup, B. (eds.) Sustaining Linguistic Diversity. Endangered and Minority Languages and Varieties, pp. 7–21. Georgetown University Press, Washington, DC (2008b)

Schulz, C., Stauffer, D., Wichmann, S.: Birth, death and survival of languages by Monte Carlo simulation. Commun. Comput. Phys. **3**(2), 271–294 (2008)

Sponsel, L.E. (ed.): Endangered Peoples of Southeast and East Asia: Struggles to Survive and Thrive. Greenwood Press, Westport (2000)

Sutherland, W.J.: Parallel extinction risk and global distribution of languages and species. Nature **423**, 276–279 (2003)

Voegelin, C.F., Voegelin, F.M.: Languages of the world: native America fascicle one. Anthropol. Linguist. **6**(6), 2–45 (1964)

Wickström, B.-A.: Can bilingualism be dynamically stable? A simple model of language choice. Rationality Soc. **17**(1), 81–115 (2005)

Wilson, E.O.: Consilience. The Unity of Knowledge. Abacus, London (1998)

Wyburn, John: Hayward, John: The future of bilingualism: An application of the Baggs and Freedman model. Journal of Mathematical Sociology **32**, 267–284 (2008)

Wyburn, John: Hayward, John: OR and language planning: modelling the interaction between unilingual and bilingual populations. Journal of the Operational Research Society **60**, 626–636 (2009)

# Chapter 11
# Conclusions

**P.-M. Binder and K. Smith**

If you made it to this point, this is what we have tried to offer you: a view of language like no other book has provided before. It is quite possible that this is one of very few books you have read on the subject; it could even be your first. Chances are that those you have read were written for a general audience. Some of them are very good, but they usually try to push a particular viewpoint of their author. Or you may have followed an introductory linguistics course at the university level. Such courses, and their textbooks, focus on technical aspects of the subject: the details of theories of phonology, syntax, semantics and so on. Knowledge of all of these areas is absolutely necessary for a deep understanding of the subject. But both single-author general books and technical textbooks have the same shortcoming: they do not give the reader a broad and complete account of how language works.

Our book is different in several respects. It was written by multiple authors, which allows experienced researchers from several areas of linguistics to not only present their knowledge to you with greater depth, but also to open a window into how they work and how they think. Secondly, this book does not have an agenda other than presenting an honest, holistic view of contemporary research on language. Most of the work presented is data-driven, and when not, the contributors make explicit their intellectual assumptions. Thirdly, this book is aimed at educated non-specialists. It is harder to read than an airport book, but at the same time far more rewarding. We have charged the authors with the task of writing in a way that allows for sophisticated arguments made in an accessible way and without being intimidating. We hope we

---

P.-M. Binder (✉)
Natural Sciences Division,University of Hawaii, 200 West Kawili Street,
Hilo, HI 96720-4091, USA
e-mail: pbinder@hawaii.edu

K. Smith
School of Philosophy, Psychology and Language Sciences, University of Edinburgh,
Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD, UK
e-mail: kenny@ling.ed.ac.uk

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2_11,

and our authors have succeeded. Finally, this book shows linguistics as a dynamical science, and the understanding of language as an evolving enterprise. This is a report of work in progress, work that is performed across many disciplines and with the use of numerous concepts and techniques.

In our description of language we have seen how it manifests itself at different scales. There is a delicate interplay between localization, distributed networks and plasticity at the brain level. The mechanisms involved in the production and understanding of language last only fractions of a second, while the rearrangements of neural connections following brain injury may take a significant fraction of a lifetime. Dialogue, which typically lasts minutes, is easy because speakers and listeners align, making language a highly cooperative activity in which the participants are constantly teaching, and learning from each other. Indeed, while most apparent during childhood, language learning is a lifelong activity. At the heart of it is the implicit identification of patterns that leads to the formation of grammatical generalizations. Although this could make the learner appear rather passive, evidence shows that the learner is in fact a very active participant, with the nature of the formal and semantic representations over which statistical analyses are made strongly influencing what is learnt.

