ARE BETTER COMMUNICATORS BETTER READERS?

AN EXPLORATION OF THE CONNECTIONS

BETWEEN NARRATIVE LANGUAGE AND READING COMPREHENSION

GABRIELA SILVA-MACEDA

A thesis submitted in partial fulfillment
of the requirements of the University of Wolverhampton
for the degree of Doctor or Philosophy

November 2013

This work or any part thereof has not previously been presented in any form to the University or to any other body whether for the purposes of assessment, publication or any other purpose (unless otherwise indicated). Except for any express acknowledgements and references cited in the work, I can confirm that the intellectual content of the work is the result of my own efforts and of no other person.

The right of Gabriela Silva-Maceda to be identified as the author of this work is asserted in the accordance with ss.77 and 78 of the Copyright, Design and Patents Act 1988. At this date copyright is owned by the author.

Signature………………………..

Date……………………………

Abstract

The association between receptive language skills and reading comprehension has been established in the research literature. Even when the importance of receptive skills for reading comprehension has been strongly supported, in practice lower levels of skills tend to go unnoticed in typically developing children. A potentially more visible modality of language, expressive skills using speech samples, has been rarely examined despite the longitudinal links between speech and later reading development, and the connections between language and reading impairments. Even fewer reading studies have examined expressive skills using a subgroup of speech samples – narrative samples – which are closer to the kind of language practitioners can observe in their classrooms, and are also a rich source of linguistic and discourse-level data in school-aged children.

This thesis presents a study examining the relationship between expressive language skills in narrative samples and reading comprehension after the first two years of formal reading instruction, with considerable attention given to methodological and developmental issues. In order to address the main methodological issues surrounding the identification of the optimal linguistic indices in terms of reliability and the existence of developmental patterns, two studies of language development in oral narratives were carried out. The first of the narrative language studies drew data from an existing corpus, while the other analysed primary data, collected specifically for this purpose.

Having identified the optimal narrative indices in two different samples, the main study examined the relationships between these expressive narrative measures along with receptive standardised measures, and reading comprehension in a monolingual sample of eighty 7- and 8-year-old children attending Year 3 in the UK. Both receptive and expressive oral language skills were assessed at three different levels: vocabulary, grammar and discourse. Regression analyses indicated that, when considering expressive narrative variables on their own, expressive grammar and vocabulary, in that order, contributed to explain over a fifth of reading comprehension variance in typically

developing children. When controlling for receptive language however, expressive skills were not able to account for significant unique variance in the outcome measure. Nonetheless, mediation analyses revealed that receptive vocabulary and grammar played a mediating role in the relationship between expressive skills from narratives and reading comprehension. Results and further research directions are discussed in the context of this study's methodological considerations.

Contents

List of Figures

List of Tables

Acknowledgements

There is a long list of people who have helped me along the way of my PhD, in ways big and small, directly and indirectly. My utmost appreciation goes to Dr Josephine Chen-Wilson for her guidance and understanding, not only in supervising my work, but in supporting and encouraging me from the beginning. Also, Prof Ken Manktelow kept me grounded and always contributed a unique and seasoned perspective to our discussions as my second supervisor.

I have the deepest gratitude for Prof Trevor Hocking, who accepted me into the program at the selection interview and supported me indirectly along the way. It was a powerful and beautiful lesson to learn that in the UK merits come before circumstances, and that will stay with me throughout my life. Within the Psychology Department, Dr Neil Morris made me think I could become a researcher, and Dr Chris Fullwood always included me whenever he has sought resources or solved administrative complications for his own postgraduate students. Their help and inspiration is completely valued.

I would also like to thank in particular the Psychology Department's technicians, who went beyond the call of duty to resolve our logistical and technical difficulties: Ian Jukes, Alex O'Rourke, Louise Davies and Phillip Oates… I appreciated every single instance of your help.

The Heads and Deputy Heads, teachers, parents and children of the schools who volunteered for participation in any of my studies deserve all the credit for making this research project possible. I was especially aware of how constrained their daily schedules are from my previous teaching experience, and they were very generous to allow me into their schools to collect my data.

At a personal level, my husband David has been the greatest single source of support during my research activities. He has always backed my work and without his help in coordinating the care of Pablo, our son, conducting my studies would not have been possible. In his own sweet way, Pablo has also unknowingly motivated me to do my best and his joy for learning is contagious. Also offering moral and, towards the end

Chapter 1

Introduction

The widespread interest in research in education seems well deserved. The multitude of benefits and other outcomes in adulthood of educational attainment are extensively recognised at the social and individual level. Among them, educational attainment has been linked with a premium in earnings (Perna, 2003), positive health status (Steenland, Henley, Calle, & Thun, 2003; Winkleby, Jatulis, Frank, & Fortmann, 1992), and life satisfaction (Meeks & Murrell, 2001).

Research has also supported the link between *early* academic success, when basic literacy and numeracy skills are learned, and *later* academic attainment: it has been found that poor academic achievement in first grade decreased the odds of graduating from high school in the USA (Ensminger & Slusarcick, 1992) and academic achievement by the end of first grade has been found to be significantly correlated with years of education at age 22 (Entwisle, Alexander, & Olson, 2005).

Literacy therefore has been paid particular attention as it is the foundation upon which many of the other skills are built. Learning to read is one of the most fundamental abilities, in its own right and more than ever as a medium to access information, particularly in a technology and knowledge driven society in which routine tasks are increasingly automatized (OECD & Statistics Canada, 2000).

However, even when having access to education, many children will still struggle to become competent readers, with prevalence rates for reading disorders ranging from 5.3% to 11.8% depending on differing cut-off criteria in children's performance on standardised assessments (Katusic, Colligan, Barbaresi, Schaid, & Jacobsen, 2001). Even typically developing children who will not meet the criteria for a reading disability diagnosis will have a variety of difficulties. Follow-up studies of children at risk of reading disorders (Snowling, Gallagher, & Frith, 2003) highlight the fact that the nature of these deficits is rather continuous.

Although the causal factors of reading acquisition are still being debated, there is some consensus that they result from the interaction between biology and the environment (Keenan, Betjemann, Wadsworth, DeFries, & Olson, 2006; Snow, Burns, & Griffin, 1998). That environmental factors play a role in the acquisition process is important, as research has shown that there is plenty which can be done to build the children's capabilities required for reading, particularly with enriched instruction in school settings (Vellutino et al., 1996). For these reasons, it is important to better understand how reading is acquired, to be able to derive useful implications for practice.

1.1 Contributors to reading acquisition

Reading is a complex cognitive task that is carried out to understand a text. It requires the ability to simultaneously coordinate visual, phonological, semantic and linguistic processes (Plaut, 2005). Although it relies heavily on language skills, other cognitive and psychological factors have been shown to be related to the task of reading, such as attention and working memory (Booth, Boyle, & Kelly, 2010), and motivation (Guthrie, Wigfield, Metsala, & Cox, 1999). To acquire their reading skill, children must learn to map sounds to letters, i.e. to link phonemes and graphemes, and retrieve word meanings. To understand words or texts, they must also connect new information to their background knowledge and draw inferences. It is not surprising then, that such great individual variability exists when children acquire reading skills.

Ample support has accrued for the notion that word reading skills contribute to reading (Gough, Hoover, & Peterson, 1996; Hoover & Gough, 1990). In turn, a vast amount of research has converged on finding phonological awareness – the ability to identify and manipulate sounds – as the single most important contributor to word reading in longitudinal and intervention studies (Adams, 1990; Bradley & Bryant, 1983; Ehri et al., 2001; Goswami & Bryant, 1990). In their watershed study, Bradley and Bryant (1983) were able to show that training children to categorise words based on their common initial, middle or ending sounds had a direct effect on the children's reading skills two years later. More studies have replicated these results: a meta-analysis of 96 comparison studies concluded that instruction designed to increase phonemic awareness – a type of phonological awareness referring to the ability to identify and manipulate *individual* sounds – had a statistically significant influence on reading ($d = 0.53$; Ehri et al., 2001).

So consistent has been this finding that it has usually been central to government recommendations for teaching reading in the USA and the UK (National Reading Panel, 2000; Rose, 2006). Specifically, the significance of the intervention methodology for arguing a causal role has meant that other skills for which there is also abundant longitudinal evidence were, for some time, comparatively neglected. More researchers have already emphasised that although essential, phonological awareness is not in itself sufficient, and have advocated for recognising the role of a more comprehensive set that includes broader language sub-skills when trying to predict reading development beyond the early primary years (Bishop, 1991; Nation & Snowling, 2004; Paris, 2005; Scarborough, 2005). Among some of these broader language skills, semantics, syntax and discourse comprehension (i.e. language comprehension) have been found to play a role in reading development (Snowling & Hulme, 2006).

Research studies that have expanded the window of measurement to include very early pre-reading development have established that oral language skills as a comprehensive construct precede reading development (NICHD Early Child Research Network, 2005; Storch & Whitehurst, 2002). Moreover, in typically developing readers, oral language skills measured in kindergarten have been shown to have an influence on primary school differences in reading skills by different socioeconomic groups, with middle class children having higher scores over disadvantaged children (Durham, Farkas, Hammer, Tomblin, & Catts, 2007). Not only are oral language skills associated with reading in typically developing children, but evidence has accrued that many children with language impairments tend to have reading difficulties (Catts, Bridges, Little, & Tomblin, 2008; Snowling, Bishop, & Stothard, 2000). Indeed, so many language deficits have been observed in children with specific comprehension problems (Nation, Clarke, Marshall, & Durand, 2004) that there is a current debate about whether Specific Language Impairment and one type of reading disorder, dyslexia, are distinct disorders (Bishop & Snowling, 2004).

The importance of other language sub-skills does not negate the importance of phonological awareness. Nonetheless, as more studies have accumulated evidence on the role of other language aspects or sub-skills, it is becoming evident that there is a need to expand the picture of how and when all of these different language sub-skills influence reading.

So why has the focus on phonological awareness taken such a strong hold if so many factors influence reading? Part of the answer lies in which part of the developmental window researchers have focused on, and derived from that, on how reading has been conceptualised.

1.2 A fine look at reading: Distinguishing between word reading and reading comprehension

Based on the premise that the sooner the deficits can be identified, the sooner an intervention can be conducted to prevent these deficits from getting worse, it has therefore been natural that studies have focused on early reading skills. However, reading cannot be regarded as a simple, unitary skill, but rather as composed of two more complex skills: reading at the *word level*, which is usually referred to as either *decoding* or *word recognition*, and reading continuous text, which is known as *reading comprehension*. For clarity, decoding refers to the identification of phoneme-grapheme correspondences (i.e. mapping sounds to letters), while word recognition refers to the identification of words without such mappings. Though these terms are not necessarily interchangeable, they are in many instances used to refer to word reading. Typically developing children are able to master word reading by 9-11 years of age (Fuchs & Fuchs, 1999), so a vast amount of early reading studies had focused on reading at the word level.

However, a narrow focus on the earliest stages of reading as a global term has also produced a narrow focus on the factors contributing to reading as well (Paris, 2005), and here is where the distinction between the two skills, word reading and reading comprehension, becomes important. The notion that the factors which directly predict reading words can differ to some extent from the ones directly predicting reading comprehension has now received empirical support in numerous studies, which have used both word reading and reading comprehension as outcome measures. These studies have converged in reporting that while phonological awareness mainly influences word reading, it is broader oral language sub-skills that more directly influence reading comprehension (Catts, Fey, Zhang, & Tomblin, 1999; de Jong & van der Leij, 2002; Demont & Gombert, 1996; Muter, Hulme, Snowling, & Stevenson, 2004; Oakhill, Cain, & Bryant, 2003; Share & Leikin, 2004; Vellutino, Scanlon, Small, & Tanzman, 1991). Moreover, many children who are considered to be fluent readers but have specific

comprehension problems seem to have parallels with children diagnosed with Specific Language Impairment (Nation et al., 2004).

A much cited study by Hoover and Gough (1990) paved the way for this distinction. These researchers conceptualised a global concept of reading as the product of two skills, both necessary, but neither sufficient in itself: a decoding skill and a listening comprehension skill. Termed the Simple View of Reading, this model highlighted the changing developmental pattern in reading beyond the very early grades: in the first two US grades, decoding skill showed the highest correlations with reading ($r = .84$ and $.80$ in 1st and 2nd, respectively) while listening comprehension had lower correlations. The pattern was reversed in the following two grades, with listening comprehension having the strongest correlations with reading ($r = .80$ and $.87$ in 3rd and 4th, respectively) while correlations with decoding were decreasing. In other words, although the pattern of contribution changed with development, the ability to read was formed by the ability to read words in isolation with the ability to understand sentences and discourse.

More studies followed with converging findings. A meta-analysis of 17 such studies showed that the contribution of decoding skills is stronger at the beginning stages of learning to read, while listening comprehension exhibits a stronger association with reading comprehension at the later stages (Gough et al., 1996). Figure 1.1 illustrates the differing patterns found in this meta-analysis.
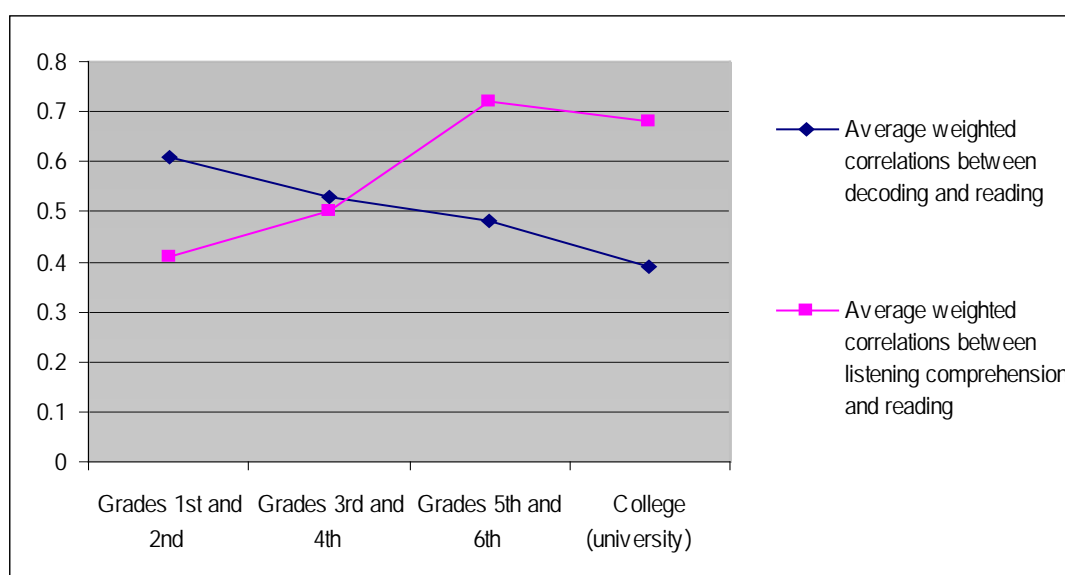
Figure 1.1 Correlations between decoding and reading, and between listening comprehension and reading from a meta-analysis of 17 studies (drawn from data in Gough et al., 1996).

Two interrelated issues were brought forward with the Simple View model: 1) changes in the pattern of development make clear the need to examine further beyond the early stages of word reading; and 2) listening comprehension (one of several broader language skills) is critical for reading comprehension and its contribution is particularly evident after decoding skills have been mastered.

In some instances this finding has been taken to mean that decoding abilities should be developed first, and broader language skills could be focused on later. Nevertheless, the research looking at pre-school language measures and later reading skills does not support this implication. Some studies have actually found support for early language also having a direct influence in decoding skills as well (NICHD Early Child Research Network, 2005; Storch & Whitehurst, 2002). A few studies, which will be described in more detail in Chapter 2, have even found links between linguistic measures at the very onset of receptive and expressive language development, and later reading skills (Scarborough, 1990b; Shapiro et al., 1990; Walker, Greenwood, Hart, & Carta, 1994).

In sum, evidence accumulated over the years portrays a more complex picture of the precursors of reading skills than previous research had established. Although undeniably phonological/phonemic awareness is still a major factor influencing decoding skills, which in turn enables reading through the later school years, it is now clear that many strands of language influence long-term reading development, i.e. reading comprehension skills. In the next section, a brief description of how these language factors are related to reading comprehension is reported.

1.3 A fine look at language: Distinguishing different language sub-skills and modalities associated with reading comprehension

Apart from phonology, other aspects of language, which are not only present in oral but also in written language, include semantics, syntax and pragmatics (Perfetti, 1985; Webster, 1994). Their relative contributions to reading, although documented, are not entirely consistent across different studies. What is consistent is the notion that all of them do play a role in word reading and reading comprehension development.

It has already been proposed that different sub-skills may play differential roles at different points of development (Scarborough, 2005; Vellutino, Scanlon, & Tanzman,

1994). In addition to their potential differences in contribution when studying different developmental windows, interpretation of studies linking reading and different oral language aspects complicates the picture with equivocal findings mainly because they employ different methodologies to measure oral language skills.

Adding one more layer of complexity, oral language skills can be measured as receptive (i.e. listening) skills, and expressive (i.e. speaking) skills. Most studies focus on the contribution of receptive skills while very few consider expressive ones. Expressive skills are important because some of the earliest precursors to reading have been very early speech measures (Bishop & Adams, 1990; Harlaar, Hayiou-Thomas, Dale, & Plomin, 2008; Scarborough, 1990b; Shapiro et al., 1990) and because of the great overlap between language and reading disorders (Bishop & Snowling, 2004; Catts, Adlof, Hogan, & Weismer, 2005). Although expressive skills can be measured using standardised tests and speech samples – and there is evidence linking both kinds of expressive measures to reading – the latter have attracted interest in terms of their potential for achieving greater ecological validity, as language samples do not penalise culture-specific forms of language (Hughes, McGillivray, & Schmidek, 1997), and for being closer to the kind of children's language that a lay person, parent or teacher, can perceive on their own.

Moreover, comprehensive studies that include measures of *all* language sub-skills in both modalities, can aid in identifying the relative importance of each one. Among the few studies which reported meta-analytical work with broader language sub-skills, was the one conducted by Scarborough (1998). Her meta-analysis of 61 research samples of kindergarten predictors of later reading included not only phonological awareness but also the wide range of linguistic, cognitive and even motor skill measures in longitudinal studies to predict reading abilities, in receptive and expressive modalities. Her findings list 16 measures, of which the first five in order of average correlations are: letter identification ($r = .52$), concepts of print ($r = .46$), phonological awareness ($r = .46$), expressive vocabulary ($r = .45$) and sentence/story recall ($r = .45$); the rest of the factors have lower but still significant average correlations. These first five skills can be distinguished into two categories: the first two are skills that are mastered within a very short period of time by typically developing children, while the last three skills are all language aspects or sub-skills.

Regarding the first two, letter identification and concepts of print, it has been argued that rather than being a genuine underlying individual difference, they reflect the moment in time where the child is learning letter names and sounds, and therefore, are not stable predictors of long-term reading (Paris, Carpenter, Paris, & Hamilton, 2005). The last three skills –phonological awareness, expressive vocabulary and sentence/story recall – are all different receptive and expressive language sub-skills, and these precede any reading skill (although there is evidence for reciprocal relationships once a child starts reading). Phonological awareness, as discussed above, has been extensively investigated and arguably its main contribution has been to word reading skills. In contrast, expressive vocabulary and sentence/story recall (which measures both language modalities) may continue to grow through a person's lifetime and could therefore be considered a more authentic source of individual differences.

Expressive vocabulary and sentence/story recall reflect higher order aspects of language: at the word level, vocabulary and semantics; at the sentence level, syntax; and finally at the discourse level, listening/language comprehension, which involves making inferences and connecting new information with background knowledge. As discussed before, their main direct contribution has been shown to be to reading comprehension.

Their contribution in relation to each other is still obscured by methodological and developmental issues. Next, some of the evidence relating receptive language and reading is introduced in Section 1.3.1. Then, in the following Section 1.3.2, the scarcer but relevant evidence for the role of expressive skills and the methodological considerations when examining this relationship will be mentioned. A more detailed literature review of the studies linking receptive and expressive measures to reading is deferred to Chapter 2.

1.3.1 Receptive language and reading

Oral language is composed of different aspects that together allow us to communicate with each other: phonology, semantics, syntax, and pragmatics. Beyond the contributions of phonology to reading which have already been established, the contributions of semantics and grammar are relatively less specified. ***Semantics*** is the study of word meanings, and is usually but not always measured with a vocabulary task. The rules for arranging the elements of language are studied by ***grammar***, which is

composed of ***morphology*** and ***syntax***: while morphology is focused on word structure, syntax is concerned with word order. Finally, although more a level than a facet of language, ***discourse-level comprehension***, also known as listening comprehension, is also considered.

Many of these language sub-skills are critical for reading comprehension. Indeed, 8-year-old children with specific reading comprehension impairments have been found to have semantic and morpho-syntactic deficits similar to those found in children with language impairment (Nation et al., 2004). Similarly, Catts and colleagues found that 13- and 14-year-old pupils with specific comprehension difficulties exhibited concurrent and retrospective deficits in vocabulary, grammar and discourse six years before (Catts, Adlof, & Weismer, 2006). Moreover, in these two studies, the pupils with specific comprehension impairments did not exhibit phonological deficits.

Studies linking a specific variable with reading are more abundant. Vocabulary in particular has received ample support showing different vocabulary measures having differing levels of association with reading (Ouellette, 2006). Furthermore, experimental comparisons between children with specific comprehension problems and controls have shown that these children with relatively intact decoding skills do have semantic processing weaknesses, when tested with tests such as synonym tasks and semantic fluency tasks (Nation & Snowling, 1998).

There is also a considerable amount of studies linking syntax and reading, by using a variety of tasks. Evidence for such associations has been found in longitudinal studies (Muter et al., 2004; Nation & Snowling, 2004) and studies comparing the syntactic abilities of 9-year-old children with specific reading comprehension deficits in younger skilled readers (Nation & Snowling, 2000; Tunmer, Nesdale, & Wright, 1987).

Finally the relationship between discourse-level comprehension, also known as listening comprehension, and reading is unquestionable, since many of the same higher-level processes required are the same, such as the ability to generate inferences and create mental models (Kintsch & Rawson, 2005). The importance of listening comprehension was already introduced in section 1.2 when the Simple View model was discussed (Gough et al., 1996). Although its contribution is more clearly seen after decoding skills have been mastered, listening comprehension skills are evidently developing in the early school years (Paris & Paris, 2003).

In sum, there is ample evidence of a strong association between semantics/vocabulary, syntax and discourse-level comprehension, and reading comprehension. Therefore, any study of comprehensive language skills should include all of these sub-skills to be able to establish their relative contribution to reading.

1.3.2 Expressive language and reading

When compared to research investigating receptive language skills, expressive skills had been until relatively recently largely overlooked, despite the connections found between early expressive language skills and reading (Harlaar et al., 2008; Walker et al., 1994), and the connections between language deficits or impairments – both receptive and expressive – and later reading difficulties (Bishop & Adams, 1990; Catts et al., 2008; DeThorne et al., 2006; Shapiro et al., 1990). Furthermore, there is some evidence that expressive deficits can have a differential impact on reading skills: it has been reported that children who have both receptive and expressive language deficits are the ones likely to exhibit the greatest reading deficits (Simkin & Conti-Ramsden, 2006).

Different studies have used different tools depending on their aims, but this variety has also made it more difficult to reach some uniformity in tasks, and consequently, comparisons and generalizations when interpreting different studies have become more cumbersome. Expressive skills can be operationalized in many different ways, but the main distinction is the one between standardised assessments and speech/ language samples. In the context of speech pathology, the analysis of language samples has consisted of the recording and transcription of a dialogue or discourse to identify productive language disorders (Leadholm & Miller, 1992), but they are now widely used for non-disordered populations as well (Hughes et al., 1997). There is evidence linking both kinds of expressive measures to reading.

Standardised assessments have by definition, good psychometric properties, while measurements from language samples tend to be considered weaker in such properties. However, even with the variety of methodological issues intrinsic to the use of language samples, there are four main theoretical and practical reasons to consider the use of measures from language samples as relevant in research of reading comprehension: 1) language samples offer the possibility of identifying more ecologically valid expressive

measures, since these measures do not penalise non-standard variations of language and they get away from the test-taking format, possibly requiring the least amount of meta-linguistic skill (Hughes et al., 1997); 2) the development of language stems from having opportunities for communicative experiences (Hoff, 2006) and providing children the space to express themselves creatively is in itself a communication opportunity; 3) language samples can be and have been used as an alternative form of language assessment that could be closer to what a lay person (parent or teacher) could perceive on their own, as evidenced by the fact that parent- perceived speech delays are the more common way that children come to receive specialist services (Bates, Dale, & Thal, 1995); 4) narratives, a specific type of language sample, lend themselves for interventions and training protocols (Cable, 2007); and 5) narratives have the potential to elicit decontextualized language (Ricard & Snow, 1990), which has been shown to be related to literacy.

Regarding the first issue of ecological validity, the use of standardised assessments to uncover language disorders has been criticised because these assessments are perceived as limited in their capacity to distinguish between disorder or simple language difference, particularly for children from non-mainstream cultures (Gillam, Pena, & Miller, 1999). Moreover, qualitative and quantitative evidence supports the notion that children from minority or socially disadvantaged communities have a literacy gap when compared with mainstream children (Durham et al., 2007; Heath, 1983).

Even with this potential, the methodological difficulties intrinsic to analysing language samples could mean that exploring its relationship to reading comprehension is not a viable option. However, there are some studies which have already found links between narrative language samples and reading comprehension in cross-sectional and longitudinal analyses (Cain, 2003; Cain & Oakhill, 1996; Griffin, Hemphill, Camp, & Wolf, 2004; Klecan-Aker & Caraway, 1997; Snow, Tabors, Nicholson, & Kurkland, 1995). The variability in study design, samples and methodologies makes it difficult to integrate these findings. Still, this evidence of a relationship with reading, coupled with the theoretical and practical reasons outlined above, make language samples, and more specifically narratives, promising candidates for investigating expressive language in relation to reading skills. Moreover, some of the methodological difficulties when using language samples can be to some extent minimised by the use of newer and more sophisticated versions of computerised language analysis software (Adolphs, 2006).

Many studies of expressive skills using narratives, tests or surveys, however, focus on early language development while neglecting language development in school-aged children. The implications from the vast majority of these early language studies and the findings that language measures are quite stable leave practitioners (teachers, reading specialists and language therapists) with the idea that little can be done after the first few years. In addition, although many studies have looked at the *receptive* language of school-aged children, few have looked at the *expressive* skills of these children, and how they relate to reading.

Given the findings that broader language sub-skills are more clearly related to reading comprehension than phonological skills, and that expressive skills are both relevant and likely to influence reading in a different way, and that narratives could offer a semi-structured language sample allowing for the production of decontextualized language, a gap in the literature was identified where the link between expressive language using narratives and reading comprehension could be examined at precisely the time in development when children start to face continuous texts. Moreover, it was important to include receptive skills in order to estimate whether they are related to reading comprehension in a differential manner and if so, to compare their relative and unique contributions.

1.4 Scope of this study

With the general idea of linking expressive skills and reading comprehension, the study was more narrowly defined to examine this association using narrative language at a particular window of development after the first couple of years of formal reading instruction have begun. In this section, the aim, limitations and assumptions of this study are described.

1.4.1 Aim and limitations

This study was designed to examine the concurrent associations between expressive language and reading comprehension that would take into account children's receptive language skills, by using narrative language samples. Narrative language samples were selected over standardised assessments given the evidence showing their potential to

elicit decontextualised language, their noted ecological validity, and the possibilities of being closer to language forms observable by practitioners, lending itself for training and opening opportunities for communicative experiences.

These relationships were evaluated at a very specific time when children start to face continuous texts. To date, considerable literature has examined the contributions of either receptive or expressive skills to reading word or reading comprehension, usually in younger children, and mostly using psychometric assessments without much consideration of performance in decontextualized or more natural language samples. In this study both modalities were brought together to evaluate whether more ecologically valid language samples and measures of expressive language could still explain some variability in reading comprehension, firstly on its own and then also after receptive skills were accounted for, in children after the first two years of formal reading instruction. As far as we can establish, this had not been attempted before in such a comprehensive way at this stage of development, using narrative language samples for the measurement of expressive language.

Given its cross-sectional design, it should be noted that the study was intended to be of an exploratory nature and not to describe precise causal relationships. In the same way, generalisations cannot be made about developmental trajectories in reading nor language.

In addition, this study does not intend to account for the many environmental variables that influence the process of learning to read, but is rather focused on the cognitive and language tasks that are more proximally related to observing reading skills at the behavioural level. Furthermore, a diverse socioeconomic sample was sought and a fairly brief socioeconomic analysis at the school-wide level was included, given the known associations between language and socioeconomic status (Farkas & Beron, 2004; Hoff, 2003), and reading and socioeconomic status (Duncan, Yeung, Brooks-Gunn, & Smith, 1998). The choice of narratives would be even more relevant for disadvantaged samples, given that such an ecologically valid measurement of expressive language does not penalise the use of non-standard variations of language.

Finally, this study does not follow a random sampling process for selecting children participants. Many schools were invited, but eventually only a small portion agreed to participate in the study. The schools and children who participated were volunteers who gave up some time to take part in this research project and, in accordance with ethical

guidelines, only children whose parents gave consent for participation were able to do so. In that sense, it is a convenience sample derived from a single region in the UK. It is argued however that these kinds of studies are necessary in order to complement the information obtained through large-scale random/stratified sample studies by presenting a more detailed and rich picture of individual differences when some of the environmental variables are somewhat uniform.

Nonetheless, within its scope, this study presents a wide-ranging set of oral language skills and their relationship with reading comprehension at the time when children are progressively facing more continuous texts, with a specific contribution in examining concurrent expressive skills using narrative language samples.

1.4.2 Assumptions

Investigating the relationships between oral language sub-skills and reading comprehension was conducted with the assumption that when measuring expressive language, a compromise could be reached between the artificiality and validity of standardised tests on the one hand, and the ecological validity but additional methodological issues of language samples on the other.

A second assumption was that expressive narrative language measures would reflect additional facets of language from the ones provided by the receptive measures in standardised tests. Dealing with expressive narrative language at an age range where there are few studies and where even fewer comparisons between semi-structured language samples and receptive measures exist, meant that only a theoretical reason was given for why narrative measures should bring additional information. To some extent this assumption was tested in Chapter 5 in the main reading study, but only for this age range.

1.5 Outline

This thesis is organised into 6 chapters. After this initial introduction, Chapter 2 summarises the literature directly linking receptive and expressive skills with reading comprehension, in order to present the context of research already carried out. Building

on that information, Chapter 3 describes the methodology devised to measure the different constructs in the three main areas: receptive language, expressive language and reading comprehension. In the same chapter, two studies are reported. First, having acknowledged the lack of standards when measuring school-age language, one study analysed the developmental variability and reliability of lexical and syntactic indices using an existing database of 60 picture-book narratives. Next, a pilot study is described that examined the feasibility of the proposed methodology at a small scale. In particular, the suitability of the expressive language measures was evaluated, and preliminary evidence for the hypothesised concurrent associations between expressive language and reading comprehension in this age group was obtained.

Although the pilot study results were more promising than expected, it was recognised that the language measures for expressive language needed to be further identified with primary data collected specifically for the purpose of eliciting the most decontextualized language possible.

Therefore the next study aimed to gather further evidence of the suitability of the indices chosen across a range of tasks, in a larger sample. This second study of narrative language, reported in Chapter 4, sought to confirm the developmental patterns using primary data, to evaluate their within-participant variation with different stimuli, with the addition of a discourse-level organisation measure.

Taking the results of both expressive language studies into account, the main research question of the relationship between both modalities of language and reading comprehension was then addressed with a larger and more socioeconomically diverse sample. Regression analyses were performed to examine the relative contributions of receptive and expressive skills. This main study is reported in Chapter 5.

Finally the major findings were interpreted in terms of existing reading research and of potential implications for practice, and this is presented in Chapter 6, along with some suggestions for future lines of work.

Chapter 2

Literature review:
Receptive/expressive language and reading comprehension

Even with so many advances in reading research in general and a growing consensus that reading comprehension in particular builds upon general language skills, the relations between language and reading are still being defined. Developmental issues and methodological difficulties for measuring both *reading* and *language skills* have resulted in a vast number of studies with many converging but also some diverging results.

One of the greatest methodological difficulties regarding reading measures was the consideration of reading as a unitary global skill. The use of global measures of reading was insufficient to account for the complexity of factors associated with it. Once reading was identified as having two main components, namely word reading and reading comprehension, evidence emerged that different factors contributed to these two components: while phonological awareness has its greatest influence over word reading, broad language skills seem to influence reading comprehension more directly (Oakhill et al., 2003).

Just as reading was divided into two components, there is now the need to further define specific language sub-skills. Evidently global language measures have been useful, particularly for very early language data. For example, it was with the use of composite language measures that early oral language has been shown to precede both word and reading comprehension skills (NICHD Early Child Research Network, 2005) and even to precede the development of phonological awareness (Cooper, Roth, & Speece, 2002). However, to better understand the contributions of language when children start to face continuous texts, and to derive implications for practice at this reading acquisition stage, it has become necessary to distinguish among specific language sub-skills, beyond phonological awareness. In this thesis, these non-phonological aspects of language will be referred to as either ***broad language skills*** or ***language sub-skills***, focusing on semantics, grammar and discourse.

Broad language skills have garnered greater attention given the evidence supporting their role in those phenomena that phonological awareness has been limited to explain: specific comprehension deficits and reading beyond the early years. Evidence confirms that broad language skills are clearly implicated in children with specific comprehension deficits (Nation et al., 2004), that is, children who despite being able to decode efficiently still have difficulties understanding what they just read (Oakhill, 1994). Moreover, enough evidence has been accumulated to support the notion that, after decoding skills have been mastered, broad language skills are the ones most related to long-term reading (Paris, 2005).

The particular contributions of these skills to reading comprehension are not yet well specified. Studies looking into language sub-skills contributions generally converge in finding semantics, syntax and discourse as important, but they disagree over their relative importance. One reason for this variation is that studies tend to employ different methodologies to measure similar language constructs which complicates their interpretation and makes generalisations more difficult. Furthermore, while a few studies have reviewed a comprehensive set of language measures, most studies of language and initial reading comprehension usually focus on just a few sub-skills.

Also, there are clear developmental issues that go beyond the methodological ones. The first developmental issue deals with the phenomenon that at the initial stages of reading, word reading (i.e. decoding) will be more strongly correlated with reading comprehension than language skills, as supported by the Simple View model (Gough et al., 1996). A second developmental issue more specifically related to language, comes from longitudinal studies which have taken language measurements from children at different times. It has been shown that there is indeed further specification for the contribution of language skills to reading *depending on the developmental window studied* (Scarborough, 1990b; Scarborough, Rescorla, Tager-Flusberg, Fowler, & Sudhalter, 1991). In this particular study, for example, Scarborough and colleagues found that children with reading deficits at 8 years had deficits in different sub-skills at the different ages at which language data was collected, when they were younger.

An additional reason for the variation in findings, which formed the basis for the research project reported in this thesis, concerns the two different modalities that oral language skills can be measured in, that is receptive (listening) skills and expressive (speaking) skills. Most studies have concentrated on receptive skills, while fewer

consider the expressive ones, even when there are considerable theoretical arguments for considering the role of expressive skills. The relevance of expressive skills is clear when considering that some of the earliest precursors of reading delay are former delays in age of attainment of expressive milestones, such as producing 4 to 6 words, using linguistic scales completed by parents (Shapiro et al., 1990). Considerable accruing evidence also points to the overlap between language disorders and a type of reading disorder, dyslexia (Bishop & Snowling, 2004), suggesting general language deficits associated with reading deficits. Also indicating the potential importance of expressive skills are the findings that 11-year-olds are more likely to experience reading disorders if they experience receptive and expressive language deficits (Simkin & Conti-Ramsden, 2006). Based on all these reasons, it could be argued that expressive skills have been relatively neglected in the literature.

It is therefore necessary to pay particular attention to the overlapping and unique contributions by each of the two modalities of language, reception and expression, apart from the focus on developmental issues and methodological differences in operationalizing language constructs. In addition, the implications of methodological differences for measuring expressive language between standardised testing and language samples are highlighted, for the purpose of linking this to reading comprehension.

A review of the relevant literature is presented here which intends to present a clearer picture of the published evidence, with an emphasis on expressive skills.

This review is organised by the two language modalities. First, significant studies supporting the role of receptive language skills are reviewed in Section 2.1, covering studies linking specific sub-skills and reading. Then, Section 2.2 reviews the relatively scarcer research describing the role of expressive language skills, focusing particularly on studies linking expressive skills using speech samples with reading.

2.1 Receptive language and reading

Apart from phonology and its central role in reading acquisition, semantics, grammar, and pragmatics are usually considered to form the basis for communication. Already introduced in Chapter 1, these concepts are again defined here for clarity. *Semantics* is

the study of word meanings. The rules for arranging the elements of language are studied in **grammar**, which is composed of **morphology** and **syntax**: while morphology is focused on word structure, syntax is concerned with word order. Finally, **pragmatics** is the use of conventions for using language in context, for expression and understanding of meaning. Among the pragmatic aspects, studies focusing on receptive language skills and their relationship to reading have mostly focused on **discourse-level comprehension**. Also known as listening comprehension, linguistic comprehension, and sometimes language comprehension, discourse-level comprehension requires the integration of new information with background knowledge to gain understanding, skills that are also required for comprehension in reading (Kintsch & Rawson, 2005).

It has already been reported that children with word reading deficits tend to have concurrent deficits in many of these language sub-skills (McArthur, Hogben, Edwards, Heath, & Mengler, 2000). Moreover, the association between these skills and reading comprehension becomes even clearer when considering studies of children with **specific reading comprehension** problems (i.e. **poor comprehenders**) that is, children who are able to read accurately and fluently but fail to derive meaning from what they have just read (Oakhill, 1994). The results from studying this group of children are particularly useful in that, whatever is found to be weak in poor comprehenders is likely to pertain to comprehension processes more directly than deficits found in the general population or in poor word readers. In a UK study, poor comprehenders tended to have concurrent receptive deficits in semantics, morphosyntax and listening comprehension skills at 8 years of age, while still possessing good phonological skills (Nation et al., 2004). Similar findings have been reported for older children in the USA. Pupils classified as poor comprehenders at 13-14 years of age have been shown to experience concurrent deficits in semantics, syntax and listening comprehension, while having good phonological skills as well (Catts et al., 2006).

Not only do poor comprehenders have concurrent broader language deficits, but also retrospective ones. In the same study by Catts and colleagues (2006), these 13-14 year old poor comprehenders were found to have had previous receptive language deficits in semantics, syntax and listening comprehension when they were 5-6, 7-8 and 9-10 years of age. These studies suggest that specific language sub-skills deficits start very early and are already measurable at the earliest stages of schooling. Interestingly, both the USA and the UK studies converged in finding that these deficits would not have attracted attention in the early primary years, since reading might be unimpaired at the

initial stages of reading acquisition, i.e. word reading. In the UK study, a third of these poor comprehenders' deficits had gone unnoticed by their teachers (Nation et al., 2004) even when these pupils could have met the language assessment requirement for a language impairment diagnosis. This is not as surprising once it is recognised that these children were decoding well, and also that teachers are not necessarily trained to attend to children's language and may lack sufficient awareness of language elements themselves (Moats, 1994; Moats & Foorman, 2003). Still, these unnoticed language deficits need to be given particular consideration when studying reading comprehension, precisely because they are not the focus of instruction.

These problems might include a substantial number of children if we take into account that, in practice, not all poor readers are poor comprehenders. Research focused on poor comprehenders have shed light on the way broad language skills are specifically related to comprehension, but in schools only around 10% of children fit the poor comprehender profile (Nation, 2005). Most children with reading comprehension deficits are likely to have mixed reading comprehension *and* decoding deficits, as illustrated in Figure 2.1 (Catts et al., 2006; Catts, Hogan, & Adlof, 2005; Hoover & Gough, 1990), but teachers might assume that the difficulties stem from decoding weaknesses. For this reason, it is important to study typically developing children, given that some might have depressed broad language skills that would normally go unnoticed but will nonetheless be related to their reading skills, both concurrently and over time.
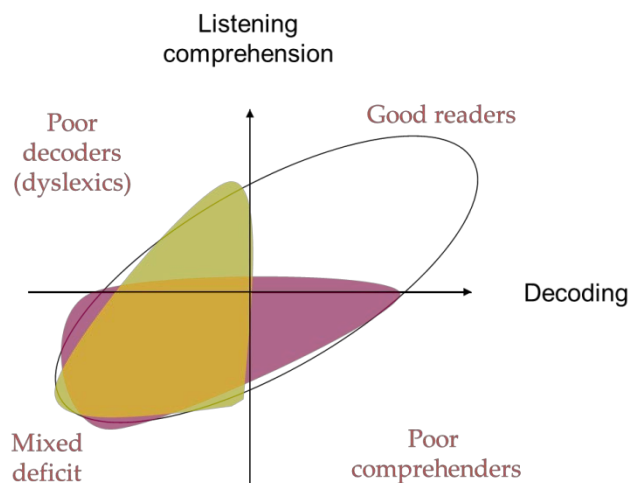
Figure 2.1 Illustration of classification of readers based on the Simple View model adapted from Catts *et al.* (2005, 2006).

These relatively comprehensive studies, which have included several receptive sub-skills at the same time, converge in emphasising the importance of all of them to reading comprehension. However the relative importance of these sub-skills remains to be further specified, giving careful consideration to methodological and developmental issues. Studies focusing on a few or even a single sub-skill, reviewed below, illustrate how methodological decisions have a critical impact on the results that are reported. Although more sub-skills have been considered in the literature, the following subsections focus on vocabulary, syntax and discourse organisation.

2.1.1 Receptive vocabulary/semantics and reading

In linguistics, *semantics* refers to how language conveys meanings, while *vocabulary* (more formally known as the lexicon) refers to the information about both meaning and pronunciation of the words (Crystal, 2008). Accordingly, semantic tasks, such as the word-association or the synonym judgement task, measure links between words while vocabulary tasks measure knowledge of single words. Abundant evidence from reading research studies supports the association between vocabulary knowledge and reading comprehension skills, as described below, while a few studies using semantic tasks have converged in finding semantic weaknesses in poor comprehenders.

The relationship between vocabulary knowledge and reading is well established, and causality has been argued to go both ways with vocabulary aiding reading acquisition and vice versa. On one hand, at the initial stages of reading, a typically developing child encountering a word will decode it, and if the word is already known, the child will understand its meaning. Therefore, a larger vocabulary repertoire should aid in developing reading comprehension skills because more words will be understood. On the other hand, if the successfully decoded word is not known by the child, the child might be able to derive meaning from context embedded in the text if their discourse-level comprehension skills are good (Cain, Oakhill, & Lemmon, 2004), and in this sense greater reading skills can increase a pupil's vocabulary knowledge. Perfetti and Hart (2002) conceptualised it as a continuous cycle in which "lexical skills allow comprehension, comprehension allows reading practice, reading practice strengthens lexical skills, etc." (p. 189) and indeed, there is empirical support for such reciprocity. In a longitudinal study, structural equation modelling was used to test how these relationships panned out between vocabulary and reading from 8-9 years to 10-11 years

of age, and the reciprocal relationships model was found to have the best fit to the data (Wagner, 2005).

While acknowledging the reciprocal nature of this relationship, it is clear that vocabulary development precedes reading instruction and acquisition. At the onset of reading instruction at least, vocabulary has long been acquired and refined (Tabors, Beals, & Weizman, 2001). Evidence from a longitudinal study using a larger stratified sample (n= 2143) starting at age 6, where all variables are allowed to influence each other, has shown that the onset for reading comprehension starting to influence vocabulary goes from reading comprehension at 7-8 years to vocabulary at 8-9 years (Verhoven & Van Leeuwe, 2008). Therefore, it may be reasonable to assume that causality goes more strongly from vocabulary to reading comprehension at least at these initial stages of reading acquisition.

In longitudinal studies of typically developing readers, the strength of the association between vocabulary and both components of reading ranges from $r = .28$ with word reading (Senechal, 2006) up to $r = .71$ (e.g. Tabors, Snow, & Dickinson, 2001) for reading comprehension, including everything in between those ranges (Muter et al., 2004; Roth, Speece, & Cooper, 2002; Snow et al., 1995). Just like in the Simple View model, where listening comprehension is more strongly related to reading comprehension than to word reading, correlations for vocabulary tend to be lower with word reading than with measures of reading comprehension, so this is a clear developmental issue that should be considered. Still, even when restricting the findings to reading comprehension measures only, the variability is wide.

One of the reasons for this variation may lie in which domain of vocabulary is being tapped, since children's knowledge of words can go from the word only being recognised superficially to its having well specified meanings and uses. These two dimensions have been termed as ***vocabulary breadth*** to refer to the number of words known, and ***vocabulary depth*** to refer to the extent that their meanings are known (Ouellette, 2006; Tannenbaum, Torgesen, & Wagner, 2006). There is evidence that these dimensions relate differently to distinct reading components, depending on the task used. Given that there is yet not agreement as to how to classify different tasks in these two dimensions, within this work the tasks will be referred to by their names, instead of using these dimensions.

A *receptive vocabulary task* only requires the testee to identify by pointing to the correct picture out of four that corresponds to a word presented orally; this task is said to tap into vocabulary knowledge since it minimises conscious effort. Slightly more complex are *naming tasks* which require the child to produce the word, adding an additional processing demand (this kind of task is also known in the literature as expressive vocabulary). Finally, in an *oral definitions* task the child needs to not only recognise a word spoken aloud by the assessor, but also construct an appropriate definition that meets the requirements of the test, making this the most cognitively challenging of the three vocabulary tasks.

A study seeking to relate these three different tasks to both word reading and reading comprehension has found evidence that vocabulary relates to both reading components, although here the focus is on reading comprehension. In an American sample of 9- to 10-year-olds, an oral definitions task was able to concurrently predict 12.1% of unique variance in reading comprehension, even after stringent controls for age, non-verbal ability and word reading skills (Ouellette, 2006). While testing alternative regression models in predicting reading comprehension, the model where the oral definitions test was placed first resulted in both receptive and naming tasks no longer adding significant variance, suggesting that any variance explained by the simpler tasks was already accounted for in the more complex oral definitions task.

Another study in a British sample of 8- to 10-year-olds replicated the previous finding that oral definitions could concurrently predict unique variance after stringent controls for age, non-verbal intelligence and word reading skills (Ricketts, Nation, & Bishop, 2007). Even though this study did not compare different vocabulary tasks and used only oral definitions, this task also predicted a sizeable unique variance in reading comprehension beyond stringent controls of either 17.8% or 30.7%, depending on whether the analysis additionally controlled for irregular word reading or not. Both studies converge in suggesting that skills required for the oral definitions task are specifically related to reading comprehension in typically developing readers between 8 and 10 years of age.

