Chapter 4
Methodology 2: Coding Bilingual Transcripts

Viniti Vaish

Abstract All coding procedures used on the approximately 23 h of video data collected for the two research studies are described and discussed in this chapter. The coding categories are amount of talk, Mean Length of Utterance (calculated for words and morphemes), contexts for translanguaging, questioning patterns, interactional patterns, lexical density, and episodic structure (story grammar). Finally, in cases where the coding category is controversial, e.g., Mean Length of Utterance, I share the challenges in using this variable and justify why this code was used on the data.

Introduction

The 19 h of video data from the baseline study and the 14 h of video data from the Proof of Concept were fully transcribed and translated. Thereafter the data bank of transcripts was coded by the team of full-time and part-time bilingual research assistants for quantity and quality of talk. The codes that measure quantity of talk were amount of talk and Mean Length of Utterance. In the latter category, i.e., quality of talk, we coded for number and types of questions in teacher and student talk and motivations and contexts for translanguaging and tried to find patterns in interaction. Finally the 22 oral retellings performed by students in the last week of the Proof of Concept classes were coded for lexical density and episodic structure, which was a more in-depth analysis of quality of talk. The entire coding took place over a span of 3 years and the phases were as follows:

In the first instance, bilingual research assistants coded the transcripts for **Quantity of Talk**:

- Amount of talk (number of English, Malay, and Chinese words)
- Mean Length of Utterance (MLU)

V. Vaish (✉)
National Institute of Education, Nanyang Technological University, Singapore, Singapore
e-mail: viniti.vaish@nie.edu.sg

V. Vaish, *Translanguaging in Multilingual English Classrooms*,
https://doi.org/10.1007/978-981-15-1088-5_4

Thereafter the transcripts were coded for **Quality of Talk**:

- Motivations and contexts for translanguaging
- Number and types of questions in student and teacher talk
- Interactional patterns

Finally we coded the oral retelling of stories by students in schools F and C. The stories, which were in English only, were coded for:

- Amount of talk (number of words)
- Lexical density (number of content words)
- Episodic structure (elements of story grammar)

Not all the 23 h of data were coded for every single code or variable. Specifically the baseline study was coded for amount of talk, MLU, questioning patterns, and interactional patterns. The Proof of Concept was coded for all these and motivations and contexts for translanguaging. The oral narratives produced by the children at the end of the Proof of Concept were coded for lexical density and episodic structure.

I will now discuss each of these coding categories in detail and conclude this section with a discussion of the challenges we faced regarding specific languages that we have coded in our dataset.

Amount of Talk

The rationale for coding amount of talk was based on research conducted by Hart and Risley (1995). In a longitudinal study of 42 families, Hart and Risley (1995) compared amount and richness of parent talk across three socioeconomic groups: professional, working-class, and welfare parents. They followed children in these 42 families for 2.5 years from the time they were 10 months of age to 36 months. Through intensive coding of parent-talk, they found that parents who were professionals uttered more words per hour on an average to their children, displayed greater lexical density (used more types of words), used more complex sentences (e.g., multiclause sentences), and displayed more affirmation (praise, encouragement, affective speech), as compared with working-class parents and parents on welfare. As a result children at the age of 3 from professional families uttered more words per hour on an average and displayed richer vocabulary and sentence structure compared to their peers from the other two groups. All these language skills, especially vocabulary, are measures of expressive language which are correlated with later reading achievement.

The Hart and Risley (1995) study has been critiqued for essentializing social class and glorifying deterministic language outcomes. From a theoretical and methodological point of view, the Hart and Risley (1995) study has been critiqued on the basis of sampling (the sample size for the low SES children was very small, and all of them except one were African American children living in Kansas City) and a pseudo-scientific method of counting words to indicate amount of language learned while neglecting more holistic aspects of language acquisition like narratives

(Michaels 2013; Dudley-Marling and Lucas 2009; Dyson 2015). Most importantly, the existing cultures of the low-income families were ignored by Hart and Risley (1995). My interest, however, is purely on adapting some of their coding methods to make sense of my data. There are no conclusions about social class on the children in this study as we do not have any data on their household income, housing type, parents' occupation, etc.