Language changes beyond the lifetime of an individual involve evolutionary processes. Some of these are of a cultural nature. The propagation of language variants over a few generations (roughly one to several hundred years) is driven by social interaction, at the level of dialogue between individuals and through speech communities in which language change is initiated and adopted. Here we see an example of interaction across linguistic scales: cultural changes, taking place over hundreds or thousands of years, can be driven by fleeting conversations which help language variations to propagate beyond the lifetime of a single speaker.

Deeper changes happen at longer scales of the order of a thousand years. These changes are again driven by language transmission in populations. Such longer-term changes affect linguistic structure itself: the transition from single-meaning (holophrastic) utterances to those composed of combinations of several utterances (such as full sentences with subjects, verbs and objects), and the creation of a split between words that convey meaning (nouns, verbs) and those that have a purely functional role (words like “of” and “the”). These longer-term processes of change again show a wide separation between linguistic timescales.

The genes that speakers carry also affect language production and processing, through coding that affects the brain and our articulatory system, but also our abilities to learn and to interact with others. This already spans several timescales. Human populations differ in the frequency of the variants of certain genes; these discrepancies are driven in part by historical processes taking perhaps tens of thousands of years, such as mass migrations. Interestingly, there appears to be a correlation between allele frequency and at least one significant linguistic feature: tonality. Can genes influence language in a population? Are language and mind co-evolving?

This leads us to another timescale, that of biological evolution: several million years ago for the emergence of hominids, and a one or two hundred thousand for that of modern humans are the relevant time scales. The origin of language is one of the most fascinating mysteries in the field, one about which much has been written. In this book evidence from linguistics, biology, the fossil record, evolutionary theory and the study of other animals is used to address this puzzle. A plausible scenario is thus constructed for the evolution of language as both a biological and cultural phenomenon. Early hominids, perhaps already possessing a rudimentary protolanguage, moved to savannah environments which drove selection for cooperative group behavior. This may have resulted in an increase in the complexity of language for social reasons, and the development of learned vocalizations. This led in turn to an evolutionary advantage, as information could now be transmitted across generations: culture was born. Through this positive reinforcement loop between biological and cultural evolution, hierarchical compositionality, or the possibility of combining words with other combinations of words to express increasingly complex meanings, was acquired. The new combinatorial ability may have contributed to other spheres of human thought such as planning and social cognition, in short, to major features of our human nature.

The next-to-last chapter in the book is integrative: starting from the position that natural selection is not enough to explain the emergence of key features of language, self-organization is proposed as a complementary force. By means of a computer model that includes vocalization, the understanding of sounds, and neural representations for both, it shows how a robust system of distinct vowels shared by all members of a population can emerge—with specific vowel systems appearing with frequencies uncannily similar to those observed in human languages. A key feature of the computer model is the existence of separate timescales for dialogue and neural plasticity.

The final contributed chapter addresses future timescales. A pressing problem is the rapid disappearance of endangered languages, especially those spoken by small populations. Languages have been born and died before but at present time an unprecedented half of the world's 6–7000 languages are in some sort of danger in a timescale of a century, much shorter than a typical language's life. Languages exist in the context of both a natural and a social environment; the mechanisms behind current language shifts are reviewed, along with computer models addressing the same issues. This sobering chapter is mildly tempered by a discussion of recent successful examples of language revitalization efforts.

This is, broadly, the picture of language this book provides. Linguistic phenomena at many time scales have been described, as well as their interactions. To get to this point tools from neuroscience, genetics, evolutionary biology, psychology, statistics and archaeology have been used. The increasing role of computer analysis and simulations is obvious, although we have taken care to not over-emphasize it. We have posed three questions to our contributors about the future of the field. One asks them to anticipate a promising direction for a breakthrough in the understanding of language, another asks about future interdisciplinary work, and the last addresses the training of future linguists. Here is what our authors have to say.

***Focusing on your own specialist discipline, what new theory, idea or method do you think will (or could, or should) contribute most to our understanding of language in the next decade?***

*Müller*: Let's call it "heteronomy" of language, in contrast to Chomskian "autonomy" (*Editors*: *heteronomy is the idea that the brain may not have areas that exclusively do language processing, as opposed to having some kind of* "*language organ*".). Language emerges from earlier developing functional abilities in children. Neuroscience provides tools to elucidate this. For example, once we really understand what "Broca's area" does and why (and how it develops in children), we will understand much better what its role in language is. A few hints have become clear: It is a multifunctional area, due to its extensive connectivity patterns. The combination of input and output connections allows it to become pivotal in language acquisition, but also in many nonverbal domains (e.g. action understanding; see literature on mirror neuron system).