A second source of variation concerns the developmental issue discussed before: results are likely to change depending on when the criterion measure and the outcome measure are collected. Also, developmental predictors might be different from concurrent associations: a longitudinal study (Tabors, Snow et al., 2001) challenges the

predominance of oral definitions over receptive vocabulary measures found in the concurrent associations studies mentioned above (Ouellette, 2006; Ricketts et al., 2007).

In the longitudinal study by Tabors and colleagues (2001), kindergarten receptive vocabulary rather than their oral definitions had the strongest correlation *across time* with reading when children were 9-10 and 12-13 years of age. Interestingly, this was not the pattern found when this sample was initially followed up after only one year: from kindergarteners' vocabulary to their first grade reading, oral definitions rather than receptive vocabulary had the strongest correlation with reading (Snow et al., 1995). It has to be noted though, that the study used an experimental task for collecting oral definitions, not a standardised test, so the comparison with concurrent associations is not straightforward. In any case, it illustrates the difficulties associated with methodological decisions not only of which developmental window is studied, but also of which dimension of reading to use and which specific task is used to assess any variable.

Both of these methodological and developmental issues are summarised in Table 2.1, which compares longitudinal studies measuring some dimension of vocabulary at time 1 and a measure of reading comprehension at time 2. Although more studies exist relating vocabulary and reading longitudinally (Kendeou, Van den Broek, White, & Lynch, 2009; Nation & Snowling, 2004), only those reporting the correlation value between vocabulary on its own (or with a close semantic assessment) and reading were included in this table.

The combination of methodological and developmental issues is clearly seen here, where the oral definitions task appears as having the strongest correlations in the short term, but the simplest receptive vocabulary task appears as having the strongest correlations in the longer term in most studies, with the exception of the one by Adlof and colleagues (Adlof, Catts, & Lee, 2010), where oral definitions also has a slightly stronger correlation than receptive vocabulary over the long term. It should be noted that even though a weighting procedure was used, over half of the population in the Adlof study was composed of language impaired children.

Altogether, the evidence provides clear support for an association between different kinds of vocabulary tasks and reading comprehension, in typically developing readers, at both early and later stages in development. Also apparent from the studies reviewed is that the more complex tasks involving formulating oral definition are the ones with

the strongest concurrent relationships with reading comprehension, while most of the evidence with reading outcomes measured at 9 years onwards supports receptive vocabulary as a stronger longitudinal predictor in the long term.

Table 2.1 Reported Pearson correlations between vocabulary tasks and either word reading or reading comprehension in longitudinal studies. The beginning of the arrow marks the vocabulary measure at time 1; the head of the arrow marks the reading measure at time 2.

Age	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14
School Year in the UK / School Grade in the US	Reception	Year 1 Kinder	Year 2 1st Grade	Year 3 2nd Grade	Year 4 3rd Grade	Year 5 4th Grade	Year 6 5th Grade	Year 7 6th Grade	Year 8 7th Grade	Year 9 8th Grade
Muter et al., 2004		.53 (rv)								
Senechal 2006		.28 (rv)								
			.67 (rv)							
Snow et al., 1995		.44 (rv)								
		.53 (od)								
Oakhill and Cain, 2012					.47 (rv)					
					.41 (od& sim)					
Tabors et al., 2001				.60 (rv)						
				.55 (od)						
					.71 (rv)					
					.51 (od)					
Adlof et al, 2010			.47 (rv)							
			.56 (od)							
						.45 (rv)				
						.53 (od)				

NB. rv=receptive vocabulary; od=oral definitions; od&sim=oral definitions and similarities composite.

Regardless of the published evidence though, it can be argued that the oral definitions task clearly places additional demands rather than just assessing vocabulary knowledge. Of all the vocabulary tasks, this one requires the most effort since it requires the

participant to think about the category that the word belongs to, and state the characteristics which make it different from other members of the category. Given that it requires such a level of analysis to articulate a definition, the oral definitions task has been described as displaying metalinguistic skills beyond the linguistic ones (Snow, Cancino, De Temple, & Schley, 1991). Since poor comprehenders tend to have metalinguistic deficits (Paris & Myers, 1981), metalinguistic tasks may not be able to distinguish whether it is the metalinguistic skill or the linguistic skill per se, the one most affecting reading comprehension. This additional issue of how a particular task measures metalinguistic more than linguistic skills is especially relevant when considering the contributions by the next language sub-skill, grammar.

2.1.2 Receptive grammar and reading

Strictly speaking, in linguistics *grammar* refers to both *syntax*, which refers to the arrangement of words to form sentences, and *morphology*, which refers to the combination of word parts to attach meaning (Crystal, 2008). In reading studies however, many researchers use grammar and syntax interchangeably. Even when there is a vast amount of literature implicating morphology in reading comprehension (Deacon & Kirby, 2004; Jarmulowicz, Hay, Taran, & Ethington, 2007; Nagy, Berninger, & Abbott, 2006), this review is focused in describing the supporting evidence for syntax to focus on the supralexical aspects of language as other researchers have done (Share & Leikin, 2004), in order to maintain the focus on the higher levels of both language and reading.

Reading comprehension requires a good grasp of syntactic structures at the sentence level, since these are the foundations for integration of information at the discourse level. Children need to understand how words are arranged in sentences in order to understand them, both in oral discourse and in written text. As with vocabulary, there is the possibility that the arrow of causality goes in two directions: having better syntactic skills could facilitate reading acquisition and having more reading practice could enhance the knowledge and application of syntactic structures. As Perfetti and Hart (2002) have described for vocabulary, these two skills could be interdependent where each facilitates the other. While acknowledging the potentially reciprocal nature of this relationship, it is reasonable to assume that, at the initial stage of reading comprehension at least, the greatest influence goes from syntax to reading, since syntax

has been developing for years facilitated by inputs from both home and school (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002).

The evidence on the relation between syntax and reading comprehension is less straightforward than the one with vocabulary. On the one hand, there is abundant evidence that supports a positive association between these two variables; on the other, there is also evidence that syntax cannot account for unique variance when controlling for other known factors, such as phonological awareness (Gottardo, Stanovich, & Siegel, 1996; Shankweiler et al., 1995) or intelligence and maternal education (Bryant, MacLean, & Bradley, 1990), or intelligence and vocabulary (Oakhill et al., 2003). Once more, methodological and developmental issues could explain some of the variability in these results.

Some of the methodological difficulties when interpreting the literature linking syntax and reading have to do with construct overlap, the specific task characteristics used for measuring syntax, and the kinds of controls employed. First, in this work, construct overlap is meant to denote the difficulty of measuring one higher sub-skill, like syntax, without measuring another lower sub-skill, such as vocabulary. In the same sense, discourse comprehension cannot be tested without consideration of the child's grammar or vocabulary. This is not unexpected, and in most children language sub-skills would grow in tandem as evidenced by the high correlations usually reported amongst them (e.g. Oakhill et al., 2003). Standardised assessments usually try to minimise this problem by employing simple vocabulary and/or simple grammar but, to some extent, this issue is unavoidable.

A second methodological issue concerns the relative differences between tasks employed. While some studies used a ***receptive grammar*** task, where the child has to choose the correct picture amongst several corresponding to an orally presented sentence, others used a ***syntactic awareness*** task, which requires the ability to identify and manipulate structural grammatical elements, such as making judgements on the grammaticality of orally presented sentences or asking the participant to repeat erroneous sentences. Evidently, the second category of tasks is more complex and cognitively challenging, and since it requires conscious manipulation of language structures could be characterised as a meta-linguistic skill (Cain, 2007).

Significant zero-order correlations with reading comprehension have been reported using a receptive grammar task, ranging from $r = .39$ to $r = .52$ between 7 and 9 years of

age (Oakhill et al., 2003), up to $r =.57$ in the 8;07 to 10;08 age range (Goff, Pratt, & Ong, 2005). Similar associations with reading comprehension have been reported in correlational studies using a syntactic awareness task, from $r =.29$ to $r =.63$, depending on the task used (Bowey, 1986; Rego, 1997; Willows & Ryan, 1986). Furthermore, using reading-matched studies, the syntactic abilities of 9-year-old poor readers were compared against those of younger skilled readers matched for decoding skills (Tunmer et al., 1987); the syntactic awareness skills of the poor readers were significantly lower than those of the younger typically developing children whose word-reading levels were the same. This would suggest a causal role for syntax in reading comprehension, because the better syntax of the younger skilled readers cannot be attributed to greater reading experience. Similarly, another experiment sought to compare poor comprehenders with typical readers as controls, matched for both age and decoding ability, and it also found that poor comprehenders had weaker syntactic awareness than controls (Nation & Snowling, 2000).

So the evidence clearly supports a positive relationship between syntax and reading comprehension, and matched-reading studies even suggest a causal role. The main debate in evaluating the role of syntax in reading comprehension however is not the existence of the relationship, but whether syntax can provide additional information independently of other known factors, such as phonological awareness, verbal working memory, intelligence, vocabulary skills or metalinguistic skills given that causal evidence has used awareness tasks, which brings us to our third methodological issue: control variables. Controls are the main difference between reporting significant or non-significant results in reading studies that consider the role of syntax.

The greatest debate in the role of syntax comes from the argument that syntactic deficits are merely a by-product of phonological deficits, because the latter are hypothesised to create a bottleneck for further higher processing (Shankweiler, Crain, Brady, & Macaruso, 1992). Using three receptive grammar tasks, a study of reading disabled children from 7;06 to 9;06 found that their syntactic abilities did not differ from those of typically developing children of the same age (Shankweiler et al., 1995). An issue with this study is that, to diagnose reading disability, it used a discrepancy criterion of low reading ability despite normal intelligence, in other words, they were poor decoders (also termed dyslexic); the problem lies in that verbal abilities were normal or high from the outset in the reading disabled group, so it is natural to also expect strengths in syntactic abilities.

A more stringent analysis placing syntax as a by-product of phonological awareness was the one carried out in a Canadian study of 8- to 9-year-olds, which found that two experimental syntactic awareness tasks could not predict unique variance beyond the variance accounted for by phonological awareness and verbal working memory (Gottardo et al., 1996). Although more compelling, the methodological issue with this study is that it used sentences for the verbal working memory task. Even when the sentences were short, up to 9 words and grammatically simple, it still required some integration of information at the sentence level. More recent and converging evidence from poor comprehender studies reviewed before (Catts et al., 2006; Nation et al., 2004; Nation & Snowling, 2000) undermines the concept of syntactic deficits resulting from phonological ones, since they have identified syntactic difficulties in these children who nonetheless had good phonological skills.

It has also been debated whether syntax can account for any additional variance in reading beyond solid vocabulary skills. Using a receptive grammar task, syntax and reading comprehension have been linked in 7- and 8-year-olds (r=.39, p<.001) and in 8- and 9-year-olds (r=.52, p<.01), but the unique contribution of syntax became non-significant (R^2 change = .018 for concurrent relations and R^2 change = .001 for longitudinal relations, both not significant) when controlling for receptive vocabulary, verbal and non-verbal intelligence (Oakhill et al., 2003). Conversely, when using a syntactic awareness task, others have found syntax to predict additional variance in reading comprehension beyond vocabulary skills (Bowey, 1986; Willows & Ryan, 1986) and beyond vocabulary skills and intelligence, in French (Demont & Gombert, 1996).

Whether and how to control for intelligence certainly has implications for how the results are reported. Another study which did not find syntax to provide additional information beyond controls was a longitudinal study that measured pre-schoolers' syntax. This study found that three syntactic measures when the children were 3 ½ years and a syntactic awareness task when they were 4 ½ accounted for unique variance in reading when they were 6 ½ years old controlling for age (Bryant et al., 1990). However, once the child's intelligence and the mother's educational level were taken into account, the contribution of these three syntactic measures was no longer significant. It is possible to argue though, that controlling for intelligence confounds the oral language variable. To control for general intelligence, the Bryant and colleagues (1990) study employed the Wechsler Preschool and Primary Scale of Intelligence,

which contains not one, but several subtests assessing oral language; thus controlling for a child's intelligence is already taking into account a considerable amount of variance in language skills. This is the reason newer studies only control for non-verbal intelligence.

The developmental issues in studies investigating syntax have documented differing results depending on which age range the measurements are taken. When a longitudinal study considered a wider age range from 4 up to 8 years of age, syntactic awareness predicted unique variance in reading comprehension, but not for word reading (Demont & Gombert, 1996). Indeed, in early reading acquisition, an association between syntax and decoding would not be expected, given the dissociation between the predictors for word reading and the ones for reading comprehension. Nonetheless, significant associations between a syntactic awareness task and reading isolated words have been reported (Willows & Ryan, 1986). One way in which strong grammatical knowledge and skills are thought to help reading words is by helping children use context to infer the meanings of unfamiliar words (Rego & Bryant, 1993).

There is evidence that the development of some of these skills may be influenced by reading instructional practices. In the Rego and Bryant (1993) study, children who were skilled in grammar were also good in a measure of contextual facilitation, which may help decoding. The authors hypothesised that this facilitation could be due to environmental factors, specifically the school's instructional methods, since children from their sample had been taught by a whole-language approach. Indeed, in a later study (Rego, 1997) with children taught in a structured phonic approach (where the emphasis is on learning grapheme-phoneme correspondences) the relationship between syntactic awareness and decoding was not supported, while a relationship to reading comprehension was substantiated. These results seem to suggest that instructional practices which emphasised contextual facilitation could play a role in whether syntactic awareness is important or not for decoding ability. Since most instructional policies in the USA and the UK now emphasise the phonic approach (National Reading Panel, 2000; Rose, 2006), it is possible that syntactic awareness might be mainly related to reading comprehension.

In this review, syntax has mostly been conceptualised as a receptive measure or as an awareness measure, but the latter might be more of a meta-linguistic than a linguistic skill. The appeal of syntactic awareness over receptive grammar tasks comes from the

literature reviewed above suggesting not only a link, but also a causal role in reading comprehension. Nevertheless, if several language sub-skills are to be studied simultaneously in a comprehensive design, it would be very likely for another task to also have a meta-linguistic component that would contribute to shared variances. In addition, there is already some evidence that some syntactic awareness tasks cannot provide additional information in reading comprehension once receptive vocabulary skills, receptive grammar and working memory are taken into account (Cain, 2007).

In sum it is still debated whether syntactic skills provide additional information beyond the one contributed by general intelligence and solid vocabulary skills, and comparison of studies addressing this debate is difficult due to methodological differences. However, even with this variability in syntactic measures, the evidence remains robust for a significant association with reading comprehension, particularly after the first few years of formal reading instruction.

2.1.3 Receptive discourse/listening comprehension and reading

Vocabulary and syntax make a necessary contribution to understanding spoken discourse. Beyond the word and the sentence levels, understanding spoken discourse also depends on how listeners process that information in order to generate inferences and create representations or mental models, based on their general knowledge (Bishop, 1997). Since these are also the same higher-level processes required in skilled reading comprehension (Kintsch & Rawson, 2005), it is natural for discourse comprehension and reading comprehension to be associated.

The relationship nonetheless is not straightforward as there are important differences between spoken and written language. While in spoken language the context is shared and language reflects this common context relying on non-verbal cues, the context in written language is different for writer and reader. Consequently, comprehension cannot take clues from the shared environment, and meaning must be built into the words and sentences (Nelson, 1988), for example replacing deictic expressions such as 'there' for a precise description such as 'on the table behind the vase'. There is spoken language however, where certain distance between speaker and the listener is assumed, such as a piece of discourse. Regardless of whether it is spoken or written, the kind of language where this separation in time and space exists has been denominated ***decontextualized***

language. Snow (1991) found that it was children's decontextualized language abilities rather than the contextualized or face-to-face language skills, the ones which were strongly associated with reading comprehension.

Therefore, associations between spoken comprehension and reading comprehension become clearer when the spoken comprehension measure is operationalized as discourse. Many of the studies which have measured spoken comprehension beyond the sentence level use a variety of terms to refer to *discourse-level comprehension*. While some call it linguistic or language comprehension, others call it listening comprehension, but all measure discourse-level comprehension with narrative or expository texts, which use decontextualized language. For simplicity, the most common term of *listening comprehension* will be used here.

As with vocabulary, listening comprehension seems to have a reciprocal relationship with reading comprehension where each one enhances the other. Initially though, listening comprehension would have been developing for years while reading comprehension is still developing, so the causal arrow should go from listening to reading at least after the first few years of formal reading instruction. Indeed, reading-matched studies where older poor readers are compared to younger good readers matched for reading comprehension skills, add support for the causal role of listening comprehension in reading comprehension after the initial years of formal reading instruction. Stothard and Hulme (1992) compared 7- to 8-year-old poor comprehenders with two matched control groups, typical readers matched for age and younger readers matched for reading comprehension ability. Their research found that poor comprehenders had deficits in listening comprehension when compared to the control group of the same age, but not when compared to the younger comprehension-matched control group, suggesting a general language comprehension deficit which manifested itself in both listening and written forms.

Converging evidence for causality going from listening to reading at this early stage comes from the Dutch longitudinal study reviewed in Section 2.1.1 before, which collected yearly measurements for both listening and reading comprehension from age six onwards (Verhoven & Van Leeuwe, 2008). Amongst all the reciprocal relationships allowed in this study, the *first* reverse influence from print to language, in other words, from reading to listening, is seen from 7-8 years reading comprehension to 8-9 years listening comprehension (Verhoeven & Van Leeuwe, 2008, Fig. 4, p. 417). Although

Verhoeven and Van Leeuwe do not propose an explanation for why these reciprocal relations are found, a possibility may be that content-specific background knowledge presented in written form may aid when encountering that same type of knowledge orally. Marzano (2004) argues that even when reading will not create as robust experiences as direct experiences, reading can create *virtual* experiences powerful enough to "significantly increase background knowledge" (p. 36). That could be a possible, albeit untested, pathway for a reciprocal relation of reading comprehension facilitating listening comprehension.

Another piece of evidence showing reciprocal relations between reading and listening comprehension came from a cross-sequential study (Berninger & Abbott, 2010), where listening comprehension explained unique variance in reading comprehension concurrently across grades 1, 3, 5 and 7, and in the same way, reading comprehension also explained unique variance in listening comprehension in the same grades. Since these regressions were done concurrently, the study by Verhoeven & Van Leeuwe (2008) is the one that offers longitudinal evidence, with reciprocal relations starting from 7-8 years of age. Therefore, even when the reciprocity of these variables is acknowledged after the ages of 7-8 years, the focus of this review is on the evidence where causality goes from listening comprehension to reading comprehension at any age.

The importance of listening comprehension was already introduced in section 1.2 when the Simple View model was discussed. This model highlighted how developmental changes affect how the contribution from listening comprehension to reading is seen more clearly after decoding skills have been mastered. Figure 1.1 illustrated how listening comprehension has almost the same relative importance as word reading during 3rd and 4th grades, and how its relative importance increases from 5th and 6th grades onwards (Gough et al., 1996). Zero-order correlations from Gough and colleagues (1996) meta-analysis are presented in Table 2.2, which had been illustrated before in Figure 1.1, but the coefficients are presented here to give a precise picture of the patterns.

Table 2.2 Average weighted correlations between reading and word reading (decoding), and between listening and reading from a meta-analysis of 17 empirical studies (Gough et al., 1996).

Grade Level	Average weighted correlations between decoding and reading	Average weighted correlations between listening and reading
Grades 1st and 2nd (6-8 years)	0.61	0.41
Grades 3rd and 4th (8-10 years)	0.53	0.50
Grades 5th and 6th (10-12 years)	0.48	0.72
College (university)	0.39	0.68

These developmental patterns suggest that once decoding has been mastered, the association between listening comprehension and reading comprehension is most visible. Even when its contribution is confounded by decoding skills from 6 to 8 years of age at the beginning of reading acquisition, children's listening comprehension skills are still being developed (Paris & Paris, 2003).

Besides the developmental changes illustrated in this meta-analysis, the findings regarding listening comprehension are also influenced by methodological issues. Listening comprehension is thought to share similar processes to reading comprehension. Specifically, many discourse processes are common to both listeners and readers. The goal of comprehension is ultimately to build a coherent mental representation or structure (Gernsbacher, 1997). This mental representation's most relevant characteristic is a connectedness or *coherence* of the information (Sanders & Gernsbacher, 2004). In understanding a text, this coherence shows how "appropriate, meaningful connections are established between elements of text and the reader's prior knowledge" (Rapp, Van den Broek, McMaster, Kendeou, & Espin, 2007). Skilled readers strive to maintain a high standard for coherence (Perfetti, Landi, & Oakhill, 2005), which allows them to identify and resolve inconsistencies between elements of the text, or between elements of the text and background knowledge.

However, besides striving for this full discourse or *global coherence,* the listener/reader also strives to achieve *local coherence* between sentences or within sentences (Clifton

& Duffy, 2001). This kind of skill, also known as ***cohesion***, refers to the syntactic and semantic connectivity of linguistic forms (Crystal, 2008), with some linguists and cognitive psychologists paying particular attention at resolving anaphoric relations (based on the work by Halliday & Hasan, 1976).

Being a higher order skill, the methodological difficulties of measuring listening comprehension are related to the issue of construct overlap. Of all the linguistic variables, skilled listening comprehension will depend to some extent on the previous two language sub-skills, vocabulary and grammar. Moreover, these sub-skills are not only overlapping but are dynamically related. For example, vocabulary and listening comprehension have been shown to influence each other over time in a structural equation model (Verhoven & Van Leeuwe, 2008, Figure 5, p. 417) where listening comprehension in grades 2 and 4 were found to be significantly related to vocabulary in grades 3 and 5, respectively.

In addition to the overlap with lower sub-skills, others have discussed the possibility that comprehension skills might be explained by working memory deficits alone (Seigneuric, Ehrlich, Oakhill, & Yuill, 2000). One study has already addressed these issues and their findings support the idea that there are indeed higher-order skills that go beyond vocabulary and memory skills in explaining reading comprehension (Cain, Oakhill, & Bryant, 2004). In this study inferencing skills, comprehension monitoring skills and knowledge of narrative structure were hypothesised to contribute to reading comprehension. These three skills were found to explain unique variance in reading comprehension skills from 7 to 11 years of age, beyond strict controls for word reading, receptive vocabulary, verbal intelligence and even two working memory tasks. However, only the inference tasks were strictly presented as listening tasks, while the other two required some reading. Although only inferencing was a purely listening task, it can be argued that the several controls for word reading would deduct the influence of decoding skills, in turn giving weight to the contributions made by monitoring and narrative structure.

A second related methodological issue with listening comprehension is not whether there is an association with reading comprehension, but whether it can account for any unique variance beyond other known factors. Further evidence has come from studies which have found support for the role of listening comprehension even after the most stringent possible control: accounting for *initial* reading comprehension ability. Nation

and Snowling (2004) found that listening comprehension at 8;6 years accounted for 14.1% of significant unique variance in reading comprehension at age 13, even after stringent controls for previous reading comprehension, non-verbal ability, non-word reading and phonological skills. These contributions have also been observed even in early reading: a Dutch study of younger children found that listening comprehension accounted for unique variance in reading comprehension in grade 3 even after strict controls of previous reading comprehension and decoding ability at the end of grade 1 (de Jong & van der Leij, 2002).

Although listening comprehension can be measured for several genres, a great amount of initial listening comprehension is usually about narratives. Narratives are a form of discourse where a sequence of events, real or imagined, is shared orally or in a written format and it is usually the kind of discourse that children are first exposed to at a very early age, usually in their home environments (Beals, 2001). Being a cognitively challenging task, it would be natural to expect different levels of ability by developmental stages, and there is evidence that narrative skills do have a clear developmental trend. Paris and Paris (2003) were able to show that narrative reasoning skills based on a picture-book task (not only comprehension but also production) tend to increase with age and with reading skill, in a cross-sectional sample, and with time in a longitudinal sample, from kindergarten through second-grade children (5 to 7-8 years of age). Evidently, as children develop they will also encounter more expository texts, and this is the main kind of prose they find when reading for other areas of the curriculum. Still, the majority of initial reading instruction is mainly focused on narratives.

Once again, as with the rest of the linguistic sub-skills reviewed here, methodological and developmental considerations such as these are crucial to making statements about the relative contribution of any sub-skill to reading in a given developmental window. Overall, the evidence for the role of listening comprehension supports the notion that comprehension processes are shared in listening and reading comprehension (Stothard & Hulme, 1992), that listening comprehension skills can be longitudinal precursors of reading comprehension (Verhoeven & Van Leeuwe, 2008) and that listening comprehension skills are particularly visible after decoding skills have been mastered (Hoover & Gough, 1990). Therefore, listening comprehension should be included in any study of comprehensive language skills.

2.2 Expressive language and reading

Compared to the wealth of evidence linking different receptive language skills and reading, the literature connecting expressive skills and reading has been relatively neglected, despite a significant number of studies linking them indirectly in atypical populations, and a few studies linking them directly in typically developing children.

The first group of studies supporting a role for expressive skills are those of atypical populations which have found a link between different types of language deficits and later word reading and reading comprehension difficulties. There is now abundant converging evidence those children diagnosed with a language impairment tend to have reading difficulties as a group (Bishop & Adams, 1990; Catts, Adlof et al., 2005; DeThorne et al., 2006; Snowling et al., 2000). Similarly, children at a high familial risk of acquiring reading disorders who eventually do develop reading difficulties have shown early expressive deficits (Lyytinen & Lyytinen, 2004; Scarborough, 1990b). Such overlap has brought forward the debate about whether language and reading impairments are truly distinct disorders (Bishop & Snowling, 2004; Catts, Adlof et al., 2005). Furthermore, some evidence has illustrated that children who have a combination of expressive and receptive deficits tend to have the worst reading outcomes (Simkin & Conti-Ramsden, 2006).

A second and relatively smaller category of studies linking expressive language and reading has been carried out in unselected samples of typically developing preschool and school-aged children, reviewed in Section 2.2.2. Several studies have now found connections between very early expressive language and later reading skills (Harlaar et al., 2008; NICHD Early Child Research Network, 2005), while a few have found links between school-age expressive language and reading (Cain, 2003; Cain & Oakhill, 1996; Chen-Wilson, 2005; Klecan-Aker & Caraway, 1997).

Given this evidence and considering that production tasks are routinely used as indices of early and school-age language development (Liles, 1993), a deeper examination of the relationship between language and reading is warranted. Below there is a brief overview of these studies that attempts to interpret the nature of this relationship. The following two sections briefly review these two categories of studies. Section 2.2.1 is focused on research about atypical populations while 2.2.2 will summarise studies carried out with typically developing children.

2.2.1 Language impairment and reading

Children whose language reception and production are impaired from an early point in life face well documented social, emotional and academic negative outcomes (Johnson, Beitchman, & Brownlie, 2010). Amongst the academic outcomes, several longitudinal studies have now established that children with language impairments are likely to face reading deficits and/or disorders (Bishop & Adams, 1990; Catts, 1993; Catts, Adlof et al., 2005; Catts, Fey, Tomblin, & Zhang, 2002; Snowling et al., 2000; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

The observation of a connection between language impairment and reading is not a recent occurrence. Practitioners had previously reported that, in some schools, many of the children who were identified for speech or language services in the first years were the ones who later on received services by the schools' reading specialist (Wallach & Liebergott, 1984). Not all language impaired children, however, become impaired at reading. Studies seeking to specify the reading outcomes for children with some form of language impairment reach diverging conclusions, and this is due in part to methodological considerations surrounding the conceptualisation of language impairments, and to the fact that children with language impairments encompass a very heterogeneous group (Law, Boyle, Harris, Harknes, & Nye, 2000).

There are three main distinctions or classifications that researchers use when studying language impairments and their related outcomes. The most common distinction in reading outcomes research concerns whether or not the language impairment is specific to language with general intelligence in the normal range, or whether it is accompanied by other more general cognitive deficits. A second common classification amongst speech-language practitioners distinguishes between speech and language impairments. This classification is different from the third and most relevant to this review: receptive and expressive language skills. Each of these classifications is reviewed next.

2.2.1.1 Specific versus non-specific language impairments and reading

A fairly common grouping amongst language researchers focuses on children with *Specific Language Impairment* (SLI), i.e. a language impairment despite normal

intelligence, hearing, emotional development and adequate opportunity to learn (Bishop, 1997). These studies usually focus on how likely these children with SLI are to develop a *specific* reading disability, that is, reading impairment despite normal intelligence, development and adequate opportunity to learn.

In the general population, SLI prevalence rates have been calculated at 7.4% in kindergarten children of 5-6 years of age in an epidemiological study (Tomblin, Records, Buckwalter, Zhang, & Smith, 1997).

In turn, estimates for how many of these SLI children develop reading disabilities vary greatly. Some reviews have placed the risk for SLI children to develop reading disabilities at a substantial 40-50% (McArthur et al., 2000). Others have criticised these estimates as very broad stating that they could confound reciprocal relations of reading into language because language impairments were identified when reading was already being taught. Indeed, more conservative estimates have been reported when SLI is identified *before* reading instruction begins. When identified in kindergarten (5-6 years), 17% to 35.8% of SLI children are likely to be diagnosed with a specific reading disability when they become 7 and 14 years of age (Catts, Adlof et al., 2005). The wide variability in Catts and colleagues' study was due to whether a discrepancy criterion was used to diagnose reading disability (i.e. higher rates if the discrepancy criterion was not used), and also to which subsets of intelligence were used for the discrepancy criterion (higher rates of reading disorder were found if only the nonverbal subset was taken into account, and lower rates if the child had both normal *verbal* and *nonverbal* intelligence).

As with previous studies, the developmental window studied in which results are reported is crucial (Scarborough, 1990b). The main developmental issues regarding the relationship of language impairments with reading have been conceptualised in two hypotheses: the critical age and the illusory recovery hypotheses. First, Bishop and Adams (1990) proposed that the issue is not whether there is a language impairment or not, but whether it is present at the critical age when children start to receive reading instruction, which in the UK is at age 5. In their study, they followed 83 four-year-old children with SLI, and those who still had language impairment at 5 years also showed reading and language deficits at 8 years; while those children whose impairment was mainly resolved by 5 years showed no reading deficits at 8 years, but were still behind in receptive grammar and listening comprehension. This hypothesis has received some

support for initial reading instruction (Nathan, Stackhouse, Goulandris, & Snowling, 2004): those children with persistent speech difficulties at 6;09 had the worst reading accuracy (a proxy for word reading/decoding) outcomes.

However a different picture emerged when the Bishop and Adams cohort was followed into adolescence: a greater proportion of these pupils had difficulties with both word reading and reading comprehension (Snowling et al., 2000), which brings us to the second hypothesis of the illusory recovery of SLI children. This hypothesis aims to account for the fact that some of the SLI children appear to go through a period when they seem to catch up with their peers in reading and language, only to display deficits later on (Scarborough & Dobrich, 1990). However, not everyone has found this pattern. A different pattern of wide but stable differences across the school years has also been reported in a long-term longitudinal study in the USA (Catts et al., 2008). In any case, what both longitudinal studies do support is the notion that preschool language impairments affect reading comprehension up to 15-16 years of age, irrespective of whether SLI children catch up during the word reading acquisition phase or not.

2.2.1.2 Speech versus language impairments and reading

The most common classification of language impairment disorders amongst language practitioners is focused on whether the disorder is in individual sounds or phonemes, which are usually referred to as *speech disorders* or speech sound disorders (SSD), or in wider aspects of language, which are referred as *language disorders* (American Speech-Language-Hearing Association, 2010). In this classification system, speech disorders can, at least in theory, be either receptive and/or expressive, and the same could apply to language disorders, although speech disorders are usually noticed because of their expressive component in articulation. In other words, this classification system does not stem from the receptive/expressive distinction.

If a child has only one type of either speech or language disorder on its own, it is still debated whether they can impair reading (Pennington & Bishop, 2009). Since speech disorders are usually noticed by their expressive component, they tend to be very visible and thus have received a great deal of attention. Given that speech disorders encompass problems of articulation and reception of *phonemes,* and since phonological awareness is clearly implicated in word reading, speech impairments could be expected to impair

word reading, but results are mixed. Empirically, while some studies have found the ability to produce sounds correctly to have a negligible influence on early reading skills between ages 6 and 8 (Bishop & Adams, 1990; Catts, 1993; Nathan et al., 2004; Scarborough, 1990b), other investigations that accounted for the severity of the speech difficulties have found that speech skills are still important (Bird, Bishop, & Freeman, 1995; Larrivee & Catts, 1999). Thus, this debate is far from settled.

For children who have both disorders however, the research has reached a consensus. In the general population, 2% to 8% of the children up to 4;06 years of age are expected to have a combination of speech and language impairments (Law et al., 2000), and reviews have found that it is this group of children with a combination of both disorders who are at a considerably higher risk of developing reading disabilities (Pennington & Bishop, 2009).

In any case, regardless of whether they can account for future reading outcomes on their own, it is reported that speech disorders tend to overlap greatly with expressive semantics and syntax in preschool (DeThorne et al., 2006). As impairments in these expressive broader language skills in kindergarten are related to subsequent reading comprehension skills (Catts et al., 2002, reviewed in the next section), it is reasonable to expect then that some children diagnosed with speech disorders could be at risk for later reading comprehension deficits. Indeed, in children at familial risk of reading disorders, it was those high children who displayed deficits in expressive syntax *and* articulation errors as early as 2:06 years of age, the ones who eventually developed reading difficulties at 8 years of age (Scarborough, 1990b).

2.2.1.3 Receptive versus expressive impairments and reading

A few studies of atypical language development have distinguished between receptive and expressive skills, and it is in the estimates of reading outcomes using this distinction where the developmental influences are most clear, due to differences between perception and production of language. Children who are identified with a language impairment before or during the first years of school are usually noticed because they tend to show deficits in expressive language, although these deficits are usually accompanied by receptive deficits (Bishop, 1997). While some children have combined receptive and expressive difficulties, some might display expressive deficits only, but

exclusive receptive deficits are rare (Bates et al., 1995). However, developmental asynchronies where one modality of language appears to be low relative to performance in another modality do exist and are documented in atypical development (Bates et al., 1995; Farmer, 1996).

How this asynchronous development is related to reading was already highlighted by the Scarborough (1990b) study reviewed above in Section 2.2.1.1 where different expressive or receptive measures were statistically different between typical and impaired readers at 8 years of age: at 2 ½ years, expressive syntax from a speech sample was statistically different, while neither expressive vocabulary nor receptive vocabulary were statistically different; at 3 ½ years, receptive vocabulary with two standardised tests was the statistically different variable in children who later became dyslexic.

Since there could be dissociations between modalities, it becomes relevant to examine both of them when evaluating the relationship between language and reading. Furthermore, the fact that mild deficits in receptive skills can be relatively unnoticed (Nation et al., 2004) makes expressive skills, which are much more visible to parents and teachers, a potentially useful tool to uncover the kind of mild deficits that can ultimately be related to reading comprehension.

Even when none of the atypical language studies investigated in this review addressed the question of whether expressive skills can account for unique variance on their own, findings of different modalities being more or less predictive of later skills emphasise the need to capture the full spectrum of language abilities, which could capture unexplained variance. One large sample longitudinal study (n=1064) actually found such patterns when comparing outcomes for expressive and receptive deficits, and it found that a child performing below the 10th percentile in any of the two modalities of a standardised preschool language test at 3 and 4;06 years of age, had significantly lower kindergarten literacy skills as rated by their teachers (Justice, Bowles, Turnbull, & Skibbe, 2009). The effect was more pronounced for expressive than for receptive deficits at 3 years of age (d=.91 v .86), while the reverse was true at 4;06 years (d=.96 for receptive and .75 for expressive deficits). Nonetheless, both were found to be related to school entry literacy skills.

Another study of atypical language (Catts et al., 2002) did not find inconsistencies between receptive and expressive skills, but nonetheless found that, for language impaired children in kindergarten at 5-6 years of age, receptive and expressive

concurrent language measures were similarly correlated with reading comprehension outcomes in second grade (7-8 years of age; r: = .56 v .52 for receptive and expressive) and fourth grade (9-10 years of age; r: = .59 v .55 for receptive and expressive), converging that impairment in any of both modalities is related to concurrent reading comprehension skills. Interestingly, the same pattern of results was found for word reading outcomes as well. In this study of school-aged children, receptive skills were consistently better predictors than expressive skills.

Further evidence from older children (10 and 11-year-olds), however, with either selective or combined impairments did find that those who have both receptive and expressive language deficits are the ones likely to also exhibit greater reading deficits (Simkin & Conti-Ramsden, 2006). In this study, children with a history of language deficits were classified into one of three groups: expressive deficits *only*, a *combination* of expressive and receptive deficits, and a *mainly-resolved* group. A main effect of group was observed for reading comprehension $F(2, 84)=22.193$, $p < .001$. While the mainly-resolved group had reading scores within a normal range, the expressive only and the expressive/receptive group averages in both word reading and reading comprehension were significantly lower in post hoc tests ($p <.001$). Comparing the two groups with current deficits, the children with expressive-only deficits were less impaired in reading comprehension than the ones having combined expressive/receptive deficits ($p <.05$). The results from these comparisons seem to suggest that, at 10 and 11 years of age, the more comprehensive deficits tend to signal a greater risk for both word reading and reading comprehension difficulties. They also highlight that receptive measures might not provide the whole spectrum of language abilities, which a combination of receptive and expressive skills might do.

Even when none of the atypical language studies reviewed directly addressed the question of whether broader expressive deficits can predict reading comprehension impairments on their own, they still support the idea that different modalities can be differentially related to reading comprehension skills, possibly due to asynchronous developmental patterns, and also that having combinations of both receptive and expressive difficulties at both preschool and school can increase the likelihood of developing reading comprehension deficits.

2.2.1.4 Overall trends in atypical language development

Taking the evidence from these three kinds of studies in atypical language, their results converge in that a significant proportion of language impaired children do become deficient in their reading, but that these deficits will depend on how both language and reading impairments are diagnosed, on whether deficits in intelligence are present, the timing of language and speech disorders, and finally, on having a combination of expressive and receptive deficits, or a combination of speech and language deficits since combinations enhance the probability of developing reading disorders.

More specifically, regarding the evidence for atypical expressive skills, it seems to support the idea that even when their relative contribution to reading comprehension is yet to be established, particularly for broader language skills, they are nonetheless related to reading comprehension, whether the expressive deficits are identified early or they persist later on. If a combination of deficits in impaired children signals a particularly detrimental prognosis, expressive skills could potentially provide a very useful means to observe the full range of language development in typically developing children. Since not every child in a classroom is likely to be tested individually for broader language skills, the visibility of expressive language skills makes them a good candidate for investigation, particularly in light of findings that after the first few years of reading instruction, some children's broad language deficits might go unnoticed, particularly if they are decoding well (Nation et al., 2004).

Assuming that both language and reading are continuous behaviours (Plomin & Kovas, 2005) in the sense the difference between the serious deficits present in disorders and the mild deficits in non-impaired children is a matter of degree, then the evidence presented thus far in atypical development might mean that the evidence for the relationship between language and reading comprehension in typically developing children might also be present but to a varying degree. From this perspective, the evidence that children with language impairments tend to have broader expressive and receptive language skills deficits such as semantics, syntax and discourse (Catts, Adlof et al., 2005) coupled with the evidence that at the earliest stages of reading comprehension seemingly typical readers also display mild deficits in broader receptive language skills which tend to go unnoticed (Nation et al., 2004), then it could be proposed that broader *more noticeable* expressive language skills in *typical* readers could potentially uncover an expressive counterpart for those hidden receptive deficits

that could ultimately influence reading comprehension. Indeed, a few studies in typically developing children have already explored the idea that broader expressive skills might be related reading comprehension, and they are ones that, if the data supports it, would provide the strongest case for identifying mild expressive deficits in typically developing children. These findings are reviewed in the following section.

2.2.2 Expressive language and reading in typically developing children

The section on atypical language provided indirect evidence for a proposed relationship between expressive language and reading comprehension in typically developing children. Compared to children with language impairments, who are likely to receive specialist support, it is reasonable to argue that there is a potentially greater proportion of typically developing children with unnoticed mild language deficits, who are unlikely to be identified, whose deficits could ultimately be related to their reading comprehension skills. For that reason, this section evaluates the findings of typically developing children, and a few of the studies about typical children who were later diagnosed with a reading disorder. Even when the literature becomes scarcer in typically developing children's reading development, there are some solid longitudinal studies and a few concurrent studies available. The following two subsections examine these relationships longitudinally, between infant and preschool language, before reading instruction starts and later reading outcomes, while the last subsection deals with children's school-age language and their concurrent associations with reading.

2.2.2.1 Infant emergent language and reading

An acid test for the linguistic bases for reading would be that the earliest measures of language were related to later reading outcomes. Indeed, receptive differences have been reported in lab studies of newborns at familial risk for reading disorders (Guttorm et al., 2005). However, most parents will first notice differences when their children start to talk. Therefore, expressive indices are a relatively noticeable and accessible first measure of linguistic development. Actually, very early expressive indices, such as vocabulary and grammar at the onset of communication, have been found to be related to reading development in typically developing children (Harlaar et al., 2008; Shapiro et al., 1990; Walker et al., 1994), reviewed in this section, as well as in children who were later diagnosed with dyslexia (Scarborough, 1990b), reviewed before, by using either standardised assessments of communication inventories or speech samples.

In a long-term longitudinal study, Shapiro *et al.* (1990) followed 240 children from birth to 7;06 years of age, and collected infant language and motor development measures in the first two years of life and later primary school reading measures to categorise children as reading delayed. Early language measures were collected using a

combination of scales and parental reports, while reading was evaluated using the reading section of the first version of the Woodcock-Johnson Psycho-educational Battery (Woodcock & Johnson, 1977). Reading delay was operationalized as being 6 or more months behind their chronological age. Whether the study measured word reading or reading comprehension was not reported. Statistically significant differences were found for age of attainment in children with and without reading delay for the following expressive milestones: a) produce 4-6 words; b) produce 7-20 words; c) produce 50 words; d) produce two-word sentences; and the following receptive milestones: e) point to 5 body parts; and f) point to 8 body parts. In other words, children with reading delay at 7 ½ years, were significantly slower to attain these particular expressive and receptive measures in their early years. Moreover, when specifying the slowest 10th percentile of language attainment, three expressive measures had the highest predictive values for later reading delay: 1-step command with gesture (50%), producing 7-20 words (47%) and producing 2-word sentences (50%). To sum up, a good proportion of the really late talkers among this unselected group of children were later found to have reading delay.

Another long-term longitudinal study supporting a role for early language production, but with a much smaller sample (Walker et al., 1994), followed children's language production from 7 months through 3 years, and reading skill in Kindergarten and the first three primary years. In this study, syntactic and vocabulary indices derived from early speech were significantly correlated with a composite measure of reading skill from two standardised tests (The Comprehensive Test of Basic Skills, CTBS, 1987; The Metropolitan Achievement Tests, Prescott, Balow, Hogan, & Farr, 1984) in Kindergarten, 1st, 2nd, and 3rd grades (it is not specified which was word reading and which was reading comprehension). The longest term correlation with 3rd grade reading, was $r=.43$, $p<.05$, for both a lexical index (types) and a syntactic index (Mean Length of Utterance). In a regression analysis, the lexical index types accounted for significant unique variance in later reading achievement in kindergarten (17%), first (17%) and particularly third grade (25%) after controlling for socioeconomic differences.

More generalizable evidence comes from a large scale longitudinal twin study in the UK, which collected parent reports on children's speech at 2, 3 and 4 years and teacher assessments on their reading skills at 7, 9 and 10 to examine the extent to which language and reading skills are influenced by the same shared genetic and environmental factors (Harlaar et al., 2008). Using this methodology, this study has found that, in addition to the common genetic and shared influences in both language

and reading, there is supportive evidence for a causal role of early language in later reading skills. By comparing different models, the best fit for the data was provided by the model that allowed a direct effect from language to reading, with a vocabulary and a grammar composite explaining about 12% of later reading variance, even after taking into account the genetic and environmental influences that affect both. In other words, even with similar biological bases and home environments, early expressive language differences would still account for significant variance in later reading skills. Moreover, their analysis suggested that it was some shared influences from the environment which facilitate the development of vocabulary and syntax, which in turn facilitate the acquisition of reading skills.

Overall, there is convergent evidence between the evidence reviewed here about typical language, and atypical early language reviewed before (Justice et al., 2009; Scarborough, 1990b) that the earliest expressive vocabulary and syntactic measures are significantly related to later reading skills. The last typical development study in particular illustrates how this relationship is underpinned by environmental and/or genetic influences (Harlaar et al., 2008), and how parents knowledge of their children's progress in expressive language can be tapped for research.

2.2.2.2 Preschool language and reading

Very robust evidence has linked *composite* language measures when the children are 3 and 4 years old with later reading skills when the children reach 8-10 years of age (NICHD Early Child Research Network, 2005; Storch & Whitehurst, 2002). In comparison, very few studies have undertaken the task to identify how specific expressive language sub-skills, such as vocabulary, grammar and discourse, are related to later reading comprehension in typically developing preschool samples. The preschool language studies reviewed here address this issue and highlight once more the need to identify the contributions of individual skills and their own developmental course.

A two-year small longitudinal study (n=39) looked at language production in kindergarten in relation to second-grade reading comprehension, when the children were 7 and 8 years old (Roth et al., 2002). The narrative task requested the preschool child to tell their favourite story, which was later analysed for number of propositions

and number of episodes. In regression analyses, even when the score for episodes added unique variance to first grade reading comprehension after controlling for vocabulary, it did not seem to add unique variance to second-graders' reading comprehension. Propositions did not add unique variance to reading comprehension at either grade. These results suggested that, at least with the two tasks used here, expressive discourse level skills are only temporarily related to reading comprehension at 6/7 years, but not at 7/8 years of age.

Another small scale study (n=32) which also extracted expressive measures from speech samples at 5 years, did not find a strong relationship to 8-year-old's reading comprehension skills from syntactic indicators, but it found that the ability to provide detailed information about a picture and the prolific use of evaluative vocabulary elements did (Griffin et al., 2004). Therefore, in this study it was expressive vocabulary, rather than syntax, that was related to later reading comprehension.

The last longitudinal study reviewed here found associations between narrative production and reading comprehension (Snow et al., 1995; Tabors, Snow et al., 2001). First, Snow *et al.* (1995) found that narrative production score in kindergarten was significantly correlated with one assessment of reading in first grade (*r* =.37, *p* <.01, Reading subtest of the Wide Range Achievement Test, Revised: WRAT-R), but not another one (*r* =.20, n. s., Gray Oral Reading Test: GORT). The narrative production task consisted of allowing the child to observe three slides and then tell a story about them. The composite score comprised structure, elements and syntax. This task was also correlated to reading comprehension at 9-10 (*r* =.47, *p* <.001) and at 12-13 (*r* =.45, *p* <.001) years of age.

Once more, the heterogeneity of tasks makes comparisons between studies rather difficult, as they did not measure the same variables. Moreover, all were American studies, where reading instruction starts at 6 years of age, so the measurements at 8 years for the first two studies (Griffin et al., 2004; Roth et al., 2002) have been taken after less than two years of reading instruction. The study by Snow and colleagues (Snow et al., 1995; Tabors, Snow et al., 2001), which used a composite measure, displayed the strongest results, suggesting that narrative skills are indeed relevant for reading in general, and for reading comprehension in particular as the associations were larger for later reading comprehension skills than for initial word reading in first grade.