Like Hart and Risley (1995), I value an increase in amount of talk though in our case we are looking at the impact of dialogic reading and translanguaging on students. My assumption is that when amount of talk for emergent bilingual increases, they are better able to demonstrate that they have learned new vocabulary, syntactical structures, and discourse features. Increasing amount of talk is also related to the Asian context in which this book is situated. Singaporean children, and, indeed, children in many Asian countries, are known to learn without talking too much in class. As the ensuing chapters will demonstrate, the Mean Length of Utterance of the children in the LSP is lower than other children of their age.

Mean Length of Utterance

Mean Length of Utterance (MLU), defined as the number of words and morphemes in spontaneously occurring utterances, is a robust indicator of language acquisition in both typically developing children and those with speech and language impairment. It was popularized by Brown's (1973) study in which he calculated the MLU of monolingual English-speaking children. In a large-scale longitudinal study by Rice et al. (2010), a sample of 306 monolingual English-speaking children, which included both typically developing and children with speech and language impairment, was examined. The procedures Rice et al. used to collect speech samples from the subjects resembled laboratory settings where children were given toys, and they were supposed to talk about the toy to the examiner. The key finding of this study was that though there was growth in MLU for both the normal children and those with speech and language impairment, the gap between these two groups remained consistent. Also, children of better educated mothers did not show more growth in MLU than their peers.

For older children, e.g., children 7–8 years of age, MLU is not considered as appropriate a measure of language acquisition as Brown (1973) found that MLU is not valid after about 4 years of age. For instance, in the study by Jiménez et al. (2006), 16 primary school (7–8 year olds) Spanish-English bilingual students were observed during shared book interactions. All except two students came from Spanish dominant families. The two exceptions were from English dominant families. I was interested to note that all utterances with code-switching were discarded by Jiménez et al. (2006). Regarding MLU Jiménez et al. (2006). found that though before the intervention parents used few dialogic reading strategies, at the end of the intervention they were making connections between the text and personal experiences and asking higher-order questions. At post intervention children used a larger

number of word types and demonstrated longer utterances when measured by the number of words they produced per turn.

The rationale for Jiménez et al. (2006) calculating MLT (Mean Length of Turn) rather than MLU (Mean Length of Utterance) was the age of the children. Since their study was on older children who had gone past the age for morpheme acquisition, the authors thought that MLU was not an appropriate measure for their subjects.

Though students in our study are 6–7-year-olds, only a year younger than in Jiménez et al.'s (2006) study, we have chosen to calculate MLU with this cohort because they have still not acquired most morphemes in English like plural and past tense. In fact, targeting the plural and past tense morphemes in English was a focus of some classes where translanguaging was used, as I will discuss in the chapters on findings.

Numerous studies (Rice et al. 2010; Hickey 1991; Brown 1973) have found a strong correlation between age and development in MLU. The 6–7-year-old children in our study are similar in their development to much younger children who are in the process of acquiring morphemes in English. In fact the MLU outcomes for the children in our study are comparable to the Singaporean preschoolers studied by Eng (1994). Eng calculated MLU in English for 59 Singaporean preschoolers with a mean age of 59 months and found outcomes ranging from 3.76 to 3.83. These numbers are similar to our outcomes for older children in the LSP class. Thus the fact that our LSP students were weak readers and had low proficiency in English was one reason we decided to calculate MLU for them despite their age.

We calculated a child's Mean Length of Utterance (MLU) using both the transcripts and the corresponding videos. The reason for using the videos also was to make sure that fillers like "ah," "um," etc. could be checked for intonation. In case the filler had meaning, it was included in the total count of words; otherwise it was discarded. The total number of words and morphemes produced by each student was counted from a random selection of 50% of all the transcripts. This was then divided by the total number of utterances produced by each child. The utterances could be in only English, only Malay, only Chinese, or translanguaged.