*Garrod and Pickering*: We think that there needs to be an explicit integration of theory and research on the relationships between comprehension and production processes for language. This will hopefully lead to a more direct study of the processes underlying successful dialogue.

*Wonnacott*: One of the most important recent developments for research in language acquisition has been the increasing contact with research in adult psycholinguistics. It is clear that there is a highly interdependent relationship between how language structures are initially learned, how they are processed and used in the mature system, and how learning continues throughout the lifespan. I think that elucidating these relationships will be of central importance in the study of language psychology. While there are many useful methodologies which will continue to be important, one approach which can directly explore relationships between learning and processing is to conduct language learning experiments with learners of different ages and which include the use various techniques developed by psycholinguistics (priming, eye-tracking etc.) to probe the outcome of learning.

*Kirby*: The most exciting development in recent years, and one that has still much to offer, is the expansion of laboratory techniques for studying the cultural evolution of language. I expect to see more papers relating individual human biases to emergent behaviors at the population level through the use of diffusion chains, micro-societies and artificial language learning in the psychology lab.

*Croft*: (i) The greater employment of quantitative methods and mathematical modeling. This is coming, but quite slowly. Most linguists do not have the relevant mathematical training, but collaboration with those who do can speed things along. (ii) More fine-grained analysis of meaning (especially lexical semantics) and of discourse function. Although a lot of research has been done in this area by typologists, semanticists, discourse analysts and computational/corpus linguists, a lot more remains to be done.

*Dediu*: This a difficult question for several reasons, one of the most important being that the pace of advancement in genetics and evolutionary biology (and allied disciplines) is not only maddeningly fast but keeps increasing almost every year. However, cheap and fast *whole genomes* and *large-scale studies* involving tens or hundreds of thousands of individuals from many different populations will surely foster advances that we can only dream of right now.

*Zuidema*: Dense corpora and test batteries. It's now technically possible to (automatically) record (and transcribe) a very substantial portion of the language input to a child, and the language produced by a child in the first years of life. Moreover, it is possible to track at an unprecedented scale cognitive development through game-like applications of cognitive test batteries on touchscreen devices that children just over 1 year old can already use. Together, these will provide a wealth of new data on the relation between language and other cognitive domains in development, to study within subject correlations and priming effects.

*Romaine*: Understanding the dynamics of language competition and evolution of linguistic diversity from the perspective of ecological niche theory and social networks. (*Editors*: *see* Gorenflo et al. (2012) *for a recent paper relating linguistics and ecology*).

*Oudeyer*: Language development shall be understood fundamentally through its complex interaction with sensorimotor/action development. My guess is that sensorimotor development is much more complex that what we understand today (for example in terms of compositional and context-dependent structures), and important parts of language may be special cases of general sensorimotor development.

*Editors*: One common theme brought up by three different researchers (not including three more who specifically included it in their chapters) is the role of quantitative and computational methods in the future progress of linguistics research, partly in conjunction with the generation and study of large linguistic data sets. This theme is echoed by several responses to the next question, as well. One further area of inquiry intrigues us: Müller's response appears to imply that language areas in the brain not only have other functions, but also that the latter are more fundamental than language functions. In apparent contradiction, Zuidema in his chapter made the suggestion that linguistic ability may have preceded the development of other high-level cognitive functions. Certainly, the full resolution of this issue may have been lost in the fog of evolutionary timescales. Regardless of the answer, the co-evolution of language and other forms of reasoning exclusive to humans appears to be a fascinating subject of inquiry.

***Thinking beyond your own discipline, what potential interactions or synergies with research in other disciplines do you think will (or could, or should) contribute most to our understanding of language in the next decade?***

*Müller*: Neuroscience. Genetics (and transcriptomics), but with great caution and appreciation for the complexities linking genotype and phenotype.

*Garrod and Pickering*: We would like to see a closer collaboration between researchers in language evolution (especially cultural evolution) and those in language processing and language acquisition.