Nonetheless, they do not identify how and when each specific sub-skill will predict subsequent reading skills.

2.2.2.3 School-age children's language and reading comprehension

The converging evidence reviewed in the previous two sections described longitudinal connections between early language and later reading skills. The premise for studying the earliest signs of individual differences in language skills is that the earlier deficits are identified, the earlier that interventions can modify developmental trajectories. While acknowledging that change can be most effective at the beginning of development (Vellutino et al., 1996), there are several reasons why school-age children's broad expressive language skills, and more particularly expressive skills from language samples, are relevant to reading skills.

First, expressive skills, such as those from language (or narrative) samples are easily observable in a lay manner by parents and teachers, and in a more systematic way by researchers. If they were related to reading skills, they would be the first clues that could be perceived by those closest to the child.

In addition, there is comparatively far less research-based evidence to support new or existing reading difficulties beyond the first couple years of instruction. Given that once decoding has been mastered and broad language skills become more relevant for reading continuous text (Paris, 2005), and since expressive deficits can have a differential impact on the reading skills in older children with language impairments (Simkin & Conti-Ramsden, 2006), it is reasonable to argue that weak broad expressive skills in general could be differentially related to reading skills in typically developing children.

Finding concurrent expressive associations is important because concurrent receptive deficits may remain hidden (Nation et al., 2004), language sub-skills may not be the focus of instruction and teachers may lack sufficient awareness of language elements themselves (Moats, 1994; Moats & Foorman, 2003). By providing a complete picture of linguistic skills, it would be possible to examine whether more visible expressive deficits can identify potential deficits in children and a comprehensive study could tease

apart whether expressive skills are redundant or indeed provide more information than receptive skills alone.

More specifically, by using language samples to measure expressive skills, a more level-playing field can be obtained where children from minority or non-mainstream communities can still exhibit some degree of complexity that might not be culture bound, which is not the case with standardised tests of expressive language (Gillam et al., 1999).

Still, more serious criticism to the examination of the relationship between school-age language and reading is the argument that preschool language measures (ages 3 and 4), *both* receptive and expressive, are even more predictive of subsequent reading skill than later kindergarten measures at 5 years of age (Scarborough, 2005). Although this is a counterintuitive finding, a possible explanation is possible when considering that the Scarborough meta-analysis stopped at the early primary years, and the developmental changes illustrated in Section 2.1 on receptive language show evidence that some longitudinal correlations grow stronger after the initial decoding years. Similar results for expressive vocabulary have been found in the small sample study by Walker and colleagues (Walker et al., 1994). Therefore, the results obtained by Scarborough (2005) are not unexpected, and they would be hypothesised to change if a longer developmental window had been studied. In addition to developmental changes in reading, it is clear that language development is not linear since it does not follow the same rate of growth throughout its course (Bates et al., 1995).

Furthermore, although it was once considered that basic linguistic maturation was reached by 3;06 (Bates & Goodman, 1999), it has also been argued that even when the basic features of language are already established by preschool, language has a longer developmental course in vocabulary, syntax and discourse organisation (Anderson & Nagy, 1993; Berman & Slobin, 1994; Tomasello, 2000). If indeed language has a longer developmental course, this has implications for how the connections between expressive language and reading are conceptualised, as different linguistic features appropriate for older ages would warrant more attention.

Indeed, the study of atypical language reviewed previously (Catts et al., 2002) found that *concurrent* associations exist between expressive language and reading comprehension that were very close to their receptive counterparts: at 7-8 years of age, the correlations were r: = .56 v .52 for receptive and expressive skills, while at 9-10

years of age, correlations were *r:* = .59 v .55 for receptive and expressive skills. Even when in this particular study of atypical language correlations with reading comprehension were slightly stronger for receptive than for expressive language, abundant evidence reviewed here has shown that many expressive deficits could have at times, a stronger level of prediction than their receptive counterparts (Justice et al., 2009; Scarborough, 1990b; Shapiro et al., 1990), or a complementary level of prediction when added to receptive deficits (Simkin & Conti-Ramsden, 2006). While it would be reasonable to expect receptive deficits to be stronger predictors *overall*, the changing course of developmental language patterns with peaks and valleys (Scarborough, 2010), along with the visibility of expressive language in general - which has lent itself to tap into knowledge from parents (Harlaar et al., 2008; Shapiro et al., 1990) or teachers (Williams, 2006) - make expressive language a strong candidate for linking this modality of language to reading comprehension.

In addition, given that most children would not undergo the one-on-one testing that could uncover mild receptive deficits related to reading comprehension (Nation et al., 2004), finding an expressive measure that gets closer to the kind of language used in natural settings by children would be ideal because these could be more noticeable. In that regard, speech samples have been routinely used as an index of language development (Merritt & Liles, 1989), and are closest to what would be considered an ecologically valid measure of children's language in that it does not penalise non-standard language variations (Hughes et al., 1997), which is particularly important for disadvantaged children or non-mainstream children (Gillam et al., 1999). Moreover, speech samples are considered to be a good alternative to the test-taking format, which would be ideal when identifying the relative contributions of receptive versus expressive language, since our review of receptive language has emphasised how methodological decisions impact the results obtained, particularly so in studies where multiple language variables are considered.

Finally, within speech samples, narratives are a semi-structured option which is likely to elicit decontextualized language (Ricard & Snow, 1990), which is the one most related to literacy in longitudinal studies, as the two previous sections have shown. In addition, narratives lend themselves to training (Cable, 2007) and could be part of the communicative experiences that children need to drive their own language development (Hoff, 2006).

For all these reasons, examining the concurrent relationship between expressive language – using speech samples or more specifically narrative language – deserve consideration in the reading comprehension research field. Indeed, a few studies have added empirical support to this hypothesised association. To our knowledge, five studies of expressive language using narratives have been carried out that *concurrently* evaluated the link between school-age language and reading comprehension *in typically developing children* (Cain, 2003; Cain & Oakhill, 1996; Chen-Wilson, 2005; Klecan-Aker & Caraway, 1997; Snyder & Downey, 1991). Once that it is considered that extracting linguistic indices from speech samples is a very labour intensive task, it is not surprising there are not more of them. Even when these studies are not directly comparable, they seem to provide some empirical support for the hypothesised associations with reading.

Here the focus is on the developmental windows, since we would expect broader expressive skills such as vocabulary, syntax and discourse organisation to play a greater role after the initial decoding acquisition years. The current evidence offers a rather fragmented picture, but does offer some support for reading comprehension's concurrent associations with different expressive sub-skills.

First, the most compelling evidence is offered by studies into the role of discourse organisation. How well a story is structured or remembered seems to be related concurrently to reading from at least 7 years of age. Concurrent associations with reading comprehension have been reported for the ability to include story grammar elements at 9-10 and 11-12 years (Klecan-Aker & Caraway, 1997), and others have found that the proportion of stories retold have the strongest links in the 8-14 age range (Snyder & Downey, 1991).

A more stringent analysis using a comprehension-matched design found that it is the 7- and 8-year-olds' ability to structure a story causally, by using a simple three-level rubric, the one that distinguishes poor comprehenders from same-age controls and from younger comprehension-matched controls (Cain & Oakhill, 1996). These findings were later replicated with another study using the same rubric (Cain, 2003). The comprehension-matched design could be considered, among these methodologies, the one that offers the most stringent test of the relationship between narrative production and reading comprehension because comparing poor comprehenders to younger comprehension-matched readers makes it difficult to explain better skill performance in

the youngest children in terms of a greater print exposure. This methodology makes it easier to disentangle reciprocal relationships. In other words, since old and young readers are at the same comprehension level, any skill differences observed could potentially be related to the younger readers' relatively good comprehension. In this manner, it is more likely that discourse organisation in narrative production plays a causal role in reading if younger normally-achieving comprehenders are better at telling stories than older poor comprehenders, which is what Cain and Oakhill (1996) and Cain (2003) found.

These findings of significant differences in overall narrative structure in a very conservative analysis reflect what Perfetti and colleagues claim is the marker of a good comprehender: striving for coherence (Perfetti et al., 2005). In other words, a good comprehender is guided by an effort to fully integrate the information read. If Cain and Oakhill's discourse-level measure using a simple three-level rubric is useful to separate skilled from poor comprehension at this young age, it could be argued that it might predict comprehension in typically developing readers.

Regarding expressive syntax, the evidence is relatively scarcer, but some evidence for its importance has been found. The complexity of syntactic skills employed in narrative tasks has been reported to be associated with reading comprehension in the 6-9 (Chen-Wilson, 2005) and the 9-10 and 11-12 age ranges (Klecan-Aker & Caraway, 1997). Cohesive ties have also been found to be used more frequently in skilled comprehenders' narratives than in those produced by less skilled comprehenders (Cain, 2003).

Finally, regarding expressive vocabulary, concurrent associations have not, to our knowledge, been documented, even when ample longitudinal evidence reviewed above related composite expressive measures including vocabulary to later reading (Griffin et al., 2004; Harlaar et al., 2008; Shapiro et al., 1990; Tabors, Snow et al., 2001; Walker et al., 1994). Although expressive vocabulary was included in the test battery in one of the four studies (as number of words/ length; Cain & Oakhill, 1996), it did not significantly differentiate poor comprehenders from any of the two control groups.

This review has illustrated preliminary evidence for concurrent links between some expressive language sub-skills in narrative samples and reading comprehension. Altogether, the literature seems to support the idea that expressive language can be concurrently and longitudinally related to reading skills in typical and atypical

development. Moreover, in light of the evidence finding that a combination of expressive and receptive deficits accounts for the greater risk for reading difficulties, these could potentially provide a very useful means to observe the full range of language skills in children, and how they are related to reading comprehension skills.

None of the five studies examining concurrent relationships (Cain, 2003; Cain & Oakhill, 1996; Chen-Wilson, 2005; Klecan-Aker & Caraway, 1997; Snyder & Downey, 1991) sought to compare the expressive versus the receptive language sub-skill counterpart, nor did they examine a reliable measure of expressive vocabulary. Therefore, the question is still open as to whether those arguably more noticeable expressive skills can account for unique variance – on their own – and whether typically developing children with mild deficits in both modalities are the ones with the worst reading outcomes.

This is what the study presented in this thesis aims to address, so that the concurrent contribution of each skill, including vocabulary, can be further specified in order to derive implications for practice. In this way, the study presented here can potentially contribute to the often overlooked role of expressive skills in general, and of narrative skills in particular, in reading comprehension.

Chapter 3

Methodology:

Database study for the selection of indices and pilot study

The extensive literature relating typical and atypical language to later reading skills reviewed in the previous chapter have brought attention to the generally neglected but potentially useful role of expressive skills. While many studies have explored expressive skills, not all have always included receptive skills as part of the study design. With the exception of the Chen-Wilson study (2005), most of the comprehensive studies studying concurrent relationships have addressed atypical language development. Therefore the literature has not yet ascertained whether expressive skills in typically developing children can add unique variance beyond receptive skills in explaining reading comprehension outcomes when examining concurrent associations in school-aged children, or what the specific contribution of each sub-skill to reading comprehension might be.

Investigating the role of expressive skills in typically developing school-aged children is not only of theoretical, but also practical importance. Given that expressive skills are very visible to both parents and teachers, exploring this modality has even greater practical implications since they could potentially provide a window into the full range of linguistic development in children whose reading comprehension skills seem fine during the initial instruction years, possibly making it easier to identify those whose linguistic deficits tend to go unnoticed (Nation et al., 2004). For that reason measuring expressive skills with language samples, with their greater ecological validity, would be likely to result in findings with clear implications for practice (Paul, 2007). This particular developmental window when mild language deficits may be hidden, at 7 and 8 years of age, is particularly important, because language elements might not be the focus of instruction (Moats, 1994).

However, most of the few studies linking expressive language using language samples and reading comprehension in typically developing children cannot be directly compared in light of the methodological differences. The scarce evidence that does exist shows mixed results for the link between school-age specific expressive sub-skills with reading comprehension skills. Even with these methodological differences, the previous review found reading comprehension links with expressive syntax (Chen-Wilson, 2005; Klecan-Aker & Caraway, 1997) and expressive discourse organisation (Cain, 2003; Cain & Oakhill, 1996; Klecan-Aker & Caraway, 1997; Snyder & Downey, 1991). So far, concurrent associations of school-age expressive vocabulary and reading have not been reported to our knowledge: within language sample analysis, a commonly used indicator for expressive vocabulary – number of words – was found unrelated to reading in the Cain and Oakhill (1996) study. It should be noted that although some researchers refer to naming or oral definitions tasks as expressive vocabulary, they are considered in this work as receptive tasks. Overall, the literature reviewed before has not yet specified what the relative contribution of each expressive specific sub-skill is to reading comprehension skills.

However, nowhere are methodological decisions more critical in how results are reported than in measuring language in older children (Scott & Stokes, 1995), so these methodological differences make it difficult to compare results directly. While some studies in the field have used *standardised assessments* to measure expressive language (Simkin & Conti-Ramsden, 2006), others have used *speech/language samples* (Cain, 2003; Cain & Oakhill, 1996; Chen-Wilson, 2005; Klecan-Aker & Caraway, 1997), and yet others have used a *retell paradigm* (Snyder & Downey, 1991). It is reasonable to expect that task characteristics affect measurements directly, and some of these differences have already been illustrated empirically (Cain, 2003; Cain & Oakhill, 1996; Masterson & Kamhi, 1991; Merritt & Liles, 1989).

In sum, the contribution of school-age children's expressive broader skills to reading ability is not yet well established and it is heavily reliant on methodological considerations. Evaluating a comprehensive picture of linguistic skills would allow progress in identifying whether expressive skills are redundant or indeed provide more information than receptive skills alone.

Just as methodological decisions are critical for measuring reading skills (Cutting & Scarborough, 2006), so are the decisions for measuring expressive skills, particularly

for school-aged children. In this chapter, the methodological implications of using different types of methods to obtain expressive measures are discussed in detail in Section 3.1, leading to the selection of the optimal methods for the purpose of linking expressive language and literacy. Then two studies are reported which aimed to add further evidence for the suitability of our methodological choices. First, to address the reliability and sensitivity shortcomings of the indices chosen for typically developing children, a corpus-based study is reported in Section 3.2. Next, these measures were tested in the field in a pilot study in order to examine the feasibility of the elicitation procedures; to provide some preliminary evidence that the indices identified as optimal in the database study were, in fact, related to reading comprehension, at least in a small sample; and to compare the size of the expressive associations relative to the receptive ones. This pilot study is reported in Section 3.3.

3.1 Measurement issues in expressive language

The scarcity and diversity of studies linking school-age expressive language and reading makes generalisations difficult. For that reason, special attention was given to the selection of expressive measures in this section. The first level of categorization among expressive tasks is concerned with either using standardised tests or eliciting speech/language samples, which will be referred to here as language samples since in school-age children the focus is less on the reception/production of phonemes and more in the wider language skills. After discussing the relative advantages of both kinds of measurements, the specific methodology considered appropriate for this study is further refined in each subsequent section.

3.1.1 Standardised testing v language samples

Expressive skills can be operationalized in many different ways, but the main distinction is the one between standardised assessments and language samples. There is evidence linking both kinds of expressive measures to reading, as illustrated in Section 2.2.2.

Evidently, standardised assessments have by definition, good psychometric properties, while measurements from language samples were likely to have greater reliability issues in the collection, transcription or analysis. Without a deeper examination, standardised tests would seem the first choice for evaluating expressive language, but they do suffer from a significant disadvantage for the purpose of this study.

While standardised tests of expressive language are designed to be reliable, objective, specialised and have norms to compare children to a standard population, they still place higher processing demands on the testee than just expressing linguistic skill (Scott & Stokes, 1995). As evidenced from the review of receptive language, these test characteristics do have a bearing on which results are reported. Standardised tests focus on a narrow range of items, can be confounded by the individual's test-taking skills and do not assess the multidimensional aspects of language (Shipley & McAffe, 2004). These in turn are precisely the advantages of language samples, since they make processing demands that are more similar to those encountered by children in their own environments; therefore they are more likely to elicit natural language that allows for multiple indices of language to be assessed at the same time. Furthermore, some subtle language weaknesses only appear in language samples which some standardised testing does not uncover. For example, late-talkers who were later examined at 5 years of age, were found to have caught up with their peers in many of the standardised assessments, but were still weaker in the syntactic complexity, use of cohesive devices, narrative discourse ability and the degree to which they were able to tell a story without support than controls (Girolameto, Wiigs, Smyth, Weitzman, & Pearce, 2001).

One of the main reasons narrative assessment has become so prevalent in the work of speech and language pathologists is because standardised assessments are not very useful for children whose language weaknesses are derived more from lack of experience or cultural differences, particularly for children from non-mainstream communities (Gillam et al., 1999). The possibilities offered by language and narrative assessment are that they may be culturally un-biased and could distinguish language difference from language disorder. Although the series of investigations reported here did not intend to focus on atypical language, they aimed to uncover those mild weaknesses that could ultimately be related to reading comprehension.

Expressive language elicited in this way could prove particularly useful when disentangling socioeconomic influences. The relationship between both expressive and

receptive language skills and socioeconomic status is already documented in the literature (Farkas & Beron, 2004; Hoff, 2003). Moreover, some evidence has supported the pathway from the family's socioeconomic status having an effect on language, which in turn has an effect on reading skills (Durham et al., 2007). It might be possible, in theory at least, that children with linguistic skills in their home environments but with vocabularies that do not correspond to vocabulary used at school could perform equally well in a creative task such as language/narrative assessment than in a bound task such as a standardised test.

In addition to concerns over language difference versus disorders, there is the issue of whether the enriched school language environments can provide the language model to provide the child with the literate linguistic forms that they might not have received from school. It is reasonable to expect that a stronger emphasis on developing children's oral language skills should have a direct effect on their reading comprehension skills. Teachers could provide a new literate language model than children could potentially be useful for their reading comprehension development.

Some indirect evidence has shown that specific interactive teacher strategies can influence academic language growth in older children. In Australia, teacher talk practices were identified which allowed students to learn technical language in the context of a high-school history lesson (Sharpe, 2008), and science terms in older primary children learning a second language (Gibbons, 1998). It should be noted, however, that these interactive strategy studies did not address the question of how the quality of a teacher's daily language influenced the rate of pupils' language development. In any case, by focusing on the assessment of narrative abilities, the aim was to obtain a glimpse of the kinds of literate language the children have already acquired and know well enough to be able to use them for communication.

Of course, language samples have their own set of disadvantages, of a theoretical and practical nature: on the theoretical side, low reliability in any part of the process, be it collection, transcription or analyses can have a detrimental effect on the validity of the variable or construct being examined; on the practical side, collecting data is expensive in terms of resources of time and skill, since recordings have to be transcribed and analysed after the language sample is collected (Scott & Stokes, 1995).

Addressing the reliability of using language samples has been minimised, to a certain extent, with the use of newer more sophisticated versions of computerised language

analysis software. Computers have now made it possible to surmount many of the problems that used to be associated with the analysis of language samples. As analyses have moved from native-speaker intuition to computerised text analyses, reliability and accuracy issues have been significantly reduced (Adolphs, 2006).

Indeed, there have been extraordinary advances in the ways computers have been used to study spoken language from the fields of natural language processing and corpus linguistics. Tools that were originally created to describe how language production emerges in very young children (MacWhinney, 1996) are now being used to characterise developmental change in older children. Different computer software programs currently exist to obtain lexical and syntactic indices (MacWhinney, 2000; Miller et al., 2005), making transcription more reliable and analyses more automatic (Heilmann et al., 2008; Long & Channell, 2001). Of course, language analysis programs have limitations if the aim is to carry out specific analyses, e.g. grammatical errors (Scott & Windsor, 2000), but they are able nonetheless to provide very powerful tools for global measures. The use of computers has also addressed the second practical issue, since they have made the time required to obtain analysis from a large corpus of samples to be virtually insignificant, though for the most part, these still need to be transcribed.

A final strength of language samples for the purpose of this study comes from its implications. In contrast to standardised assessments, where implications for practice are difficult to be drawn from, language samples lend themselves to identifying areas for intervention (Paul, 2007). Although in research the aim is to study language in a systematic way, a narrative sample will be closer to the kind of discourse language that teachers can observe from their pupils without specialised training. Language samples are routinely used in the areas of speech and language therapy (Hughes et al., 1997), but not in the educational context, where these tools have not yet been fully exploited.

Even when language samples still entail time- and labour-consuming data collection and transcription, the recent gains in reliability, in speed and versatility of analyses, together with the possibility of using a measurement which gets away from the test-taking format and arguably requires the least amount of meta-linguistic skill, makes them more promising candidates for investigating expressive language from speech samples in relation to reading skills. Moreover, this last characteristic becomes especially relevant

when using a design that already includes standardised assessments for measuring receptive skills.

3.1.2 Language samples: genres

There are mainly three kinds of genres that can be elicited to obtain language samples for children: conversations, narratives and expository discourse. Even when conversational samples are good indicators of a child's early language development, it has been argued that they do not reflect an older child's true syntactic growth (Nippold, 1988; Scott, 1988), and are unlikely to be related to reading performance as conversation uses contextualised language (Snow, 1991) as opposed to the decontextualized language that has been associated with literacy (Ricard & Snow, 1990). In discourse, a separation between speaker and listener is assumed, so that meaning must be built into the words (Nelson, 1988). Therefore, conversational samples might not be ideal for measuring school-age language.

Narrative and expository genres, on the other hand, require from the children not only to be able to use vocabulary and grammar and to establish some degree of linguistic cohesion, but also to be able to maintain discourse congruence (Hickmann, 2004). In addition, children are likely to be familiar with these two genres, as they also appear when children encounter continuous texts.

Narratives are a form of discourse where a sequence of events, real or imagined, are shared orally or in a written format, while expository discourse is the use of language to convey information (Bliss, 2002). Although expository discourse is ideal for eliciting complex syntax, it has been used for children from 10 years of age onwards in the literature (Nippold, Mansfield, & Billow, 2007; Nippold, Mansfield, Billow, & Tomblin, 2008; Scott & Windsor, 2000) with a few exceptions (Nippold, Hesketh, Duthie, & Mansfield, 2005). Comparatively, it seems that narratives are likely to produce more linguistic output than expositions in younger school-age children: in one study where children were asked to produce both kinds of genres (Scott & Windsor, 2000), spoken narratives were longer than spoken expositions in children between 9;10 and 12;11 (years; months), which is an older age range than the developmental window considered in this study of reading comprehension at 7/8 years of age. In this study,

children with typical language produced on average 573 tokens in narratives, but only 341 in expositions.

Moreover, narratives have several characteristics that make them an appropriate choice for measuring language in the developmental window we are focused on. First, producing narratives requires from the children to be able to convey perspectives, express whether the information is reality or fiction, and to construct characters (Pan & Snow, 1999). Second, this is the kind of discourse that children are first exposed to, usually in their home environments (Beals, 2001), and later on in school settings as well (Crais & Lorch, 1994), so they are considered to have ecological validity. Third, narrative reasoning increases with age: Paris and Paris (2003) were able to show that narrative skills based on a picture book task (comprehension and production) tend to increase with age and with reading skills, concurrently and longitudinally, from 5 to 7/8 years of age. Fourth, narratives lend themselves to training (Cable, 2007) and children often enjoy the opportunity to express themselves. Fifth, narratives allow for a variety of higher-level discourse measures to be obtained (Peterson & McCabe, 1983). Last, narratives have been widely used to document later language development up to adulthood in several languages (Berman & Slobin, 1994), although it must be noted that such work usually focused on in-depth analysis of specific linguistic forms, not on the general linguistic markers of vocabulary and syntax that are relatively easily obtainable using automated language analysis.

For all these reasons, narrative samples were elicited in this study. Since task characteristics are usually reflected in the results in narrative samples, the following section evaluates the main characteristics of different elicitation procedures for the selection of the most appropriate task for this study.

3.1.3 Elicitation methods for narratives

Having chosen a narrative sample, there are still a variety of tasks, stimuli and procedures available for elicitation. Task characteristics have a bearing on the language data obtained (Masterson & Kamhi, 1991), particularly when the phenomenon studied is its relationship to literacy (Cain, 2003; Cain & Oakhill, 1996; Chen-Wilson, 2005). Therefore, it is also important to weigh the advantages and disadvantages of each specific task, in order to determine the most appropriate one for this study.

Spoken language has been conceptualised in a continuum, with oral language at one end and literate language at the other. An oral language style could be described as the kind of language children learn when learning to talk; meanwhile, a literate language style would be described as the kind of language used to reflect about previous experiences or predict future ones (Westby, 1985). One of the most relevant differences between these two styles has to do with the level of contextualisation, with the oral style being very contextualised in that it relies on non-verbal cues, while the literate language style being highly decontextualized with all the cues needed for comprehension already contained in the verbal message (Paul, 2007). With this distinction in mind, the aim of elicitation in the context of this study would be to elicit a narrative sample which is the closest to the literate end of the spoken language spectrum. This aim guided the selection of elicitation methods.

There are three main types of elicited narratives: script, personal and fictional (Hughes et al., 1997). Since the aim is to maximise the elicitation of literate language, script narratives, which account routine events, might not be optimal. Personal narratives seem to offer several advantages. For example, the use of personal narratives has been well documented, especially for measurements at the level of discourse organisation (Peterson & McCabe, 1983), and they tend to integrate particularly interesting pragmatic elements. However, fictional stories have also been documented and they offer their own set of advantages. A comparison of personal and fictional narratives from a picture book in children from 4 to 8 years of age, showed that picture narratives were more complex in terms of complete episodes and using multiple episodes, as well as containing more action sequences than personal stories (Allen, Kertoy, Sherblom, & Petit, 1994). It has also been found that, when comparing fictional and personal narratives in the first (6-7 years) and fourth grades (9-10 years), younger children and children from disadvantaged backgrounds were less capable of producing fictional narratives, but not when producing personal narratives (Shiro, 2003). Therefore, fictional narratives seem to be more challenging and elicit a wider range of skills, from low to high, than personal narratives, which even young children can produce (Peterson & McCabe, 1983). For these reasons, fictional narratives will be used in this study.

The next choice concerns using a retell or a self-generated story. Although there is evidence of retell formats being linked to reading (Snyder & Downey, 1991), the retell where a story is verbally presented for a child to repeat would give the participant linguistic input that would influence how they construct their own story. If the aim is to

obtain a narrative sample which is most reflective of the pupil's true linguistic repertoire, then having no verbal input from the examiner should aid in that purpose.

The last methodological elicitation decision concerns whether a verbal prompt or a controlled stimulus such as a single picture or sequence or pictures, a picture book or a silent film should be used. Cain and Oakhill (1996) and Cain (2003) noted that a verbal prompt was the most difficult condition and gave the greater range of responses, while the stories elicited from a given picture sequence were better structured. In particular, these results seem relevant because both studies investigated possible links between narrative skills and reading comprehension. Regarding the comparison between prompts and stimulus, the narratives from poor comprehenders differed significantly from the younger comprehension-matched controls' narratives only in the prompt condition, but not in the picture sequence condition. While acknowledging that structural elements of a narrative might be more uniform if supported with a stimulus, thus giving a more narrow range of responses in terms of global structure, the main focus of the present study was to elicit the greatest amount of literate language in typically developing children, in order to distinguish individual differences in linguistic analyses. Both verbal prompt and even single-picture elicitations tend to elicit shorter narratives, while summaries of 19-minute videos tend to elicit the longest samples (Scott & Windsor, 2000). In between, a picture book with some substantial change of events in its plot should elicit enough linguistic data for this study, as evidenced by a previous investigation (Chen-Wilson, 2005), balancing the need for productivity of individual linguistic output with the need for efficiency in collecting and transcribing a large amount of narratives.

Given the ample documented evidence in the literature of narrative development using picture books (Berman & Slobin, 1994; MacWhinney, 1996), a picture book was selected for this study to elicit language samples from typically developing children.

In addition, providing a semi-structured controlled stimulus would make it easier to reduce variability from differences coming from their individual experiences, and to establish a level playing field from which linguistic differences can be identified. Moreover, it is important that the pupil has constant access to the story while narrating to be able to make the task a linguistic one and not a memory task.

3.1.4 Linguistic indices in narrative samples

Once the elicitation procedure was selected, there was a need to identify the most useful linguistic indices. Although many indices exist, here the focus is on those linguistic categories whose importance for reading comprehension has been extensively documented in the previous literature review in Chapter 2: vocabulary, syntax and discourse. For vocabulary and syntax, the review was centred on those automated analyses which could be obtained from computer software. The discourse measures reviewed were all manually obtained.

3.1.4.1 Lexical indices

Vocabulary measures have been characterised by a wide range of lexical measures that are not equivalent to each other, even if they are related. Despite more research being carried out, few studies focus on establishing developmental patterns in typically developing children, which complicates interpretation of results using different indices in different populations. Given the evidence that different receptive vocabulary measures are differentially related to literacy (Ouellette, 2006), it would be reasonable to expect these different indices of expressive vocabulary from speech samples to show different degrees of association with literacy as well, with more complex measures having stronger relationships with reading comprehension. This section describes these different vocabulary measures, and those few studies addressing vocabulary growth in typically developing children.

Studies in the speech and language literature most commonly use *tokens*, i.e. total number of words, and *types*, i.e. number of different words. Tokens have been reported to have inadequate reliability, while types have been found to be highly reliable in preschool children (Gavin & Giles, 1996). Of their potential relation to literacy, types in emerging language have been used in studies finding significant links between early expressive vocabulary and literacy (Walker et al., 1994). Although the types measure is considered reliable, it is still highly correlated with language sample size as measured by tokens (Justice et al., 2006).

Another measure which is intended to be less influenced by sample size is a measure of lexical diversity, known as Type-Token Ratio (TTR; Templin, 1957). TTR is a measure

of the proportion of the types of words over the total number of words. It aims to account for the frequency of such words in a single speech act. Being a proportion, it represents an improvement over tokens and types, but as further studies showed, TTR scores still tend to be a function of the total amount of words (Richards, 1987).

Building upon TTR, a new measure of lexical diversity was created which sought to measure lexical diversity while taking into account each full language sample (McKee, Malvern, & Richards, 2000). The Parameter "D", more commonly known by the program used to obtain it, VOCD, is calculated from a mathematical model that takes into consideration the number of words in a particular language sample.
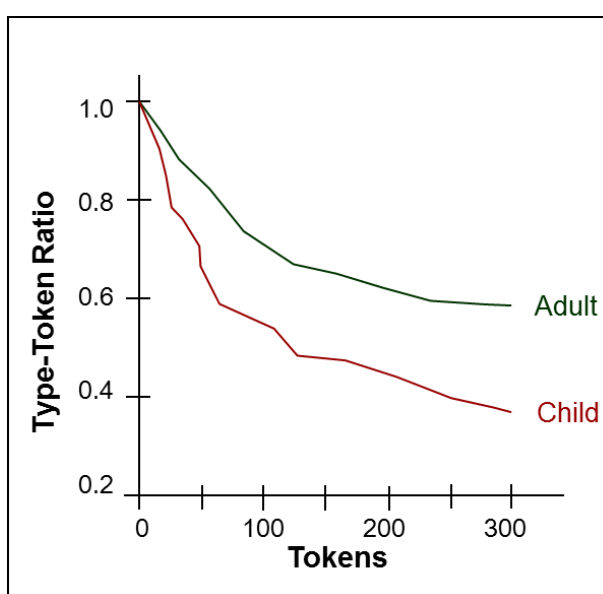
Fig. 3.1 TTR by tokens for speech of a two-year-old and academic writing for an adult (redrawn from Malvern, Richards, Chipere, & Duran, 2004, figure 2.1, p. 23)

Although TTR is considered unreliable, its dependency on sample sizes deserves further description in order to understand VOCD, because the latter is built precisely upon the relationship between this ratio and the amount of tokens in a given sample. Fig. 3.1 illustrates how the dependency on tokens occurs: as a language sample becomes longer, the TTR score tends to become stable because as more tokens are produced, then it gets progressively more difficult for new types to be introduced (Malvern et al., 2004; McKee et al., 2000; Richards, 1987). In other words, it is relatively easy to introduce

new types at the beginning when there are few tokens, making it more likely to produce a high TTR at the start, than at the end, where many tokens already exist, producing a comparatively low TTR. Therefore, not only are these two indices related, but more importantly, the relationship is not a linear one.

Empirically, it has been shown that a language sample's size had more impact on TTR than individual differences in typically-developing children. Studying American children of 9-12 year of age, researchers segmented 600-token language samples into 50-, 100- and 200-token sizes, and then compared TTR and other mathematical variants at the different sample sizes (Hess & Haug, 1989). Remarkably, individual variation was not detected when sample sizes were the same, but variations were found when sample sizes changed. Therefore, if TTR remains fairly stable after a certain amount of tokens, then TTRs are unlikely to show a developmental pattern of growth, and this issue is virtually independent of skill or age. This dependency on tokens might also be the reason why TTR has not been shown to be developmentally sensitive for school-age children (Pearson, 2002) and it also may be the reason why was not found to be related to reading comprehension (Chen-Wilson, 2005).

In search for a valid measure of lexical diversity, there have been two main approaches. The first one has been to choose a cut-off point to make all transcripts the same size, or the size of the smallest sample of a set, and use a TTR based on the reduced samples. The disadvantage to this method is that some valuable information is wasted, and also that the decision of which part of the sample to discard, i.e. the beginning, end, etc. is still an arbitrary one, and portions of texts may differ in their diversity within the same sample (McKee et al., 2000).

A second approach has been to develop mathematical transformations of TTR which aim to avoid its dependency on sample sizes, such as taking the square root or the log of the amount of tokens. The most common of these transformations, including the Guiraud, Herdan and Uber indices, are described and exemplified in Vermeer (2000), but none of these emerged as a clear preferred option in terms of validity, or empirical evidence of developmental sensitivity.

In a similar vein, VOCD is also built upon TTR but goes further than just doing a transformation. It tries to overcome the sample size dependency issue by actually using this dependency to derive its measure. VOCD works by taking a random sample of tokens, and observing how much TTR changes as the number of tokens increases in that

specific sample (McKee et al., 2000). The program then compares the observed TTR-tokens curve from the random sample to that of a family of computer-modelled TTR-tokens curves. As illustrated in Fig. 3.2, the higher the observed TTR-tokens curve is compared to the family of computer-modelled TTR-tokens curves, the larger the VOCD score would be. In this manner, it does not matter – to a certain extent – how different in size language samples are, and therefore a more independent measure of lexical diversity is arguably obtained, which would in turn make it the most appropriate vocabulary measure in linguistic analyses.
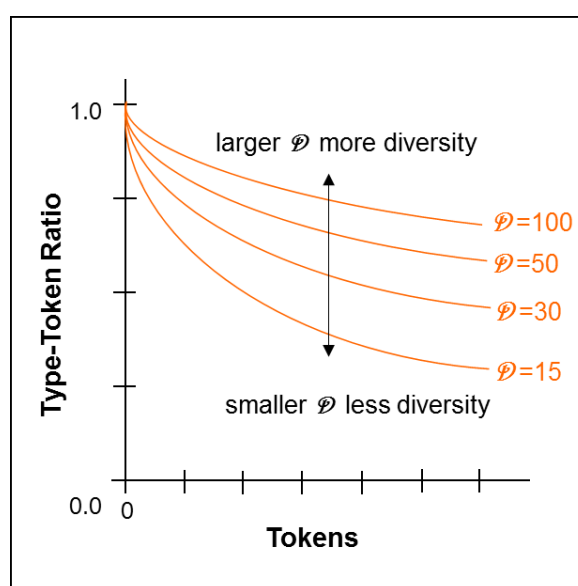
Fig. 3.2 Ideal TTR-versus-token curves showing increasing diversity with increasing scores for Parameter D (redrawn from Malvern et al., 2004, figure 3.5 p. 52).

When comparing the performance of VOCD against the other mathematical transformations of TTR, a modified version of VOCD with full, instead of random sampling was indeed the most accurate of all the measures, describing 98.19% of the texts at this range from 140 to 285 tokens in written English narratives by Finnish, Swedish and native English speakers (Jarvis, 2002). It should be noted though, that one of the mathematical variations of TTR known as the U index came in closely behind it, describing 97.83% of the texts.

Still, these were written narratives, and lexical quantitative measures are generally larger for written than for spoken samples (Stromqvist et al., 2002). Additionally, the procedure involved full, instead of random sampling. However, based on a large corpus with several spoken and written genres of up to 2000 tokens, McCarthy and Jarvis (2007) found that random sampling, the default in VOCD, outperforms full sampling. Although VOCD is still affected by sample sizes when applied to large texts (correlation with tokens $r = 0.22$), this study found it behaves in a reasonably stable way within certain ranges, from 100 to 400 tokens (Table 10, p. 482, McCarthy & Jarvis, 2007).

Empirical evidence for VOCD's developmental sensitivity exists for early childhood speech (Malvern et al., 2004), and for school-aged children's writing in English (10-14 years of age in Jarvis, 2002; 7-14 years of age in Malvern et al., 2004). Since written narratives tend to display higher lexical measures than their spoken counterparts, it is still debatable whether VOCD can detect differences in the spoken narratives of the 7-8 age range that is the interest of this work.

The only empirical evidence of developmental analysis using VOCD in spoken samples by school-aged children, to the best of our knowledge, was in Swedish (Stromqvist et al., 2002). In this study, VOCD displayed slight differences in spoken narratives/expositions between 10- and 13-year-olds, but these differences were not significant in post hoc comparisons, while the VOCD scores between the 13- and the 17-year-olds were significantly different.

On the basis of the literature, it seems that VOCD could be helpful in discerning performance in lexical diversity between ages and between children, and in turn, having a more valid lexical index could allow for a more robust test of a relationship between expressive vocabulary from language samples and literacy.

However, the empirical evidence for the developmental sensitivity for VOCD that exists to the best of our knowledge does not include English speech in the 7-8 age range which is the developmental window this study is interested in for evaluating the links between narrative language and literacy. Moreover, as VOCD is not included in one of the most ubiquitous commercially available software programs, SALT (Miller, 2008), it is therefore absent from studies evaluating a comprehensive set of measures, like the previous ones mentioned above (Justice et al., 2006; Scott & Windsor, 2000), and from the literature linking expressive language from speech samples and reading.

For the purpose of this project, it was deemed necessary to examine how VOCD behaves in a developmental window that includes this age range, in spoken narratives in English, by means of a picture-book elicitation procedure that provides enough linguistic data for its proper application, and which allows drawing comparisons to previous studies.

3.1.4.2 Syntactic indices

To measure grammatical development from oral language samples, there is a very widely-used quantitative measure for preschool children. The average number of morphemes per utterance or, more formally, the **Mean Length of Utterance** (MLU) has been found to vary consistently with age in young children, and is useful in the identification of normal language development (Brown, 1973; de Villiers & de Villiers, 1973; Wells, 1985). To clarify, an *utterance* is an act of speech bound by silence or a change of speaker (Crystal, 2008), so in conversational language samples, an utterance has been identified with a turn. After Brown's (1973) detailed observations of the language development of a few children, more empirical evidence supports the notion that MLU varies consistently with age in young children ($r=.88$, between 1;05 and 4;11 years) with normal language development (Miller & Chapman, 1981).

However, the utility of MLU beyond the early years has been contested (Klee & Fitzgerald, 1985; Scarborough et al., 1991). Brown himself had already warned against the use of MLU beyond 4.0 morphemes because, he argued, utterance growth would then be related to the nature of the interaction with the adult in a conversational setting (Brown, 1973). It is not surprising then, that the application of MLU in school-aged children is less common (Hughes et al., 1997).

It has been described previously how conversational samples are not optimal for school-aged children, so the utterance itself cannot be used for narrative samples. When trying to adapt the MLU for school-aged children, other measures have been used such as *clause* length, *sentence* length, subordination indices and multi-structure indices (Scott, 1988).

A very common metric for segmenting spoken narratives has been the *clause* (Berman & Slobin, 1994), which has been defined as a unit containing a unified predicate, in the

sense that it expresses a single situation or event. This unit, however, was not intended to be used as an alternative to MLU; most of the Berman and Slobin narrative analyses had been focused either at the level of discourse or at the level of very specific linguistic features. Nonetheless, using clauses Chen-Wilson (2005) was able to document associations between the Mean Length of Clauses (MLC) and reading comprehension in 6- through 8-year-olds. Therefore, it might be possible that clauses could still be useful up to age 8, if they are not yet differentiated from other multi-clause measures.

A better option perhaps is the sentence, which is the largest structural unit in terms of grammar (Crystal, 2008). The sentence has been commonly used for segmenting school-age samples (Scott & Stokes, 1995). Some reviews have evaluated the use of sentences (Kemper, Rice, & Chen, 1995) with mixed evidence for its correlations with age in children between 5 and 10 years of age: one study based on a prompt found non-significant associations with age (Sutton-Smith corpus, $r = .32$, n.s.), while a second study based on a silent film found significant ones (Hicks corpus, $r = .34$, $p < .05$). Evidently, the magnitude of these correlations is quite similar, so the difference in statistical significance is due to the Hicks corpus being larger, n=38 than the Sutton-Smith corpus, n=24.

However, an analysis of differences including the sentence and other syntactic measures in the same study (Kemper et al., 1995) found that there is little developmental change in any of these measures after the six years of age, and almost none after eight years of age, when comparing adjacent groups using ANOVA. However, there are two reasons to contest this conclusion: first, it has been argued that after the early school years, growth in school-age children's language is gradual so linguistic development should be analysed by comparing between non-adjacent groups (Nippold, 1988); second, there is some evidence that individual differences in syntactic growth after the age of four could sometimes be greater within the same age group than the differences between the averages of adjacent age groups (Lee, 1974: see graph in p. 16). This has implications for identifying statistically significant differences that compare within-group variance to between-group variance.

A specific and frequent problem in transcribing children's narratives occurs with clausal chaining, that is when 'and' is overused, which brings the challenge of trying to distinguish when a conjunction is being used appropriately or whether it is being used only as a linguistic crutch; this issue is not clearly addressed with the sentence.

Precisely to address this issue, the T-Unit (Hunt, 1970) was developed originally for written samples, and it has been more commonly applied than sentences in measuring grammatical development from language samples (Hughes et al., 1997). A T-Unit is formed by one main clause and any subordinate clauses; compound sentences are treated as separate units. For example, the sentence "the child went back and the mum was preparing dinner", would be considered as *one* single sentence, but it would be segmented into *two* T-Units. By using T-Units, the issue of whether 'and' was used meaningfully or unintentionally becomes irrelevant because in both cases they will be kept in separate units. Of their relation to literacy, a variant of T-Units, clauses per T-Unit has been used in one of the few studies linking expressive language from narrative samples and reading in 9- to 12-year-olds (Klecan-Aker & Caraway, 1997), while T-Units is yet to be linked to reading comprehension.

There is some evidence that adults tend to use longer T-Units than children (Verhoven et al., 2002). Studies describing more detailed developmental windows using T-Units are rare, but a few do exist. The first study to use T-Units examined children in three age groups, 8/9, 10/11 and 12/13 using an 8-minute film, and found T-Units to increase with age (O'Donnell, Griffin, & Norris, 1967). Another study found differences in T-Unit length between children of 6-7 and 8-9 years of age (Klecan-Aker & Lopez, 1985); yet another found significant differences between children of 12-13 and 15-16 years of age (Klecan-Aker & Hedrick, 1985). It should be noted however, that the ***Mean Length of T-Units in words*** (***MLT-w***) was similar for the 8-9-year-olds and the 12-13-year-olds at 9 words per T-Unit, but the elicitation procedures were different, with a film used for the first study with the children, and a prompt used for the second one with the adolescents (Klecan-Aker & Hedrick, 1985; Klecan-Aker & Lopez, 1985).

Another study with a large and representative sample analysed different quantitative syntactic measures in 250 children from 5-12 years of age (Justice et al., 2006). Although this research study did not intend to compare between different indices, but to combine several of these measures into one single linguistic "Index of Narrative Microstructure", it showed that, between several grammatical variables from speech samples elicited for standardisation of the Test of Narrative Language (Gillam & Pearson, 2004), the ones more strongly and significantly correlated with age were Mean Length of T-Units in Words (MLT-w; $r = 0.27$, $p < 0.01$, two-tailed) and the Proportion of Complex T-Units (PROPCOMPLEX; $r = 0.30$, $p < 0.01$, two-tailed). From these results, proportion of complex T-Units seems slightly more attractive as a measure of

grammatical development. However, a second analysis in Justice and colleagues' study (2006) indicated MLT-W to be a more valid indicator of grammatical complexity, rather than productivity: by using a factor analysis, MLT-W loaded more strongly with finer-grained manual syntactic analyses.

The disadvantage of this study, as described previously, is that it contains very short narratives. Independently of the developmental sensitivity of MLT-w, an additional methodological issue which has been explored in preschool children (Gavin & Giles, 1996), but not in school-age children is the stability of this measure at different sample sizes. This issue has been addressed extensively for lexical measures, as described here, and touched upon for the syntactic measures in preschoolers using sentences in the Gavin and Giles (1996) study.

The main issue with the T-Unit is that it may be too coarse a measure to identify more subtle grammatical changes. In that sense, more fine-grained measures like multi-structure indices such as Developmental Sentence Score (DSS; Lee, 1974) or Index of Productive Syntax (IPSyn; Scarborough, 1990a), could also be potentially useful. There are some drawbacks however to the use of these measures. For example, IPSyn needs more than 50 utterances from each speech sample, which goes beyond the tasks used in studies using a wide range of measures in a wide developmental window. The Justice and colleagues study (Justice et al., 2006), averaged 11.3 and 15.8 T-Units in ages 7 and 8, respectively. On the other end, the narratives generated from 19-minute videos in another study (Scott & Windsor, 2000), did reach a mean of 55.8 T-Units, but this kind of task is considered excessively long for the purpose of the associations between language and reading comprehension. Moreover, some of the grammatical forms required for the measurement of DSS and IPSyn, such as questions and negations (Lee, 1974; Scarborough, 1990a) are not usually contained in narratives.

In sum, there is published evidence that MLU, when using T-Units, can display some growth when comparing non-adjacent age groups (Nippold, 1988), whether using films, verbal prompts or a single picture for elicitation, which make it the most feasible and desirable option for the purpose of linking expressive narrative language and literacy. However, major gaps still exist around this measure. First, it is not clear whether T-Units are indeed better than the alternatives. Second, great discrepancies exist in the MLT-w values obtained with different elicitation procedures. Finally, it is not yet

established how T-Units are affected by sample sizes. These issues will be examined in some detail in the study of narrative language reported in section 3.2.

3.1.4.3 Discourse-level measures from narrative samples

Here, a ***discourse-level measure*** will refer to a measure of the ability of a child to organise a discourse's elements into a whole. Often labelled as coherence, this discourse organisation construct has been measured using mainly three methodologies: a) ***narrative stages*** (Applebee, 1978), focusing on the relationships among the events, and between the events and a common theme; b) ***story structure***, measured either as episode structure (Stein & Glenn, 1979) containing a set elements universal to all stories (also known as story grammars) or measured as high-points analysis (Labov & Waletzky, 1967); and finally c) using ***information checklists*** or propositional lists.