Unlike Jiménez et al. (2006), we did not discard the translanguaged utterances; instead we calculated their MLU. In fact Yip and Matthews (2006) and Bedmore et al. (2010) also discarded code-switched utterances from their data set. Though all these studies were on MLU, the authors reported MLU in two languages separately. In our data analysis translanguaged utterances were treated exactly as if that utterance was in one language. One reason for this was that we had practically no utterances in only Malay or only Chinese. The ones we did have were extremely brief consisting of one or two words. Because the children in the LSP class produced very few utterances, were very similar in terms of their output, and each utterance was extremely brief, the MLU of all the children in one class was averaged to one number that represented the MLU of that class. Thus for schools F and C, we have 2 MLU values for each class: one for English and one for translanguaged utterances. We do not have an MLU value for only Malay or only Chinese.

Since our focus was on Malay-English and Chinese-English bilingual children, we tried to find all the papers on MLU in Malay and Chinese. Though Chinese was

fairly well represented in the literature on MLU, we found only one paper in which MLU was calculated for Malay. Razak et al. (2016) calculated the MLU and discourse complexity of 130 Malay children ranging in age from 1 to 7 in Malaysia. The subjects were bilingual and came from homes where Malay was used 80% of the time. After analyzing nearly 9585 Malay utterances from these 130 children, the authors created five stages of language development and described the profile of children in each of these stages. They also calculated the MLU of these subjects which ranged from 1.05 for the 1-year-olds till 2.62 for the 6–7-year-olds. Data were collected through free conversation, story retelling, and narrating personal experience. The results reported in Razak et al. (2016) are only for Malay as bilingualism was not a focus area for the authors.

Though calculating the number of words and morphemes in Malay did not pose a problem, doing the same for Chinese was a challenge. Yip and Matthews (2006), who calculated the MLU of five Cantonese-English bilingual children, in order to make decisions on their language dominance, discussed issues of comparability between Cantonese and English. They commented that agglutinating languages, like Turkish, which have numerous morphemes attached to a stem, will result in a higher MLU when compared with isolating languages like Cantonese. Though comparing MLU across languages is fraught with problems, Yip and Matthews (2006) contend that this measure can be used to compare the two languages within one bilingual.

Cheung (n.d.) in an unpublished paper from the National Taiwan University calculated the MLU of five Taiwanese children who were Chinese dominant on the basis of words and syllables. He found a high correlation between MLU counted as words and MLU counted as syllables with the age of children. While counting words, Cheung did come across a few bound morphemes, which he counted as separate words. Cheung counted repeated words, names of places and persons only once, which is a practice we have followed.

Hickey (1991) made a similar point in her study of young children learning Irish. She recommended: “It seems advisable to regard MLU as purely intralinguistic device, allowing comparisons of the same child’s language over time, and between children acquiring the same language” (pg. 569). In keeping with these recommendations, our study does not focus on a comparison between the MLU of the Chinese children vs the Malay children. Rather our focus is on what happens to the MLU of two groups of bilingual children (Malay-English and Chinese-English) when the pedagogy changes from monolingualism to translanguaging.

In our coding, the MLU, counted for both words and morphemes in English, Malay, and Chinese, is reported as an aggregate for the entire class of approximately ten students. Our goal in reporting MLU values is not to show language dominance, as Yip and Matthews (2006) have done for their five Cantonese-English bilingual subjects, but to investigate if the children in our study talk more when opportunities for translanguaging are made available to them.

In concluding this section on MLU, I want to emphasize that we were fully cognizant of the limitations of MLU as a variable to measure how bilingual a child is. The shortcoming regarding age has already been justified. The specific shortcoming

regarding pairs of languages with different attributes needs further clarification. I have already discussed Yip and Matthews' (2006) concern regarding calculating MLU for Cantonese and English. A similar concern was raised by Otwinowska et al. (2018) who calculated MLU for Polish-English bilingual preschoolers through oral narratives. Otwinowska et al. (2018) pointed out that Polish is a pro-drop language, and in grammatically correct utterances, the MLU for Polish-English bilingual children can be lower for Polish than for English. Thus morpho-syntatic differences between languages can inflate or deflate MLU results.