*Wonnacott*: In general, I think it's great when language researchers from a range of backgrounds (developmental, psycholinguistic, theoretical, evolutionary) using a range of methods (e.g. experimental, observational, computational) begin to recognize that they are tackling the same problem from different perspectives. Bringing these lines of research together is what will lead us to a coherent picture of human language.

*Kirby*: I'd like to see a stronger bridge built between cognitive science and the field of cultural evolution. Cultural evolutionists have much to gain from including more sophisticated models of individuals in their models, and cognitive scientists have much to gain from taking cultural evolution seriously. Language is the perfect domain for applying this approach.

*Croft*: (i) Working with psychologists such as Herb Clark, Mike Tomasello and others on social cognition and interaction. (ii) Working with mathematical modelers (from mathematics, physics, evolutionary biology, computer science, etc.) to model linguistic behavior and linguistic patterns.

*Dediu*: The two directions that I think are extremely important are *development* (language acquisition) and *diversity*. We need to understand the incredibly complex interplay between genes and environment during the first, while the latter will allow a better appreciation of how our evolved capacity for language works.

*Zuidema*: Collaborations with cognitive psychologists and game developers.

*Romaine*: Related to my answer to the previous question (*understanding the dynamics of language competition and evolution of linguistic diversity from the perspective of ecological niche theory and social networks*)—this would require collaboration between linguists, geographers, ecologists and those in other disciplines like physics to build more sophisticated computational models to investigate language competition and evolution of linguistic diversity and make these models useful and comprehensible to social planners and language planners.

*Oudeyer*: Language, especially through interaction with embodied action, is inherently a complex dynamical system, and this requires strong knowledge in the mathematics/physics of dynamical systems, as well as strong experimental skills to explore these dynamics with computers, robots, but also humans (such as experimental semiotics).

*Editors*: The replies to this question point more at integrative, trans-timescale research, linking various individual-level and population-level processes. The connection between all biological ingredients of language–neuroscience, genetics and language production (as brought up by Oudeyer in the previous question) is a good example. In our view an additional element that may help elucidate this connection will be biochemistry, as gene expression, brain, ear and vocal cord development and brain processes can be ultimately seen as biochemical phenomena.

***Clearly, the study of language is an increasingly interdisciplinary enterprise. What would your advice be for a scientist starting out in this field? In particular, what blend of skills and knowledge would you suggest them to seek out?***

*Müller*: There are a lot of linguists who tinker with neuroscientific half-knowledge. A generation of scientists having solid knowledge in linguistics and neurobiology (ideally including genetics) would promise true advances.

*Garrod and Pickering*: Understanding cognition, communication, neuroscience, and evolution.

*Wonnacott*: There is so much and we can't all be expert in all of it. However, I do notice that researchers working in language psychology who come from backgrounds in Psychology versus Linguistics continue to have quite distinct knowledge bases, and I think that there may be gaps in both. A general knowledge of cognitive psychology and neuroscience provides vital insights for understanding problems in the psychology of language (for example there are many problems in vision which are directly analogous to those in language). On the other hand researchers who lack any basic training in linguistics may underestimate the complexity of the system they are investigating. In terms of skills, I constantly wish I had a better grasp of statistics. I also don't know how anyone manages to do this job without being able to program.

*Kirby*: I think everyone should try and learn how to build simple computational models. It is a lot less challenging than it sounds, and has the potential to change the kinds of idealizations we might naturally make when approaching the problem of understanding language. Beyond that, I personally wish I knew more math. Barely a day goes by when I'm trying to understand language evolution that I don't run up against the limits of my mathematical expertise.

*Croft*: Although I believe very strongly in the need for mathematical and quantitative skills to be acquired by linguists, I also believe that "traditional" (humanities-style) qualitative scholarship is necessary as well. I think that apart from a few rare exceptions, mastering both quantitative and qualitative linguistics skills in depth is probably beyond most individuals. A linguist needs to figure out the balance—as long as some basic skills in both are acquired.

Another important category of knowledge is typological knowledge, that is, knowledge of the details of linguistic diversity, so far as it has been documented by typologists. Even those who specialize in just English or their native language or their field word language needs to have a grasp of linguistic diversity and where "their" language fits into it. Also connected to this is knowledge of function as well as structure of language in some detail, as well as the dynamic aspects of language (acquisition, social variation, change). Many programs of study do not (yet) give this sort of breadth.