Information or propositional checklists used in standardised tests like the Expression Reception and Recall Narrative Instrument (ERRNI; Bishop, 2004) seem a good option since they are more likely to be reliable, and norms exist for older school-aged children. However, this standardised measure lacks an index of the overall degree of coherence or the causality that ties the story together, or a hierarchy of which propositions are essential to the story and which ones are secondary. Meanwhile, narrative stages seems to be best suited for evaluating younger children's narratives (Liles, 1993).

Story structure has been measured using many methodologies, but two have been applied the most: ***episode structure*** and ***high-point analysis***. Moreover, they have been used across a larger developmental window including school-age children (Peterson & McCabe, 1983; Stein & Glenn, 1979). Developed by Stein and Glenn (1979), episode structure might be the most often used way of measuring the macrostructure. In their initial study, these authors sought to identify the underlying schema or story structure in children's narratives, and found the following to be the main elements: setting, initiating event, internal response, internal plan, attempt, direct consequence, reaction. This methodology has been used in the studies of narratives of monolingual children (Peterson & McCabe, 1983), bilingual children (Munoz, Gillam, Pena, & Gulley-Faehnle, 2003) and in one study linking oral narratives and reading (Merritt & Liles, 1987). This could probably be the most widely used methodology for measuring macrostructure that gives an account of how elements are interrelated.

High-point analysis, on the other hand, is composed of the following elements: a sequence of events, which leads to a crisis or 'high point', a personal evaluation of such crisis, and a resolution (Labov, 1972). Proposed by Labov and Waletzky (1967), high-point analysis represented a departure from cognitive-based organisation systems, which recognised the pragmatic elements of a story. Based more on a sociolinguistic rather than psycholinguistic framework, these authors recognised that personal narratives are shared to fulfil social functions and therefore, possess an innate subjectivity which is meaningful in itself. They represented the structure of a story as anchored around emotional high points. These high points are relevant for the story teller, and the sequence of events is interrupted to elaborate on such importance. Having established this high point in the story, the story teller resumes describing the events with a resolution. In this methodology, the narrator's personal perspective, which Labov terms evaluation, is a key functional element to analyse narratives. With its emphasis on personal evaluation, it is not surprising to notice that high-point analysis has been mostly used for the analysis of personal narratives (McCabe, Bliss, Barra, & Bennett, 2008; Peterson & McCabe, 1983).

As discourse measures have been used with older children, they are not examined empirically in the database study, but are included in the pilot study relating language and reading comprehension to give a complete picture of expressive skills using narratives. Given that the discourse-level measure was the only one carried out manually, a relatively efficient procedure was sought. For the purpose of a pilot study, a rubric for global narrative was considered to be simple enough to be reliable and efficient, yet descriptive enough to include story elements. A global narrative score used for studying the link between writing and reading in 10-year-old poor comprehenders (Cragg & Nation, 2006), was adapted for the pilot study. This global narrative score from written narratives differentiated poor comprehenders from controls, so the question was still open as to whether this relationship could also be observed in typical readers and using spoken narratives. A six-point rubric was constructed to obtain a discourse-level measure, based on the presence or absence of three elements: problem, attempt (to resolve the problem) and outcome, with a subtle change intended to capture the child's ability to generate inferences. Within each element, two points were awarded: one point was awarded if the essential element as it appeared in the picture sequence was verbalised (fact), and one point was awarded if a conclusion was drawn about such an element (inference). In this manner, the discourse-level measure could

capture whether children were able to generate inferences beyond the facts established in the drawings.

Although this global structure measure does not strictly follow the elements in the episode structure, nor the high-point analysis, it is still loosely based on both measures of story elements, in that an initial problem is resolved.

A similar, but simpler three-point rubric (Cain, 2003; Cain & Oakhill, 1996) has been used to characterise causally-linked narratives. Its relevance comes from the findings that this simple rubric was able to differentiate poor comprehenders from younger comprehension-matched controls, suggesting that it was the better structure of the younger readers what aided in their comprehension of the text, since they were unlikely to have benefited from more print exposure.

Given that a simple three-point rubric was related to the good comprehension skills in these younger children, it could be argued that the wider six-point rubric used in Cragg and Nation's (2006) writing study could offer more variability in terms of finding individual differences in performance, which in turn could be related to reading comprehension skills in typically developing readers.

Since discourse measures have already been used extensively for older, school-age children, and some of these measures have even been found to be linked to reading comprehension skills (Cain, 2003; Cain & Oakhill, 1996), a discourse narrative measure will not be examined in secondary data study reported in the next section, which will focus exclusively on the more contentious lexical and syntactic narrative indices.

3.2 Database study: The search for the ideal lexical and syntactic indices

Findings that expressive measures at a younger preschool age, 3-4 years, tend to be equal or stronger predictors of literacy than the measures collected later at age 5 (Scarborough, 2005), highlight not only the developmental issues of measuring reading beyond the initial decoding-focused instruction years, but also the methodological problems inherent in capturing developmental growth with valid linguistic indices from language samples in school-aged children. Having acknowledged that there are no agreed standards in the linguistic and discourse measures from narratives by typically developing children (Scott & Stokes, 1995), then our selection of indices, and consequently our results, would be open to criticism in regards to their reliability and validity.

Of course, the greater strength of association between earlier preschool measures and reading than those from 5 years could also be due to a number of different factors, which may or may not include the validity of measuring expressive language at age 5. Weaker longitudinal associations from expressive language at age 5 may be reported because these associations are in fact weaker, or perhaps because older children's expressive language is truly less descriptive of their overall language abilities. Nonetheless, it may be the case that the methodology used in the literature is less sensitive to capture authentic language growth in school-aged children because they are relatively untested extrapolations from preschool measures. If the methodological validity of the linguistic indices derived from language samples remains unaddressed, then none of the alternative explanations can be either ruled out or further explored.

Indeed, a common critique of such expressive measures from language samples is that they lack developmental sensitivity (Kemper et al., 1995). For example, MLU and type-token ratio were measures originally intended for pre-school children which have been commonly extrapolated in studies involving school-age children (e.g. Chen-Wilson, 2005; Pearson, 2002) with debatable results, since the developmental patterns for these or alternative indices is not well established. It should be noted that discourse measures have been relatively better established (Peterson & McCabe, 1983), and their links to literacy have been found and replicated (Cain, 2003; Cain & Oakhill, 1996).

It is therefore pertinent to ask whether expressive language measures in narratives can capture growth and individual differences in school-aged children, and if so which of

them does the best job. Since we intend to ask the question of whether expressive language and literacy measures are linked, the weaknesses of the linguistic indices used here need to be further addressed.

Given the agreement that language still grows in school-aged children and the lack of consensus on which indices to use, it was decided that further exploration was needed to address reliability and developmental sensitivity issues in narrative language samples. To do so, a study was conducted to examine secondary data using an existing database of narratives produced by typically developing preschool and school-aged children, which is reported in this section.

Studies focused on evaluating linguistic indices from narratives in school-aged children are relatively scarce when compared to the studies in early childhood. Some studies, already described before, have looked at a wide selection of indices in a narrow range of ages (e.g. Scott & Windsor, 2000), while others have examined a few indices in a wide range of ages, usually focusing on grammar (e.g. Kemper et al., 1995; Nippold et al., 2005). However, none of the comprehensive studies have included the vocabulary measure identified here from the literature, VOCD, as being the least sensitive to differences in sample size, and therefore, the most reliable.

It is necessary to have studies using both a wide range of measures and also to collect them at a wide developmental window. The necessity of including a wide range of measures comes from the documented interactions between different sub-skills in typical and impaired children (Crystal, 1987), where children displayed greater skill in one linguistic aspect at the expense of another. In addition, it is relevant to include a wide developmental window to provide evidence of the suitability and developmental sensitivity of such measures, particularly for school-age children.

Relatively few studies have considered a wide selection of indices in a wide range of ages in typically developing children. Possibly the most comprehensive of such studies, in terms of ages and measures, looked at several of the most commonly used vocabulary and grammatical indices in narratives produced by children between 5 and 12 years of age (Justice et al., 2006). Using factor analysis, the data supported two factors coinciding with the theoretical constructs generally used in the speech and language literature: a *productivity* factor grouped together types, tokens and number of T-Units, while a *complexity* factor loaded heavily on mean length of T-Units in words. In such analysis, all automated lexical measures correlated with tokens, while the syntactic

79

measures loaded on a separate factor. Since the vocabulary measure VOCD, as a diversity index rather than a productivity measure, was not included, it is not yet known where this construct would load.

Justice and colleagues' (2006) study offered some evidence that developmental patterns can increase up to the age of 10, used a representative sample in the USA, and was based on a standardised elicitation procedure (Test of Narrative Language; Gillam & Pearson, 2004). Still, it had the following disadvantages. First, example stories were modelled by the examiner, and it is unclear whether and how much linguistic output by the examiner had an effect on the linguistic output produced by children. Second, children generated a story from a single picture, which resulted in very short narratives, with a range of means of 68 to 148 tokens for ages 5 to 12, respectively, with wide standard deviations of tokens relative to the tokens mean. Length of narratives matters because "[w]ithout a reasonable incidence of the variable under investigation, the data are susceptible to individual and situational variation, thus restricting statistical analyses and interpretation" (Liles, 1993, p. 877). Replicating these results with longer samples, improved elicitation procedures and inclusion of the lexical measure VOCD would help support the notion that these indices are indeed valid for school-aged samples.

For these reasons, even when the indices identified from the literature showed potential for finding significant associations with reading comprehension, it was deemed necessary to empirically examine further which of the many linguistic indices available in narrative samples show the greatest developmental sensitivity and reliability in school-children, in order to better address the original aim of the study of linking expressive language and reading comprehension. Section 3.2.2 describes the investigation of how lexical indices behaved when applied to an existing corpus. Section 3.2.3 is focused in the application of syntactic analyses. Finally, Section 3.2.4 brings both kinds of indices together to examine how synchronised their trajectories are.

It is worth noting that an evaluation of the discourse-level measure is outside the scope of this particular study for it was focused on the automated linguistic analyses that can be obtained in a comparatively efficient way and with a potentially higher degree of reliability. The use of discourse measures is also deferred to Section 3.3 because this kind of measures has already been used widely in school-aged children, so their validity is less contested.

3.2.1 Aim and Method

The aim of this study was to bring together for examination a set of vocabulary and syntactic indices from narrative samples identified from the literature and apply them to an existing corpus of narrative data, to find empirically which of them showed the greatest reliability and developmental sensitivity.

First, lexical indices were evaluated for both developmental sensitivity and for independence from differences in language sample sizes. Another goal was to compare the indices obtained from the corpus to the ones published in the literature in the context of the specific task characteristics of each study.

Then, syntactic indices identified from the previous review were also compared for developmental sensitivity and independence from the differences in sample sizes.

Selection of the database

At first, an attempt was made to use a publicly available corpus from the CHILDES database (MacWhinney, 1996, 2000), which would facilitate replication of results from independent researchers in the future. This database contains three corpora of English narratives using a similar approach to elicitation in a similar sample, i.e. picture stories produced by typically-developing monolingual school-aged children (Miranda, Camp, Hemphill, & Wolf, 1992; Pearson, 2002, the monolingual portion of the corpus; Wetherell, Botting, & Conti-Ramsden, 2007). A main disadvantage of this set of narratives was that they were collected from different age ranges, with little or no overlapping across the studies. It was considered that the differences between corpora would make it difficult to draw any conclusions regarding developmental patterns as they clearly came from three different populations, not to mention the inevitable differences in elicitation and transcription methods.

Therefore, the selection turned instead to an unpublished corpus of 60 children's narratives which suited the research question about developmental patterns, since it included the 7-8 years of age range (Chen-Wilson, 2003). As there was continuity from the same cross-section of the population from 3 to 9 years of age, even if cohort differences could not be ruled out, at least the environmental and task-specific characteristics would be the same.

Child participants were typically-developing native speakers of English from mixed socioeconomic backgrounds, 12 from each of the following ages: 3, 4, 5, 7 and 9. An additional set of 12 narratives by adults was available for comparison, with a mean age of 37 years, who were also from a varied socioeconomic status. The narratives were elicited from an original 13-page picture book. The corpus came from the English portion of data collected for a previous study focused on specific aspects of language development in English and Mandarin (Chen-Wilson, 2003). The elicitation procedure had involved a child previewing the picture book and then having to tell the story to a naïve peer listener, who was selected by the child. A naïve elicitation procedure should also aim in maximising decontextualized language and minimising the use of nonverbal cues. Rapport had also been established as the researcher had spent time as a classroom visitor prior to the administration of the task.

The transcripts in the Chen-Wilson corpus were already in CHAT (MacWhinney, 2000) format and, with the exception of a few coding notations that a newer version of CLAN software (MacWhinney, 2000) did not recognise, it was ready for language analysis with such software. A sample of such narratives can be found in Appendix F1. Transcripts contained morpheme segmentation by hand, but as these were done consistently in CHAT format, they could be turned on and off with the appropriate CLAN instructions.

Analysis of lexical and syntactic measures

Linguistic indices were obtained using the set of programs in Computerized Language Analysis or CLAN (MacWhinney, 2000) in December 2008. First, three lexical quantitative measures were obtained from the children's narratives: tokens, types, and VOCD. Vocabulary analyses were carried out excluding repetitions and ignoring the morphological divisions in the transcripts.

For the syntactic analyses, four indices were based on Brown's Mean Length of Utterance (MLU; 1973) but adapted for use in the narratives, rather than conversation. Two segmentation procedures were used: clauses and T-Units. Analyses were run for both clauses and T-Units, and each of these was in turn segmented into whole words and in morphemes, to see whether this would account for any differences in these children's narratives.

The original corpus was already segmented into clauses, so the first set of analyses was performed on the corpus for both words and morphemes as it was. Then, a copy of the whole corpus was re-segmented into T-Units, and the same analyses were conducted in words and morphemes.

Consequently, four quantitative variants of MLU were obtained: 1) Mean Length of Clauses in words, 2) Mean Length of Clauses in morphemes, 3) Mean Length of T-Units in words and 4) Mean Length of T-Units in morphemes. The MLU command in CLAN ignores repetitions by default. Analyses were conducted in January 2009.

Reliability

As the narratives were already transcribed and audio recordings were not available, reliability of transcription was not carried out. For T-Unit segmentation, a second examiner reviewed 15% of the sample (9 narratives) to compute inter-rater reliability. Mean reliability was 94.8% (range 73% to 100%).

3.2.2 Lexical indices results and discussion

The following two subsections describe and discuss the results for tokens, types and VOCD.

3.2.2.1 Results

Descriptive statistics for each of the raw vocabulary measures by age are presented in Table 3.1. Normality was examined for all variables. Types did not meet the kurtosis assumption while VOCD did not meet the skewness assumption. Tokens met both assumptions. An outlier was present in the tokens by a single 4-year-old. Although transformations were considered, it was thought that non-transformed variables would produce more interpretable results. In this way, tokens, types and VOCD can be compared directly with previous studies, while a transformation of scores would have not allowed such comparisons.

Table 3.1 Vocabulary measures from corpus narratives by age group (n=12 per group).

Measure	Age 3	Age 4	Age 5	Age 7	Age 9	Adults
1. Tokens						
Mean	91.00	134.33	166.00	164.25	203.33	386.50
SD	30.96	82.33	66.01	39.99	53.03	99.69
Range	55-158	70-380	86-282	82-230	130-288	260-568
2. Types						
Mean	34.00	58.08	74.50	72.50	88.25	150.83
SD	19.83	17.76	24.90	14.40	16.66	27.95
Range	9-70	41-101	46-120	41-92	65-116	108-200
3. VOCD						
Mean	14.53	26.80	34.90	32.89	38.35	56.02
SD	13.06	5.80	9.73	9.04	7.92	14.39
Range	1.07-37.38	15.30-33.39	21.12-49.55	18.98-49.61	22.75-50.52	32.54-91.13

Table 3.1 shows that there is a trend from all indices to grow from 3 years of age through adulthood, with the exception of the change between 5 and 7 years, which shows a slight decrease in all lexical measures.

To illustrate the differences between these indices, Figures 3.3a, 3.3b, and 3.3c display graphically the median, 10th and 90th percentiles by children's age for each of the different lexical indices. Adults are not shown because differences with children are so large they would obscure the variability amongst children in the graphs. Descriptively, as shown in Figures 3.3a, b and c, children talked more as they got older. However, there were 4-year-olds who spoke as much as some of the older children. Types exhibited growth until age 7, where it had a dip, and started to grow again at 9 years of age. Finally, VOCD showed a similar pattern to types, though the changes in the school-age years from 5 to 9 are even more subtle.
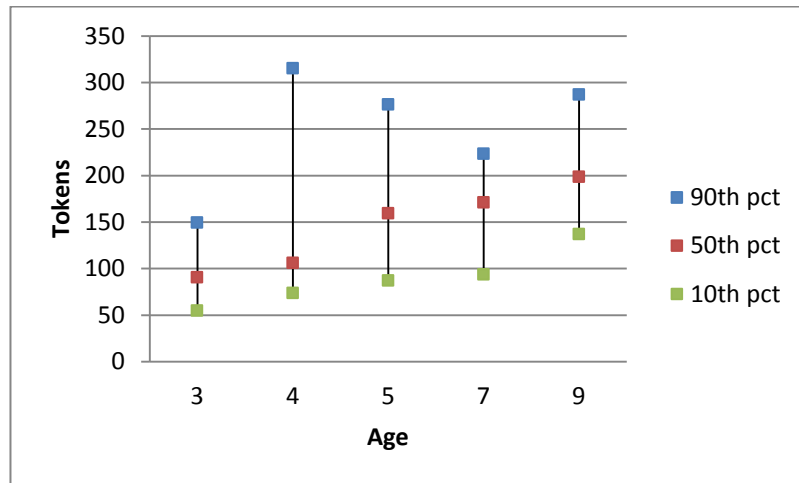
a)
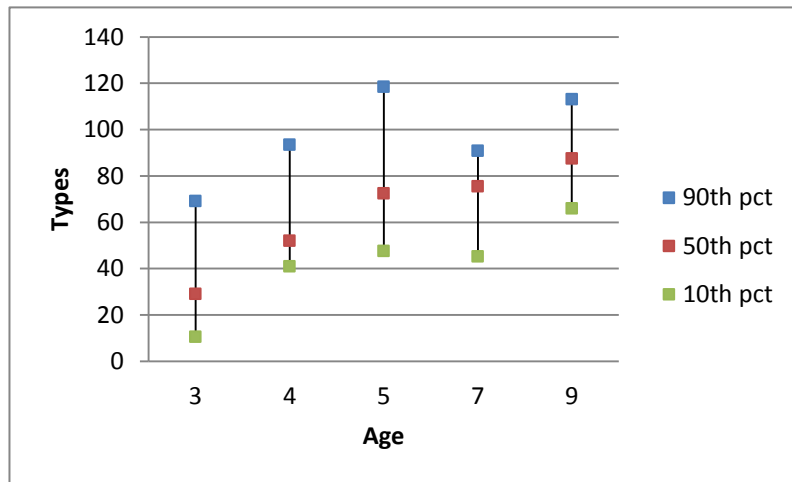
b)
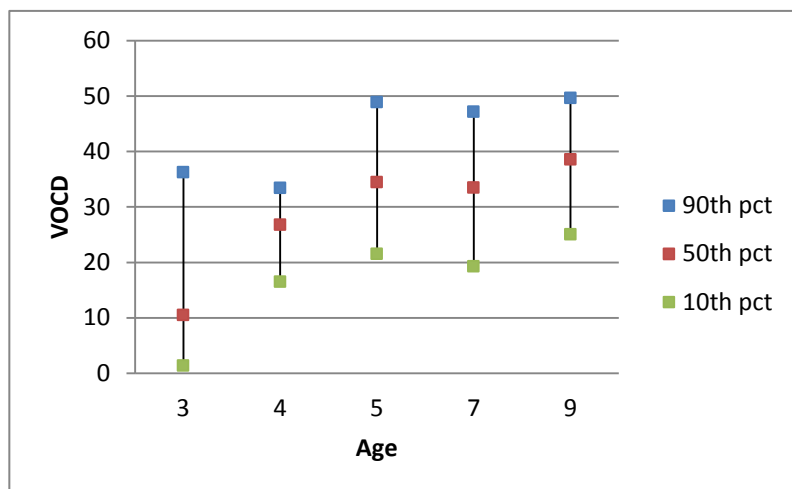
c)

Figures 3.3. Median, 10th and 90th percentile scores
in a) tokens, b) types and c) VOCD by age groups.

To evaluate differences in the three lexical measures by age group, one-way ANOVAs and planned comparisons were carried out. A separate one-way ANOVA for each index was chosen because, since the indices are constructed differently, they are inherently on different scales, as described in the previous Section 3.1.4.1. Moreover tokens measure mainly productivity and VOCD specifically aims to get away from a measure of productivity.

Planned comparisons were chosen over post-hoc tests because they have slightly greater power to detect small differences, as were the ones expected in school-aged children's language. The homogeneity of variance assumption was only met by tokens and types, but not by VOCD using Levene's test.

For the measure tokens, there was a statistically significant effect of Age ($F(4,55) = 6.37$, $p < .001$), indicating that group differences existed between age groups, with older children displaying higher scores, as shown in Table 3.1. Planned comparisons were only carried out for contiguous age groups: contrast 1 (c1) compared ages 3 and 4, c2 compared ages 4 and 5, c3 compared 5 and 7, and finally c4 compared 7 and 9. These planned comparisons revealed that no significant differences existed in tokens between any of the contiguous age groups examined. In other words, the developmental differences seen for tokens in Table 3.1 were not significant from one age group to the next one, although a developmental trend does exist, as seen in Figure 3.3a, suggesting the significant differences were between non-contiguous age groups.

For types, Age was also a significant effect ($F(4,55) = 14.01$, $p < .001$), and planned comparisons revealed significant differences between ages 3 and 4 ($t = 3.10$, df = 55, $p < .005$), between 4 and 5 ($t = 2.11$, df = 55, $p < .05$), and between 7 and 9 years of age ($t = 2.03$, df = 55, $p < .05$), while the differences between ages 5 and 7 were not significant. For all the significant differences, older children produced higher scores of types, and these patterns can be clearly seen in Table 3.1.

Finally, Age was a significant effect for the measure VOCD ($F(4,55) = 11.85$, $p < .001$). The planned comparisons with equal variances not assumed showed that significant differences existed between ages 3 and 4 ($t = 2.97$, df = 55, $p < .01$) and 4 and 5 ($t = 2.48$, df = 55, $p < .05$) only, but not between the contiguous school-age groups between 5 and 9. Once more, for all the significant differences, older children produced higher

scores of VOCD, but the differences between contiguous age groups were not significant, as seen in Table 3.1.

Correlations with age as well as between lexical measures were obtained. Correlations with age were carried out using a non-parametric correlation coefficient, Spearman's rho, given that age was measured in years. Table 3.2 shows that all measures were significantly correlated with age, with types displaying the strongest correlations, followed by VOCD. However, when addressing the question of how likely the narratives are to be influenced by sample sizes, VOCD displayed an anticipated lower correlation with tokens, signalling greater independence than types. It is worth noticing that even when lower, this correlation is still significant, so total independence from sample size was not achieved in this sample and at this range of responses.

Table 3.2 Pearson correlations between lexical measures and Spearman correlations with age (n=60)

Lexical measure	1	2	3
1. Tokens	-		
2. Types	0.90**	-	
3. VOCD	0.56**	0.83**	-
Age	0.64**	0.68**	0.59**

**$p < .01$; 2-tailed

The results of the non-parametric correlations with age measure the extent to which the variables are related. Just like the results from the ANOVAs, these correlations suggest that types have a slight advantage over VOCD in terms of developmental sensitivity but VOCD offered a much greater advantage in terms of independence from tokens.

Finally, a comparison of tokens and types in the Chen-Wilson corpus to those published in the study selected as reference and described at the beginning of this section (Justice et al., 2006) is shown in Table 3.3, for the ages that were included in both studies. VOCD was not analysed in the benchmark study, so VOCD results are not included in this table. The comparison is only illustrative as these data are from different populations and obtained with slightly different elicitation methods.

Table 3.3 Comparison of tokens and types with published data

	Tokens				Types			
	CW corpus		Justice *et al.*		CW corpus		Justice *et al.*	
Age	M	SD	M	SD	M	SD	M	SD
5	166.00	66.01	68	47	74.50	24.90	39	20
7	164.25	39.99	96	74	72.50	14.40	52	28
9	203.33	53.03	162	96	88.25	16.66	79	30

NB. CW = Chen-Wilson corpus; Justice et al.= (Justice et al., 2006)

Table 3.3 shows that in the Chen-Wilson corpus children told narratives that were longer both in terms of tokens and types at all ages. Narratives in the Chen-Wilson study also displayed lower standard deviations for ages 7 and 9, but not for the youngest children at age 5.

3.2.2.2 Discussion

Lexical analyses of the Chen-Wilson corpus with narratives from preschool and school-aged children showed that even when differences with adults were large, the differences amongst the children were more subtle. As expected, tokens did not reliably differentiate children in this age range. Results showed that even some 4-year-olds were able to talk as much as their oldest counterparts. Meanwhile types and VOCD did show greater developmental sensitivity, though not in the way it was expected: comparisons for types indicated significant differences between almost all age groups, except between the 5- and 7-year-olds, while VOCD showed no significant differences between the school age years collected here, from 5 to 9. Correlations with age showed concurring evidence: while VOCD was expected to show greater developmental sensitivity, types in fact had the strongest correlations with age.

Addressing the issue of relative independence from sample sizes, VOCD fared comparatively better than types, as anticipated. Correlations with tokens showed that types were very much influenced by tokens, while VOCD was less so. Nevertheless, in the Chen-Wilson corpus, VOCD was still significantly correlated with tokens ($r = .56$),

compared to the $r = .22$ reported previously (McCarthy & Jarvis, 2007). Therefore, the length of the narratives in this particular sample might still have affected the VOCD scores, although this influence was much less than for types.

The length of the samples in the Chen-Wilson corpus was from 55 tokens to 380 in the 3 through 9 age groups. Excluding preschool children however, the range was from 82 to 288, which is slightly below and in the lower end of the 100-400 token range recommended for the application of VOCD (McCarthy & Jarvis, 2007). Therefore, narratives with these elicitation methods in the school age years are likely to produce enough linguistic data for the proper application of VOCD in most, but not all, children. It is possible that using a longer picture book should ensure greater productivity that would display a better fit for the most appropriate application of VOCD.

To address the question of how these elicitation procedures fare in relation to published data, tokens and types were compared to those in the study that included similar measures and some of the same ages (Justice et al., 2006). To reiterate, the Justice and colleagues narratives were elicited using a single picture, to an adult, with modelling from the examiner. However, the Chen-Wilson corpus was elicited with a 13-page picture book, to a naïve peer, without modelling. Increased linguistic output at the ages 5, 7 and 9 for both tokens and types in the Chen-Wilson corpus could be due, amongst other reasons, to the semi-structured naïve elicitation procedure with a picture-book stimulus.

Moreover, since the standard deviations in both types and tokens are reduced in the Chen-Wilson corpus at least for the school-age years, 7 and 9, the structure provided by the pictures may aid in this regard. Since the large within-group variability in all measures poses a concern for reliability, the fact that the elicitation procedures in the Chen-Wilson corpus seem to be related to increased linguistic output and reduced variability in the school-age years, makes them more technically promising for further work. In addition to the fact that the results come from two different populations (Chen-Wilson, 2003; Justice et al., 2006), it is acknowledged that these results could come from other factors such as task and environmental differences. Nonetheless, replication could help in identifying whether elicitation methods using a picture book instead of a single picture do in fact facilitate larger linguistic output.

In sum, types and VOCD showed some developmental sensitivity and reduced within-group variability compared to tokens, as had been anticipated, with types showing

stronger sensitivity for the school age group, and VOCD showing stability in the early school age years. Although it was expected to see growth in the school age years, the dip at 7 years in types and the plateau from 5 to 9 in VOCD illustrated a more complex developmental pattern.

It remains to be seen, if the patterns were replicated, whether indeed lexical stability occurs at age 7, as suggested by the more reliable VOCD, or whether it is peculiar to characteristics of the sample. One concern in interpreting these results is that, given the cross-sectional nature of the data, a factor which cannot be ruled out is that cohort differences exist, such as having a relatively quiet cohort of 7-year-olds, whose low productivity could still have affected their VOCD scores. This is a particular concern in the context of the main study since this is the age group that the reading study will be focused on.

An additional concern from the Chen-Wilson corpus stems from the elicitation with a peer listener, which is the only characteristic of the Chen-Wilson corpus that could constrain the linguistic output. Originally intended to maximise the naïveté feature of the elicitation procedure, the children telling the narratives spoke to peers who clearly had not seen the picture book before. Listener-speaker interactions have been reported where children will change their language style and linguistic output when addressing a child compared to the one they use when addressing an adult (Hansson, Nettelbladt, & Nilholm, 2000; Shatz & Gelman, 1973).

If the stability of vocabulary growth at 7 years is genuine, an option that can be explored is whether this plateau is accompanied by growth at other levels, such as syntactic development, as in the trade-offs predicted by Crystal's theory (1987) .

In any case, at least from the results in this single corpus, the slight advantage of types over VOCD in terms of developmental sensitivity is offset by the loss of independence from tokens, with its loss of reliability. Replication should address many of the issues raised in this discussion.

3.2.3 Syntactic indices results and discussion

The following two subsections describe and discuss the results for the syntactic index length comparisons between the four variants of MLU achieved by segmentation and morphological division.

3.2.3.1 Results

The four measures obtained were as follows: Mean Length of Clauses in Words; Mean Length of Clauses in Morphemes; Mean Length of T-Units in Words, and Mean Length of T-Units in Morphemes. The within-subjects analysis compared the syntactic index length achieved by the two Segmentation procedures (clauses versus T-Units), and by the occurrence or absence of Morphological division (morphemes versus whole words). The between-subjects analysis looked at the five Age groups in children only.

Two measures of syntactic productivity were also included, number of clauses, and number of T-Units. Descriptive statistics are shown for the four grammatical indices for all age groups in Table 3.4. Normality assumptions were met for both indices in clauses, but not for indices in T-Units. No outliers existed in any of the four measures. Again a decision against transformation was made to favour interpretability of results.

Table 3.4 shows that change in clauses is very slow, predictably more so for words than for morphemes. It also shows that change in T-Units is more pronounced, and more so for morphemes than for words.

Table 3.4. Syntactic measures for narratives from corpus by age group (n=12 per group)

Measure	Age 3	Age 4	Age 5	Age 7	Age 9	Adults
4. Mean Length of Clause in Words						
Mean	4.66	5.44	5.41	6.23	5.96	6.26
SD	1.05	0.76	0.59	0.52	0.29	0.23
Range	2.50-6.00	3.73-6.48	4.33-6.23	5.42-7.20	5.53-6.36	6.01-6.85
5. Mean Length of Clause in Morphemes						
Mean	5.36	6.35	6.89	7.54	7.50	7.83
SD	1.31	1.52	0.75	0.57	0.35	0.33
Range	2.75-7.37	2.31-7.81	5.66-7.85	6.72-8.80	6.83-7.96	7.37-8.50
6. Mean Length of T-Unit in Words						
Mean	4.78	5.72	6.79	7.89	8.20	10.46
SD	1.06	0.95	1.27	1.25	1.43	1.17
Range	2.50-6.00	3.83-7.29	5.29-8.63	6.17-10.10	6.45-11.00	8.70-12.38
7. Mean Length of T-Units in Morphemes						
Mean	5.48	7.14	8.67	9.92	10.29	13.07
SD	1.28	1.19	1.62	1.55	1.61	1.43
Range	2.75-7.37	4.95-8.79	6.43-10.90	7.65-12.47	8.25-13.40	11.45-15.85
8. Number of clauses						
Mean	16.91	24.33	29.58	27.67	34.67	61.50
SD	4.88	13.87	11.57	7.31	9.59	15.96
Range	12-25	15-66	16-52	13-39	21-51	41-90
9. Number of T-Units						
Mean	16.92	23.58	23.67	21.08	25.75	37.75
SD	5.21	11.56	8.63	5.63	8.02	11.03
Range	12-28	14-57	14-41	13-32	15-43	21-58

The following graphs illustrate how means start to diverge at age 5. Figure 3.4a groups the two indices in words, MLC-w and MLT-w. This figure comparing clauses and T-Units both in words shows that while T-Units seem to grow up to age 7, for clauses, stability is reached much sooner.

Meanwhile, Figure 3.4b groups the two indices in morphemes, MLC-m and MLT-m. This figure shows virtually the same pattern than the previous Figure 3.4a for words.
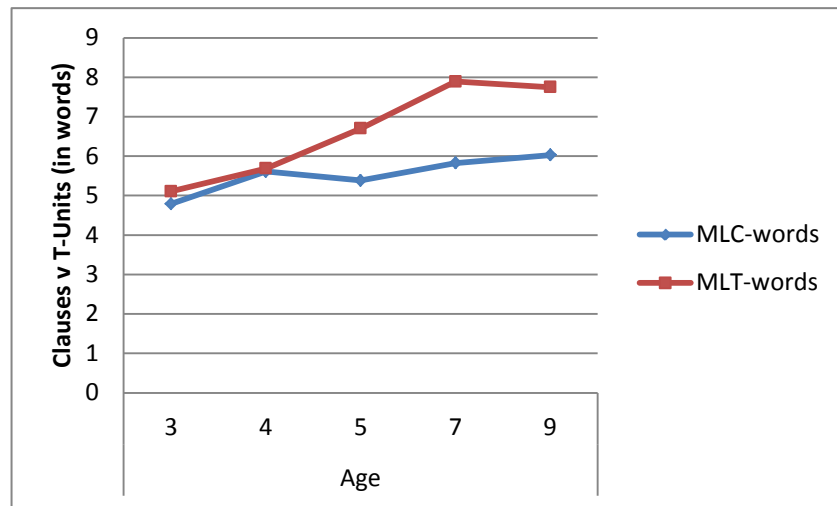
a)
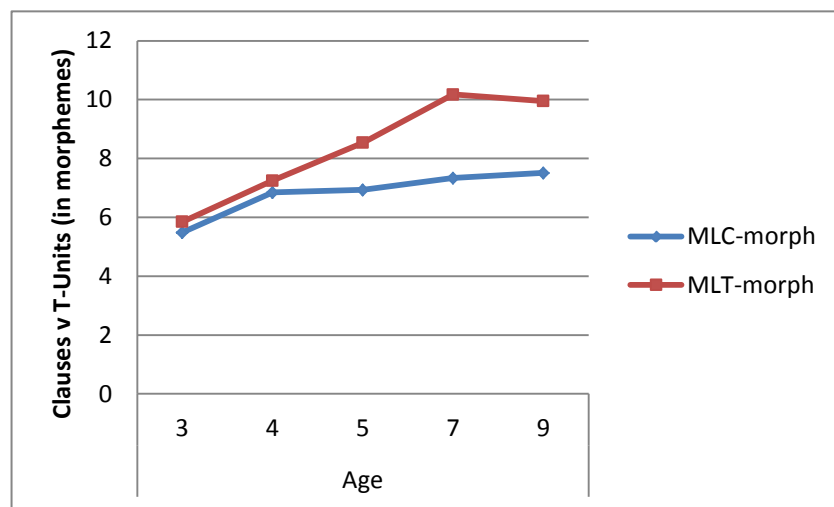
b)

Figure 3.4a Mean Length of Clauses in words v Mean Length of T-Units in words.
Figure 3.4b Mean Length of Clauses in morphemes v Mean Length of T-Units in morphemes.

To address the comparisons between Segmentation procedures (clauses versus T-Units), occurrence of Morphological division (morphemes versus whole words) and between the five Age groups in children, a 2 x 2 x 5 mixed ANOVA was performed.

First, the between-subjects effect of Age was significant: $F(4,55) = 20.58$, $p < .001$, partial $\eta^2 = .60$, indicating general age group differences, with older children displaying longer syntactic measures, as seen in Table 3.4 and Figures 3.4a and b. The main within-subjects effect of Segmentation procedure was significant: $F(1,55) = 84.85$, $p < .001$, partial $\eta^2 = .61$, reflecting that length in T-Units was higher than for Clauses, also shown in Figures 3.4a and b. The main within-subjects effect of whether there was Morphological division or not, i.e. whether the words were left intact, was also

93

significant: $F(1,55) = 791.54$, $p < .001$, partial $\eta^2=.94$, indicating that morphemes showed higher length than whole words, as seen in Table 3.4.

Also significant was the interaction of Segmentation by Age ($F(4,55) = 9.39$, $p < .001$, partial $\eta^2=.41$), indicating that segmentation effects did not occur evenly across age groups, as can be seen in Figure 3.5, where older children displayed higher disparities by segmentation procedures than younger children.
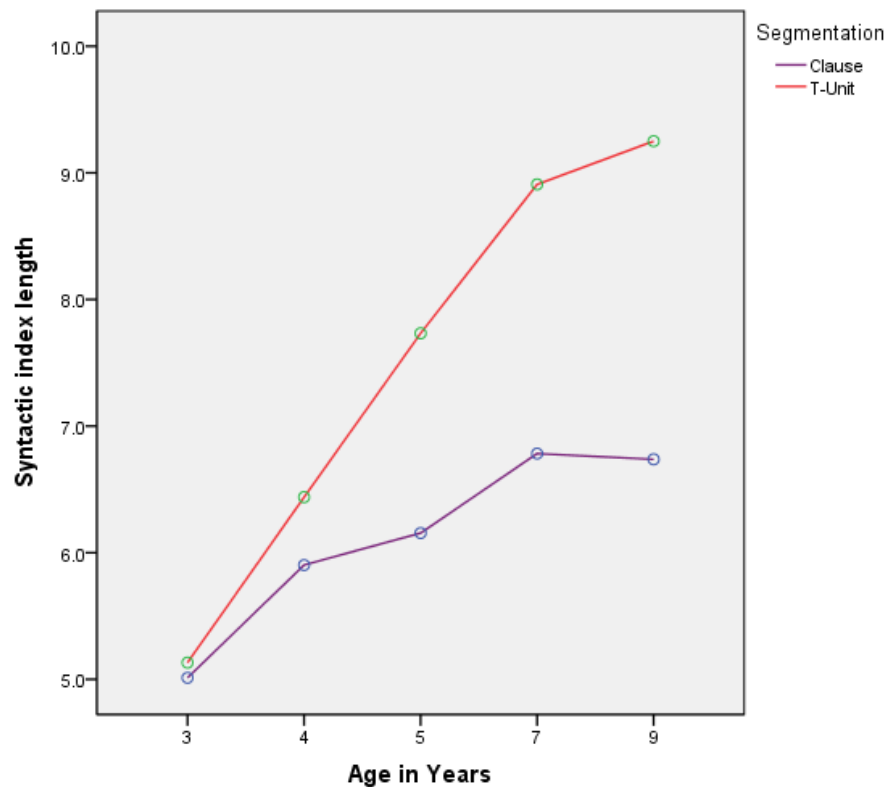
Figure 3.5 Mean syntactic index length by Segmentation procedure (clauses v T-Units)

The interaction of Morphological division by Age was also significant ($F(4,55) = 18.20$, $p < .001$, partial $\eta^2 = .57$), as seen in Figure 3.6, which also indicates that morphological division effects were also more pronounced for older children.
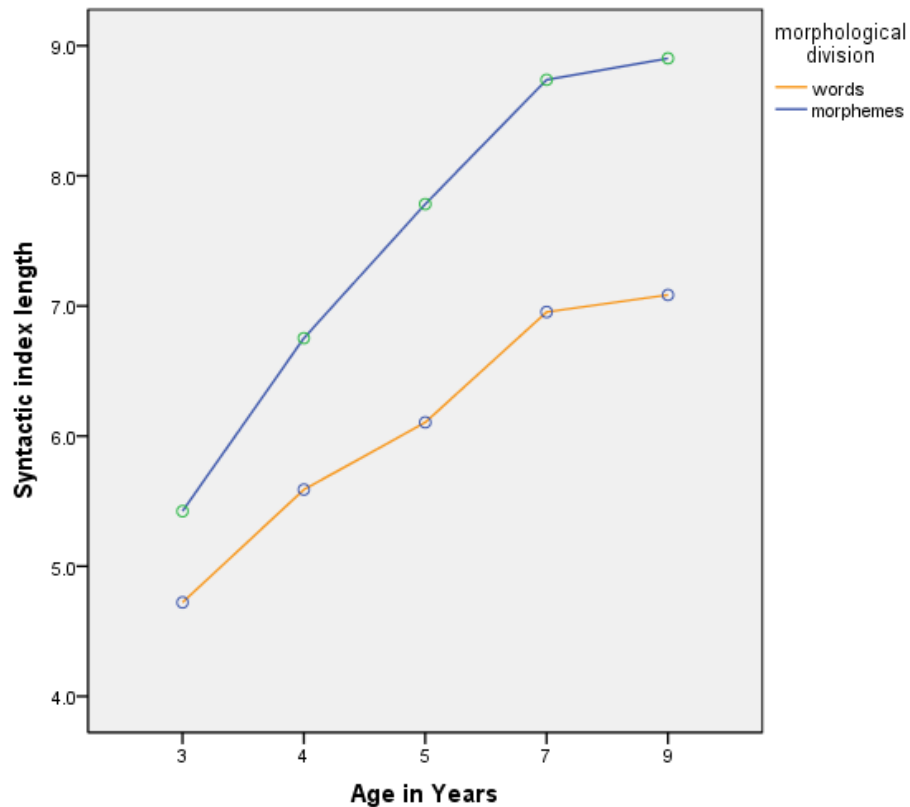
Figure 3.6 Mean syntactic index length by Morphological division
(words v morphemes)

In addition, the interaction of Segmentation by Morphological division was also significant ($F(1,55) = 14.13$, $p < .001$, partial $\eta^2 = .20$), reflecting that segmentation effects were different according to levels of morphological division: as seen in Figure 3.7, differences between words and morphemes tended to be larger in T-Units than in Clauses.
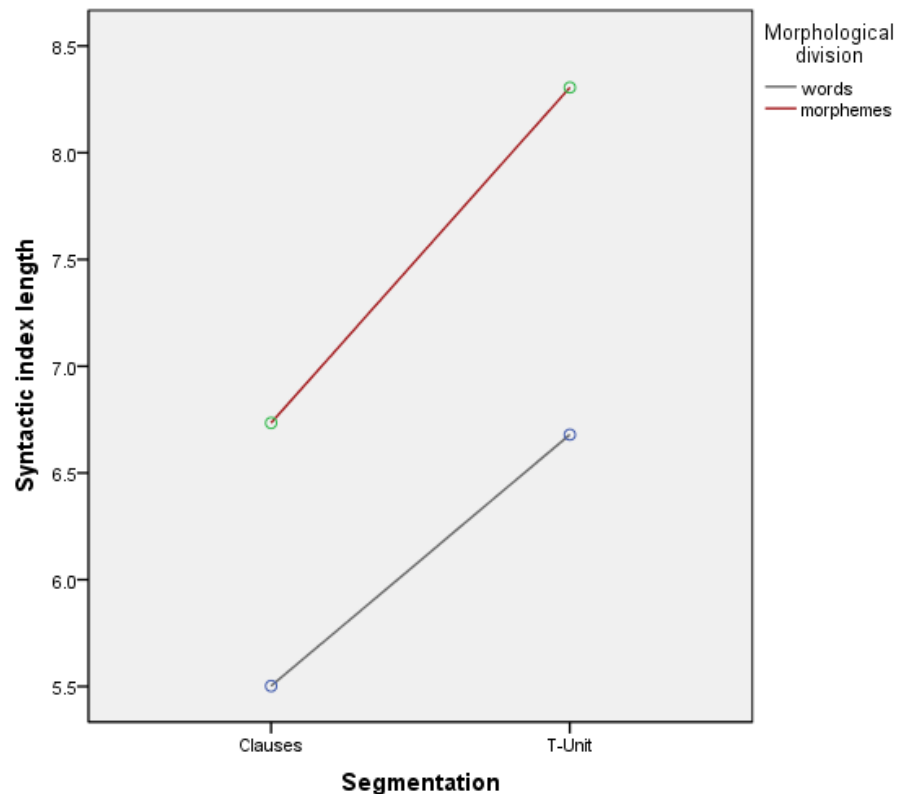
Figure 3.7 Mean syntactic index length by Morphological division
and by Segmentation procedure

Finally, the interaction between Age, Segmentation procedure and Morphological division was not significant ($F(4,55) = .92$, $p > .05$, partial $\eta^2 = .06$), indicating that there were no variations in one factor that depended on the level of the other two factors simultaneously.

To examine developmental sensitivity, planned comparisons were carried out for the between-subjects factor of Age for each syntactic index in a similar way to the developmental analysis for lexical indices, with each age group compared to the contiguous one. The homogeneity of variance assumption evaluated using Levene's statistic was not met by both measures in clauses, but it was met for both measures in T-Units; hence, the contrasts are reported accordingly. For Mean Length of Clause in Words (MLC-w) there were significant differences between ages 3 and 4 ($t = 2.09$, df $=$ 20.06, $p < .05$, equal variances not assumed) and between 5 and 7 ($t = 2.63$, df $= 21.62$, $p < .05$, equal variances not assumed). Developmental comparisons in MLC in

morphemes only showed significant differences between the 5 and 7 age groups (t = 2.38, df = 20.58, $p < .05$, equal variances not assumed).

The pattern for Mean Length of T-Units in words (MLT-w) was a barely non-significant difference between ages 3 and 4 (t = 1.92, df = 55, $p = .06$), but significant differences between ages 4 and 5 (t = 2.17, df = 55, $p < .05$) and between 5 and 7 (t = 2.22, df = 55, $p < .05$) were found. The pattern was similar for MLT in morphemes, except that the first comparison was also significant: between 3 and 4 (t = 2.78, df = 55, $p < .01$), 4 and 5 (t = 2.54, df = 55, $p < .05$), and between 5 and 7 (t = 2.10, df = 55, $p < .05$). None of the four syntactic indices examined captured growth between ages 7 and 9.

These results reflect what Figures 3.4a and b, and Figure 3.5 show, that segmentation into T-Units exhibits a more clear developmental pattern than clauses, particularly from age 5 onwards, and within T-Units, morphological division displays the best developmental variability when the whole corpus is considered. However, for school-aged children, T-Units in both words and morphemes display the same developmental sensitivity.