However, despite these shortcomings, both Yip and Matthews (2006) and Otwinowska et al. (2018) use MLU as one of the measures in their studies. Similarly MLU is one of the many measures used in this book to estimate whether children during translanguaging talk more than when they use English only.

In this book the MLU of monolingual vs bilingual classes have been compared using a within-schools methodology. LSP classes in the three schools where the Proof of Concept (W, F, and C) was conducted were observed during monolingual and translanguaging pedagogy. The MLU comparison is thus for the same students in each school during monolingual vs translanguaging pedagogy.

Motivations and Contexts for Translanguaging

In the process of ascribing motivations for translanguaging, we ran into a basic methodological problem: is it at all possible, in the utterances of a bilingual, to ascribe a specific motivation to every minute switch? Though this might be possible while analyzing teacher talk, because teachers are purposeful talkers, we could only speculate about the reasons for this in the utterances of emergent biliterates who are weak in English. The current movement in the literature on bilingualism from "code-switching" to more holistic concepts like translanguaging, "codemeshing" (Canagarajah 2011), and metrolingualism (Otsuji and Pennycook 2010) helped us solve this problem. In keeping with this current movement in the literature, we have analyzed "Exchanges" which refers to a section of continuous utterances in the transcript (a more detailed explanation of what is an Exchange is given under the heading "Interactional Patterns"). We have not analyzed utterances which have been taken out of context. Thus, methodologically, we moved from a code-switching approach, which typically analyzes data utterance by utterance, to an approach influenced by translanguaging, which analyzes data in a more holistic manner.

The challenges we faced during coding mirror the experience of Sayer (2013) who reported that he started data analysis, in an ethnographic study of a grade 2 class where the teacher and students translanguaged between Spanish and English, by initially trying to figure out which language was used by the interlocutors for which purpose. "This approach proved problematic because, with the exception of a few functions, most interactions were unconstrained in that participants could (for the most part) freely choose from across their linguistic repertoire" (Sayer 2013). Thus Sayer found it more constructive to use the concept of translanguaging because

this new concept did not restrict the researcher into the straitjacket of ascribing labels to utterances. Rather, translanguaging is conducive to a holistic analysis of bilingual transcripts where the researcher can analyze large chunks of discourse according to meaning and not divide it up on the basis of languages or separate utterances.

Questioning Patterns

Transcripts were coded using Myhill's (2006) typology for the form of a question: factual, speculative, procedural, or process. In choosing exactly what constitutes a question, Burns and Myhill's (2004) methodology was used in which an utterance which required a response was coded as a question. In other words, the difference between a question and a statement is that a question requires a response (whatever the grammatical form of that utterance might be) and a statement does not. Factual questions, also called closed questions, are those which require a predetermined answer and usually elicit recall of information already provided to students. Speculative questions, also called open-ended questions, elicit "opinions, hypotheses, imaginings, ideas" (Myhill 2006). Procedural questions relate to the management of a lesson, for instance, can you all see? Finally, process questions ask students to explain their thinking, for instance, how did you work that out? In summary, four main codes were used to categorize the data bank of questions asked by teachers and students: factual, procedural, speculative, and process.

In comparing questioning patterns between monolingual and bilingual classes, an across- schools method has been used in this book. The monolingual classes have been taken from the baseline study, and the bilingual classes are from the three schools where translanguaging was attempted. Though the schools, students, and teachers are different, all the classes are from the LSP program where students are streamed after a nationwide test. Thus all the students are similar in terms of their low proficiency in reading skills in English.