*Dediu*: Sophisticated quantitative and computational skills will be a must, and I would enumerate here advanced statistics, phylogenetics and computer programming, with an accent on analyzing large and very large complex datasets. Of paramount importance will be a proper understanding of the foundations of (post-genomic) evolutionary biology and neurobiology, with an ability to transfer between domains and find informative parallels and metaphors.

*Zuidema*: Every linguist should know some basics about experimental psychology, statistics, programming and philosophy of science, in addition to the standard linguistic curriculum.

*Romaine*: A combination of descriptive and theoretical approaches to language, knowledge of statistics and GIS (global information systems) skills.

*Editors*: One of our chapter authors is right in pointing out that we cannot be all experts on everything. In fact, when it comes to language, it is more accurate to say that nobody can be an expert on everything. There are just too many (potentially important) details to the language phenomenon.

According to our contributors, the most desirable individual areas of knowledge, other than the implicit standard training in linguistics are neuroscience and evolutionary theory. For several of our writers knowledge of brain processes at some level is central to the understanding of language as a whole. When it comes to skills, those related to mathematics and computation are most often invoked by our contributors, with a strong showing for statistics. But interestingly, “softer” knowledge of a humanistic or philosophical nature is recommended several times.

It may be that there’s no one best blend of skills: if we can’t all master everything, some diversification is inevitable and indeed desirable. But we also believe that thinking about language in the broader context, as provided by the book you are holding, is of considerable importance to present and future linguists: the best explanation for one particular linguistic phenomenon (e.g. a structural feature of the language of a certain individual) might be in terms of processes acting at another level of explanation (e.g. in terms of the organization of language in the brain, or the way in which languages are transmitted and used), and we should always be aware of this in building our theories of how language works.

### ***Final statement***

Language is a fundamental facet of being human. It originated and evolves through a constant interaction of biological and cultural processes. The feedback loop between language and culture is well established. Language is a part of culture: it is acquired through social learning along with a suite of other skills, technologies and beliefs, and it changes as a result of its transmission and use, just like other aspects of culture. Our ability to communicate also shapes and possibly even enables the development of a rich non-linguistic culture, facilitating collaboration and diffusion of ideas within human populations and setting the scene for the development of complex science, complex technology, and complex societies. These material cultures in turn feed back in to language: the development of alphabets and printing presses would not have been possible without language, and these technologies in turn provide a mechanism which allows language to persist at timescales that exceed human lives and even cultures; at the same time, texting, instant messaging and the internet facilitate rapid communication between previously disparate groups.

The loop between culture in general and biology is also well-established: our culturally-acquired and culturally-evolved technologies and lifestyles change the environment we alter and therefore directly modify the selective pressures acting on our genes: see Laland et al. (2000) for review. The case for a direct impact of language on human biology is less clear-cut, beyond the obvious case of adaptation of the human apparatus for speech. It may be that our minds have also evolved in response to pressures from language, for instance evolving to build equipment for learning and using language. Closing the loop between biology and culture, and explaining what (if any) impact possession of language has had on our *capacity* for language, is a major challenge (see also Fig. 9.3 in this volume and its associated discussion).

To sum up: in this book our contributors have examined how language manifests itself at different scales, and how it is shaped by interactions between processes acting at those different scales. Our hope is that an understanding of such interactions can form a general guiding principle in our search for a fuller understanding of the Language Phenomenon.

## References

Gorenflo, L.J., Romaine, S., Mittermeier, R.A., Walker-Painemilla, K.: Co-existence of linguistic and biological diversity in biodiversity hotspots and high biodiversity wilderness areas. Proc. Nat. Acad. Sci. USA **109**(21), 8032–8037 (2012)

Laland, K.N., Odling-Smee, J., Feldman, M.W.: Niche construction, biological evolution, and cultural change. Behav. Brain Sci. **23**(1), 131–146 (2000)

# Index

P.-M. Binder and K. Smith (eds.), *The Language Phenomenon*,
The Frontiers Collection, DOI: 10.1007/978-3-642-36086-2,