In order to appreciate the patterns for syntactic growth in T-Units, two graphs below, Figures 3.8a and b illustrate the median, 10th percentile and 90th percentile for T-Units in words and in morphemes, respectively. Both graphs illustrate steady growth up to 7 years of age.
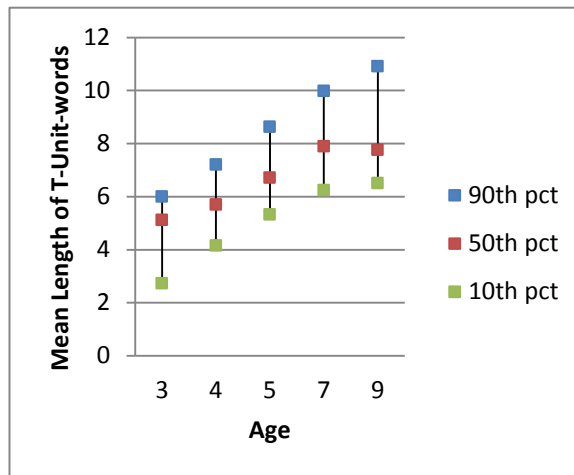
a)
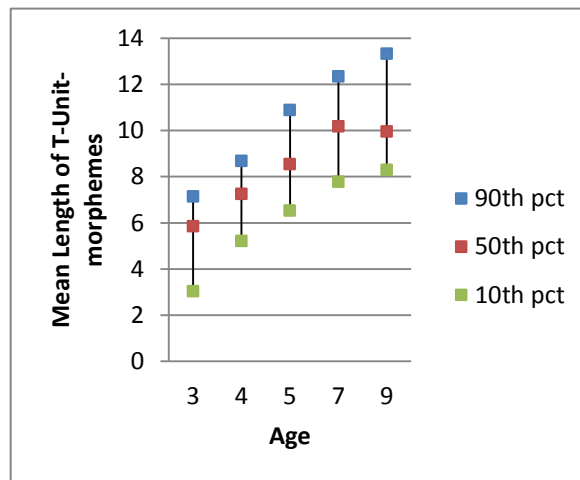
b)

Figure 3.8 Median syntactic index length, 10th percentile and 90th percentile
for (a) T-Units in words and (b) T-Units in morphemes.

For converging evidence, all four indices and two more usually used to measure syntactic productivity, number of clauses and number and T-Units, were correlated with age, as shown in Table 3.5. Again, a non-parametric correlation coefficient, Spearman's rho, was used, as age was measured in years, not months. These confirmed what the graphs showed, that T-Units in both words and morphemes displayed strong, significant correlations, while clauses in words and morphemes displayed moderate, though still significant correlations.

Table 3.5 Pearson correlations between syntactic measures and tokens
and Spearman correlations with age

Syntactic Measure	1	2	3	4	5	6
1. MLC in Words	-					
2. MLC in Morphemes	0.80**	-				
3. MLT in Words	0.72**	0.68**	-			
4. MLT in Morphemes	0.71**	0.68**	0.98**	-		
5. Number of Clauses	0.25	0.31*	0.46**	0.47**	-	
6. Number of T-Units	0.14	0.19	0.12	0.12	0.89**	-
Age	0.54**	0.64**	0.76**	0.78**	0.60**	0.35**
Tokens	0.41**	0.42**	0.56**	0.55**	0.96**	0.85**

*$p < .05$; **$p < .01$; 2-tailed

To address the issue of whether any of the variants is dependent upon language sample size, correlations with tokens, also displayed in grey in Table 3.5, were also examined. These correlations suggest that the first four syntactic indices are still related to sample size, though the magnitude of these correlations with tokens is similar to the ones displayed by what the previous analyses results showed to be the most independent lexical measure, VOCD: r correlations for these syntactic indices with tokens ranged from .41 to .56, while the correlation between VOCD and tokens was $r = .56$. Meanwhile, as Table 3.5 shows, both number of clauses and number of T-Units were highly correlated with tokens.

A final comparison from the indices obtained in the Chen-Wilson corpus with those in the benchmark study (Justice et al., 2006), is described in Table 3.6.

Table 3.6. Comparison of three syntactic indices with published data

Measure	Age	C-W corpus		Justice *et al*	
		M	*SD*	M	*SD*
MLT in words	5	6.79	1.27	6.8	1.7
	7	7.89	1.25	8.5	3.8
	9	8.20	1.43	8.4	1.4
MLT in morphemes	5	8.67	1.62	7.6	1.8
	7	9.92	1.55	9.5	4.3
	9	10.29	1.61	9.4	1.6
Number of T-Units	5	23.67	8.63	8.5	5.4
	7	21.08	5.63	11.3	9.1
	9	25.75	8.02	17.3	9.6

Comparing the results from the Chen-Wilson corpus with those reported in the study by Justice et al (2006), indicates that although the stories in the Chen-Wilson corpus were longer in number of T-Units produced in all ages, the overall Mean Length of T-Units was comparable. As seen in Table 3.6, standard deviations appear smaller for the 7 years of age group in the Chen-Wilson corpus, while it is very similar for the ages of 5 and 9. Once more, the comparison is illustrative only, since the populations are different, and so are the elicitation procedures.

3.2.3.2 Discussion

Results add empirical evidence to previous reports that segmentation procedures have an impact on MLU syntactic measures (Scott & Stokes, 1995). Using a semi-structured elicitation procedure, findings from this corpus illustrate that at five years of age children's clauses start to be significantly linked into increasingly longer T-Units. This growth seems to reach stability between 7 and 9 years of age, but resumes in adulthood, a pattern not displayed by clauses.

Morphemes displayed significantly more growth than words, and morpheme differences seemed larger when T-Units were also longer. Planned comparisons for Age revealed that T-Units in morphemes were developmentally sensitive from 3 through 7, while their equivalent in words failed to reach significance for distinguishing between pre-schoolers. In other words, morpheme segmentation is still essential for preschool children. If the same patterns found here for school-aged children were to be substantiated with replication, then the selection between morphemes and words would not rest on developmental sensitivity, at least for the school-age range.

However, there is a reliability price to pay for this increase in morphemes, since morpheme segmentation needs to be done by hand, as was done in this corpus, or by using computer software (Sagae, Lavie, & MacWhinney, 2005), which is still being developed to approach reliability levels by human coders. Given than the general pattern is virtually the same for primary pupils, Mean Length of T-Units in words as a linguistic index is more likely to be reliable, and consequently more likely to be valid than its equivalent in morphemes.

Regarding the issue of independence from sample sizes, although MLT in words is not completely independent from productivity as measured in tokens, the correlations are similar in magnitude to those of the most independent lexical measure from the vocabulary analyses, VOCD. These correlations were similar for both clauses and T-Units. For their part, number of clauses and number of T-Units were so highly correlated with tokens, that they seem redundant if the count for tokens is known.

A comparison of the results with those from previous literature appears to confirm that MLT in both words and morphemes is quite stable across populations, regardless of amount of linguistic output, a result made more interesting when considering that these are samples from two different countries, at two different times, using different elicitation procedures. At first glance, the study by Justice and colleagues (2006) would seem a more efficient elicitation procedure to obtain a measure of MLT in either words or morphemes; nonetheless, standard deviations in the 7-year-old cohort appear noticeably smaller in the Chen-Wilson study, suggesting more variability with the one-picture elicitation procedure used by Justice and colleagues. In any case, even if the elicitation procedure of a single picture were appropriate for syntactic analyses, our previous lexical analyses showed that it was unlikely to produce enough linguistic

output for the appropriate application of VOCD. Once more, replication should aid in clarifying some of these patterns.

In sum, the pattern of growth in T-Units seems developmentally sensitive up to 7 years of age. Although it can only function as a global syntactic measure, Mean Length of T-Units describes a pattern of growth similar to that displayed between 3 and 6 years of age examined with more fine-grained measures, such as the Developmental Sentence Score (Lee, 1974), with a decreased rate of change at the later developmental stages. It remains to be seen whether this pattern can be replicated with similar or enhanced elicitation methods, with the narratives having an adult as the listener, as described in the discussion for lexical indices.

3.2.4 Relationships amongst the variables.

A final descriptive analysis brought together the lexical and syntactic indices to explore whether there are trade-offs between different linguistic indices by comparing the developmental patterns in vocabulary to those in grammar.

In order to evaluate the developmental trends between lexical and syntactic measures, these were converted to z-scores based only on the children's sample. These standardised measures were then charted together. Two lexical indices were selected, types and VOCD, while one syntactic index was chosen, MLT in words. The results plotted against age are shown in Figure 3.9.
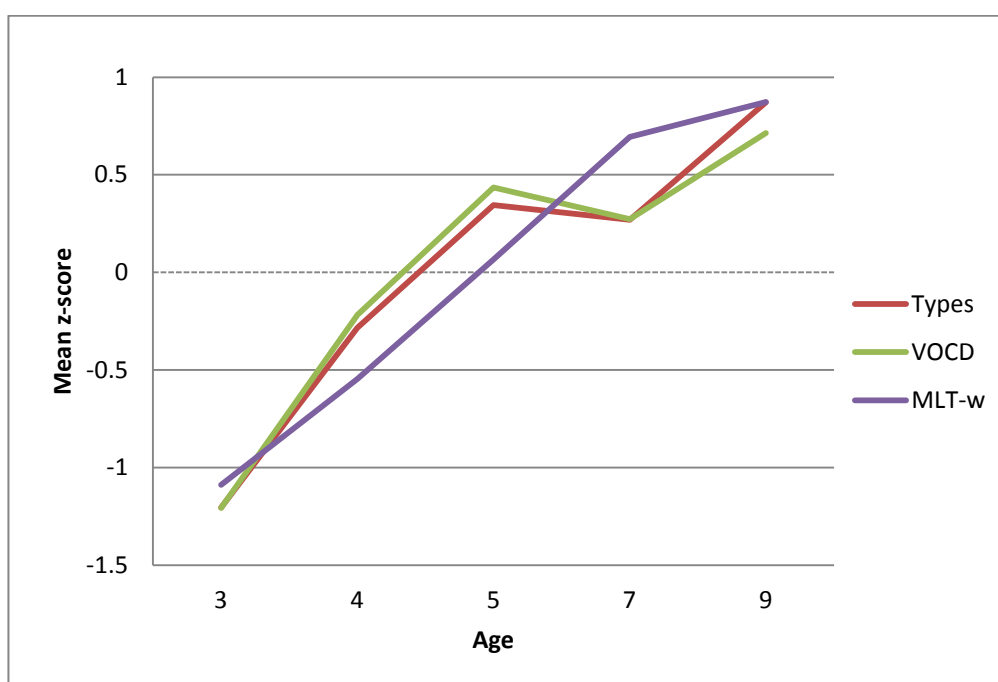
Figure 3.9 Mean z-scores for three linguistic indices by age

Figure 3.9 illustrates that, at least in narratives from the Chen-Wilson corpus by using these elicitation methods, syntactic growth as measured by T-Units in words continues to occur between ages 5 and 7 while lexical development goes through a plateau. These jagged patterns are not inconsistent with early language development descriptions in the literature (Brown, 1973) and may even help explain why there are differences in the

measured contribution of language according to developmental window (Scarborough, 1990b, 2010).

A factor analysis would have statistically clustered these three and the previous analysed indices into constructs, and would have allowed us to a certain extent, to replicate the factor analysis described before (Justice et al., 2006) to see how VOCD would load in relation to lexical and syntactic measures, since the previous factor analysis placed all lexical indices as productivity measures. However, the sample size of 60 independent narratives in this corpus was deemed insufficient for such an analysis (Costello & Osborne, 2005; Tabachnick & Fidell, 2007).

If the patterns found here could be replicated, the lexical plateau does coincide neatly with the first few years of reading instruction, when the instructional emphasis is on learning decoding skills based on relatively simple words. In any case, replication should aid in recognising whether these patterns are genuine, or a result of sampling (i.e. cohort) effects.

3.3 Pilot study: Narrative indices and reading comprehension

The literature surveyed led to the selection of the task procedures and the database study allowed for a stronger basis in the identification of the linguistic measures to capture expressive language. Narrative samples elicited from a wordless picture book as a stimulus were chosen to be the most appropriate tools for observing language in a sample of 7- and 8-year-olds. Moreover, armed with the empirical evidence supporting the suitability of our lexical and syntactic indices for the age group of interest, after the first couple of years of formal reading instruction, it was considered necessary to examine whether these identified indices were, in fact, related to reading comprehension albeit in a small sample.

A study was designed to examine the concurrent associations between school-age expressive narrative language and reading comprehension that would also consider the role of children's receptive skills. Since it has also been established that different sub-skills may play differential roles at different points of development (Scarborough, 2005; Vellutino et al., 1994) it was important to focus on a narrow developmental window. A specific window was considered to be especially useful, between 7 and 8 years of age, right after the first few years of decoding-focused instruction, and when there is already some evidence of school-age expressive skills being related to reading (Cain, 2003; Cain & Oakhill, 1996). This is also the age when mild receptive language deficits tend to go unnoticed (Nation et al., 2004), and so the more visible narrative skills could prove more useful. Furthermore, this is also the developmental period when causality from reading to receptive language, i.e. the onset of reciprocal relationships, has been reported (Verhoeven & Van Leeuwe, 2008); in other words, studying older children with a cross-sectional design, the relationships found would also be the product of reciprocal relations between language and reading.

3.3.1 Aim and hypotheses

This pilot study was intended to provide preliminary evidence of the narrative language and reading association. The purpose of this pilot was to inform the design of the actual larger-scale study, and the practical experience of administering the tests was useful as a familiarisation with the materials that would be used in the main study. The main study would then be able to answer the question of whether expressive narrative measures add

unique variance to reading comprehension beyond the variance already explained by receptive language.

In sum, the aim of this pilot study was to examine the feasibility of using narratives as a medium to investigate the relationship between expressive language skills and reading comprehension at a developmental window after the first few years of formal reading instruction, when mild language deficits might be harder to notice. To do so, we had the following objectives: to examine whether automated lexical and syntactic automated indices could be derived from narrative samples using a wordless picture book; whether a manual discourse measure could be also obtained; and finally, to examine whether order of administration of the standardised language tests had an effect on the results.

On the basis of previous research and the former database study, our proposed hypotheses were the following:

a) Expressive measures from narratives would be significantly associated with reading comprehension skills in 7- and 8-year-old children.
b) A more reliable vocabulary expressive measure from narratives, VOCD, would be significantly associated with reading comprehension skills in 7- and 8-year-old children.

3.3.2 Method

Participants

Twenty monolingual children, aged 7 and 8, who were attending Year 3 at two local schools in the West Midlands region of the United Kingdom were asked to participate. Children involved in the study did not have a diagnosed reading, language, or developmental disorder, or behavioural problem. Consent was obtained from the head teacher at one of the schools, and from both the head teacher and the children's parents at the other one.

Testing materials

a. ***British Picture Vocabulary Scale***, 2nd edition (BPVS-II; Dunn, Dunn, Whetton, & Burley, 1997).

This standardised instrument measures the vocabulary level of the child without any speaking, reading or writing. It only requires the child to recognise a word and choose a picture out of four that best corresponds to the word heard.

b. ***Test of the Reception of Grammar***, 3rd edition (TROG-2; Bishop, 2003).

This test appraises a progressive understanding of morphology and syntax, and like the BPVT, does not require any reading, writing or speaking, only pointing to the picture that matches the information presented orally.

c. Understanding Spoken Paragraphs subtest of the ***Clinical Evaluation of Language Fundamentals***, 4th UK edition (CELF-4 UK; Semel, Wiig, & Secord, 2006).

This standardised test evaluates attention, focus and comprehension of orally presented narratives, including the ability to obtain the main idea and generate inferences. Its administration is segmented by age groups. For this study, the corresponding 7-8 years section was used.

d. ***Neale Analysis of Reading Ability***, 2nd Revised British edition (NARA-II; Neale, 1997).

Reading comprehension was measured using the Neale Analysis of Reading Ability–II, which is the most widely used standardised test used for reading research in the United Kingdom. Although this instrument has its critics for its format (it is a test of open questions and requires oral reading), it has also been argued that it is precisely the format of oral administration with corrections from the test administrator that makes it is less dependent on decoding skills (Cain & Oakhill, 2006). In terms of measuring comprehension with open questions, it has also been argued that reading comprehension tests in multiple-choice formats, such as the Gates-MacGinitie Reading Tests (MacGinitie, MacGinitie, Maria, & Dreyer, 2000) are likely to influence responses, as the options interact with the text (Cain & Oakhill, 2006).

Because NARA asks the comprehension questions orally, we argue that it also reduces the burden on the child's test taking skills, which in this sense, allows us to obtain a truer measure of reading comprehension. In addition, because it does not offer multiple choices, the opportunities for guessing are minimised. Finally, NARA offers a measurement for reading accuracy, which can, to a certain extent, be used as a proxy for decoding skills.

e. Matrix Reasoning subtest of the ***Wechsler Abbreviated Scale of Intelligence*** (WASI; Wechsler, 1999).

Intelligence was measured as a control variable using a short and valid test of intelligence: the Wechsler Abbreviated Scale of Intelligence. Its validity is comparable to that of the Wechsler Intelligence Scale for Children (WISC). Like similar studies in this field, only non-verbal intelligence was measured, since verbal intelligence has indeed been used as a vocabulary measure (Ricketts et al., 2007) and therefore could confound receptive vocabulary's contribution. Of the two non-verbal subtests, the Matrix Reasoning subtest was used, as in previous studies using non-verbal intelligence only as a control variable (Cragg & Nation, 2006).

f. A wordless picture book "The Sweets story" and prompting instructions were used to elicit narratives (see Appendix A for the picture sequence, and Appendix B for the prompting instructions).

Hand-drawn illustrations were created for investigating language development in a separate study (Chen-Wilson, 2003), and good-quality photocopies of the original drawings were bound together to appear as a children's book. Although a widely used picture book exists for narrative elicitation (Mayer, 1969) used in the studies compiled by Berman and Slobin (1994), an original book was chosen for this task over a commercially available picture book to ensure the children would produce narratives from material none of them could have encountered before.

g. A digital voice recorder Olympus WS-210S was used to record the narratives produced by the children, for later transcription and analysis.

Procedure

Children were seen outside their classrooms on an individual basis, and were administered one or two tasks at a time. To address the research question regarding whether order of administration of the standardised tests had an effect on reading results, half of the children at each school were administered the standardised tests from easiest to most complex in terms of verbal processing demands (non-verbal intelligence, receptive vocabulary, receptive grammar and listening comprehension) and half were administered the same tests in the exact opposite order.

The narrative production task was collected at the next stage for both groups, to allow for greater rapport between the investigator and the child, so that the pupil would be willing to produce the largest possible amount of linguistic output. First, the child was asked to preview the picture book "The Sweets Story". Then they were asked to tell the story in their own words while looking at the pictures, and their stories were audio recorded. Finally, the reading test was administered. After data collection, all audio recordings were transcribed and analysed as described next.

Transcription

The elicited stories were transcribed into CLAN software using CHAT conventions (Computerized Language ANalysis & Codes for Human Analysis of Transcripts, respectively; MacWhinney, 2000). False starts and repetitions were excluded. The only departure from CHAT conventions was in the segmentation procedures. Instead of segmenting speech into clauses following the conventions by Berman and Slobin (1994), lines were segmented into T-Units (Hunt, 1965; 1970). All transcriptions were performed by the author of this thesis and they were checked over again after the initial transcription. Some examples of the narratives produced by the children can be found in Appendix F2.

Analysis

Transcripts were analysed using CLAN (MacWhinney, 2000) to obtain the two first linguistic measures: expressive vocabulary and expressive grammar.

First, for expressive vocabulary, a quantitative measure of *lexical diversity* "Parameter D" (McKee *et al.*, 2000), most commonly known as VOCD was calculated using the VOCD program in CLAN. Then, for expressive grammar, Mean Length of Utterance (sentence length) of T-Units in words (MLT-w) was computed using the MLU program in CLAN as a quantitative measure of *syntactic complexity*. Both measures excluded false starts and repetitions. Finally, the expressive discourse measure was manually scored following the rubric already described above in Section 3.1.4.3.

3.3.3 Results

Three objectives were addressed in this study. First, it aimed to evaluate the feasibility and success of specific elicitation procedures for obtaining narrative samples in this age group of children of 7 and 8 years of age. It also examined the feasibility of obtaining specific linguistic indices from computer analyses. All children were able to produce enough linguistic output in their stories when hearing the prompts and after previewing the picture book used as stimulus, to allow for computerised and manual analyses with a mean of 200 tokens or number of words (without repetitions) and a range of 131-338 tokens, considered adequate for the proper application of VOCD.

Second, in addition to the lexical and syntactic measures from narratives, a discourse measure was derived manually. Given that all scores obtained from the narrative sample do not have corresponding norms, raw scores were used for both standardised assessments and expressive indices from the narrative samples for further analyses. Descriptive statistics for all variables are shown in Table 3.7. Mean standardised scores for standardised measures are also included for reference.

An additional objective of this study was to determine if order of administration of the standardised test had an effect on reading comprehension, in order to define whether this was a nuisance variable that should be controlled for in a future larger-scale study. For that purpose, one group of ten children was administered these tests in the following order: nonverbal ability, receptive vocabulary, receptive grammar and listening comprehension. The other ten children were tested in the opposite order. There were slight differences between these two groups in reading comprehension. However, an independent t-test showed that the difference between groups was not statistically significant ($t = 0.81$, df $= 18$, $p = 0.43$, two-tailed).

Reading comprehension scores between schools were compared. Although minor differences were observed, they were not statistically significant ($t = 0.01$, df $= 18$, $p = 0.79$, two-tailed).

Table 3.7   Descriptive statistics for raw receptive scores and expressive indices, including tokens (number of words) as a measure of the size of narrative samples (n=20); mean standardised scores are provided for those tests with norms.

	Mean raw score	SD	Range	Mean standardised score
Vocabulary: BPVS-II[a]	88.45	13.93	63-119	100.68
Grammar: TROG-2	12.70	3.28	7-18	91.05
Listening Comp: CELF-4 [UK]	11.50	2.50	5-15	14.53
Expressive Vocabulary: VOCD	37.83	9.68	25.83-57.66	-
Expressive Grammar: MLT-w	8.65	1.13	6.74-11.50	-
Narrative Production Congruence Score	4.10	1.21	2-6	-
Reading Comprehension: NARA-II	17.65	8.52	3-36	100.89
Tokens	211.25	56.28	131-338	-

NB: [a] Form 1 of NARA II

As can be seen from the standardised scores in Table 3.7, these pupils were typically developing in reading comprehension, with typical vocabularies and slightly lower receptive grammatical skills. It should be mentioned that although it was not the focus of this pilot study, the mean standardised score for reading accuracy, a proxy measure of word reading, was 101.68, confirming these were typically developing in both reading skills.

Also, the study aimed to explore how expressive as well as receptive measures of language relate to reading comprehension skills at this particular stage of development.

As shown in Table 3.8, four of the six bivariate zero-order Pearson correlations between the language variables and reading comprehension were statistically significant: receptive vocabulary, receptive grammar, expressive vocabulary and expressive grammar, as measured in this study. Indices for Listening comprehension and for Narrative production, as measured here, failed to reach significance in this small sample.

Table 3.8 Pearson bivariate correlations of raw language scores with reading comprehension, non-verbal ability, and age (n=20).

	1	2	3	4	5	6	7	8
1. Reading Comprehension: NARA-II	-							
2. Vocabulary: BPVS-II	.70**	-						
3. Grammar: TROG-2	.50*	.52*	-					
4. Listening Comp: CELF-4^{UK} (U.S.P. Subtest)	.39	.51*	.55*	-				
5. Expressive Vocabulary: VOCD	.62**	.43	.40	.01	-			
6. Expressive Grammar: MLT-w	.53*	.43	.29	.50*	.12	-		
7. Narrative Production Congruence Score	.26	.24	.15	-.04	.20	.25	-	
8. Non-verbal ability: WASI (Matrices Subtest)	.29	.40	.49*	.36	.18	.22	.09	-
9. Age in months	-.51*	-.10	-.23	-.18	-.18	-.32	.18	.06

**$p < 0.01$ (2-tailed)
*$p < 0.05$ (2-tailed)

Interestingly, at this sample size only receptive grammar was significantly associated with non-verbal ability. More remarkable, however, is the fact that age showed a significant *negative* association with reading comprehension, where slightly older pupils were less skilled comprehenders than their younger classmates. It should be noted that age was, by design, restricted to ages 7 and 8, so this strong association was unexpected. Nonetheless, there was also a trend for age to be negatively correlated with language scores, with the exception of the narrative production congruence score, which had a weak non-significant association.

In order to eliminate any differences related to age, the correlation coefficients were recalculated partialling out the pupil's age in months at the time of testing, and results are shown in the second column of Table 3.9. Non-verbal ability was also collected as a background variable and was also controlled for in the third column of Table 3.9.

Table 3.9 Zero-order and partial correlations with Reading Comprehension controlling for age, and simultaneously for age and non-verbal ability (n=20).

	Zero-order correlation with Reading Comprehension (NARA-II)	Controlling for age	Controlling for age and non-verbal ability (WASI)
Vocabulary: BPVS-II	.70**	.75**	.71 **
Grammar: TROG-2	.50*	.46*	.34
Listening Comp: CELF-4^{UK}	.39	.35	.25
Expressive Vocabulary: VOCD	.62**	.62**	.60**
Expressive syntax MLU-T	.53*	.46*	.41
Narrative P. Congruence Score	.26	.41	.42

**$p < 0.01$ (2-tailed)
*$p < 0.05$ (2-tailed)

The results of the second analysis in Table 3.9 show that after controlling for both age and non-verbal ability, only the receptive and expressive vocabulary measures were still significantly associated with reading comprehension. Expressive vocabulary using VOCD seemed quite unaffected by both the age and non-verbal ability controls, while receptive vocabulary and the narrative congruence score partial correlations seemed to strengthen after the age control, although not enough to make the narrative congruence score's association significant at a conventional *p* level in such a small sample. Finally, it can be observed that while receptive grammar and discourse comprehension became weaker after controls, the reverse happened for their expressive counterparts, suggesting that expressive narrative skills might be able to offer a unique window into a child's repertoire of linguistic abilities.

3.3.4 Discussion

Before addressing the main results of this study, group comparisons are discussed, since those results affect the main analyses. Comparing the reading performance of one school against the other, there were no statistically significant differences between the samples, which was expected due to both schools having similar socioeconomic environments, and allows us to consider all the children as a single group for the purposes of this pilot study. Regarding whether there would be order effects in groups with different order of administration, the absence of a statistically significant difference between these two groups suggests that this variability is not likely to affect the reading comprehension assessment.

Therefore, the results from both schools and from both administration orders are considered here together. The finding of no significant differences between orders of administration also suggests that, in a future larger-scale study, order effects are unlikely to become a nuisance variable.

To test the first hypothesis, we explored the contribution of receptive and expressive oral language skills to reading in a specific developmental window between 7 and 8 years of age, with a special focus on expressive skills, which are far less researched. Findings from this pilot study suggest that expressive vocabulary and expressive grammar from narratives are strongly associated with reading comprehension at this age and, more interestingly, are almost as strongly related to reading comprehension as their receptive equivalents, by using specific measures identified from the literature. These significant relationships were maintained even when age was controlled for.

For comparison, these correlations were equal or of greater magnitude than those reported before with other indices with the same stimulus and similar elicitation conditions in 6- to 8-year-olds (Chen-Wilson, 2005). To compare equivalents, in both comparisons ages are controlled for. In regards to expressive vocabulary, the association between reading comprehension and VOCD was significant and greater ($r = .62$, $p < .01$) than the one reported previously using tokens ($r = .34$, n.s). Regarding expressive grammar, the association between reading comprehension and MLT in words was significant and greater ($r = .46$, $p < .05$) than the one reported for Mean Length of Clauses in morphemes ($r = .16$, n.s.), but comparable when measuring grammar in the

personal story ($r =.48, p <.05$). Therefore, our results seem to suggest that, when using a semi-structured stimulus such as wordless picture book, the methodological choices do seem to represent an improvement on the magnitude and significance of these associations.

Our second hypothesis focused on finding a significant correlation in particular for expressive vocabulary in narratives using VOCD, which had not been reported before. Finding a significant association with reading comprehension after the lexical analyses reported in this chapter confirmed our hypothesis.

However, once the influence of age and non-verbal ability was accounted for, only the expressive and the receptive vocabulary measures were still statistically significant. Even when non-significant after both controls, there was a trend for an association with expressive grammar and expressive narrative ability, which could be clarified with a larger sample in a subsequent study.

In addition, the correlations between the expressive language variables themselves were much lower than the ones amongst the receptive measures. This could be useful for conducting a multiple regression analysis in a further study, since low correlations between predictor variables would satisfy the assumption of non-collinearity, which the receptive measures do not normally meet. Correlations were also moderate/low and non-significant between each receptive test and their expressive counterpart, suggesting that they are not necessarily redundant.

Even when promising, these results have to be interpreted in the context of the study's methodological considerations. First, the sample was composed of a small sample of 20 children who, although they came from two different schools, belong to a similar middle-class socioeconomic area. Language and reading performance have been shown to be heavily influenced by socioeconomic status (Charity, Scarborough, & Griffin, 2004; NICHD Early Child Research Network, 2005; Raz & Bryant, 1990; Walker et al., 1994), so inclusion of children from disadvantaged areas would make the sample more representative of the general population.

Second, there are several measurement considerations that could have influenced the results. The researcher administering the test is not a native speaker of English, which could have affected one of the standardised tests which involved a greater amount of speaking on the researcher's part: the Understanding Spoken Paragraphs subtest of the

Clinical Evaluation of Language Fundamentals (CELF-4 UK; Semel et al., 2006). However, results for this particular test suggest that it might have been possible that the researcher's accent had a negligible effect since there was actually a trend towards a ceiling in performance. This, in turn, is a more serious methodological constraint, which should be addressed in a larger-scale study. One way this could be done would be the inclusion of a more diverse socioeconomic sample, which could be assumed, would show a greater range of scores and not only ceiling ones. Regarding the issue of the researcher's accent, although it cannot be changed for subsequent studies the researcher has to be mindful of its potential as a nuisance variable.

Finally, there are the validity issues regarding our experimental measures for the expressive variables. The Parameter D, or as is more commonly referred to, the index VOCD to measure expressive vocabulary, enjoys the greatest validity of the expressive variables used here, and most importantly, is less sensitive to language sample size in terms of number of words, but developmental sensitivity is not well established. The measure used for grammar was considered appropriate for this developmental age, but how it is affected by sample size has not been established in previous studies. Concerning the validity of the narrative score, it needs to be established with more studies, though its reliability would be the least affected by sample size, since the structure of the narrative was provided by the picture sequence.

Although they were selected on the basis of the literature and our analysis of secondary data, there is the possibility that different expressive or syntactic measures are differentially related to literacy, just as is the case for receptive measures (Ouellette, 2006; Ricketts et al., 2007), and the ones used here might not be the optimal measures. This validity could become into question given the large standard deviations due to differences in the narrative samples sizes, and due to the contested evidence of developmental sensitivity for the specific indices (Kemper et al., 1995), using the specific elicitation methods used here.

Section 3.2 presented the results of a study based on an existing corpus of children's narratives from 3 to 9 years of age, to examine lexical and syntactic indices for reliability and developmental sensitivity in this age range. However, several questions emerged from the results. First, the findings from lexical measures were mixed, with two indices showing differing desirable properties: the most developmentally sensitive

was found to be types, while the most reliable was found to be VOCD. It remains to be seen whether these results would withstand replication.

In addition, the characteristics of language sampling in general, and of a cross-sectional design, in particular, also opened up more questions than were answered. Despite being the only corpus available that included the early school age years, the fact that these indices were only estimated from a single story using a single elicitation method, does not address how these indices would fare in within-child comparisons with different stimuli. It is quite possible that some linguistic output, particularly the lexical indices which seem to be more influenced by tokens, might change as the picture books become longer, which would be more desirable for the measure VOCD, as some of the 7-year-olds still had lower levels of the required range of tokens for the appropriate application of the measure. Although it is known that task characteristics affect these linguistic indices, it is not clear how these would change according to stories or whether they would show strong reliability across stories.

An additional concern already described was that, in order to maximise the naïveté of the elicitation procedure, these narratives were told to other children, but listener-speaker interactions could also play a role in the amount and quality of linguistic output (Hansson et al., 2000; Shatz & Gelman, 1973). If naïve elicitation could be created or approximated while talking to an adult, then it is possible that such enhancements would aid in maximising the production of literate language. Although a direct comparison with the results of the pilot study reported here cannot be made across a range of ages, the tokens and VOCD scores for 7-8-year-olds in the pilot were larger than in the Chen-Wilson results for 7- and even 9-year-olds, raising the possibility that adult-directed speech might be more ideal.

Overall, a new study which included more than one story, focused on school-aged children, had a larger sample per group, and featured naïve elicitation to an adult, could address whether the patterns found here are quite stable, or whether they are subjected to sampling errors. Such a study was carried out to further explore the appropriateness of the measures identified here as the optimal linguistic indices in terms of sensitivity and reliability.

With the caveat of these methodological considerations, results from the pilot study suggest that in this age group where decoding skills are still being mastered, expressive vocabulary skills are almost as strongly associated with reading comprehension as are

receptive vocabulary abilities, even when age and non-verbal abilities are taken into account. Of course, these are only correlational results and no cause and effect relationship can be established, though this age range was chosen to minimise this issue.

Still, the results of this pilot study are taken as offering preliminary empirical evidence to pursue the examination of expressive skills in a larger sample. A larger more diverse sample would also allow the examination of whether expressive skills can explain unique variance on their own, after accounting for receptive skills. The finding of null significant differences due to order of administration suggested that this was an unlikely confounding variable. In that sense, the main study did not try to control for this variable. The null findings regarding significant differences between schools of similar socio-economic status emphasised the need for greater social variability in participant recruitment. Finally, the success of the narrative task in eliciting enough linguistic output for automated analyses was taken as evidence that it was an appropriate tool for the purpose of this study. The only particular issue that still remains to be addressed for such a larger study is that, as the previous chapter described, methodological choices have a direct bearing on the results, and the reliability and validity of the expressive measures selected here should be further examined.

The risks inherent in any study using secondary data collected for another purpose are that task characteristics have some sort of influence in the results. In particular, the nature of the narratives told to a naïve peer may have an effect on some of these measures.

If the validity of the expressive measures could be better substantiated with primary data collected for the purpose of identifying developmental patterns, and with that information, the preliminary evidence presented here relating narratives and literacy were to be confirmed with a larger, more socially diverse sample, the potential implications could be that at the very least, expressive vocabulary skills from narrative samples are good indicators of reading abilities.

Chapter 4

Narrative study based on original data

The previous chapter described the empirical evaluation of several lexical and syntactic measures by using an existing corpus of children's narratives, in order to identify the most reliable and developmentally-sensitive lexical and grammatical indices. Even when this analysis provided some partial answers to these issues and illustrated how these indices behaved in school-aged children, several shortcomings were acknowledged: a) the school-aged sample was only a portion of the Chen-Wilson corpus, and the main focus of this work is in finding developmental variability in school-aged children; b) within each age group, the small sample limits the variability of the narratives; c) cohort effects cannot be ruled out from the cross-sectional design; d) a single narrative from a picture book was elicited, so reliability across several stories or task-specific differences could not be examined; e) the elicitation method asked the children to tell the story to a peer which can have an effect on the quality of linguistic output, and d) a measure of expressive discourse was not analysed.

These issues can be addressed to a certain extent, by replicating the analysis using new data collected specifically for this purpose, and by adding a developmental analysis for discourse measures. Replication could help in distinguishing how these linguistic indices behave in a developmental window which focuses exclusively on school-aged children, in a larger sample from a different population, with a variety of stimuli, while the story is being told to an adult. In addition to the within-child comparisons in each linguistic index, replicating the previous database study with the same stimulus would allow for comparing the school-aged children's scores in both studies. Furthermore, by using a standardised assessment, the collection and transcription methods, as well as the linguistic indices, could be compared in relation to published norms. Finally, by collecting primary data, reliability analysis can be also carried out for transcription, in addition to the reliability for segmentation and, in the case of this study, the coding of the manual discourse measures.

Two issues regarding the design of the study and the elicitation methods, however, warrant additional consideration since replication would only address them to some extent. First, although cohort effects cannot be ruled out in either cross-sectional study, by comparing data coming from two different populations, the similarities in developmental patterns would provide more robust support for their validity.

A second issue concerns the change in elicitation from a peer to an adult as the listener to the story, with their relative advantages and disadvantages. The database used in the previous study using secondary data (Chen-Wilson, 2003) instructed the children to tell the story to a naïve peer selected by them to maximise the naiveté feature, as a friend of theirs could not have possibly seen the picture book before. However, listener-speaker interactions occur where language production is adapted according to the audience (Clark & Marshall, 1981), and narratives told to an adult contain more complex linguistic features (Shatz & Gelman, 1973).

Naïve child listeners (Masterson & Kamhi, 1991) as well as naïve adult listeners (Gazella, 2003; Liles, 1985) have been used before with confederates chosen by the investigator. In particular, typically developing children in the naïve listener situations were able to produce more complete episodes than when they knew the listener shared the same knowledge (Liles, 1985). However, an informal comparison of narratives told to an adult non-naïve listener with narratives told to a naïve peer chosen by the child (pilot study narratives v. database narratives in Chapter 3) seems to suggest that stories told to adults, even non-naïve, display greater lexical diversity using VOCD, though no greater syntactic complexity using MLT-w. The question is still open as to how these linguistic indices would behave in narratives told to naïve adults.

With the aim of overcoming some of the limitations identified in the previous study, the present study was designed in order to see how the developmental trends identified in the database study would behave in a different population cross-section, with the addition of an analysis of global discourse measures. For lexical analyses, developmental trends and reliability issues are examined in a similar fashion to the database study. For the syntactic indices, meanwhile, the present analysis aimed to extend the findings of the database study, not in a direct replication, but rather focusing on the robustness of a single measure across different stimuli, for the following reasons: first, T-Units had already shown great robustness, measured both in words and morphemes, in both the database and the Justice and colleagues study (2006) with

different elicitation methods in different populations; second, regarding the distinction between words and morphemes, our study mirrored the findings in the Justice study where they were so strongly correlated with each other ($r = .99$ in the Justice study, n=250; $r = .98$ in our database study, n=60), that morphemes seem to offer no additional information in developmental sensitivity to compensate for the loss in reliability; and third, regarding the distinction between clauses and T-Units, the findings that the differentiation between clauses and T-Units emerges relatively early from age 5 onwards, suggest that even if these findings suffered from cohort effects, the ages of interest of 7-8 years represent a much later time in development, and syntactic growth is therefore quite likely to be better indexed by T-Units. Therefore, the present study is focused on replicating developmental trends for MLT in words only in a different cohort with different stimuli.

Finally, for a full analysis of narratives, measures of global congruence or structure were also examined for developmental patterns. However, instead of using an experimental measure like the one used for the pilot study, the focus of this study using primary data was to use tools that are currently being used in school aged populations. The literature about discourse measures has been described in Section 3.1.4.3. Briefly, the three main methodologies to measure how children are able to organise a discourse's elements into a whole are the following: *narrative stages* (Applebee, 1978), *story structure* (Labov & Waletzky, 1967; Stein & Glenn, 1979) and *information checklists* (Bishop, 2004). While narrative stages has been considered best suited for evaluating the narrative development of younger children (Liles, 1993), story structure and information checklists provided a good opportunity to evaluate the developmental patterns best suited for school-aged children. Given that story structure and information checklists seem best suited for school-age language, one scoring system was selected for each of these two discourse-level measure types.

First, within story structure methods, two scoring systems were identified that were related to existing narrative standardised tests. One of them, the Index of Narrative Complexity or INC (Petersen, Gillam, & Gillam, 2008), is based on Stein and Glenn's story grammar (1979), Labov's high-point analysis (1972) and further work by Peterson and McCabe (1983), and it was developed to correlate with the standardised scoring by the Test of Narrative Language (Gillam & Pearson, 2004). Another story structure method that is widely used is the Narrative Scoring Scheme or NSS (Heilmann, Miller,

Nockerts, & Dunaway, 2010), which is included in the language analysis software SALT (Miller, 2008).

INC was chosen over NSS (Heilmann et al., 2010) because the latter has a certain degree of ambiguity in the scoring system. For example, in some guidelines, the rater is asked to observe whether there is 'inconsistent mention', 'excessive', or 'minimal' quantities of certain elements. Likewise, a phrase such as 'not all conflicts and resolutions [are] critical' (Heilmann et al., 2010, p. 165), could be interpreted in different ways by different raters. Meanwhile, INC is rather specific on how to score the different elements by offering precise guidelines, such as those for the element 'Plan', where 0 points are awarded for 'No overt statement is provided about the character's plan to act on or solve the event or problem', 1 point is awarded for 'One overt statement about how the character might solve the complication or problem', and so on, up to 3 points (Petersen et al., 2008, pp., p. 123).

In addition, INC was designed to be sensitive to change with intervention, it is relatively efficient to administer and it provides categories to rate the complexity of different narrative elements. Only those elements of INC that referred to the story structure were used. The rest of the elements referred more to cohesive markers and story conventions. Although specific linguistic cohesion markers have been found to distinguish good from poor readers, story conventions have not (Cain, 2003), and the aim was to maintain the focus on the macrostructure elements.

Then, within second type of discourse-level measure, informational checklists, the Information Score from the Expression, Reception and Recall Narrative Instrument or ERRNI (Bishop, 2004) was selected to evaluate developmental patterns, as it is a narrative instrument specifically normed with British school-aged children. Even when the psychometric properties of this instrument make it more likely to be reliable, the downside of this method is that it does not capture the overall degree of coherence that ties the story together, which the story structure measures arguably do.

Therefore, it was interesting to use two different types of discourse-level measures to examine developmental patterns in school-aged children's narratives. Since these measures are already being used in school-age populations, it is not within the scope of this particular study to compare the two methods using the same story, but rather to use them with the closest type of story, or the specific story (in the case of the Beach story) for which they were designed.

4.1 Aim

The aim of this study of linguistic and discourse indices in narratives was to add further evidence of developmental variability in school age children in a larger sample, and with a series of methodological considerations aimed at maximising the elicitation of decontextualized language.

These methodological considerations include a larger sample per group than in the database study, with specific instructions aimed to maximise the production of literate language, with elicitation to an adult, and by using three different stories featuring different characteristics which provide specific advantages: the Sweets Story (Chen-Wilson, 1997) was used in the database study, the Frog Story (Mayer, 1969), which is commonly used for the collection of narratives in a wide range of ages (Berman & Slobin, 1994), and finally the Expression, Reception and Recall of Narrative Instrument (ERRNI; Bishop, 2004) which is a commercially available instrument for narrative language assessment. In addition, a visual barrier was used to recreate to some extent a situation where the listener could not see the stimuli, in order to help in maximising literate language.

With all these methodological considerations for maximising decontextualized language and for controlling to a certain extent the story content, we hypothesized that differences between age groups would be statistically significant. For lexical indices, significant differences were expected between all the primary groups, possibly more so for types than for VOCD. Meanwhile, for the syntactic index of Mean Length of Utterance in T-Units in words (MLT-w) we once more expected to find significant differences between age groups, including differences with the oldest group of pupils. Regarding the discourse-level measures, the analyses were exploratory, but it was still expected that these two methods would exhibit different developmental patterns, given that the Information Content measure from ERRNI scores propositions from throughout the entire story, while the Index of Narrative Complexity scores up to 2 or 3 points per element, but many elements tend to cluster at the beginning or the end of the story. In other words, a high Information Content score would mean that the narrative is richer in the events told, with most critical to the story's plotline; by contrast a child scoring

higher in INC would mean that this child is knowledgeable of specific narrative structural elements. Nonetheless, developmental patterns were examined for significant differences between age groups within each measure, the same way it was done for lexical indices in the previous database study.

As for the within-child differences, the hypotheses were specific for each of the indices. For types, we expected that they would be different according to story because this particular index is more sensitive to narrative length, and the picture books differ in the number of pages; in turn, we expected VOCD to be more consistent across stories (i.e. no significant differences) because of its better independence from tokens.

Based on the findings from the database study reported in Chapter 3, we also expected MLT-w to be quite robust across tasks, given the relative resilience of this index to different populations and elicitation methods.

Results of no differences or non-significant differences between stories in any lexical index or the single syntactic index would signal that such an index is reliable across tasks with different elicitation methods.

Finally, for the discourse measures, a within-child comparison of the two methods was not done, since the measures were applied to different stories.

4.2 Method

Participant recruitment and selection

Several state schools across the West Midlands area of the UK were contacted which represented a mix of socioeconomic backgrounds and a mix of lay and parochial schools. Five schools in diverse socioeconomic neighbourhoods volunteered for participation, and parents were sent an invitation letter where they were asked to grant their permission for their children to take part in a narrative language study. Only those children whose parents gave their permission took part in the study, and their consent was obtained verbally at the beginning of assessments.

At schools in disadvantaged areas, the language in the invitation letters was simplified at the suggestion from two Head teachers at different schools, who identified the formal language in the letter as a likely barrier to obtaining volunteers. A brief questionnaire about parental education had been planned to be included in the invitation letter, to serve as a proxy for socioeconomic status. However, after suggestions that any requests for information requiring the parents to disclose their possibly lower education levels would greatly reduce the chance of obtaining volunteers, the questionnaire was not included in the request for consent.

At the four volunteer primary schools, all children in Years 1, 3 and 5 were invited to participate, corresponding to ages 5-6, 7-8, and 9-10. At the single secondary school agreeing to participate, pupils in Year 8 were asked to take part, corresponding to ages 12-13. Volunteer rates were low at the secondary school, therefore secondary pupils are only included in descriptive analyses. Given the labour-intensive nature of data collection and processing, children in Year 3 were also invited to participate in a subsequent main study relating narrative language and reading, to make full use of their transcribed stories.

In total, after excluding ineligible children who had an additional language at home, or those with an identified reading, language or behavioural disorder as reported by the teachers, 111 pupils took part in the study. Of those, three children were excluded as outliers after transcription was completed, as described later in Section 4.3. The final sample was distributed as follows: 29 in Year 1, 31 in Year 2 and 33 in Year 3, plus 15 from Year 8 in secondary school.

An attempt was made at recruiting the same number of participants from both sexes given reported differences in early language output by gender (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1992). However, volunteer rates were higher for girls than for boys. As gender differences were not the goal of the study, this was not considered an issue. The full final sample included 49 (45%) males and 59 (55%) females.

Materials

a. The wordless picture book 'The Sweets story' (Chen-Wilson, 1997) and prompting instructions were used to elicit narratives (see Appendix A for the

picture sequence, and Appendix C for the prompting instructions). This 13-page booklet is the same picture book used in the previous database and the pilot study, which would allow comparisons.

b. The popular 24-page frog story 'Frog, Where Are You?' (Mayer, 1969) was used with similar prompts (Appendix C). This wordless picture book, along with others by the same author with the same characters, has often been used for eliciting narratives (Berman & Slobin, 1994). For the pilot study the idea had been not to use a commercially available book, to ensure that no child has seen the pictures before; however, it was found that this particular book is not available in bookshops. Therefore, it was unlikely that children could have seen it before.

c. 'The Beach story' within the Expression Recall and Reception of Narrative Instrument (ERRNI, Bishop, 2004), a standardised narrative production test that contains UK norms for T-Units and information content, was chosen for comparison. ERRNI has norms for 4 years to adulthood. The instructions from ERRNI were followed with one exception: the manual directs the test administrator to guide the pupil over the pictures to ensure the child perceives the salient features. The decision to omit this instruction was taken to maximise the potential for decontextualized language even at the cost it might have on the structure of the story expressed by some children.

d. Within the scoring rubric for the Index of Narrative Complexity (INC; Petersen et al., 2008) the structural elements portion was used to obtain a score for story structure as a measure of expressive discourse.

e. A blank A4 hardcover notebook served as a visual barrier/screen.

f. A digital voice recorder Olympus WS-210S recorded all narratives produced by the children.