Interactional Patterns

The 19 h of video data from the baseline study and 14 h from the Proof of Concept resulted in a huge data bank of transcripts. While these transcripts were being coded for minute details, e.g., number of Malay words vs number of English words in teacher talk, we also coded holistically to see if there were Exchanges, or specific units of larger transcripts, which addressed our research questions. The resulting Exchanges are a measure of quality of discourse. In using the term "Exchange," I am following the work of Nystrand (1997) whose book analyzes an equally large data set of transcripts drawn from secondary school classrooms in the USA. Nystrand used the term "episode" where an episode is a section of a transcript with distinct

boundaries which can be categorized under a specific code. Though Nystrand's "episode" is useful, I prefer "Exchange" as this term is less literary than "episode." The main codes used to mine the data for this book and cull a set of Exchanges were:

- Translanguaging to teach vocabulary in English
- Translanguaging to teach comprehension in English
- Translanguaging to teach grammar in English
- Using higher order questions (speculative and process questions)
- Translanguaging in student talk that displays higher MLU/better comprehension

As these are high inference codes, we met regularly as a team to check for agreement on slotting exchanges into specific categories. Though we did not calculate an inter-rater reliability figure as this would not be meaningful for such high inference categories, we focused on discussing why a particular Exchange fit or did not fit a specific code. The easiest code was "translanguaging to teach vocabulary," and the most contested was "translanguaging in student talk that displays higher MLU/better comprehension." Despite the fact that the latter is a high inference code, this final category was the most important as it was a way of discussing the outcomes of translanguaging, as no standardized testing was conducted in these research projects. The only way the researchers could make some judgments about the outcomes of translanguaging pedagogy was through evidence displayed in student talk.

Our data set is unusual in that it contains transcripts in three languages: English, Malay, and Chinese. Unlike Sayer (2013) who counts TexMex as a separate language I have not listed Singlish as a language separate from English. The reasons for this are both theoretical and methodological. Theoretically Singlish and English are on a diglossic continuum with Singlish being the L variety of Standard English which is the H variety (Ferguson 1959). Thus I do not consider it to be a separate language just as the low varieties of Tamil and Arabic are not separate languages. Also, considering Singlish to be a separate language was not efficient methodologically as it was impossible to decide where, within an utterance, Standard English gives way to Singlish or vice versa. In other words the point at which languages are switched within an utterance or even within a word is easy to identify when the researcher is dealing with disparate languages. However, in the case of Singlish and English, the researcher would be forced to make extremely high inference judgments which are not only inadvisable but also not in keeping with the spirit of translanguaging which discourages the researcher from pointing to junctures of switches.

Translanguaging scholars make a similar point when they distinguish between monolingual speakers who speak two varieties of the same language, e.g., Singlish and English, and bilinguals. "Bilingual speakers use language differently from multidialectal, monolingual speakers. Although all speakers use language differently, bilinguals have more choices to make because their language repertoires include many more language features. Language features include, for example, phonemes (sounds), words, morphemes (word forms), nouns, verbs, adjectives, tense systems, pronoun systems, case distinctions, gender distinctions, syntactic rules, and

discourse markers (e.g. marking transitions, information structure)" (Garcia et al. 2017, pg. 18). Thus a variety like Singlish cannot carry the weight of a separate language and must be seen as such in the coding of bilingual transcripts. At the same time, the L variety of Standard English, as marker of identity and social class, and for all the symbolic and affective weight it carries for Singaporeans, is an integral part of the linguistic ecology of Singapore. For the reasons discussed above, and given the research focus of the present study, only English, Chinese, and Malay languages are coded in the transcripts.

In order for the entire team to understand all the transcripts, each audio recording had to be transcribed and translated. For the English-Malay transcripts, this was not a problem as Malay is written in the Roman script; however, for transcripts with Chinese, the Chinese words also had to be transliterated. We decided not to use Chinese characters because Chinese orthography was not a focus of either teaching or learning.

Lexical Density and Episodic Structure

A total of 22 oral retellings by eight students were coded in this section. In the first instance, the number of words uttered by each student in each performance was documented as a measure for "amount of talk." Thereafter we calculated lexical density on the basis of the number of verbs, adjectives, nouns, and adverbs, considered to be content words, uttered by each individual. The term "lexical density" is also referred to as "vocabulary diversity" by some scholars, e.g., Price et al. (2009). However both the terms refer to richness of vocabulary. In this phase of the coding, the researchers also kept in mind the uptake of new vocabulary items which were being targeted by the teacher for each of the eight books. Thus we also looked out for whether or not the children were using the new vocabulary they had recently. The total number of content words in each performance was taken to be a measure of lexical density.