Procedure

Testing took place outside the classroom in different settings as provided by each school, including libraries, landings, special education rooms, etc., on an individual

basis. Even when avoided as much as possible, the level of noise/interruptions varied across schools and times of day. Although order effects were considered a potential nuisance variable, a decision was made to elicit the narratives in the same order to reduce the between-participant variability so that each kind of narrative was directly comparable with one another. The Sweets and the Frog stories were administered first to be able to follow our own specific elicitation procedures aimed at maximising the production of decontextualized language. Between the first two stories, the Sweets story was chosen to be elicited first because being smaller any issues with the narratives could be more easily addressed. The Beach story, from the standardised assessment ERRNI, was administered last.

After a brief conversation to establish rapport, the three narrative tasks were administered, usually in a single 30- to 45-minute session in the same fixed order for all pupils: the Sweets story was administered first, then the Frog story, and finally the Beach story from the ERRNI. The elicitation procedure involved asking the child to preview the picture book each time. Then, the pupil was asked to tell a good story "like a story you would read in a book". Since the use of a confederate was precluded by limited resources, instead of a formal naïve procedure, a semi-naïve elicitation procedure was attempted. In order to achieve some degree of naïveté so that the best possible linguistic output from each child's own linguistic repertoire could be obtained, a visual barrier was placed between the child and the researcher whilst the instructions emphasised how the researcher, i.e. the listener, could not see the pictures (see Appendix C for full elicitation procedure). Then, the child was asked to tell the story while looking at the pictures. In the few instances where a pupil started to tell the narratives without looking at the pictures, they were gently reminded to do so. Pupils were allowed to tell a story for as long as they wished to. The three stories were audio recorded for later transcription.

On a few occasions testing was stopped if there were any definite interruptions, and narratives were restarted afterwards from the beginning. Pupils who missed a story due to time constraints were followed-up as necessary to collect all three narratives. There were no missing data.

Regarding the administration and scoring of the discourse measures, several differences with the original scoring methods should be noted. First, the Index of Narrative Complexity story coding form was not used in its entirety. Only those eight elements

corresponding to the story structure were scored to obtain a measure of global structure: character, setting, initiating event, internal response, plan, action/attempt, complication and consequence. Therefore, the elements of formulaic markers, temporal markers, causal adverbial clauses, knowledge of dialogue and narrator evaluations were excluded from the final score, as these were more related to linguistic cohesion or story conventions. Following the guidelines set by Petersen et al. (2008), four of the INC elements can be scored from 0 to 3 points, while the five others can be scored from 0 to 2. In total, the highest score possible is 22 points.

It should be noted that the INC scoring method has been used with similar elicitation methods where the pupil previews the pictures and has access to the pictures throughout the production of the narratives, and it has also been used with other similar stories from the frog series authored by Mercer Mayer (Petersen et al., 2008). One story from this series, Frog Where Are You?, was the picture-book story used here.

Regarding the Information Content score, it was calculated as described in the ERRNI Manual and the corresponding Answer Form for the Beach story. However, the administration of the test differed from the instructions of the manual in that the children were not aided by having the examiner point to the pictures, with the specific purpose of preserving the distance created for the first two stories when the visual barrier was used. The descriptive and inferential results for both measures are described next.

Transcription

Three-hundred and thirty-three stories were transcribed using CHAT conventions and entered into CLAN software (MacWhinney, 2000). Following these conventions, false starts and repetitions were coded to be excluded, as well as unintelligible words or phrases. Segmentation was done in T-Units, which contain an independent clause and all its subordinate clauses (Hunt, 1965, 1970). After transcribing the first 15 narratives, many issues arose that required consistency for better transcription at the word and at the T-Unit level. A brief manual was created to address these issues in a consistent way, and it is found on Appendix D.

Briefly, at the word level, the most common difficulties were related to the coding of onomatopoeia and handling compound words. All onomatopoeic expressions were coded following CHAT conventions to be excluded from lexical analyses. Meanwhile, compound words were checked against the Shorter Oxford English Dictionary (Trumble, 2007), following the same guidelines offered in the ERRNI Manual (Bishop, 2004).

In turn, the most common difficulties for segmenting T-Units were related to the coding of direct speech and fragments. Direct speech was included with its main clause, unless it had several clauses, in which case each clause after the first one was segmented separately. Fragments were coded so that they could be excluded from syntactic analysis.

As some of the rules shown in Appendix D were further refined as the transcription continued, the whole corpus was checked once more for accuracy of transcription, and for adherence to the detailed segmentation rules. To get measures of inter-rater reliability, a second examiner evaluated a random subsample of 17 participants producing 3 narratives each (51 transcripts, about 15% of the total sample). The second examiner listened to the digital audio files while looking into the corresponding CLAN transcript for transcription and segmentation errors. Mean word-by-word reliability for main body words transcribed (excluding fragments, repetitions, reformulations and task-related comments) was 98.8% with a range from 98% to 99%. Mean reliability for T-Unit segmentation was 98.6%, with a range from 95% to 100%.

Reliability of the discourse measures were carried out for the Index of Narrative Complexity in the Frog Story, and for Information Content in the Beach Story, for the same subset of participants. A second examiner independently scored the stories and the scores were compared. For the measure INC, mean score reliability was 82.3% (range 71% to 94%). For Information Content, mean score reliability was better at 92.6% (range 83% to 100%).

A few of the narratives collected are found as examples in Appendix F3.

Analysis

Tokens were computed as a measure of narrative length. Types and VOCD were obtained as lexical measures, while MLT in words was obtained as the syntactic measure. The Index of Narrative of Complexity (INC) was computed for the Frog story, as this was considered the one where practice or fatigue effects were minimised. The Information Content score was computed for the Beach story as part of the ERRNI standardised test. All measures were analysed for the presence of outliers and normality.

Due to the difficulties in attracting a larger sample of secondary pupils, multivariate tests of differences were used only for comparing groups in Years 1, 3 and 5. While the Year 8 tokens, types and VOCD data were included for illustration purposes, the small and mainly male secondary school sample makes for a very unequal sample size in relation to the rest, making the Year 8 data inappropriate for ANOVA analyses. In addition, the secondary sample came from a single school, which makes their errors likely to be highly correlated. For these reasons, Year 8 is not included in the following analyses where inferential statistics are used for group comparisons.

Group comparisons for each of the lexical and syntactic indices were carried out using a 3 * 3 mixed-design ANOVA. The first factor is the between-subjects factor of Age, with three groups; the second factor is the within-subjects factor of Story with three levels. Post-hoc comparisons with Gabriel's procedure for equal variances, or Games-Howell when the equality of variances cannot be assumed, were used to be able to correct for the slightly unequal sample sizes in the different Year groups (Field, 2009, pp.374-375); post-hoc comparisons were also chosen over planned comparisons to allow all possible pair-wise comparisons between stories.

Correlations examined associations with Years 1, 3 and 5 only, given that correlations can also be affected by having fewer data in Year 8, in the sense that fewer data could mean a truncated sample (Tabachnick & Fidell, 2007).

Results and discussion are organised into lexical indices (Section 4.3), grammatical indices (Section 4.4), discourse-level measures (Section 4.5), and finally, a comprehensive grouping of all variables (Section 4.6). A general discussion is offered in Section 4.7.

4.3 Lexical indices

All participating children were able to produce the three narratives elicited with the stimuli provided. Given that some children spoke for so long that their stories could considerably bias the rest of the data, outliers in narrative length (i.e. tokens) were identified separately for each story for data grouped by school year, to allow for the examination of developmental trends between age groups.

Given the high variability characterising narrative data a liberal cut-off point was chosen to identify outliers. Cases with z scores > 3.29 in tokens, which is significant at $p < .001$ (Field, 2009) were considered to be extreme outliers. Three stories belonging to different participants had such extreme scores. Since only three participants produced a narrative with a standardised score over 3.29, and they were scattered across the primary age groups, a decision was made to eliminate all narratives from these three participants, thus obtaining a final sample of 108 pupils telling 324 stories.

For the ANOVA analyses, excluding the secondary school participants resulted in a sample of 93 participants or 279 stories.

4.3.1 Results

Although not considered here as a lexical measure, descriptive statistics and graphs for tokens are included in the lexical indices section, as a measure of narrative sample size.

Descriptive statistics are shown in Tables 4.1, 4.2 and 4.6 for tokens, types and VOCD respectively, for all age groups, including the secondary school group in Year 8. A more detailed illustration of these patterns is found in Figures 4.1, 4.2 and 4.4, displaying the 10th percentile, median and 90th percentile for each story and by Year groups.

Table 4.1 Tokens for narratives from each picture book by Year group.

School Year	1	3	5	8
Ages	5-6	7-8	9-10	12-13
	n=29	n=31	n=33	n=15
Sweets Story				
Mean	162.76	171.03	241.09	217.53
SD	57.20	54.09	94.05	91.41
Range	77-270	90-317	114-556	142-442
Frog Story				
Mean	262.52	317.74	353.73	328.33
SD	69.29	115.53	114.44	116.38
Range	134-436	168-755	135-726	185-594
Beach Story				
Mean	171.24	233.55	273.76	245.00
SD	61.02	102.68	77.81	65.67
Range	73-292	99-542	160-462	150-402

As shown in Table 4.1, there was a steady increase in narrative length measured in tokens, excluding repetitions and reformulations, in the three stories except for the secondary school group. Once more, as children grew older, they talked more using these elicitation methods up to Year 5, that is, up to 9-10 years of age, as shown by the mean increases. Figure 4.1 illustrates how, in general, variability in narrative length had a tendency to increase with age, with the exception of the last story. Differences between groups were not statistically examined, since tokens were not the focus of the analysis, but are included for illustration given that they influence both types and VOCD.
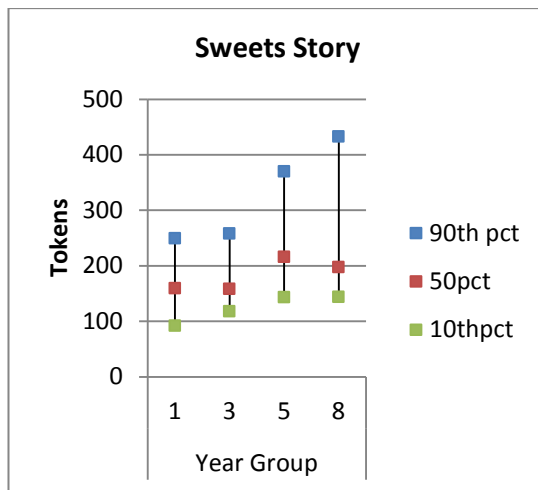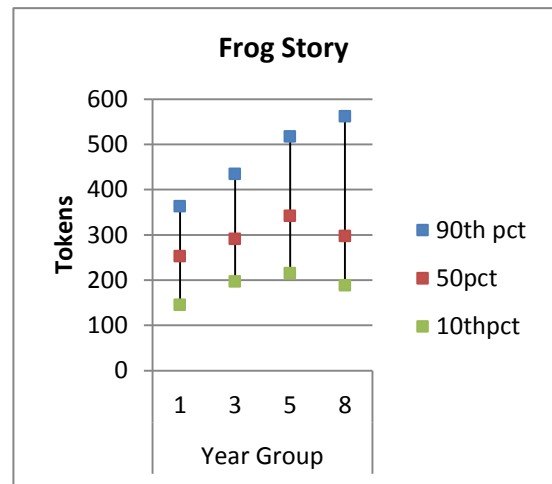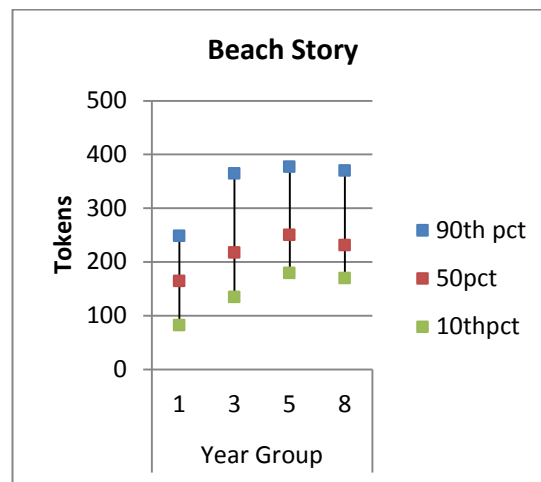
Figure 4.1 Median, 10th and 90th percentile
scores in tokens for each story by Year groups

Group comparisons between age groups and between stories examined differences in types and VOCD only, as these were the indices identified in the previous study (Chapter 3) as the ones more likely to characterise developmental change.

4.3.1.1 Types results

Table 4.2 shows the means and standard deviations for types by story and age group.

Table 4.2 Types for narratives from each picture book by Year group.

School Year	1	3	5	8
Ages	5-6	7-8	9-10	12-13
	n=29	n=31	n=33	n=15
A. Sweets Story				
Mean	68.62	76.58	99.21	94.47
SD	19.03	20.07	30.04	34.04
Range	29-105	43-131	57-170	62-187
B. Frog Story				
Mean	91.48	110.32	125.82	124.07
SD	21.32	27.64	36.21	38.60
Range	45-151	67-188	65-208	84-218
C. Beach Story				
Mean	70.03	92.52	106.85	101.13
SD	19.45	26.88	24.90	24.45
Range	35-112	61-166	71-155	78-169

Descriptively, Table 4.2 shows that types displayed a similar pattern to tokens in the sense that mean increases were observed up to Year 5, or 9-10 years of age. The oldest group of secondary school students showed a slight decrease in types.

For a better understanding of the variability of the data within each age group, Figures 4.2a, 4.2b and 4.2c illustrate, for each story, the median, 10th and 90th percentile by Year group.

Graphically at least, the within-group variability seems to also increase with age in the first two stories (Figures 4.2a and b), while it seems to remain stable for the last one (Figure 4.2c).
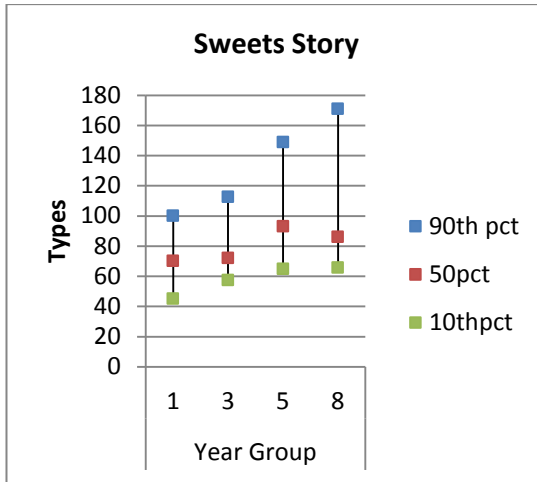
a)
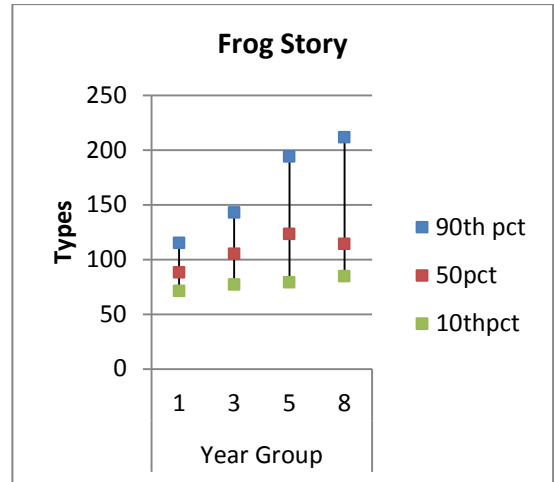
b)
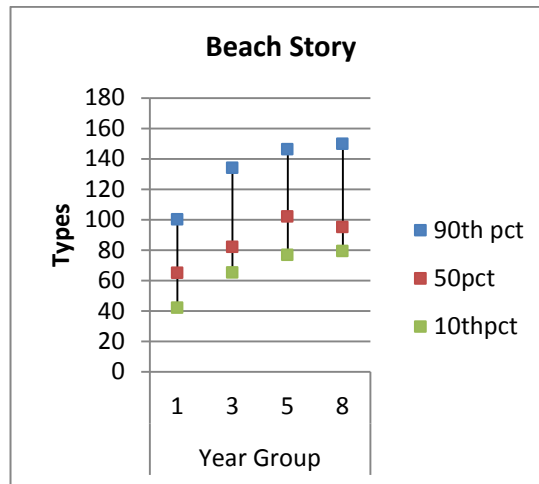
c)

Figures 4.2 Median, 10th and 90th percentile
scores in types for each story by Year groups

As described before, the very unequal sample size of the secondary school sample made it inappropriate for inclusion in the ANOVA analyses. A two-way mixed-design ANOVA compared performance between Year groups and within-child performance between stories. Reiterating our predictions, regarding age, the hypothesis proposed was that differences would be statistically significant. As for the within-child differences, we had hypothesized that the differences for types *would be* statistically significant given the sensitivity of this index to narrative length.

The normality assumption for the use of inferential statistics was examined by age group and story. Kurtosis and skewness z-scores indicated that a few groups did not meet the normality assumption. The following groups had skewness z-scores with positive values over 1.96 ($p < .05$), a cut-off point suggested for small samples (Field, 2009): Year 3 Sweets, Frog and Beach *types*. In contrast, all groups met the kurtosis assumptions. In sum, all Year 1 and 5 groups met the normality assumption, while none of the Year 3 story groups did.

As relatively few groups did not meet the normality assumption, variables were not transformed to favour interpretability.

Homogeneity of variance between age groups was tested using Levene's test; equality of the differences (i.e. sphericity) was tested using Mauchly's test for the within-group comparisons of story. Maulchy's test for the within-subjects main effect of story was not significant, $\chi^2(2) = 0.16$, $p > .05$, indicating that the sphericity assumption was met. However, Levene's test indicated that the assumption of equality of error variances was only met for types in the Beach story, but it was not met for types either in the Sweets or in the Frog stories. Therefore, post-hoc analyses are reported for a procedure that does not assume equal variances.

There was a significant main effect of Age group $F(2,90) = 15.52$, $p < .001$, partial $\eta^2 = .25$. In line with our expectations, in general, the amount of types produced was significantly different by age group. Figure 4.3 shows that older children produced more types than younger children.

There was also a significant main effect of Story $F(2,180) = 129.81$, $p < .001$, partial $\eta^2 = .59$. Also in line with our expectations, in general, types produced were significantly different depending on the story. As seen in Figure 4.3, the Frog story produced the highest types scores, the Sweets story produced the lowest, and the Beach story was right in the middle of both.

The Age by Story interaction was also significant $F(4,180) = 3.02$, $p < .05$, partial $\eta^2 = .06$. This result suggests that the number of types produced by age was significantly different depending on the story; however, the effect size for this interaction was rather small. Figure 4.3 shows that the patterns change depending on the story told, particularly with the Sweets story, where the rate of change for types was slower from Year 1 to 3 than in the other two stories.
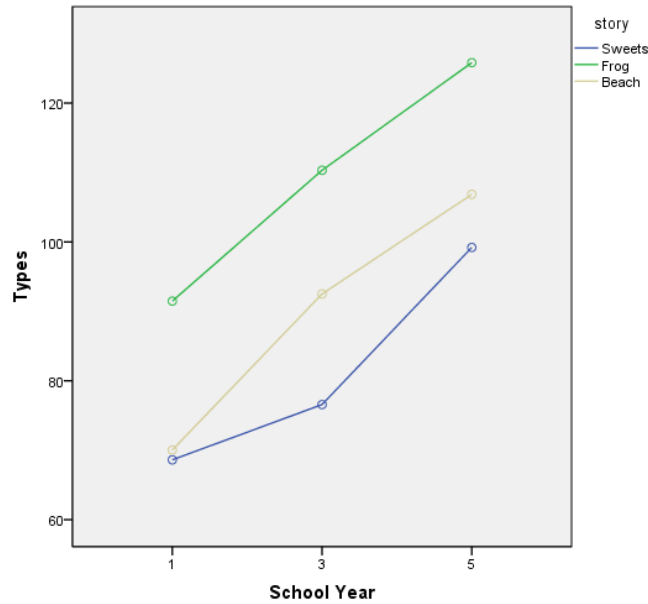
Figure 4.3 Types means by Year group and story

For post-hoc comparisons between age groups the Games-Howell procedure was used, since it does not assume equal variances and can cope with differences in sample sizes. These comparisons indicated that all pairwise comparisons were significantly different: types in Year 1 were significantly lower than those in Year 3 ($p < .01$), and in Year 5 ($p < .001$); and finally, types in Year 3 were significantly lower than in Year 5 ($p < .05$).

Using the Bonferroni correction for multiple comparisons, pairwise comparisons between stories indicated that each story was significantly different from any other in types: as shown in Figure 4.3, Sweets was significantly lower than both Frog ($p < .001$) and Beach ($p < .001$), while Frog was also significantly higher than types in Beach ($p < .001$). In other words, the index types changed depending on the stimulus characteristics.

To evaluate independence from tokens, correlations were computed between types in each story and tokens in each story. Age in months is also included to provide a more fine-grained distinction of the developmental patterns in children. All correlations are shown in Table 4.3.

Table 4.3 Correlations between types and tokens for every story, and with age

	Types in Sweets Story	Types in Frog Story	Types in Beach Story	Tokens in Sweets Story	Tokens in Frog Story	Tokens in Beach Story
Types in Sweets Story	-					
Types in Frog Story	.83[**]	-				
Types in Beach Story	.79[**]	.85[**]	-			
Tokens in Sweets Story	.95[**]	.79[**]	.73[**]	-		
Tokens in Frog Story	.76[**]	.92[**]	.78[**]	.78[**]	-	
Tokens in Beach Story	.74[**]	.82[**]	.95[**]	.71[**]	.82[**]	-
Age in months	.49[**]	.46[**]	.53[**]	.44[**]	.36[**]	.45[**]

**p <.001. Light-gray shaded areas highlight correlations with tokens in the same story; dark-gray shaded areas highlight types' correlations with age.

The relationship between the types and tokens within each specific story was highly significant (Table 4.3, light-grey areas), showing that the more a child talked the more types he/she produced. In other words, the index types in a story is very dependent on the length of that narrative. Types in any story were also related to tokens in the other two stories, suggesting that the index types tends to be influenced by how talkative a child is across stories.

Regarding the relationship with age (Table 4.3, dark-grey areas), types showed moderate correlations with age in months ranging from $r = .46$ in the Frog story to $r = .53$ in the Beach story.

To relate our findings to previous data, two comparisons were made. First, since both our data and the Chen-Wilson corpus used the same picture book as stimulus with slightly different elicitation procedures, a descriptive comparison of the Sweets story from both datasets can be made (see Table 4.4). Then another descriptive comparison is made between a composite of all three stories in our corpus from the present study (listed as the Silva corpus, and also shown in Table 4.4), and both the Chen-Wilson corpus and published data from Justice and colleagues (Justice et al., 2006).

Table 4.4 Comparison of Sweets types and all-stories composite types with previous data

Age	Sweets story Silva corpus		Composite Silva corpus		Age	Sweets story CW corpus		Justice data	
	M	*SD*	M	*SD*		M	*SD*	M	*SD*
5/6	68.62	19.03	76.71	18.19	5	74.50	24.90	39	20
7/8	76.58	20.07	93.14	22.85	7	72.50	14.40	52	28
9/10	99.21	30.04	110.62	28.86	9	88.25	16.66	79	30

NB. Silva composite= types from three stories; CW= Chen-Wilson corpus; Justice et al.= (Justice et al., 2006).

Evidently, both comparisons cannot be subjected to inferential analysis, given that our data did not restrict ages within Year groups, but rather took all ages in each Year group. For example, while the Chen-Wilson corpus had a group of five-year olds, the present study data had a group of Year 1 students, which included five- and six-year-olds.

Regarding the first comparison from the two data sets using the Sweets story, shown in the first and third columns in Table 4.4, there is a trend for the two older groups of children to produce more types with the elicitation procedures used in this study (adult audience, semi-naïve elicitation), than those used in the Chen-Wilson study (peer audience, naïve elicitation) at a descriptive level at least. With the caveats mentioned before, it might be possible that because our data has older children, these higher scores in older children reflect developmental rather than elicitation differences.

Nonetheless, the youngest group produced fewer types in our data than those in the database study, even with relatively older children (Year 1 includes some 6-year-olds), which could suggest that ages might not be as mismatched as they appear.

For the second comparison, shown in the second and fourth columns in Table 4.4, a composite mean was obtained from all three stories, and it was compared against the database study and the Justice and colleagues' study (2006). In this comparison, composite types in the Silva corpus appear larger than in the Chen-Wilson and the Justice et al. corpora in all age groups, not just the last two. This is likely the reflection of using a longer story (the Frog story) which tends to produce longer narratives, which

in turn influences the types scores. However, the possibility that these differences might reflect age differences between the datasets, rather than differences due to our elicitation methods, should still be considered as it cannot be ruled out.

4.3.1.2 VOCD results

Table 4.5 shows means and standard deviations for VOCD by story and age. Descriptive statistics for VOCD illustrate a similar though much more subtle pattern of developmental change: while the mean increases up to Year 5 are less obvious, the changes between Year 5 and Year 8 show a slight decrease as seen in Table 4.5.

Table 4.5 VOCD for narratives from each picture book by Year group

School Year	1	3	5	8
Ages	5-6	7-8	9-10	12-13
	n=29	n=31	n=33	n=15
Sweets Story				
Mean	28.92	34.82	39.4	38.47
SD	8.40	7.73	10.79	9.79
Range	11.51-47.12	22.15-60.50	20.25-63.22	26.29-61.12
Frog Story				
Mean	26.47	32.3	38.35	37.17
SD	8.16	7.37	13.6	10.26
Range	12.81-46.20	17.54-45.76	18.94-76.94	23.81-59.95
Beach Story				
Mean	28.41	36.6	40.94	39.65
SD	7.03	7.64	10.9	10.32
Range	14.25-44.54	20.88-52.08	26.18-69.55	26.22-66.05

On the following page, Figures 4.4a, 4.4b and 4.4c illustrate, for each story, the median, 10th and 90th percentile by age, to get a glimpse of the variability by each Year group. It is worth noting that, when using the median in Figure 4.4, the pattern is almost the

same as the one shown with the means in Table 4.5, with the exception of the change between Year 5 and secondary Year 8, where instead of a slight decrease a slight increase is shown.

Figures 4.4a, b and c also serve to show graphically how the within-group variability seems to increase from Year 3 to Year 5, but seems more stable from Year 5 towards Year 8.
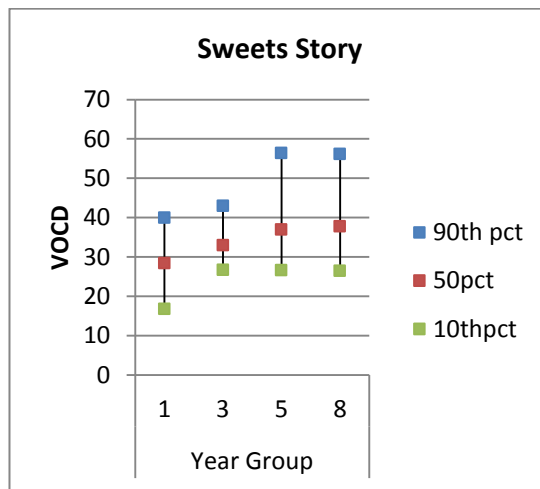
a)
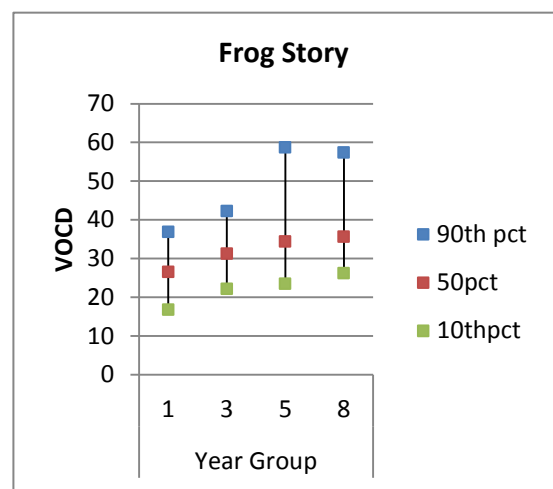
b)
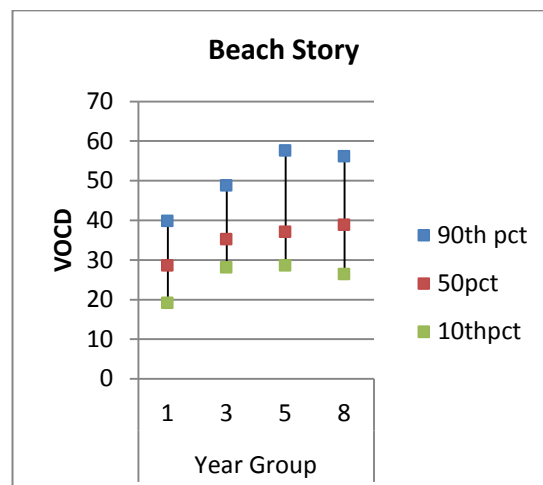
c)

Figures 4.4 Median, 10th and 90th percentile scores in VOCD for each story by Year groups

Another two-way mixed ANOVA again compared performance between age groups and within-child performance between stories. We expected to find age differences, but no differences between the stories.

Normality assumptions were tested by looking at skewness and kurtosis z-scores. The following groups had skewness z-scores in VOCD with positive values over 1.96 ($p <$ .05): Year 3 Sweets, and Year 5 Frog and Beach. Except for Year 3 Sweets which displayed a leptokurtic or 'peaked' distribution, all the rest of the groups met the kurtosis assumption. Once more, since deviations from normality were few and mild, variables were not transformed.

Levene's test indicated that the assumption of homogeneity of variance was met for VOCD in the Sweets story, but not for either the Frog or the Beach stories; therefore, post-hoc procedures were chosen that did not assume homogeneity. Mauchly's test indicated that the assumption of sphericity was met $\chi^2(2)=.69$, $p >.05$.

A significant main effect of Age was also found: $F(2,90) = 14.28$, $p < .001$, partial $\eta^2 = .24$. As expected, VOCD scores were different in the different age groups: Figure 4.5 shows older children producing higher VOCD scores than younger children across stories.

A significant main effect of Story was found: $F(2,180) = 9.22$, $p < .001$, partial $\eta^2 = .09$. Contrary to expectations, VOCD was significantly different when using different picture books; however, the effect size was rather small. Figure 4.5 shows how stories elicit different responses across children depending on the stimuli and the index, since the Frog story produced the lowest scores for VOCD compared to the other two stories, while it had produced the highest scores for types.
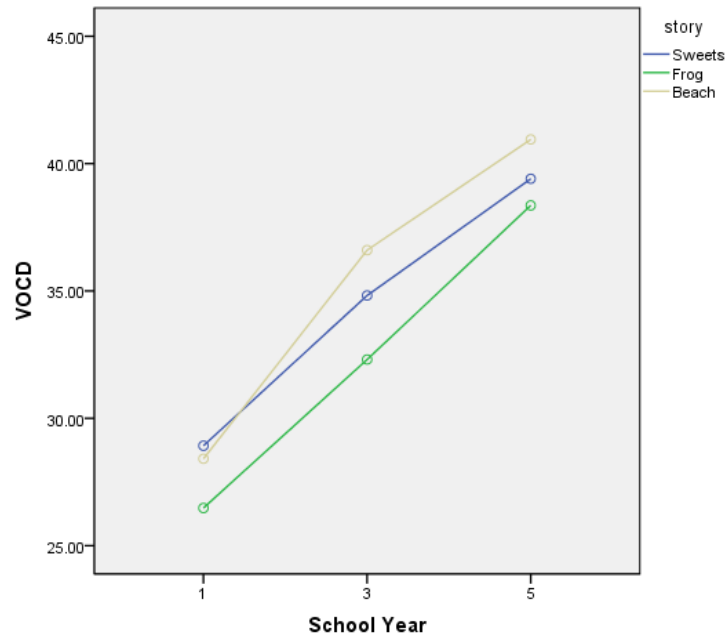
Figure 4.5 VOCD means by Year group and story

With VOCD, the Age by Story interaction was not significant $F(4,180) = .84$, $p > .05$, suggesting the change in VOCD produced by age did not change depending on the story; Figure 4.5 shows how VOCD patterns are more parallel across stories, reflecting the lack of a significant interaction.

Since equal variances could not be assumed, post-hoc comparisons between age groups were carried out using the Games-Howell procedure. While Year 1 VOCD scores were significantly lower than both Year 3 ($p <.01$) and Year 5 ($p <.001$), Year 3 and Year 5 were not significantly different ($p >.05$). Figure 4.5 shows how even when Year 3 and Year 5 appear different, this difference is smaller than the one between Years 1 and 5.

Pairwise comparisons within groups using Bonferroni indicated that VOCD in the Frog Story was significantly lower than both in the Sweets ($p <.05$) and in the Beach stories ($p <.001$). However, VOCD in the Sweets and Beach stories were not significantly different from each other ($p >.05$), also clearly seen in Figure 4.5.

In sum, VOCD displayed still significant variation according to the story and stimulus used, but this variability was lower than for types. On the other hand, it found variability between Year group 1 and 3, but failed to find significant differences between Year group 3 and 5.

143

Correlations with tokens and age in months were performed to examine independence from tokens and another facet of the developmental picture. These correlations are shown in Table 4.6.

Table 4.6 Correlations with tokens and age in months

	VOCD in Sweets Story	VOCD in Frog Story	VOCD in Beach Story	Tokens in Sweets Story	Tokens in Frog Story	Tokens in Beach Story
VOCD in Sweets Story	-					
VOCD in Frog Story	.79**	-				
VOCD in Beach Story	.78**	.81**	-			
Tokens in Sweets Story	.66**	.56**	.53**	-		
Tokens in Frog Story	.55**	.49**	.43**	.78**	-	
Tokens in Beach Story	.64**	.57**	.53**	.71**	.82**	-
Age in months	.45*	.46**	.51**	.44**	.36**	.45**

** $p < .001$. Light-grey shaded areas highlight correlations with tokens in the same story; dark-grey shaded areas highlight types' correlations with age.

Correlations with tokens (Table 4.6, light-grey areas) showed varying degrees of relationships depending on the story, ranging from the lowest correlation $r = .49$ in the Frog Story to the highest in the Sweets story $r = .66$ (both $p < .001$). It is worth noting that Frog stories had a tendency to be the longest while Sweets stories had a tendency towards being the shortest. This could suggest that VOCD is more independent as the stories become longer.

The moderate correlations with age in months (Table 4.6, dark-grey areas) showed a similar pattern to types, ranging from $r = .45$ in the Sweets Story to $r = .51$ in the Beach story.

Finally, VOCD values were compared to those obtained in the previous chapter from the Chen-Wilson corpus. Reiterating, these are not subjected to inferential analysis, but

are shown to illustrate the behaviour of VOCD in different elicitation contexts and populations.

Table 4.7 shows VOCD scores from the Sweets story, the composite VOCD score from the three stories and the VOCD obtained previously from the Chen-Wilson corpus analysed in Chapter 3.

Table 4.7 Comparison of Sweets VOCD and all-stories composite VOCD with previous data

Age	Sweets story Silva corpus		Composite Silva corpus		Age	Sweets story CW corpus	
	M	SD	M	SD		M	SD
5/6	28.92	8.40	27.93	7.50	5	34.90	9.73
7/8	34.82	7.73	34.57	6.24	7	32.89	9.04
9/10	39.40	10.79	39.57	11.09	9	38.75	7.92

NB. CW= Chen-Wilson corpus

Comparing VOCD between both single elicitations of the Sweets story (first and third columns in Table 4.7), the present data shows higher scores for the 7- and 9-year-olds, but lower scores for the 5-year olds, even when our data includes children of 5 and 6 years of age.

Also interesting is the comparison between the single Sweets elicitation and the composite VOCD from all stories, both in our corpus, which we have termed the Silva corpus (first and second columns in Table 4.7): in VOCD, there is not much change, which could suggest that a single efficient elicitation of the Sweets story could provide for a robust measure of VOCD.

4.3.2 Discussion of lexical indices

On average, all stories grew larger in tokens up to Year 5, or ages 9/10. As pupils became older, at least in primary school, they talked more with the elicitation procedures used here. Types and VOCD also grew up to Year 5. For the small Year 8 secondary sample, narratives became shorter and there was a decline in both lexical

indices; however this should be interpreted in the light that it was a small single-school, mostly male sample.

Inferential statistics and comparisons with previous data were only used for those groups with sufficient data, that is, Years 1, 3 and 5. The proposed hypothesis was that developmental differences would be found in both types and VOCD. Although effect sizes were remarkably similar (partial η^2 of .25 for types, and .24 for VOCD), and correlations with age in months showed similarly moderate associations, post-hoc comparisons revealed that, for types, all Year groups were significantly different from each other, while VOCD failed to find significant differences between Year 3 and Year 5, that is between 7/8 and 9/10 years of age.

When considering intra-child performance to examine each index's reliability, it had been hypothesised that while types would differ by story, VOCD would be more consistent (i.e., no significant differences would be found) across stories. The results showed that actually both lexical indices were influenced by the kind of stimulus used, but the effect size for the main effect of Story was much lower for VOCD (partial η^2 =.09) than for types (partial η^2 =.59), somewhat in line with our expectations. Post-hoc comparisons provided a partial confirmation of our hypothesis: while every story was significantly different from each other in types, when using VOCD only the Frog story was significantly different from the rest. In other words, VOCD in the Sweets and the Beach stories were not statistically different. Therefore, the age differences found in each index were related to the kind of stimulus in types, and to a lesser extent in VOCD, suggesting, once more, that the latter offered a more consistent picture of the child's lexical repertoire.

When comparing how these two indices relate to productivity, types were highly correlated with tokens, while VOCD was still significantly but more moderately correlated with tokens. This pattern applied not only when correlating each index within a story, but also between the index in one story and tokens in the other two stories. As with the analyses presented in Chapter 3, correlations with tokens were much higher than those reported before ($r = .22$) by McCarthy and Jarvis (2007).

On balance, from both ANOVAs and correlational analyses, it seems that the slight advantages in developmental sensitivity gained by types are offset by a considerable loss in consistency when using different stimuli, while the opposite seems to be the case for VOCD.

Of course the proper application of VOCD depends on having a range of tokens adequate for the kind of sampling performed by this index. McCarthy and Jarvis (2007, p. 482) had proposed an optimal range of 100-400 tokens. Although the full range of the data sampled here goes from 73 to 755 tokens, the vast majority of the narratives fall in the optimal range proposed by these researchers, as shown in the Figure 4.1 which exclude the lowest and highest 10% of the data. With the exception of the Frog story, which tended to have some narratives longer than 400 tokens in Years 5 and 8, almost all narratives in the Sweets and Beach stories fell in that 100-400 range for the ages sampled here.

Finally the comparison of both indices with previous data from Chapter 3, and previous published data (Justice et al., 2006), illustrated how these indices behave at a descriptive level, using different populations and elicitation methods. A caveat for these comparisons is that participants selected for the present study were not selected in chronological age ranges, as in the other two studies used for comparison, but rather were chosen from school year cohorts. For simplicity they are referred here by the age group, rather than the Year group.

For types, two comparisons were made: a) one pairing the Sweets data obtained here with the Sweets data in the Chen-Wilson corpus examined in the previous chapter; and b) another where a composite measure across stories was compared against the measure in the Chen-Wilson corpus and published data. Using the same picture book as stimulus, but different elicitation methods, the youngest group of 5- and 6-year-olds in the present study (adult listener, semi-naïve elicitation) produced on average fewer types than 5-year-olds in the Chen-Wilson corpus (peer listener, naïve elicitation). For the older groups, the opposite pattern occurred: 7-8-year-olds and 9-10-year-olds produced on average more types than 7-year-olds and 9-year-olds in the Chen-Wilson corpus. Nonetheless, when a composite for types across stories is used, all groups produced on average more types in the present study, than in the Sweets story in the Chen-Wilson corpus, and in the published Justice and colleagues' (2006) data. Although these descriptive comparisons seem to favour the elicitation methods used here, no definitive conclusions can be reached as the ages are not an exact match and cohort effects cannot be completely ruled out.

For VOCD, the youngest group also produced on average lower scores in the present study than in the Chen-Wilson corpus, while the older two groups produced on average

higher scores than their counterparts in the Chen-Wilson corpus. This happened regardless of whether the comparison was with the single Sweets story or the composite of VOCD across the three stories collected for the present study, suggesting a reliable pattern using this index. This developmental pattern in VOCD also explains why significant differences were found in the present study between ages 5 and 7, which had not been found in the database study from the Chen-Wilson corpus.

In sum, the overall picture from the different analyses reveals that VOCD displays developmental sensitivity when using these stimuli and these elicitation procedures up to age 7 and 8. Although types can display more variability in older children, VOCD seems to be a much more stable measure as reflected in our within-subject analyses across stories, VOCD's independence from tokens and from the comparisons across different populations.

4.4 Grammatical indices

Although grammar is strictly composed of morphology and syntax (Crystal, 2008), results reported in Chapter 3 suggested that distinguishing morphemes was not particularly useful for school-aged children in terms of showing a different developmental pattern from that obtained just from words. For that reason, the analysis presented here is focused exclusively on syntax, without morphological divisions. Developmental differences and within-child differences were examined using a single syntactic index calculated using CLAN: Mean Length of Utterance in T-Units (Hunt, 1965, 1970) in words.

4.4.1 Syntax results

Means, standard deviations and ranges for MLU in T-Units in words (MLT-w) are shown in Table 4.8 for all Year groups.

Table 4.8 MLU in T-Units in words for narratives from each picture book by Year group.

School Year	1	3	5	8
Ages	5-6	7-8	9-10	12-13
	n=29	n=31	n=33	n=15
Sweets Story				
Mean	7.74	9.01	10.01	10.52
SD	1.42	1.89	1.84	1.44
Range	5.50-10.91	6.00-14.70	6.84-15.05	8.81-14.00
Frog Story				
Mean	7.29	8.50	9.08	10.30
SD	1.18	1.49	1.22	1.23
Range	5.58-10.08	6.37-12.00	7.31-13.16	8.00-12.04
Beach Story				
Mean	7.48	8.96	9.40	11.36
SD	1.44	1.33	1.47	1.88
Range	5.21-11.27	6.18-11.78	7.16-14.60	8.25-14.50

From the table, there is subtle growth in MLT-w in every Year group, even in Year 8.

More detailed illustrations of these patterns using the median, 10th and 90th percentiles are shown in Figures 4.6a, b and c below. The variability from the graphs shown in these figures displays a similar general pattern to the one shown in the database study described in the previous chapter. In contrast to lexical indices, variability also seems more uniform across age groups, with the exception of Year 8 narratives of the Beach story, seen in Figure 4.6c.
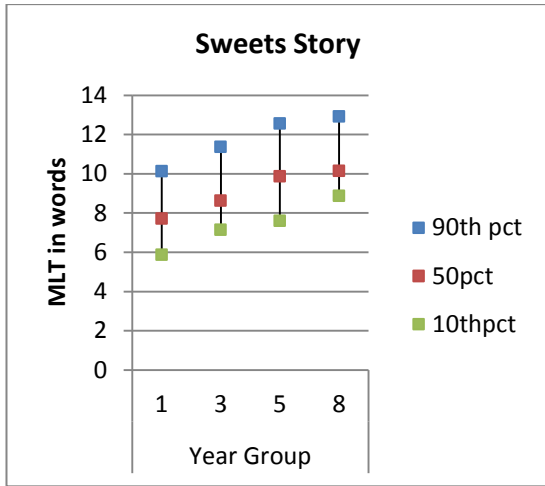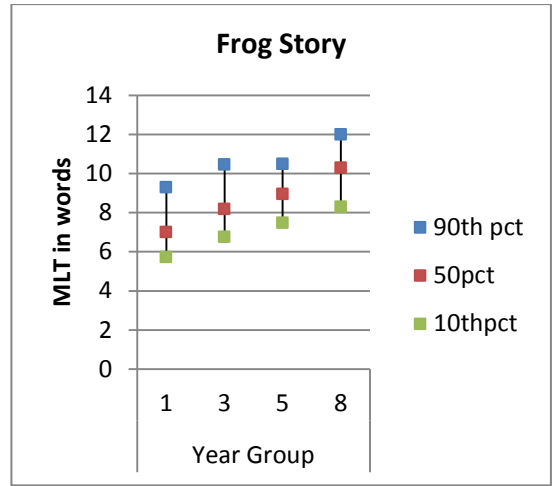
a)   b)
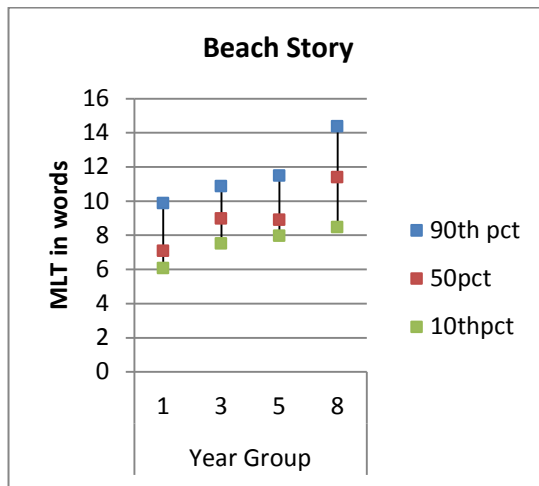
c)

Figures 4.6 Median, 10th and 90th percentile
scores in MLT-w for each story by Year groups

For inferential statistics, the narratives from Year 8 are excluded, for the reasons described in Section 4.2. A two-way mixed-design ANOVA compared the intra-child performance and the scores in this syntactic index by Year group. From the previous study results, where MLT-w seemed a fairly robust method across populations and elicitation methods, we predicted that no significant differences would be found between stories. Regarding comparisons between age groups, even when the previous database study did not find differences between ages 7 and 9 we expected that, with our different elicitation methods aimed at maximising literate language, we would be able to find significant differences with older children producing longer T-Units.

Normality was tested by age group and story. Kurtosis and skewness z-scores indicated that a few groups did not meet the normality assumption, with z-scores over 1.96 in skewness or kurtosis. The following groups showed significant positive skew in MLT-w: Year 1 Beach story, Year 3 Sweets story, and Year 5 Frog and Beach stories. These specific groups in Year 3 and 5 did not meet the kurtosis assumption either. As before, transformations were not carried out to favour interpretability of results.

Levene's test showed that for MLT-w equality of variance could be assumed. However, Mauchly's test was significant, $\chi^2(2) = 13.35$, $p < .01$, indicating that the assumption of sphericity was not met. Therefore, the Huynh-Feldt correction is reported for the within-subject main effects.

A significant main effect of Age was also found, $F(2,90) = 18.18$, $p < .001$, partial $\eta^2 = .28$, suggesting across stories, performances in the Year groups were significantly different. Figure 4.7 shows that older children produced longer T-Units in words than younger children in all stories.

A significant main effect of Story was found, Huynh-Feldt $F(1.82,164.46) = 11.55$, $p < .001$, partial $\eta^2 = .11$. Independently of age, the syntactic complexity as measured by T-Units was significantly different by story narrated. Here, Figure 4.7 also shows that some stories produced longer T-Units in words than others. The interaction between Story and Age was not significant, Huynh-Feldt $F(3.65,164.46) = 1.11$, $p > .05$, indicating that the developmental change in syntactic performance did not change depending on the story, also clearly seen in Figure 4.7, where the developmental patterns are quite parallel.
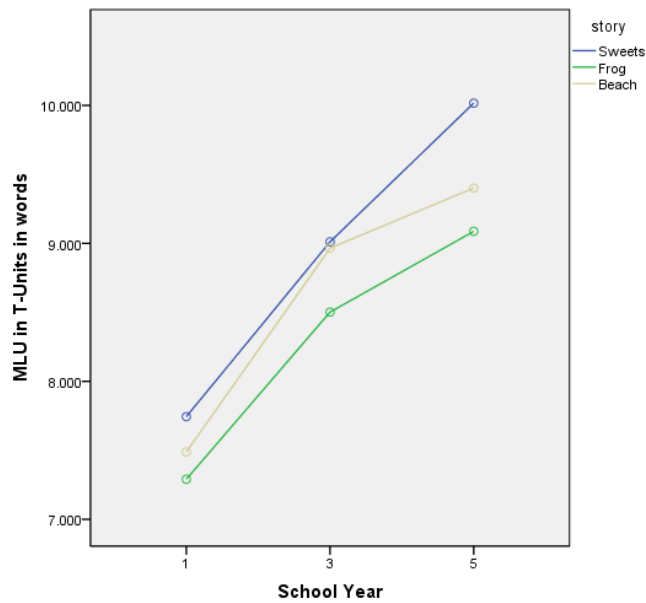
Figure 4.7 MLT-w means by Year group and story

To find which ages and which stories were significantly different post-hoc tests were performed. For differences between Year groups, Gabriel's procedure is reported which can cope with slight differences in sample sizes (Field, 2009). Post-hoc tests showed that Year 1 syntactic complexity was significantly lower from both Year 3 ($p < .01$) and Year 5 ($p < .001$). However, even with the specific procedures used in this study, performance in Year 3 and Year 5 were still not statistically different from each other ($p > .05$). These differences are clearly shown in Figure 4.7, where growth was slowing down in Year 5.

Bonferroni post-hoc correction was used as it is more robust with deviations from sphericity (Field, 2009). Comparisons between stories with Bonferroni correction for multiple comparisons showed that MLT-w in the Frog story were significantly different from those in the Sweets ($p < .001$) and the Beach stories ($p < .01$), but that these last two were not significantly different from each other ($p > .05$). Figure 4.7 displays how T-Units in words were longer in the Sweets and the Beach story than in the Frog story.

Correlations were used to examine how the syntactic complexity was independent from tokens and to corroborate its association with age in months, both shown in Table 4.9.

Table 4.9 Correlations between MLT-w and tokens in every story, and age in months.

	MLU-T in Sweets Story	MLU-T in Frog Story	MLU-T in Beach Story	Tokens in Sweets Story	Tokens in Frog Story	Tokens in Beach Story
MLU-T in Sweets Story	-					
MLU-T in Frog Story	.73**	-				
MLU-T in Beach Story	.70**	.79**	-			
Tokens in Sweets Story	.40**	.32**	.30**	-		
Tokens in Frog Story	.31**	.41**	.39**	.78**	-	
Tokens in Beach Story	.33**	.38**	.43**	.71**	.82**	-
Age in months	.49**	.49**	.46**	.44**	.36**	.45**

$^{**}p$ <.001. Light-grey shaded areas highlight correlations with tokens in the same story; dark-grey shaded areas highlight types' correlations with age.

Correlations with tokens (Table 4.9, light-grey areas) showed that MLT-w was moderately related to tokens, but this association tended to be fairly consistent across stimuli. Although slightly higher, correlations with age in months (Table 4.9, dark-grey areas) were also moderate in size, but once more, relatively consistent across stories.

Finally, a comparison of syntactic indices contrasted the values obtained in this study to those from the previous chapter from the Chen-Wilson corpus and from published data (Justice et al., 2006). Table 4.10 shows the MLT-w values obtained from the single story Sweets, a composite of this corpus (Silva corpus), and scores from the Chen-Wilson corpus and from the Justice data.

Table 4.10 Comparison of Sweets MLT-w, composite MLT-w, Chen-Wilson MLT-w and Justice *et al.* MLT-w

Age	Sweets story Silva corpus		Composite Silva corpus		Age	Sweets story CW corpus		Justice data	
	M	SD	M	SD		M	SD	M	SD
5/6	7.74	1.42	7.50	1.20	5	6.79	1.27	6.8	1.7
7/8	9.01	1.89	8.82	1.31	7	7.89	1.25	8.5	3.8
9/10	10.01	1.84	9.50	1.39	9	8.20	1.43	8.4	1.4

NB. CW= Chen-Wilson corpus

In descriptive terms, comparing both single elicitations of the Sweets story, the average T-Unit produced by children in the present study was longer than those from the Chen-Wilson corpus reported in the previous study using the same stimulus, and it was also longer than the average reported from the Justice data from a single picture elicitation (all shown in Table 4.10). Even the composite MLT-w score (second column in Table 4.10) tended to be higher in the present study than in both previous studies. Of course, the limitation to this interpretation is that our study was not restricted to chronological ages but collected data from school year cohorts.

A much more matched comparison was possible with the MLT-w data provided in the ERRNI Manual (Bishop, 2004). Normative data in the Manual showed percentiles in different age groups, so mean scores were not available. A narrow interval of scores rather than a single number is provided for each percentile, so the closest number to the 50th percentile was obtained. Then, the median age for each of the Year groups in the present study was compared to the corresponding age bracket's median MLT-w in the Manual. Both sets of descriptive scores are listed in table 4.11.

Table 4.11 Comparison of Beach story median MLT-w and Beach story median MLT-w from ERRNI Manual

Year Group	Median Age	Beach story Silva corpus	Age bracket in ERRNI Manual	Beach story ERRNI data
		Median		Median
1	6;03	7.07	6;00-6;05	7.75-7.99
3	8;01	8.96	8;00-8;05	8.50-8.74
5	10;04	8.88	10;00-10;11	9.25-9.49

Table 4.11 shows slightly higher values for Year group 3 than those provided in the ERRNI Manual, while it shows lower values in Years 1 and 5, possibly signalling fatigue effects in the youngest and oldest groups.

4.4.2 Syntax discussion

On average, the length of T-Units grew with age, even in Year 8, or ages 12-13, in contrast to the lexical indices. Therefore, at a descriptive level, the elicitation procedures used here were successful in obtaining progressively more complex T-Units.

Only the groups where we had pooled data were subjected to inferential statistics. For the examination of developmental patterns, we had expected that differences would be statistically significant even between Year groups 3 and 5. Although we had not found significant differences in the previous database study between the ages 7 and 9, we still expected that our elicitation procedures could have facilitated the production of more complex T-Units than those produced for a peer listener.

Finding significant differences between Year groups 1 and 3, corresponding to ages 5/6 and 7/8 replicated the previous results. However, even when we had expected to find significant differences with older pupils in Year 5 with our emphasis on producing decontextualized language, results indicated that between Year groups 3 and 5, or ages 7/8 and 9/10, the differences were not significant, replicating the results from the database study. These results suggest that this pattern is quite robust, and growth in T-Units does taper off at these ages, at least with the elicitation procedures used here and those used for the Chen-Wilson corpus.

Interestingly, the effect size for the main effect of developmental differences in MLT-w (partial η^2 = .28) was quite similar to that found in both lexical indices examined, namely types and VOCD. MLT-w also showed similar correlations with age in months to those displayed by the lexical indices. These correlations with age in months were lower than those reported in the database study, but it should be considered that the previous study included the preschool stage of accelerated language growth at 3 and 4 years.

On the other hand, regarding intra-child variability, we had expected to find no significant differences between stories. The main effect of story was, contrary to expectation, significant, but its effect size was quite low (partial η^2 = .11), again suggesting that the choice of picture-book and elicitation procedures had a slight influence on the syntactic performance of the children. Our no-difference hypothesis was partially confirmed for the Sweets and Beach stories, where syntactic performance was statistically equivalent, but performance in the Frog story was significantly lower than both. Given the ubiquity of the Frog story in narrative research, it is interesting to note that this might not be the ideal stimulus for obtaining a reliable measure in T-Units, if used on its own. Although some children managed to produce quite complex T-Units, for most children this was not the story that brought out the most sophisticated syntactic performance, with the Sweets and the Beach stories eliciting the higher values for T-Units.

In addition, developmental differences were not affected by the kind of stimulus used, a characteristic that also makes MLT-w a fairly consistent index.

When examining how these indices are affected by narrative length, the correlations with tokens observed were slightly lower (r = .40 to .43) than those obtained in the database study (r = .56). However, these correlations were still significant; in other words, T-Units in a narrative are still influenced by how much school-aged children talk.

Finally, the first comparison with previous studies could not address whether our elicitation procedures were capable, at least in descriptive terms, of actually enhancing the production of literate language, given that the ages did not correspond to an exact match, and our scores were higher but so were the ages of the children in the current study. Whether the discrepancy of a full T-Unit between the data presented here and the previous studies can be explained by the age mismatch or by the specific elicitation procedures used here is an issue that could be explored further in future research. Of course, cohort effects cannot be ruled out, and these difficulties highlight the complexity of measuring expressive language.

In a second comparison with data from a UK nationally representative sample, a better match of ages was achieved, and it showed that Year 3 reached higher median scores than the normative data while Years 1 and 5 scored lower median values. Of course, there were important differences in the elicitation of normative data and our data. First

of all, even when the elicitation instructions from the ERRNI Manual were followed, one specific instruction was not followed: the test administrator did not point to every single drawing as instructed. This was done because doing so would have eliminated the distance which had been created with the screen in the previous two stories for the purpose of eliciting the most decontextualized language.

However, a most important reason gathered from observations while collecting data could have been the occurrence of fatigue effects. The Beach story was always collected as the last of three, so children could have become tired after producing two stories already. This pattern arose particularly for Years 1 and 5, also noticeable in Table 4.1 showing how long children talked. From our observations, some Year 1 children became actually weary by the last story, while for some Year 5 pupils the task might have become predictable and they might have lost interest at the end.

It is also possible that some order effects occurred in the opposite direction, benefiting the Beach story with some residual effects of the semi-naïve elicitation setup, but this was not directly examined: as children were asked to produce narratives behind a screen for the first two stories, they might have tried to produce clearer narratives than if they lacked the experience of an arguably unaware listener. This possibility was not validated, but it is still feasible, given that some pupils expected the test administrator to 'set the screen up' for the last story.

Given these elicitation procedures' differences with the ERRNI Manual, and still finding that results were to some extent similar, it could be argued that MLT-w is quite robust across tasks and populations.

Overall, MLT-w seems to offer good developmental sensitivity up to ages 7/8, certain independence from narrative sample size, and some consistency when using different stimuli and different elicitation procedures. Even when a significant difference was not found between ages 7/8 and 9/10, the replication of this finding from the previous database study suggests that this deceleration of growth in terms of T-Units is more likely to happen in the school age population.

4.5 Discourse measures

The focus of the database study reported in Chapter 3 was on those purely linguistic indices that were retrievable from language analysis software because they were the ones whose use was more contentious in terms of validity. In this study of primary data however, two of the three narratives produced by each child were also analysed at a global level, to obtain a measure for expressive discourse skills. First, the second narrative, the Frog story, was analysed using the Index of Narrative Complexity (Petersen et al., 2008), a measure of the story's structure. Then, the third and last story, the Beach story, which came from a standardised assessment, the Expression, Reception and Recall Narrative Instrument (Bishop, 2004), was analysed using an information checklist, Information Content, which is not in itself a structure measure but is rather a measure of how much of the story's elements listed in a standardised form are present in the child's narrative.

4.5.1 Results

Before addressing the results, it is worth restating the differences between the original measures and the application of these measures in this study. The INC was coded only for the eight elements corresponding to the story structure. Meanwhile, a difference in the administration of ERRNI was that children were not aided by pointing, as the manual requires, to maintain some degree of the distance created for the first stories with the visual barrier.

4.5.1.1 Index of Narrative Complexity results

Table 4.12 shows descriptive statistics for the 108 Frog stories coded using the story structure part of INC, by each Year group.

Table 4.12 Index of Narrative Complexity for the Frog story by Year group.

School Year	1	3	5	8
Ages	5-6	7-8	9-10	12-13
	n=29	n=31	n=33	n=15
Mean	11.83	12.61	13.61	13.00
SD	1.67	2.21	2.51	1.89
Range	8-14	8-17	9-18	10-17

This table displays continuous but slow growth up to Year 5. In other words, older children up to ages 9 and 10 showed higher scores in INC. Variability as shown by standard deviations was highest for Years 3 and 5. The secondary group Year 8 showed a slight decrease when using this index.

Figure 4.8 shows the Median, 10th and 90th percentiles for the four Year groups, to better illustrate the variability in each age group.
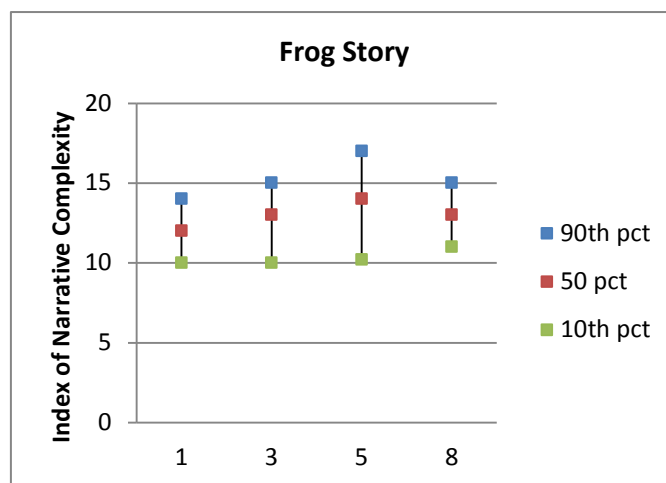
Figure 4.8 Median, 10th, and 90th percentile scores in
Index of Narrative Complexity for the Frog story by Year group.

The patterns in Figure 4.8 illustrate that while the median keeps growing up to Year 5, low performers in story structure using the Index of Narrative Complexity are at virtually the same level in Years 1, 3 and 5.

For the reasons described before, Year 8 pupils were excluded from inferential analyses. Normality was examined with skewness and kurtosis z-scores by Year group. All groups met both normality assumptions.

Levene's test was used to test for equality of variances, and it showed that this assumption was met. In other words, the variances were not significantly different.

A one-way between-subjects ANOVA was carried out to compare the performance of Year groups 1, 3 and 5. Since the data from Petersen et al. (2008) did not provide developmental patterns, but rather the effects of intervention, this analysis was only exploratory.

A significant main effect of Age group was found, $F(2, 90) = 5.20$, $p < .01$, partial $\eta^2 = .10$. As Table 4.12 shows, older children produced higher INC scores than younger ones. Post-hoc comparisons using the Gabriel procedure for unequal sample sizes (Field, 2009) indicated that only one significant difference existed: Year 5 was significantly higher than Year 1 ($p <.01$), also shown in Table 4.12. Neither of the other two comparisons with the middle Year 3 was significantly different (both $p > .05$), reflecting the very subtle changes in development.

We were also interested in how this discourse-level measure was affected by narrative length in tokens. Pearson correlations indicated that the Index of Narrative Complexity was significantly associated with tokens ($r =.60$, $p <.001$). In other words, the measure INC was sensitive to narrative length, with children producing longer stories also scoring higher on the INC.

Correlations with a finer age measure, age in months, showed a small, though still significant association ($r =.36$, $p <.01$), converging on the small effect size of the Year group effect from the ANOVA analysis.

Since only the macrostructure portion of the INC scoring rubric was used, and data on the Petersen et al. (2008) paper was not disaggregated by age, no comparisons were made with the previous published data.

4.5.1.2 Information Content results

Table 4.13 offers descriptive statistics for the measure of Information Content, calculated on the Beach story by Year group.

Table 4.13 Information Content for the Beach story by Year group.

School Year	1	3	5	8
Ages	5-6	7-8	9-10	12-13
	n=29	n=31	n=33	n=15
Mean	17.34	19.00	26.64	26.53
SD	7.28	6.02	4.06	5.75
Range	4-35	10-34	18-36	16-39

The table shows continuous growth up to Year 5. The range of variability was greatest in the youngest group, with some pupils scoring very low and also very high, with some managing to express most of the content in the story. Figure 4.9 illustrates the median, 10th and 90th percentile for this score.
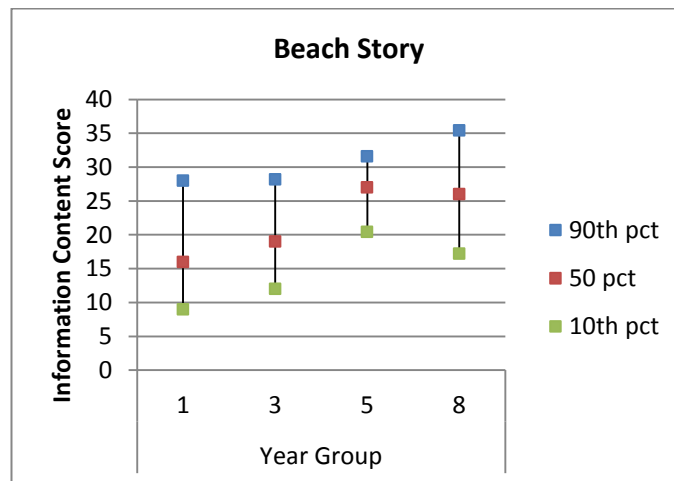
Figure 4.9 Median, 10th, and 90th percentile scores
in Information Content for the Beach Story by Year group

Inferential statistics were carried out for Year groups 1 through 5 only. Normality was assessed with z-scores for skewness and kurtosis by age group. In this measure, like in the Index of Narrative Complexity, all groups met the normality assumption.

The assumption of equal variances was examined using Levene's test, which indicated that this assumption was not met, so post-hoc analyses were chosen accordingly.

A one-way between-subjects ANOVA was conducted to examine developmental differences. As no precedent existed for this analysis, there was no hypothesis and the analysis is purely exploratory.

There was a significant main effect of Age group: $F(2,90)=22.56$, $p <.001$, partial $\eta^2 = .33$. Games-Howell post-hoc analyses are reported, which do not assume equal variances. Year 5 was significantly higher than either Year 1 ($p <.001$) or Year 3 ($p <.001$); Years 1 and 3 were not significantly different from each other ($p >.05$).

Pearson correlations examined the association with narrative sample size in tokens and age in months. Information content was significantly related to both tokens in the Beach story ($r =.67$, $p <.001$) and age in months ($r =.53$, $p <.001$). Just like the indices examined before, this proxy for discourse skills was sensitive to narrative length, but it was also positively related to age, with older children producing higher scores than younger children.

Pearson correlations were also evaluated to look into the relationship between the Index of Narrative Complexity applied to the Frog story and the Information Content measure applied to the Beach story. This association was also significant ($r = .43$, $p < .001$).

A comparison with ERRNI's (Bishop, 2004) normative data for Information Content is presented in Table 4.14. For each of the Year groups median ages were obtained, and for each of the median ages in our dataset, median values from the manual are reported from those age brackets that more closely matched each of the median ages in the present study.

Table 4.14 Comparison of Median Beach Story Information Content
and Median Beach story Information Content from ERRNI

Year Group	Median Age	Beach story Silva corpus	Age bracket in ERRNI Manual	Beach story ERRNI data
		Median		Median
1	6;03	16.0	6;00-6;05	16
3	8;01	19.0	8;00-8;05	21
5	10;04	27.0	10;00-10;11	25

Table 4.14 shows that Information Content median raw scores were quite similar to those from the normative sample (Bishop, 2004). Year 1 scores were identical in the current study and the norms. Meanwhile, Year 3 scores were slightly lower than norms and Year 5 scores were slightly higher than norms.

4.5.2 Discussion of discourse measures

Two measures of expressive discourse were examined. First, the Index of Narrative Complexity or INC was used to identify developmental patterns in school-aged children from ages 5 to 10. Even when this tool was designed to capture discourse-level organization with several levels of complexity based on pivotal story structure work (Labov, 1972; Peterson & McCabe, 1983; Stein & Glenn, 1979), results suggested this measure was only able to differentiate between Year 1 and Year 5. Although the analysis was exploratory, the unclear developmental trend observed suggests that even if this tool is sensitive to the effects of very specific intervention efforts when applied in

163

its entirety (Petersen et al., 2008), and it showed great variability within cohorts of children as shown here, it might not be optimal for identifying patterns of development, at least in typically developing children.

The other measure based on an informational checklist, Information Content, applied to the Beach story from the standardised test ERRNI showed significantly higher scores for pupils Year 5, or ages 9/10, relative to those in Years 3 or 1, although a great deal of variability exists in the youngest group. Since this measure awards points based on the entirety of a story, it required the pupil to describe each picture and make the corresponding inferences. A possible interpretation of this pattern of large initial variability and later reduced variability, but increased measures of central tendency, is that some children appear to be more receptive than others at the beginning of formal schooling, but most tend to catch up in their ability to perceive the important information and details in a story by the time they reach ages 9/10, as significant differences between the age groups appear here. These results would suggest that an information checklist such as the one offered in ERRNI might be more sensitive to developmental change in typically developing children than Petersen et al's INC (2008). However, the evidence presented here is inconclusive in that regard, as we sought to apply the measures with the kind of stimulus they were designed for. A future investigation outside the scope of the present work, focused on discourse measures where both methods are applied to the same story, could directly examine whether Information Content is indeed more sensitive to developmental change.

Interestingly, both measures were highly and significantly correlated with narrative length as measured in tokens. Moreover, we expected Information Content to show higher correlations with tokens than the INC because it requires details from each picture in the sequence, while the latter requires up to two or three mentions at most of the number of elements: while this was indeed the case, ($r = .67$, $p < .001$ for Information Content and $r = .60$, $p < .001$ for INC), the difference in magnitude was not very large. Evidently, the comparisons are not straightforward since the measures were applied to different stories. Nonetheless, this evidence suggests that both measures were, to some extent, related to how much children are willing to talk, and the slight difference in the degree of association is not enough to favour one measure over the other based solely on the independence from tokens criterion.

A possibly more relevant issue is that the INC had comparatively lower reliability than the Information Content measure (82% versus 92%, respectively). Reliability is crucial to reduce measurement error, and to be able to replicate the results. From this perspective, based on the reliability analyses performed in these data, Information Content from ERRNI fares much better than Petersen's INC.

The INC and the Information Content were significantly correlated between them (r =.43, p <.001). Although this correlation was not as high as would be expected of two discourse measures, this correlation still suggests they are related, but not redundant.

From the comparison with the normative data our Information Content results were close to those found in the ERRNI manual. However, it is not possible to ascertain whether this similarity is due to having a somewhat representative sample of children in the present study or due to this index being quite robust.

This comparison needs to be interpreted in light of a decision taken for the elicitation procedure not to point to the relevant information in the pictures to preserve the distance, and with it, the aim of maximising literate language, even if it might have a cost in the structure or the content of the story. Therefore, these Information Content scores are from what the pupils were able to notice on their own, without any help. In some way, we could argue that by not pointing, an even more authentic measure of the child's true ability was achieved, since no support was provided for the production of the narrative.

In fact, of the three stories, the ERRNI beach story was the only one where there were three pictures simultaneously presented on the same page. From our observations, some children's sight wandered between the pictures, and some even skipped essential pictures, particularly if they were placed to the left of the page. This did not happen when children were telling the Frog story, where the pupil needed to turn the pages to continue narrating their story. Nonetheless, median scores for Information Content in the present study were fairly close to normative data, suggesting that our elicitation differences might not have mattered as much. In any case, the fact that results are similar in spite of possible fatigue effects (the Beach story was always administered last) could signal that Information Score does have some degree of reliability, something that would have been difficult to achieve from the more rudimentary discourse-level exercise carried out in the pilot reported in Chapter 3.

In sum, in spite of the marginally greater sensitivity to narrative length, Information Content does seem to exhibit greater developmental variability, although this change seems to be significant between the two oldest groups sampled here: Year 3 and Year 5 cohorts. For this reason, Information Content was included in the analyses described in the next section where all variables are grouped together, and in the next chapter, where the main reading study is carried out.

4.6 Relationship between variables

Even when a factor analysis of all the variables examined here would have been desirable, the nature of narratives, with broken assumptions of normality, non-linear relationships between variables and a tendency to show outliers, coupled with a small sample size of only 93 independent narratives, make such analysis inappropriate (Tabachnick & Fidell, 2007, p. 613). However, developmental patterns could still be compared in a descriptive way.

To compare variables with one another, all indices were standardised. In order to increase reliability, composite scores were created for both the optimal lexical measure, VOCD, and for the syntactic measure, MLT-w by computing an average from the three stories. For a discourse-level measure the Information Content score from the Beach story was chosen over the INC in the Frog story as this displayed greater developmental variability in the previous analysis and was similarly sensitive to tokens. Then, these linguistic composite scores and the single measure of Information Content were converted into z-scores and plotted by Year group. This graph is shown in Figure 4.10.
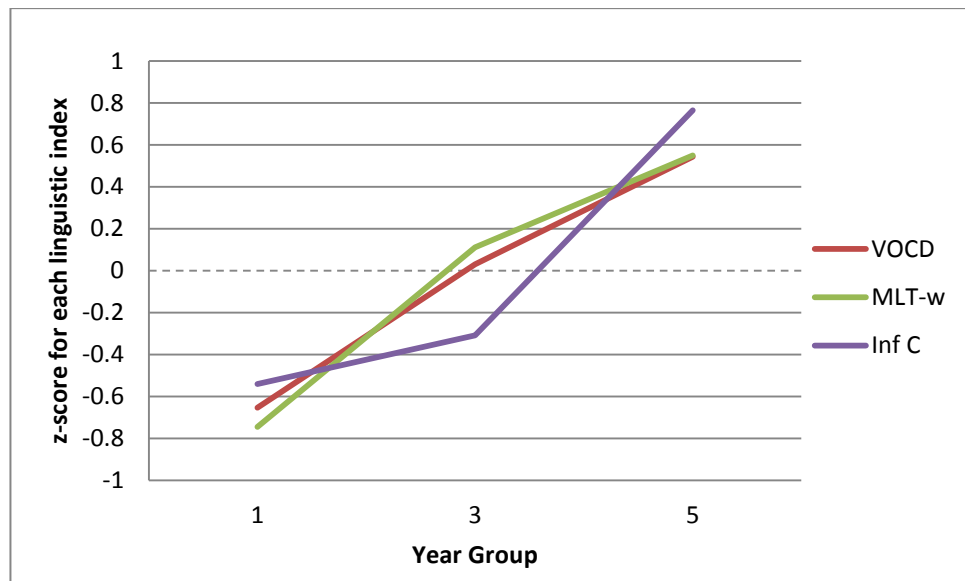
Figure 4.10. Mean z-scores for the three indices by Year group.
NB. Inf C = Information Content

From the sample collected here, the growth in lexical and syntactic indices is continuous from ages 5/6 through 9/10. In addition, they exhibit quite similar patterns. The developmental pattern for Information Content however, shows stability between Years 1 and 3, to be followed by a marked acceleration in Year 5.

Compared to the results in the previous study shown in Figure 3.9 in Section 3.2.4 illustrating change from 3 to 9 years of age, there is no lexical 'dip' in the current 7/8 age group in Figure 4.10. Since the lexical data for the present study is drawn from 3 narratives for each child, and the sample per group is larger, it could be argued that the pattern shown here might be more reliable than the one found in the previous database study. Nonetheless, given that all of these measures are still related to the length of the narratives in terms of sample size, these developmental patterns are bound to be influenced by child characteristics, beyond the influence specific to the task and elicitation procedures.

In addition it should be noted that a main difference in the retrieval of lexical and syntactic analyses from CLAN is that while lexical analyses included fragments, in obtaining the T-Units in words fragments were excluded, so that they would not bias the results.

As a final consideration, the lexical analyses as retrieved here using full words are not technically independent of grammatical indices in terms of morphological units. In the present study, *skip* and *skipped* were counted as distinct words, while grammatically they only differ by one morpheme. Still, while some authors choose to use root words for running VOCD (Vermeer, 2000), in order to exclude morphological influences in lexical diversity, the reliability cost in conducting a morphological division would potentially counter the benefits of such an analysis.

In addition to plotting the developmental patterns using z-scores, an additional issue that can to some extent be addressed with the data is whether each measure is dependent on the other, given that they are derived from the same stories. This is an important issue to consider since the regression analyses that are intended to answer the question of the separate contributions of word, sentence and discourse-level language skills to reading comprehension need to meet the assumption of the absence of multicollinearity.

To address this issue, the composite lexical and syntactic scores, as well as the Information Content score, were correlated to evaluate the degree of association between them.

Table 4.15 Pearson correlations for composite VOCD,
Composite MLT-w and Information Content from ERRNI

	Composite VOCD	Composite MLT-w	Information Content
Composite VOCD	-		
Composite MLT-w	.35**	-	
Information Content	.44**	.47**	-

**p <.001.

Table 4.15 shows the correlations between these different levels of language skills to be significant, but moderate, ranging from r= .35 to .47, p<.001, two-tailed. In other words,

even these levels of analysis are related, they do not appear to be redundant, as seemed it seemed to be the case for tokens and types, where correlations were greater than $r=$ .90, which usually indicate multicollinearity (Tabachnick & Fidell, 2007).

4.7 General Discussion and Conclusions

The present study aimed to further examine the developmental sensitivity and intra-child stability of the lexical and syntactic indices previously selected in Chapter 3 using specific elicitation procedures aimed at maximising decontextualized language in school-aged children with a larger sample per group. It was also intended to examine how two kinds of discourse-level measures behaved developmentally. Regarding overall developmental patterns, both lexical indices, types and VOCD, produced greater evidence for developmental variability in the population sampled and with the elicitation procedures used here when compared to the database study, while the syntactic measure replicated the previous developmental findings. Discourse measures were analysed in an exploratory way, but some developmental variability was also found.

Each of the lexical and syntactic indices and the discourse measures was examined separately for developmental variability using univariate/multivariate and correlational analyses, and degree of association with tokens to examine sensitivity to narrative length. In addition, intra-child variability was examined to evaluate reliability across stimuli for those indices that were collected from several narratives per child, namely, lexical and syntactic indices.

Regarding developmental variability, each of the two lexical indices was examined separately, as in the database study. Results for the index types replicated the previous findings of significant differences between 7 and 9, but the present analysis also found significant differences between ages 5 and 7. The index VOCD, which had not found significant differences between any of the school-aged groups in the previous database study, actually found significant differences between 5 and 7, but not between 7 and 9, in post-hoc comparisons in this larger sample. Comparing the actual trajectory from the database study and the present study, growth in lexical indices is continuous and the

'dip' found in age 7 in the database study is absent from the present data (see Figures 3.9 and 4.10). Given that the current data had a larger sample it is possible that the dip found in the previous data might have been a cohort effect, but the evidence is not conclusive. In any case, the correlational analyses with a more fine-grained age measure, age in months, provided converging evidence that VOCD was significantly associated with age in this sample.

For the syntactic index MLT-w, a similar developmental pattern emerged to the one found for VOCD. In MLT-w, the findings from the database study were replicated exactly: significant differences were found between 5 and 7, but not between 7 and 9. Converging evidence came from the correlations with age in months, which were significant. These results add more weight to the notion that MLT-w is a quite robust measure, which might be to some extent unaffected by population-specific characteristics. In fact, the general trajectory in z-scores is quite similar in both studies (see Figures 3.9 and 4.10).

In addition to the linguistic analyses, the present study also analysed two different discourse-level measures. First, the Index of Narrative Complexity or INC (Petersen et al., 2008), was a modernised version of Stein and Glenn's story grammar (1979), which intended to capture the complexity of the elements used by a child when telling a narrative. The second was an informational checklist type of measure denominated Information Content, which is designed to capture how many of a set of very specific predetermined story elements an individual is able to express. While these analyses were exploratory, we expected to find some developmental variability. The INC only found significant developmental differences between ages 5 and 9, but not with the intermediate children aged 7. On the other hand, Information Content from ERRNI found significant differences between ages 5 and 9, and between ages 7 and 9, an almost opposite pattern to the one found for VOCD and MLT-w. Correlations with age in months converged in finding a much smaller correlation coefficient for INC than for Information Content (r=.36 v .53, respectively). Interestingly, when comparing z-scores of this measure to those from a lexical and syntactic index (Figure 4.10), the trend implies that while the purely linguistic indices begin to stabilise by age 9/10, the organisation of ideas the child is able to perceive on their own, if it has not started yet, actually starts to accelerate at this stage.

As a whole, the results add further to the body of evidence that some developmental variability in lexical and syntactic indices still exists up to age 7/8 when examining non-adjacent age groups, as suggested by Nippold (1988), although it must be noted that groups were not as distinct as we would have hoped for given that we used school year cohorts instead of age ranges. Nonetheless, these results also counter the argument that expressive measures from language samples are inadequate for school-aged children (Kemper et al., 1995), at least up to this age range. In addition, these findings emphasise that measures developed for early language, such as Type-Token Ratio, should not be extrapolated directly to school-aged children, as they are unlikely to find meaningful developmental differences (e.g. Pearson, 2002), particularly if they are highly influenced by narrative length. For discourse measures, these findings suggest the possibility that discourse organisation might be more developmentally sensitive after age 7.

One of the main challenges of any index derived from narrative measures is to get some degree of consistency even when narrative data is by nature highly variable. Therefore all indices were correlated with tokens to evaluate their independence from narrative length. VOCD and MLT-w behaved in a comparable manner: both were still associated with tokens (r= .49 to .66 ranges for VOCD; r= .40 to .43 for MLT-w), but this association was much lower than the association between types and tokens (ranges between r= .92 and .95). Meanwhile, both discourse measures were similarly and significantly related to tokens (r= .60 for INC and .67 for Information Content). That Information Content was highly associated with tokens came as no surprise, since this informational checklist by its nature requires the child to tell details of every single picture. However, it was still somewhat unexpected that the correlation with tokens was so high for INC, since many of the elements to be scored were found at both the beginning and the end. In other words, it is theoretically possible, at least, that two children with strong beginnings and endings, but differing in their level of detail in the middle of the story, to have similar INC scores. These correlations suggest that this is not the case, and the structural elements in INC and the propositions in Information Content, were both highly dependent on how talkative a child was. By contrast, the optimal lexical and syntactic measures were less dependent on tokens, with the exception of one correlation between VOCD and tokens in the Sweets story, in which the correlation was r= .66.

Another way to examine the consistency of narratives was to look at the intra-child performance across three varied stimuli, for lexical and syntactic indices only, since those were the indices for which we had multiple measures. Discourse measures were not examined across stories since they were applied to different picture books. The within-subjects analyses provided some support for the reliability of two measures: VOCD and MLT-w. Both indices performed statistically similarly when using either the Sweets or the Beach stories, but not the Frog story. Types were influenced by any task, as all scores were significantly different depending on the story. Looking at these patterns in the Frog story, even when it produced the highest average scores for types, it also produced the lowest average scores for both VOCD and MLT-w, reflecting that Frog was the longest story produced by children on average.

It is noteworthy that even when the Frog story gave the most independent measure of VOCD (r= .49), it also gave the lowest MLT-w scores. On the one hand, the Frog story seems ideal to measure lexical diversity, but on the other, it seems to elicit the least complex T-Units. This needs to be given further consideration since this is one ubiquitous tool for the elicitation of narratives (Berman & Slobin, 1994; Petersen et al., 2008), and might not be the ideal stimulus when used on its own for syntactic indices.

In sum, this study has provided additional evidence for the developmental variability and intra-child reliability of two linguistic indices, VOCD and MLT-w, up to ages 7 and 8, while providing some exploratory evidence of developmental variability of a standardised measure that could be considered as a coarse measure of discourse level organisation in the school age sample, the Information Content measure in ERRNI (Bishop, 2004). It was also able to show that even with the most selective choices in methodology, these measures are still going to be related to some extent to how talkative these children are, but that choosing those least influenced by sample size is likely to make such measures consistent across stories. The data presented here in Table 4.15 has also shown that it is possible to obtain measures that are only moderately inter-correlated to measure different levels of analysis obtained from the same narratives, something that suggests that vocabulary, syntactic and discourse measures, even when extracted from the same narratives, are not redundant.

Of course, several limitations can be identified. Cohort effects cannot be ruled out completely, given that this is still a cross-sectional design. A longitudinal design would

be better able to discern real developmental trajectories, and that may be something that can be achieved in future studies.

An important limitation is that this sample was, after all, a convenience sample. The generalizability of the findings would have been much more robust had the sample been completely random. This limitation, however, is shared with most studies in education, and we sought to minimise this by inviting schools from differing socio-economic backgrounds.

An additional potential limitation was that the specific elicitation procedures used here might not have had the same effect at all ages. In particular, it is quite possible that 5-year-olds might not have understood that the test administrator could not see the pictures, and such possibility was not examined. Research devoted to the success rate of young children in false-belief tests (Baron-Cohen, Leslie, & Frith, 1985) pose the question of whether the youngest children in the present study were able to understand that the test administrator was unable to see through the screen. However, a meta-analysis of false-belief studies has shown that across different variants of these kinds of tasks, over 74% of children by 4;06 of age are able to pass such tests (Wellman, Cross, & Watson, 2001), suggesting the possibility that at age 5 children might have a foundation for understanding the specific instructions used in the present study to maximise decontextualized language, although this possibility was not evaluated directly. In the same manner, but at the other end of the developmental spectrum, older children might not have believed that the test administrator had never, in fact, seen the picture books before, so results have to be interpreted in light of this consideration.

Another consideration that could be addressed in future studies concerns the selection of pupils to take part on the study. With more access to schools, more resources and more time, it would be advisable to be more selective in terms of narrowing the age ranges of participants, in order to make the developmental differences more clear-cut and to truly fulfil the suggestion by Nippold (1988) to examine non-adjacent age groups. This will likely increase the distinctions among the age groups, and possibly increase the developmental variability seen in this study.

Finally, there were possibly some order effects taking place, which would have affected mostly the administration of the last story. For future considerations, it would also be advisable to partition the collection of the stories to two sessions, instead of one. For the youngest group in Year 1, this could reduce the possibility of fatigue effects, while for

Year 5 it might allow for older pupils to re-engage with the third story, which might have become predicable towards the end. Nonetheless it should be noted that even when we omitted from the instructions the help to the pupil by pointing to essential features or drawings of the story, and the Beach story from ERRNI was the last for all pupils, children were still able to produce very comparable measures of Information Content scores.

While waiting for further confirmation of the validity and reliability of these measures from longitudinal studies, language researchers could use the composite measures VOCD and MLT-w for lexical and syntactic analyses, with the elicitation stimuli and procedures used here, with some confidence that they have received some empirical support. Regarding the discourse measure, further investigations should still examine how these methods behave in a different population, but in the meantime, Information Content can be used as the one displaying the greater developmental variability and greater reliability for the age range examined.

Even when not all hypotheses were supported, and the best identified measures are still affected by narrative length and possibly the kind of stimulus and elicitation procedures to some extent, the documented changes provide clear information about the rate of change across the primary school years. In addition, an original picture has been presented of how these two linguistic skills and one discourse-level organisation skill develop in relation to each other.

Although the degree of complexity involved in quantifying changes in the primary years is great, we consider that the findings presented here provide enough evidence of reliability and developmental variability to be useful in quantifying narrative indices for the purpose of linking expressive language and reading abilities. Moreover, since the developmental variability is supported for both VOCD and MLT-w up to the 7/8 years of age, it is argued that at least at this stage, these indices are both developmentally valid and reliable for the reading study's objective. Furthermore, since the focus of the reading study is on this 7/8 age range, the data from three stories in the present study at this particular age all fall within the 100-400 optimal range suggested in the literature (McCarthy & Jarvis, 2007) for the proper application of VOCD, a novel index in studies linking narrative language and literacy.

Finally, regarding discourse-level measures, the developmental analyses found greater developmental variability for the measure of Information Content than the Index of Narrative Complexity. When considering that their sensitivity to narrative length was quite similar, even if high for both, then Information Content could present itself as the optimal measure to characterise development in the school-age years. It could be that this proxy measure could, after all, provide a glimpse into the mild discourse organisation weaknesses that might be ultimately related to reading comprehension in typically-developing children.

The next chapter describes such a study using these carefully chosen expressive indices and the previously selected reading measures, along with some receptive measures to examine if indeed school-age narrative abilities are associated concurrently with reading comprehension after the few initial years of reading instruction, precisely when these mild deficits might go unnoticed.

Chapter 5

Main Reading Study

This chapter intends to address the original research question posited at the beginning of this thesis about the relationships between expressive language and reading comprehension in school-aged children. To recap, in light of the methodological heterogeneity of the literature linking expressive language and reading comprehension, and in particular, of the literature linking expressive narrative language and reading comprehension, a considerable amount of attention was given to the selection of expressive measures that were appropriate for school-aged language. In addition, findings that earlier measures of expressive language were more predictive of reading outcomes than later expressive measures (Scarborough, 2005) in the context of a lack of methodological consensus for measuring school-age language samples (Scott & Stokes, 1995) left the question open as to whether this was a case of methodological constraints, or one of an actual phenomenon, where later spoken language is actually less predictive of reading abilities. Therefore, a study of an existing corpus of narrative language was carried out to identify the ideal lexical and syntactic indices derived from narrative samples, which found one index, types (Scott & Windsor, 2000), with some developmental variability in pre-school and school-aged children, but a high correlation with narrative length, and another, VOCD (Malvern et al., 2004), with developmental variability only in pre-school children but with a comparatively more moderate correlation with narrative length. The corpus study also found that Mean Length of T-Units was able to find significant differences up to 7 years of age, whether it was on words and morphemes. Since morphemes have a reliability cost and the basic pattern for school-aged children was similar, MLT in words was selected as an optimal syntactic measure.

Then, a pilot study was carried out seeking to examine the relationship between these expressive narrative indices, VOCD and MLT-w, and one experimental expressive discourse measure with reading comprehension skills. This pilot study found almost

equivalent correlations for these three expressive measures than for the receptive counterparts (Chapter 3).

The need to strengthen the methodological foundation for appropriately addressing the relationship between school-age narrative language and reading, led to an additional study of narrative language seeking to replicate the database study findings regarding the optimal linguistic indices, and to analyse two types of discourse measures for developmental patterns.

Reported in the previous chapter, results from this primary data study, using a semi-naïve elicitation method, and collecting three narratives per child, found that VOCD was indeed able to find significant differences between ages 5 and 7, while for MLT-w, the results were replicated exactly, also finding significant differences between 5 and 7 years, but none beyond this age. Additional exploratory discourse-level measures, the Index of Narrative Complexity or INC (Petersen et al., 2008), and the Information Content from ERRNI (Bishop, 2004) were also applied and analysed to two different stories. Developmental analyses of these two measures provided some support for the use of Information Content as it showed some developmental variability among ages 7 and 9, while Petersen and colleagues' INC did not find developmental variability in the school age range.

Summarising, results have provided some empirical evidence that VOCD and MLT-w, have some developmental variability up to ages 7/8, which is the developmental stage that is the focus of the main reading study, while Information Content has some variability between 7/8 and 9/10. Building on such evidence, it was then possible to resume the original objective of assessing the relationship between reading comprehension and expressive language in a larger, more socially varied sample, with the knowledge that the methodological issues involving the measurement of expressive skills had been given considerable attention.

Based on more reliable and developmentally sensitive measures, this study further examined how expressive skills are related to reading comprehension at a very particular developmental window, at 7 and 8 years of age, or Year 3 in the UK school system, after the initial couple of years of decoding-intensive instruction, when mild receptive deficits tend to go unnoticed particularly if children are decoding well (Nation et al., 2004) and expressive narrative skills could be a more visible window into these skills. Studying younger children by comparison, would likely be more influenced by

phonological than broad language aspects, as decoding skills are likely to place a limit to comprehension skills (Hoover & Gough, 1990). This age group was also selected because at this developmental stage causality still has been reported to go from language to reading and not yet in the other direction in structural equation models where all directions are explored (Verhoeven & Van Leeuwe, 2008) and we wanted to minimise reciprocal relations. Finally, and more importantly, this was also chosen because there is already evidence linking school-age narrative skills to reading comprehension (Cain, 2003; Cain & Oakhill, 1996). Given that different sub-skills may play differential roles at different points of development (Scarborough, 2005; Vellutino et al., 1994) it was also important to focus on a narrow developmental window.

As in the pilot study, receptive language measures were included to be able to distinguish their relative contribution to reading comprehension. However, for the expressive measures, this time the narratives would be prompted using the protocols designed in the previous study of narrative indices, featuring a semi-naïve elicitation procedure intended to maximise literate language, the elicitation of more reliable indices obtained from three different stories, and the inclusion of Information Score from the standardised narrative test, ERRNI (Bishop, 2004), as a proxy for discourse-level production. However, given the preliminary nature of these results, both variables were included for analyses in this study as measures of expressive discourse skills.

5.1 Aim

The aim of the study was to further examine the contribution of receptive and expressive oral language skills to reading comprehension at a developmental window after the first couple of years of initial reading instruction, in a varied socioeconomic sample in the West Midlands area of the United Kingdom.

The proposed hypotheses are:

a) Expressive measures from picture-book narratives would predict reading comprehension.

b) Expressive measures from elicited picture-book narratives would contribute unique variance after controlling for both receptive language and non-verbal ability.

c) Based on pilot study findings, we would expect vocabulary in both receptive and expressive modalities to have a primary role amongst all sub-skills; in other words, we would expect vocabulary skills in both receptive and expressive modalities to have the strongest standardised coefficients in a regression analysis, when compared to grammar and discourse in both modalities.

d) Pupils with a combination of expressive and receptive language weaknesses will have the poorest reading outcomes; if expressive measures add unique variance, then those children with poorer expressive and receptive skills will show lower reading comprehension skills compared to those with only receptive weaknesses.

An additional objective, focused on the language variables, was to further examine the relationship between receptive standardised measures and their expressive counterparts for this at this particular age sample, to potentially add further support for the validity of the use of these indices.

Given that early expressive language seems to be affected by gender (Huttenlocher et al., 1992), and that language seems to mediate the effects of socioeconomic status (SES) on reading (Beron & Farkas, 2004; Durham et al., 2007), another objective was to examine whether these differences are affected by a school-wide SES measure and by gender. An exploration into the influence of SES is also important considering that some of the research showing the strongest indicators for language has been conducted on disadvantaged populations (Storch & Whitehurst, 2002). Although conceptually SES is considered to precede language differences that ultimately affect reading in longitudinal studies, in the cross-sectional data examined here the aim is to find, through moderator analysis, if the effects of language on reading comprehension depend on SES or gender.