Finally, the transcripts were coded for episodic structure, also called elements of story grammar in the literature. Though scholars like Shrubshall (1997), Gutierrez-Clellen (2002), Fiestas and Pena (2004), Pearman (2008), and Kim et al. (2011) use similar methods in terms of coding the broad concept of the elements of story grammar, they differ somewhat on exactly which elements they choose to code. Some researchers, like Schick (2015), do not mention exactly what they coded in elements of story grammar. Keeping in mind the categories used by scholars in the review of literature and the data collected in this project, we coded for the following categories of story grammar: setting, description of action, identification of problem, resolution, and, finally, the motivation and reaction of characters. An expository book, *A Butterfly is Born*, was also taught in school C. The oral retellings of this book were coded differently as an expository book does not have the same elements of story grammar that a story has. Oral retellings of the expository book were coded for amount of talk and lexical density just like the oral retellings of story books.

However, regarding episodic structure, we coded for only two elements: description of action and articulation of sequence of events. These categories of story grammar were both a result of grounded analysis and inspired by the review of literature.

Moss (1997), who also assessed oral retellings qualitatively, used a different method of scoring the discourse of first graders. Moss' focus was on exposition and not narrative story books. In a study of 20 first graders retelling an information book on the birth and development of kittens, Moss scored the retellings on the basis of 5 levels. Titled "richness of scale" the 5th, or most successful level, was one in which:

> Student includes all main ideas and supporting details; sequences properly; infers beyond the text; relates to own life, understands text organization; summarizes; gives opinion and justifies it; MAY ask additional questions; very cohesive and complete retelling.

The most unsuccessful retelling was one in which:

> Student gives details only; poor sequencing, irrelevant information, very incomplete retelling. (Moss 1997, pg. 4)

Due to the fact that all the students in the LSP were weak readers, I did not think it advisable to have a scoring system like the one provided by Moss (1997) which might place most of the students at the bottom of the scale. Thus comparing each student against his/her own previous performance was a more constructive way of approaching the coding.

Conclusion

Thus the transcripts have been mined for amount of talk, Mean Length of Utterance, motivations for translanguaging, questioning patterns, lexical density, and episodic structure. What do all these coding categories have to do with translanguaging or translanguaging pedagogy? All these categories or codes are indicators of enhanced language learning. For instance, a higher MLU means that the child produced more words and morphemes on an average per utterance. Similarly a higher lexical density means that the child had a richer vocabulary. One of the premises of the book is that translanguaging provides a better language learning environment, for the specific type of learner we encounter in the LSP, than a monolingual class. If translanguaging is indeed so beneficial, then there should be a significant increase in MLU, amount of talk, lexical density, and episodic structure for the children in the Proof of Concept classes. This book will demonstrate there were, indeed, improvements in some of these categories. However, there was no change in others, and in a few, the quantitative numbers actually show a decrease. These mixed results are discussed in the ensuing chapters with reflections on possible reasons for these outcomes.

References

Bedmore, L. M., Pena, E. D., Gillam, R. B., & Ho, T.-H. (2010). Language sample measures and language ability in Spanish-English bilingual kindergartners. *Journal of Communication Disorders, 43*, 498–510.

Brown, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.

Burns, C., & Myhill, D. (2004). Interactive or inactive? A consideration of the nature of interaction in whole class teaching. *Cambridge Journal of Education, 34*(1), 35–49.

Canagarajah, S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *The Modern Language Journal, 95*(3), 401–417.

Cheung, H. (n.d.) *Utterance length and the development of Mandarin Chinese*. http://ntur.lib.ntu.edu.tw/bitstream/246246/2006092712313094524 7/1/000000000009.pdf

Dudley-Marling, C., & Lucas, K. (2009). Pathologizing the language and culture of poor children. *Language Arts, 86*(5), 362–370.

Dyson, A. H. (2015). The search for inclusion: Deficit discourse and the erasure of childhood. *Language Arts, 92*(3), 199–207.

Eng, A. L. S. (1994). Language use among Singaporean preschool children. *Singapore Journal of Education, 14*(2), 541–565.

Ferguson, C. A. (1959). Diglossia. *WORD, 15*(2), 325–340. https://doi.org/10.1080/00437956.1959.11659702.

Fiestas, C. E., & Pena, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*(2), 155–168.

Garcia, O., Johnson, S. I., & Seltzer, K. (2017). *The Translanguaging classroom: Leveraging student bilingualism for learning*. Philadelphia: Calson.

Gutierrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13*(2), 175–197.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes Publishing.

Hickey, T. (1991). Mean Length of Utterance and the acquisition of Irish. *Journal of Child Language, 18*, 553–569.

Jiménez, T. C., Filippini, A. L., & Gerber, M. M. (2006). Shared reading with Latino families: An analysis of reading interactions and language use. *Bilingual Research Journal, 30*(2), 431–452. https://doi.org/10.1080/15235882.2006.10162884.

Kim, Y. S., Kang, J. Y., & Pan, B. A. (2011). The relationship between children's spontaneous utterances during joint bookreadings and their retellings. *Journal of Early Childhood Literacy, 11*(3), 402–422.

Michaels, S. (2013). Déjà Vu all over again: What's wrong with Hart & Risley and a "Linguistic Deficit" framework in early childhood education? *Learning Landscapes, 7*(1), 23–39.

Moss, B. (1997). A qualitative assessment of first graders' retelling of expository text. *Reading Research and Instruction, 37*(1), 1–13.

Myhill, D. (2006). Talk, talk, talk: Teaching and learning in whole class discourse. *Research Papers in Education, 21*(1), 19–41. https://doi.org/10.1080/02671520500445425.

Nystrand, M. (1997). *Opening dialogue: Understanding the dynamics of language and learning in the English classroom*. New York: Teachers College Press.

Otsuji, E., & Pennycook, A. (2010). Metrolingualism: Fixity, fluidity and language in flux. *International Journal of Multilingualism, 7*(3), 240–254. https://doi.org/10.1080/14790710903414331.

Otwinowska, A., Mieszkowska, K., Białecka-Pikul, M., Opacki, M., & Haman, E. (2018). Retelling a model story improves the narratives of Polish-English bilingual children. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2018.1434124.

Pearman, C. J. (2008). Independent Reading of CD-ROM storybooks: Measuring comprehension with Oral retellings. *The Reading Teacher, 61*(8), 594–602.

Price, L. H., Kleeck, A., & Huberty, C. J. (2009). Talk during book sharing between parents and preschool children: A comparison between storybook and expository book conditions. *Reading Research Quarterly, 44*(2), 171–194.

Razak, R. A., Jin, L., & Hui, L. (2016). Profiling Malay Children's syntactic development: A Malay-LARSP. In P. Fletcher, M. J. Ball, & D. Crystal (Eds.), *Profiling grammar: More languages of LARSP* (Vol. 15, pp. 135–171). Bristol: Multilingual Matters.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language and Hearing Research, 53*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183).

Sayer, P. (2013). Translanguaging, TexMex, and bilingual pedagogy: Emergent bilinguals learning through the vernacular. *TESOL Quarterly, 47*(1), 63–88.

Schick, A. (2015). Wordless book-sharing styles in bilingual preschool classrooms and Latino children's emergent literacy skills. *Journal of Early Childhood Literacy, 15*(3), 331–363.

Shrubshall, P. (1997). Narrative, argument and literacy: A comparative study of the narrative discourse development of monolingual and bilingual 5–10 year old learners. *Journal of Multilingual and Multicultural Development, 18*(5), 402–421.

Yip, V., & Matthews, S. (2006). Assessing language dominance in bilingual acquisition: A case for mean length of utterance differential. *Language Assessment Quarterly: An International Journal, 3*(2), 97–116. https://doi.org/10.1207/s15434311laq0302_2.