5.2 Method

Several state schools in different socioeconomic areas were invited to take part in this study. Administrators at six primary schools in the West Midlands area in the United Kingdom agreed to participate. Co-operation was agreed with teachers to pull children out from the classrooms to perform the tasks on an individual basis.

5.2.1 Participants

All parents at the Year 3 classrooms were sent invitation letters and consent forms. From those who returned their consents, children whose native language was other than English, or who had a diagnosed reading or behavioural disorder, were excluded. Eighty children, 30 boys and 50 girls, participated in all tasks. Their ages ranged from 7;01 (years; months) to 8;09. Median age was 7;09.

A measure of school-wide socioeconomic status (SES) was obtained by consulting a poverty index for each school's postcode, produced by the UK National Statistics Publication Hub (Office for National Statistics, 2008). This poverty index estimates the percentage of households below 60% of the UK median income after housing costs for a local area. While it is not a proper SES measure, it is indicative of the relative level of poverty of the local authority in which the school is found. Three schools had a relatively low poverty index ranging from 14% to 19%. By comparison, the median proportion of households in poverty in the West Midlands area is 21% (Fry, 2010). Three other schools had a comparatively higher index ranging from 29 to 33%. Analyses using the SES variable will group the first schools as middle-class and the latter as disadvantaged, composed of 35 and 45 children respectively.

5.2.2 Materials

Receptive language, non-verbal intelligence and reading skills were measured using standardised tests and one experimental task. Expressive language was obtained using the stimuli and procedures developed in the narrative study, along with one standardised test for the assessment of narrative language, also used in the previous study of narrative

language. With one exception noted below, all tests were the same ones used in the pilot study, described at length in Section 3.3.2. For clarity, all materials are listed next.

Receptive Language

a. **British Picture Vocabulary Scale**, 2nd edition (BPVS-II; Dunn et al., 1997).

b. **Test of the Reception of Grammar**, 2nd edition (TROG-2; Bishop, 2003).

c. Understanding Spoken Paragraphs subtest of the **Clinical Evaluation of Language Fundamentals**, 4th UK edition (CELF-4 UK; Semel et al., 2006).

Expressive language

Three wordless picture books and prompts from the previous study were used to elicit the narratives from each child:

d. **The Sweets Story** (Chen-Wilson, 1997)
e. **Frog, Where Are You?** (Mayer, 1969),
f. The Beach story from **Expression, Reception and Recall of Narrative Instrument** (ERRNI; Bishop, 2004).
g. The story structure portion of the INC scoring form (Petersen et al., 2008).
h. Prompting instructions (Appendix C) for the first two stories. Prompting instructions for the Beach story in ERRNI came from the Manual.
i. Within the scoring rubric for the Index of Narrative Complexity (INC; Petersen et al., 2008) the structural elements portion was used to obtain a score for story structure as a measure of expressive discourse.
j. A blank A4 hardcover notebook, as a visual barrier.
k. A digital voice recorder Olympus WS-210S was used to record the children's narratives.

Control and outcome variables

Non-verbal Intelligence as control and Reading Comprehension as an outcome measure were also collected with standardised tests. Reading Accuracy was also obtained from the same reading standardised test as a proxy for decoding ability.

l. Matrix Reasoning subtest of the ***Wechsler Abbreviated Scale of Intelligence*** (WASI; Wechsler, 1999).

m. ***Neale Analysis of Reading Ability***, 2nd Revised British edition (NARA-II; Neale, 1997).

5.2.3 Design

Following the design of the pilot study, a cross-sectional correlational design was used.

First the relationship between receptive measures and their expressive counterparts was examined to see if expressive narrative measures had the potential to provide additional information to that provided by the receptive standardised measures.

Then zero-order correlations were obtained. Four different hierarchical regression analyses were used to predict reading comprehension as the outcome variable. Reading comprehension was measured using raw scores from Form A in NARA-II. The predictors for each of the four regression analyses were the following:

1. Expressive measures: vocabulary (composite VOCD), expressive grammar (composite MLT-w). Expressive discourse from either Information Content or INC had been originally considered in the design, but as it is reported in Section 5.3.3, correlations were not significant and were therefore dropped from further analyses.

2. Expressive measures after controlling for non-verbal ability only.

3. Expressive measures after controlling for receptive language (BPVS-2, TROG-2 and CELF-4[UK]) only.

4. Expressive measures after controlling for both non-verbal intelligence and receptive language.

Finally, a mediation analysis was carried out to examine possible mediation pathways for receptive variables, and a moderation analysis was used to evaluate possible moderator effects by SES and gender.

5.2.4 Procedure

Data was collected in two waves. The first wave of 27 children was from the 2008-2009 school cohort and the second wave of 53 children was from the 2009-2010 cohort. The narratives of the first cohort were also used as part of the narrative indices study reported in the previous chapter.

Procedures were similar to the pilot study for the standardised tests with the exception that the tasks did not follow a strict order, based on the null findings on the order effects in the pilot. In addition, being flexible on the test administration made the most efficient use of the time with the child and minimised interruptions for teachers. Pupils were pulled out of their classrooms for three individual sessions, each lasting between 30 and 45 minutes. Additional verbal consent was obtained from pupils at the beginning of the first session.

Procedures for collecting expressive language were the same as the ones followed for the study of narrative indices in Chapter 4, aiming to elicit the most decontextualized language possible from children. With a few exceptions, all narratives were produced in one single session. All stories were audio recorded for later transcription. Data was collected for all assessments for all children.

CHAT conventions (MacWhinney, 2000) were used for transcription, and the segmentation procedures derived in the previous study were used for T-Unit delimitations (Appendix D).

A second revision of transcription was carried out to ensure accuracy and for adherence to the segmentation rules in Appendix D. Then, inter-rater reliability was conducted for a random subsample of 12 participants producing 3 narratives each (36 transcripts, 15% of the total sample). Just like in the narrative study, the second examiner listened to the digital audio files while looking into the corresponding CLAN transcript for transcription or segmentation errors. Mean word-by-word reliability for main body words transcribed (excluding fragments, repetitions, reformulations and task-related

comments) was 98.8% (range 98% to 99%). Mean reliability for T-Unit segmentation was 98.6% (range 95% to 100%).

A sample of the children's narratives can be found in Appendix F4.

Next, linguistic analyses were performed using CLAN software (MacWhinney, 2000), to obtain VOCD and MLT-w, to represent expressive vocabulary and expressive grammar, respectively. Onomatopoeic expressions were excluded when obtaining the vocabulary index VOCD, while fragments were excluded to obtain the syntactic index MLT-w. As per CHAT conventions, all analyses excluded false starts and repetitions. Then, to minimise collinearity in the regression analyses, the scores from the three stories for each participant were averaged to create composite scores for VOCD and MLT-w.

Finally, the Frog story was scored for a measure of expressive discourse using Petersen's INC scoring form as outlined in the materials section before, while the Beach story was scored using the procedure outlined in the ERRNI Manual. Since the measures were different, no composite was created, and their relationship with reading comprehension was analysed for each one.

Discourse-level reliability analyses were carried out for the Index of Narrative Complexity in the Frog Story, and for Information Content in the Beach Story, for the same participants. Just as in the narrative study, a second examiner independently scored these narratives and the stories were compared. For INC, mean reliability was 84.3% (range 60% to 94%). For Information Content, mean reliability was 90.6% (range 82% to 100%).

5.3 Results

As the expressive indices have not been normed, raw scores were chosen over standard scores for all variables. Wherever available, the published norms are provided for comparisons.

5.3.1 Descriptive statistics

Prior to analysis, non-verbal intelligence, receptive vocabulary, receptive grammar, discourse comprehension, composite VOCD, composite MLT-w, Information Score from ERRNI, Index of Narrative Complexity, Reading Accuracy and Reading Comprehension were examined for normality, and the presence of univariate and multivariate outliers.

To check for normality, skewness and kurtosis z-scores were calculated, and those values in excess of 3.29 were considered to be significant. Receptive grammar (TROG-2) showed significant negative skew (z =-3.55, p=.001). Although transformations were considered for TROG-2, the resulting measures of associations with reading comprehension were not different, so the original measure was retained. Discourse comprehension using CELF showed a moderate negative skew, but it was not significant at this cut-off. All kurtosis values fell within the expected scores.

To identify univariate outliers, raw scores were transformed into z scores. For the purpose of this study, an outlier was defined as a case with a standardised score greater than 3.29 in absolute terms (Tabachnick & Fidell, 2007). No outliers were identified. Multivariate outliers were examined with Mahalanobis distance and none were identified. Therefore all 80 participants' scores were retained for analysis.

Table 5.1 displays descriptive statistics, including the range of raw scores, mean raw scores and standard deviations.

Table 5.1 Descriptive statistics for raw scores in all variables

Variable	Mean	SD	Range
Receptive Vocabulary: BPVS-2	79.21	13.86	46 - 108
Receptive Grammar: TROG-2	13.33	3.22	5 - 18
Discourse Comprehension: CELF-4[UK] (U.S.P. Subtest)	11.66	2.25	5 - 15
Expressive Vocabulary: Composite VOCD	31.50	7.20	16.66 - 51.31
Expressive Syntax: Composite MLT-w	8.51	1.21	6.19 - 11.78
Information Score: ERRNI	19.84	5.89	9 - 35
Index of Narrative Complexity	12.40	2.20	8-18
Non-Verbal Intelligence: WASI (Matrices Subtest)	13.16	6.42	3 - 26
Reading Accuracy: NARA-II[a]	46.40	23.70	8 - 93
Reading Comprehension: NARA-II[a]	17.40	8.29	1 - 39

NB: [a] Form 1 of NARA II

To determine how these figures related to existing data, Table 5.2 shows the mean standardised scores for measures in the standardised tests, along with the mean standardised score for each test, except for the listening comprehension assessment, the CELF-4[UK] Understanding Spoken Paragraphs Subtest, for which no standard score means are available. For the expressive vocabulary, no published norms were available. For expressive grammar, ERRNI offers medians for MLT-w, abbreviated in the Manual as MLUw; instead of using the composite MLT-w, the value for the Beach story is the one that is compared to the published norm so that the comparison is equivalent. For expressive discourse, the Information Score was only obtained for the Beach Story in the present data, so that is compared to the norms as well. For the INC score, as published scores are not dissagregated by age, no scores were available for comparison.

Scores were slightly lower for receptive grammar (TROG-2) and non-verbal ability, and higher for the reading scores in the current sample than those of the populations from which the norms were derived; the rest of the measures were similar to the published norms.

Table 5.2 Mean/Median normative scores

	Mean/Median Reading Study	Mean/Median test/ subtest
Mean Standard Score BPVS-2	100.06	100
Mean Standard Score TROG-2	96.31	100
Median MLT-w [a,b] ERRNI (Beach Story only)	8.62	8.50-8.74[c]
Median Information Score [a] ERRNI (Beach Story)	19.50	20[c]
Mean Standard Score WASI (Matrices)	48.46	50
Mean Scaled Score Reading Accuracy NARA-II	103.38	100
Mean Scaled Score Reading Comprehension NARA-II	102.08	100

[a] There were slight differences in the elicitation methods. [b] There were slight differences in the protocol of segmenting sentences from the ERRNI Manual (cf. Appendix D). [c] Medians indicated here belong to the 7;06-7;11 bracket in the ERRNI Manual, corresponding to our mean age of 7;09.

Before reporting the main analysis, the relationship between standardised receptive measures and the expressive measures from the narratives is described next. Since this particular analysis could add further evidence of the relative suitability of the narrative indices and measures analysed developmentally before, the following analyses look at how types, VOCD, MLT-w, Index of Narrative Complexity and Information Content are related to their receptive counterparts, at this particular developmental window, at 7 and 8 years of age.

5.3.2 Relationship between receptive measures and their expressive counterparts

The data from receptive language sub-skills offered the unique opportunity to assess how the expressive variables identified in the linguistic studies relate to standardised receptive tests.

Here the correlations between receptive standardised tests and their expressive counterparts are examined by story and by composite score.

Lexical indices were correlated with receptive vocabulary measured by the BPVS-2. Types and VOCD by story, and VOCD composite, are listed in Table 5.3 for comparison.

Table 5.3. Correlations between receptive and expressive vocabulary

	Receptive Vocabulary: BPVS-2
Types in Sweets Story	-.05
Types in Frog Story	.10
Types in Beach Story	-.04
VOCD in Sweets Story	.23[*]
VOCD in Frog Story	.29[**]
VOCD in Beach Story	.20
Composite VOCD Score	.27[*]

NB: *p<.05, **p<.01 (2-tailed)

Types by story comparisons showed non-significant associations with the standardised receptive vocabulary test administered to these children. Meanwhile, the VOCD by story comparison showed a moderate and significant relationship with such test. These results add further evidence to the advantages of using VOCD, while suggesting that types, in addition to the reliability problems discussed in the previous chapter, is also less likely to be valid. Interestingly, when considering each story individually, the longest story was the one most related to receptive vocabulary. In other words, the VOCD score from the Frog story, which the previous study had considered the most independent from narrative length, was also the one to show the strongest association with the receptive vocabulary test.

The VOCD composite was also significantly related to receptive vocabulary, although slightly less so than the correlation with VOCD from the Frog story alone. Nonetheless, since this VOCD composite is more reliable than the scores by individual stories, it is more likely to represent a more authentic measurement of expressive skills.

A similar analysis was carried out for MLT-w. Grammatical indices are shown in Table 5.4, by story and in a composite score.

Table 5.4 Correlations between receptive
and expressive grammar

	Receptive Grammar: TROG-2
MLT-w in Sweets Story	.28[*]
MLT-w in Frog Story	.33[**]
MLT-w in Beach Story	.31[**]
Composite MLT-w score	.36[**]

NB: *p<.05, **p<.01 (2-tailed)

Although these correlations vary less by story, compared to the variation in lexical indices, once more a surprising result is shown: when considered individually, the one with the highest correlation to the receptive measure of grammar is the one from the Frog Story, which the previous study on narrative indices had identified as producing the least complex T-Units.

Moreover, the composite MLT-w measure showed an even stronger correlation with its receptive counterpart than when considering each of the scores individually suggesting the composite for this syntactic index is both more reliable and closer to a true measure of syntax.

Finally, the correlations were also explored between the receptive discourse measure, CELF-4UK, and the two discourse-level expressive narrative measures, the Index of Narrative Complexity and the Information Content. It should be noted that since the two

discourse measures were applied to different stories, no composite was calculated. Correlations are shown in Table 5.5.

Table 5.5 Correlations between receptive
and expressive discourse measures

	Discourse comprehension: CELF-4[UK]
Index of Narrative Complexity in Frog Story	.21[a]
Information Content in Beach Story	.18

NB: [a]p=.057 (2-tailed)

Correlations between expressive discourse measures from narratives with their receptive standardised measures were lower than those seen for lexical and syntactic indices. Although both were non-significant at conventional p levels, the INC was almost significant for this sample size. Since the difference in the magnitude of correlations is minimal, this analysis is inconclusive as to whether the INC should be favoured over Information Content in terms of external validity.

Therefore, it would seem that even if the discourse measures did not correlate significantly with this receptive standardised test, the lexical and syntactic expressive narrative measures selected on the basis of developmental variability and intra-child reliability do seem to have some external validity, while at the same time seem to capture some other facet of the children's language development that is not yet measured by those standardised receptive assessments used here.

In particular, since in the case of the lexical and syntactic indices the associations with receptive standardised measures seem stronger or similar when using composites than when using single stories, the use of composites would seem optimal.

5.3.3 Correlations

Table 5.6 shows the correlations between all receptive and expressive language variables in their raw scores with reading comprehension, as well as with reading accuracy.

Table 5.6 Pearson's correlations for receptive and expressive language variables with Reading Comprehension and Reading Accuracy (n=80)

	1	2	3	4	5	6	7	8	9	10
1. Reading Comp: NARA-II	-									
2. Reading Accuracy: NARA-II	.88**	-								
3. Age in months	.12	.15	-							
4. Non-verbal Intelligence: WASI	.47**	.36**	.06	-						
5. Receptive Vocabulary: BPVS-2	.51**	.33**	.10	.50**	-					
6. Receptive Grammar: TROG-2	.64**	.54**	.05	.52**	.44**	-				
7. Discourse Comp: CELF-4^{UK}	.37**	.18	.09	.45**	.58**	.46**	-			
8. Expressive Vocabulary: Composite VOCD	.32**	.25*	.07	.19	.27*	.29**	.51**	-		
9. Expressive Grammar: Composite MLT-w	.41**	.34**	.16	.38**	.36**	.36**	.18	.20	-	
10. Information Score: ERRNI	.10	-.04	-.18	.05	.20	.17	.18	.31**	.08	-
11. Index of Narrative Complexity	.03	.02	-.05	.20	.15	.02	.21	.48**	.01	.42**

** $p < 0.01$; * $p < 0.05$ (2-tailed).

For this sample, reading comprehension and reading accuracy were highly correlated. In other words, even after focusing on a developmental window after the first couple of years of initial reading instruction, decoding as measured by reading accuracy still places some limit in reading comprehension skills.

Significant correlations with reading comprehension were found for all receptive language variables: vocabulary, grammar and discourse comprehension. Significant correlations with reading comprehension were also found for the two composite expressive measures, vocabulary and grammar, and for the control variable of non-verbal intelligence. Neither the Index of Narrative Complexity, nor the Information Score from ERRNI, showed significant associations with reading comprehension.

All language correlations were stronger for reading comprehension than for reading accuracy, as would have been expected for broader language skills. As accuracy measures the ability to read out loud, it can function as a measure of decoding. As such, its relationship with broad aspects of expressive and receptive language was weaker.

The non-significant correlation of age in months with reading comprehension (and with the rest of the variables as well) reflects the restriction on age, since the study is only investigating 7- to 8-year-olds.

5.3.4 Multiple regression analyses

In order to assess the relative contribution of receptive and expressive language factors to reading comprehension at this developmental window, a hierarchical regression analysis was performed on the data. Given the high and significant correlation between reading accuracy and reading comprehension, the former could be placed to predict the latter, but that would lead to breaking the assumption of multicollinearity in the regression analyses. For that reason, reading accuracy is not included as a predictor.

As the associations of age, Index of Narrative Complexity, and Information Score from ERRNI, with reading comprehension were low and not significant, these three variables were excluded from further analysis.

First, the contribution of expressive measures on their own was examined, without controls. Then, the contribution of expressive measures was re-examined while controlling for non-verbal ability. The third analysis examined the contribution of expressive measures after receptive language was accounted for. The fourth and final analysis controlled for both receptive language and non-verbal abilities.

Regression Analysis 1 was performed on reading comprehension scores as the dependent variable, and composite VOCD and composite MLT-w as the predictor variables to represent expressive vocabulary and grammar, respectively. This regression is reported in Table 5.7

Table 5.7 Regression Analysis 1:
Expressive language variables as predictors of Reading Comprehension

	B	$SE\ B$	β	R^2 (adjusted R^2)
Constant	-13.16	6.37		.236 (.216)
Composite VOCD score	.29	.11	.25*	
Composite MLT-w score	2.51	.69	.36**	

NB: *p<.05, **p<.01

Results show that, when considered on their own, expressive measures can explain 23.6% (21.6% adjusted) of the variability in reading comprehension. The coefficients for each variable, also shown in Table 5.7, suggest that the contribution of expressive grammar is comparatively higher than that of expressive vocabulary, as indicated by β coefficients.

Table 5.8 Regression Analysis 2:
Non-verbal ability and expressive language as predictors of Reading Comprehension

		B	$SE\ B$	β	R^2 (adjusted R^2)	$ΔR^2$
Step 1	Constant	9.34	1.88		.225 (.215)	
	Non-Verbal Intelligence	.61	.12	.47***		
Step 2	Constant	-10.36	6.05		.333 (.306)	.10**
	Non-Verbal Intelligence	.43	.13	.33**		
	Composite VOCD Score	.24	.11	.21*		
	Composite MLT score	1.68	.70	.24*		

NB: *p<.05, **p<.01, p<.001

The second regression analysis, shown in Table 5.8, aimed to examine how much of this contribution remained after controlling for non-verbal intelligence. Non-verbal intelligence on its own was able to account for 22.5% of the variance (21.5% adjusted)

in reading comprehension. Having considered non-verbal intelligence, then expressive language can only explain a further 10.8% of unique variance in reading comprehension. Comparing standardised coefficients, when non-verbal intelligence is entered first, its contribution seemed greater than the contributions by any of the expressive indices.

The third analysis looked at whether an expressive language contribution appears when controlling for receptive language abilities. A sequential regression was performed with receptive language as the first step and expressive language scores as the second step. The results for this regression appear in Table 5.9.

Table 5.9 Regression Analysis 3:
Receptive and expressive language as predictors of Reading Comprehension

		B	SE B	β	R^2 (adjusted R^2)	ΔR^2
Step 1	Constant	-13.19	4.34		.481 (.460)	
	Receptive Vocabulary	.19	.06	.31		
	Receptive Grammar	1.35	.24	.52		
	Discourse Comprehension	-.22	.38	-.06		
Step 2	Constant	-20.46	5.621		.513 (.480)	.03
	Receptive Vocabulary	.17	.06	.28**		
	Receptive Grammar	1.23	.25	.47***		
	Discourse Comprehension	-.42	.42	-.11		
	Expressive vocabulary	.16	.11	.14		
	Expressive grammar	.92	.62	.13		

NB: **$p<.01$, $p<.001$

A full model with both receptive and expressive variables can explain 51.3% (48.0% adjusted) of the variability in reading comprehension, which is a considerable increase from the results of the expressive variables alone. However, out of that figure, receptive skills on their own are able to explain 48.1% (46.0% adjusted) of the variability in reading comprehension. Hence, after controlling for receptive language, expressive skills only contribute a non-significant 3.3 % ($p = .09$) of unique variance.

Looking at the relative contribution of the individual receptive variables, receptive grammar and receptive vocabulary, in that order, are the most significant predictors to reading comprehension skills.

The final analysis controlled for both non-verbal abilities and receptive language. Non-verbal abilities were placed as the first step, receptive language as the second and expressive language as the third step. Results appear in Table 5.10.

Table 5.10 Regression Analysis 4:
Non-verbal ability, receptive and expressive language as predictors of Reading Comprehension

		B	SE B	β	R^2 (adjusted R^2)	ΔR^2
Step 1	Constant	9.34	1.88		.225 (.215)	
	Non-Verbal Intelligence	.612	.12	.47***		
Step 2	Constant	-11.80	4.56		.487 (.460)	.26**
	Non-Verbal Intelligence	.13	.13	.10		
	Receptive Vocabulary	.17	.06	.29**		
	Receptive Grammar	1.26	.26	.49***		
	Discourse Comprehension	-.27	.39	-.07		
Step 3	Constant	-18.73	5.98		.518 (.478)	.03
	Non-Verbal Intelligence	.11	.13	.09		
	Receptive Vocabulary	.15	.06	.26*		
	Receptive Grammar	1.16	.26	.45***		
	Discourse Comprehension	-.49	.43	-.13		
	Expressive vocabulary	.17	.11	.15		
	Expressive grammar	.80	.63	.11		

NB: *$p<.05$, **$p<.01$, $p<.001$

A full model with non-verbal abilities, receptive skills and expressive language can explain virtually the same 51.8% (47.8% adjusted) variance as a model excluding non-verbal intelligence (see Tables 5.6 and 5.7). Receptive language coupled with non-verbal ability can explain 48.7% (46.0% adjusted) of the variability in reading comprehension. Therefore, the variance explained by non-verbal intelligence can be explained by receptive language.

When controlling for receptive scores and non-verbal intelligence, expressive language as measured in this study is able to contribute 3.1% of unique variance, which was not significant at $p = .10$.

In the same manner to non-verbal ability, most of the variance explained by expressive skills is already accounted for by receptive language.

For the last hypothesis proposed, it had been anticipated that those pupils with a combination of expressive and receptive weaknesses could have had the poorest reading outcomes. However, given that the unique variance explained by expressive skills as measured here beyond that explained by receptive skills is rather small and non-significant at a conventional p level, it is unlikely that pupils with a combination of expressive and receptive deficits will have the worst reading outcomes, at least in this sample.

Given that the expressive variables were not able to account for significant unique variance to reading comprehension skills once receptive skills were added to the model, a possibility was considered that maybe receptive skills, particularly receptive vocabulary and grammar, mediated the relationship between expressive skills and reading comprehension. Although this was not an original aim, it could certainly help in clarifying the relationship between the expressive variables and reading comprehension. Mediation analyses were carried out to examine this possibility.

5.3.5 Mediation analyses

In addition to evaluating the effect of the expressive variables via its effects on a mediating variable of receptive language, it was noticed in the previous analyses that the contribution of non-verbal intelligence was no longer significant when taking into account receptive skills; therefore, it was added as another variable whose relationship to reading comprehension could potentially be explained by the receptive variables as well. Since discourse comprehension showed a comparatively minor contribution to reading comprehension skills in this age range, it was added to the variables whose relationship to reading comprehension could also be mediated by the two strongest receptive skills.

In sum, the relationships between reading comprehension and four variables, Expressive Vocabulary, Expressive Grammar, Non-verbal Intelligence and Discourse Comprehension, were examined in two mediation analyses to see whether those

relationships were mediated by Receptive Vocabulary (Mediation Analyses 1, 2, 3 and 4) and by Receptive Grammar (Mediation Analyses 5, 6, 7 and 8).

To carry out these mediation analyses, ordinary least squares mediation analyses with bias-corrected intervals derived from bootstrapping resampling procedures (Hayes, 2013; Preacher & Hayes, 2008) were conducted, as they have increased power while controlling for the Type I error well. In addition, these procedures also allow for a test of significance of the indirect effect (Hayes, 2013). In this work, Preacher and Kelley's κ^2 (2011) and the 95% confidence interval is used to report effect sizes for indirect effects; the number of bootstrap resamples was set at 5000 (Preacher & Hayes, 2008).

The software tool developed by Hayes (2013) called *PROCESS* allows only for the computation of one independent variable *X* and one dependent variable *Y*. Even when this method is less parsimonious than the Baron and Kenny (1986) approach for these many variables, the increase in power and the possibility of obtaining an effect size for the indirect effect were more relevant factors for choosing this tool.

Therefore, eight simple standard mediation models are reported separately, four for each one of the two proposed mediating variables, receptive vocabulary and receptive grammar. The results of each model are reported and illustrated in turn.

Mediation analysis 1: *Receptive vocabulary as mediator between non-verbal intelligence and reading comprehension*

Non-verbal intelligence was found to indirectly influence reading comprehension through its effect on receptive vocabulary. As shown in Figure 5.1, children with higher non-verbal intelligence scores also displayed higher receptive vocabulary skills (path a: $b = 1.08$, $p <.001$), and children with higher receptive vocabulary skills showed higher reading comprehension scores (path b: $b = 0.22$, $p <.001$). A bias-corrected bootstrap confidence interval for the indirect effect (path ab: $b = 0.24$) based on 5000 bootstrap samples was entirely above zero (0.085 to 0.470). The effect size for this indirect effect was $\kappa^2=.19$, 95% bias-corrected CI (0.068 to 0.323), which represents a medium to large effect.

However, there was evidence that non-verbal intelligence still had an influence on reading comprehension independent of its effect on receptive vocabulary (path c': $b = 0.37$, $p<.01$)

Figure 5.1 illustrates paths a and b, as well as the direct and indirect effects.

Figure 5.1 Model of Non-verbal Intelligence as a predictor of Reading Comprehension, mediated by Receptive Vocabulary. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 2: *Receptive vocabulary as mediator between discourse comprehension and reading comprehension*

Discourse comprehension as measured by CELF-4[UK] was also found to have an indirect influence on reading comprehension by its effect on receptive vocabulary. As Figure 5.2 shows, children with higher discourse comprehension skills also had higher receptive vocabulary skills (path a: $b = 3.56$, $p <.001$), and children with higher receptive vocabulary skills also had higher reading comprehension scores (path b: $b =0.27$, $p <.001$). The indirect effect (path ab: $b = 0.97$) had a bias-corrected bootstrap confidence interval, based on 5000 bootstrap samples, which was completely above zero (0.484 to 1.583). The effect size for this indirect effect was $\kappa^2=0.24$, 95% bias-corrected CI (0.118 to 0.363), indicating a large effect.

Moreover, there was no evidence that discourse comprehension influenced reading comprehension independent of its effect on receptive vocabulary (path c': $b=0.39$, $p=.371$).

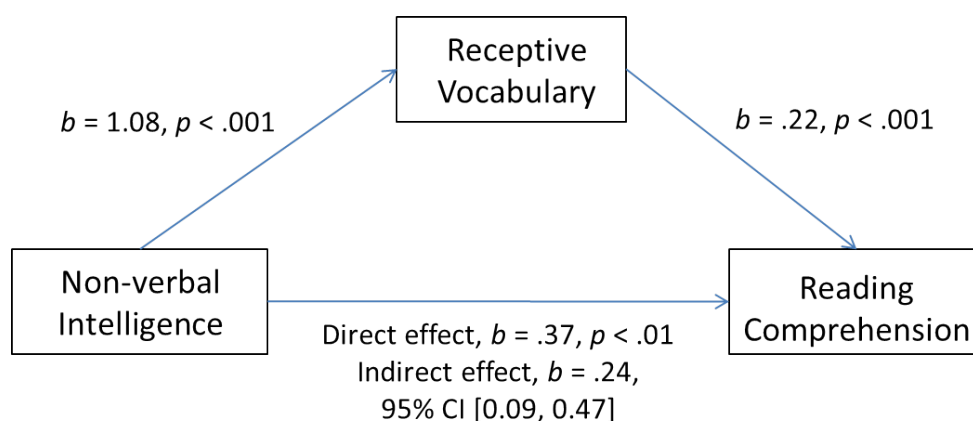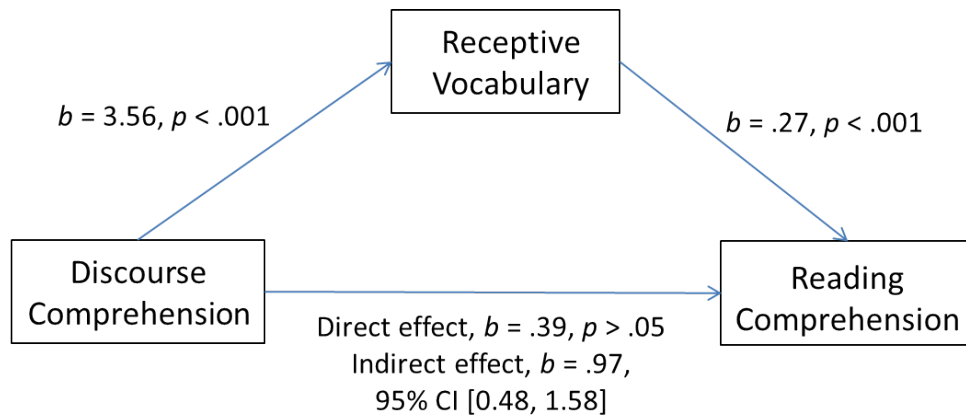All paths and effects are shown in Figure 5.2 below.

Figure 5.2 Model of Discourse Comprehension as a predictor of Reading Comprehension, mediated by Receptive Vocabulary. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 3*: Receptive vocabulary as mediator between expressive vocabulary and reading comprehension*

In this analysis, expressive vocabulary was found to indirectly influence reading comprehension mediated through its effect on receptive vocabulary as well. Figure 5.3 shows how children with higher expressive vocabulary measured by the composite VOCD, also had higher receptive vocabulary scores (path a: $b = 0.54$, $p = .012$), and pupils with higher receptive vocabulary scores had higher reading comprehension skills (path b: $b = 0.28$, $p < .001$). A bias-corrected bootstrap confidence interval for the indirect effect (path ab: $b = 0.15$) based on 5000 samples was completely above zero (0.044 to 0.294). The effect size for this indirect effect was $\kappa^2 = 0.14$, 95% bias-corrected CI (0.039 to 0.249), reflecting a medium effect.

The direct effect of expressive vocabulary on reading comprehension was just below the conventional significance level (path c': $b=0.23$, $p=.049$), suggesting the evidence was barely significant for expressive vocabulary to still have an influence on reading comprehension skills independent of its influence on receptive vocabulary. In other words, the direct effect was greatly, although not entirely diminished.

All paths and effects for Mediation Analysis 3 are illustrated in Figure 5.3.

Figure 5.3 Model of Expressive Vocabulary, measured as composite VOCD, as a predictor of Reading Comprehension, mediated by Receptive Vocabulary. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 4: Receptive vocabulary as mediator between expressive grammar and reading comprehension

The last analysis looking at receptive vocabulary as a mediator also found that it played a role between expressive grammar and reading comprehension. Indicated in Figure 5.4, pupils with higher expressive grammar in narratives, as measured by the composite MLT-w, also displayed higher receptive vocabulary scores (path a: $b = 4.13$, $p =.001$), and pupils with higher receptive vocabulary scores also displayed higher reading comprehension scores (path b: $b = 0.25$, $p <.001$). Testing the significance for the indirect effect (path ab: $b = 1.04$), a bias-corrected bootstrap confidence interval based on 5,000 samples was found to be completely above zero (0.417 to 1.987). The effect size for this indirect effect was $\kappa^2=.16$, 95% bias-corrected CI (0.064 to 0.278), which represents a medium to large effect.

There was evidence that expressive grammar influenced reading comprehension independent of its effect on receptive vocabulary (path c': $b = 1.83$, $p <.01$).

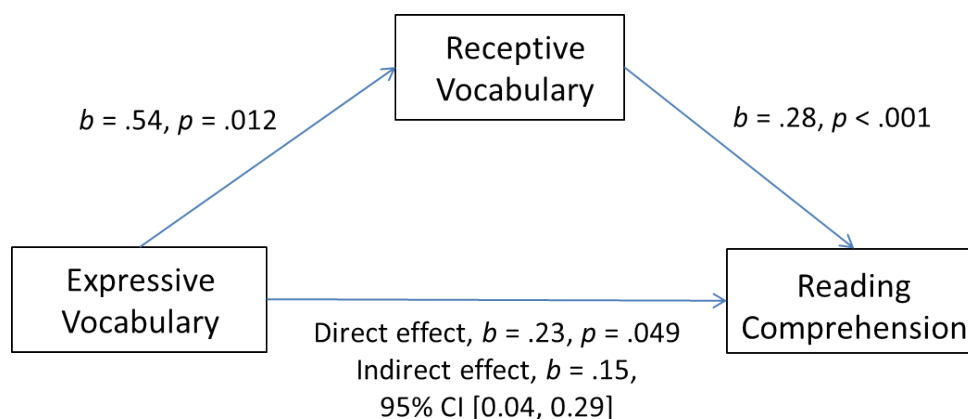Figure 5.4 illustrates the path coefficients as well as these effects.
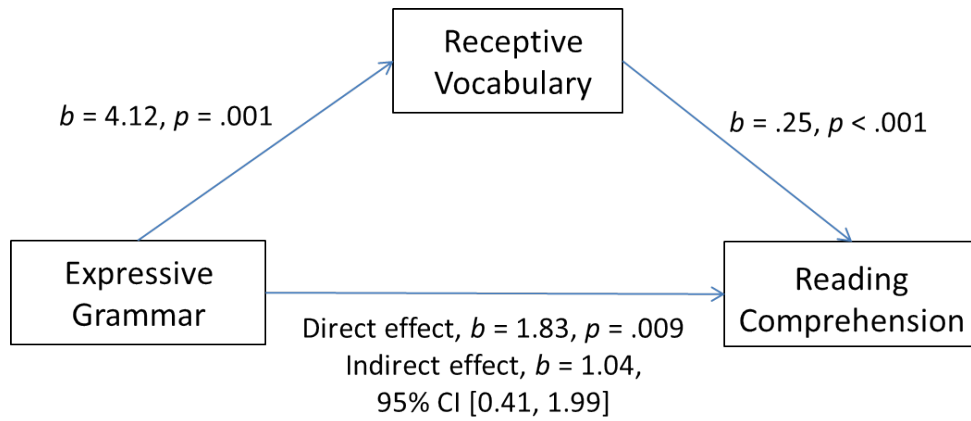
Figure 5.4 Model of Expressive Grammar, measured as composite MLT-w, as a predictor of Reading Comprehension, mediated by Receptive Vocabulary. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 5: *Receptive grammar as mediator between non-verbal intelligence and reading comprehension*

Non-verbal intelligence was also found to indirectly influence reading comprehension skills through its effect on receptive grammar. Shown in Figure 5.5, children with higher non-verbal intelligence scores displayed higher receptive grammar scores (path a: $b = 0.27$, $p <.001$), and those with higher receptive grammar scores had higher reading comprehension skills (path b: $b = 1.40$, $p <.001$). The significance of the indirect effect (path ab: $b = 0.37$) was tested using a bias-corrected bootstrap confidence interval based on 5,000 samples, which was completely above zero (0.219 to 0.581). The effect size for this indirect effect was $\kappa^2=.28$, 95% bias-corrected CI (0.178 to 0.396), indicating a large effect.

Furthermore, there was no evidence that expressive vocabulary had an effect on reading comprehension independent of its effect on receptive grammar (path c': $b = 0.24$, $p= .066$) at conventional significance levels.

Figure 5.5 illustrates all paths and these effects.

Figure 5.5 Model of Non-verbal intelligence as a predictor of Reading Comprehension, mediated by Receptive Grammar. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 6: *Receptive grammar as mediator between discourse comprehension and reading comprehension*

Discourse comprehension was also found to influence reading comprehension skills through receptive grammar. Shown in Figure 5.6, pupils with higher discourse comprehension skills had higher receptive grammar skills (path a: $b = 0.67$, $p<.001$), and pupils with higher receptive grammar skills had higher reading comprehension scores (path b: $b=1.54$, $p<.001$). A bias-corrected bootstrap confidence interval for this indirect effect (path ab: $b=1.03$) based on 5,000 bootstrap samples was completely above zero (0.547 to 1.624), and the effect size for this indirect effect was $\kappa^2=.28$, 95% bias-corrected CI (0.156 to 0.412), representing a large effect.

Just like it happened in Mediation Analysis 2, there was no evidence that discourse comprehension had an influence on reading comprehension skills independent of its influence on receptive grammar (path c': $b=0.34$, $p=.357$).

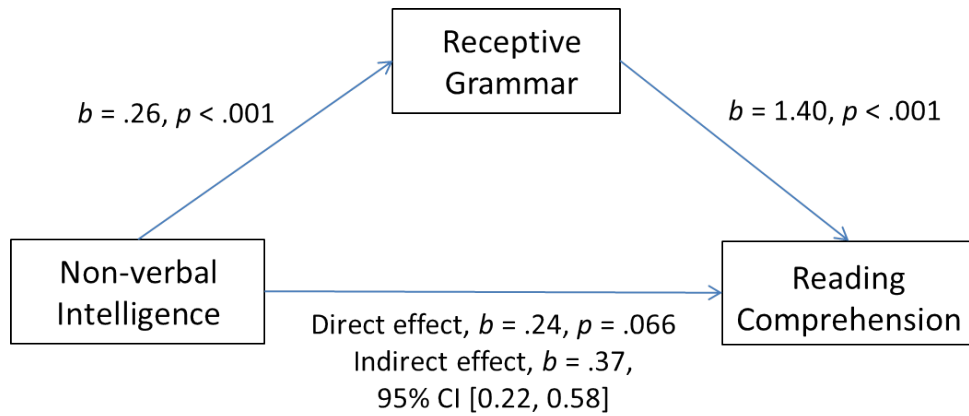These effects are shown in Figure 5.6.

Figure 5.6 Model of Discourse Comprehension as a predictor of Reading Comprehension, mediated by Receptive Grammar. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 7: *Receptive grammar as mediator between expressive vocabulary and reading comprehension*

Expressive vocabulary was found to indirectly influence reading comprehension through its effects on receptive grammar. Seen in Figure 5.7, pupils with higher expressive vocabulary scores, showed higher receptive grammar scores (path a: b=0.13, p<.01). Also, children with higher receptive grammar scores displayed higher reading comprehension skills (path b: b=1.54, p<.001). The indirect effect (path ab: b=0.20) was tested for significant using bias-corrected confidence intervals based on 5,000 bootstrap samples, and it was found to be entirely above zero (0.058 to 0.386). The effect size for this indirect effect was κ^2=.19, 95% bias-corrected CI (0.054 to 0.334), indicating a medium to large effect.

In this mediation analysis, there was no evidence that expressive vocabulary influenced reading comprehension beyond its influence on receptive grammar: the direct effect was not significant (path c': b=0.18, p=.092).

All coefficients and effects are shown in Figure 5.7.

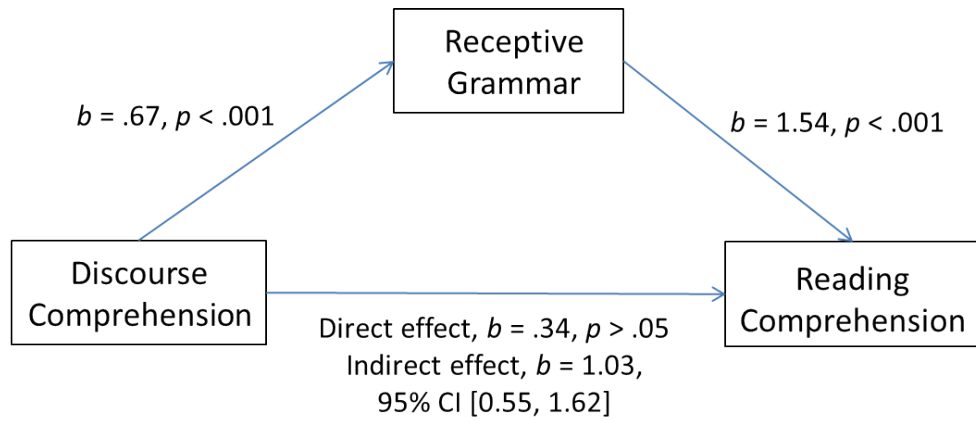Figure 5.7 Model of Expressive Vocabulary, measured as composite VOCD, as a predictor of Reading Comprehension, mediated by Receptive Grammar. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

Mediation analysis 8*: Receptive grammar as mediator between expressive grammar and reading comprehension*

The final mediation analysis found evidence to support the mediating role of receptive grammar between expressive grammar and reading comprehension. As displayed in Figure 5.8, children with higher expressive grammar skills measured with MLT-w composite had higher receptive grammar skills (path a: *b*=0.98, *p*<.001), and those with higher receptive grammar skills had higher reading comprehension scores (path b: *b*=1.45, *p*<.001). A bias-corrected bootstrap confidence interval for the indirect effect (path ab: *b*=1.42) based on 5,000 samples was completely above zero (0.647 to 2.382). The effect size for this indirect effect was κ^2=.22, 95% bias-corrected CI (0.105 to 0.335), which represents a medium to large effect.

However, there was still evidence that expressive grammar influenced reading comprehension skills independent of its effect on receptive grammar, as the direct effect was significant (path c': *b*=1.45, *p*=.022).

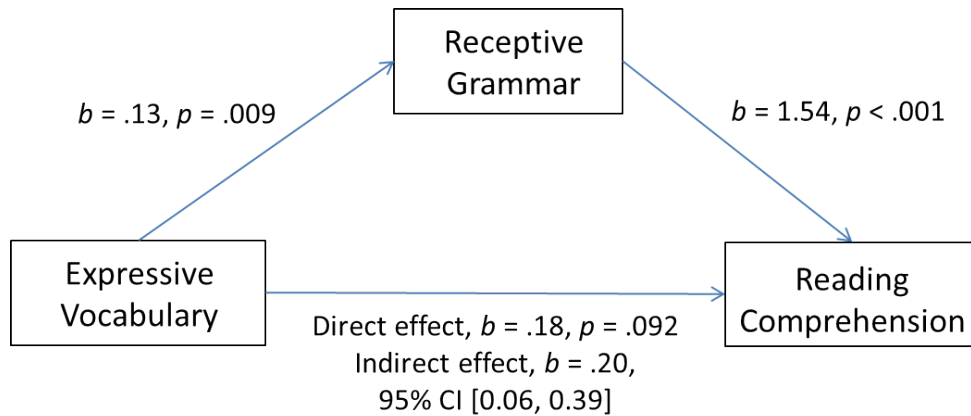Figure 5.8 illustrates path coefficients and these effects.
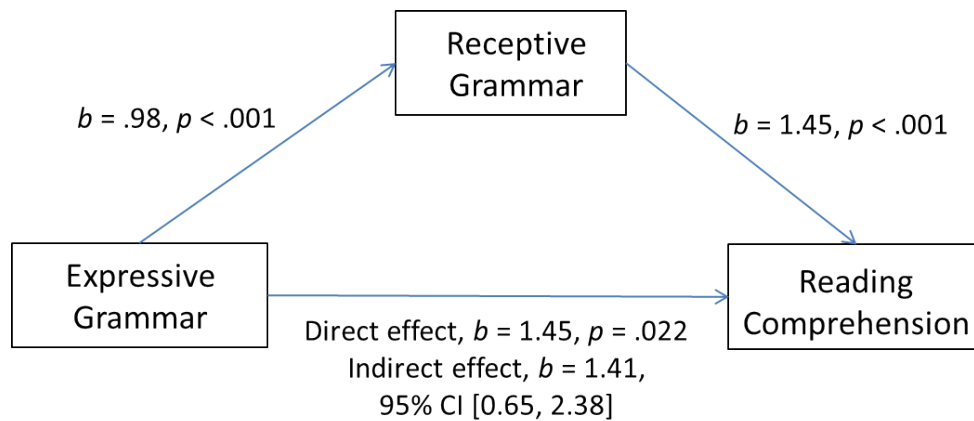
Figure 5.8 Model of Expressive Grammar, measured as composite MLT-w, as a predictor of Reading Comprehension, mediated by Receptive Grammar. The confidence interval for the indirect effects is a bias-corrected bootstrap CI based on 5000 samples.

In summary, all mediation analyses found significant indirect effects of differing effect sizes as measured by Preacher and Kelley's kappa-squared index (2011). Discourse comprehension was found to have a large indirect effect on reading comprehension through both receptive vocabulary and receptive grammar, in separate analyses. In addition to the large effect sizes, the mediations were found to be complete as the direct effects independent of the influence of both mediators were not significant.

For non-verbal intelligence, the largest effect size was found for the mediating role of receptive grammar, and to a lesser degree, for receptive vocabulary. The direct effects also reflected these effect sizes: the mediation was found to be complete for receptive grammar, while it was found to be partial for receptive vocabulary, where direct effects were still significant.

The main aim of these mediation analyses was to determine whether these receptive variables had an effect on the expressive variables. Indeed, there was evidence for expressive vocabulary's influence on reading comprehension to be mediated by both receptive vocabulary and receptive grammar skills. Interestingly, the mediation through receptive vocabulary was found to be partial, as the direct effect from expressive vocabulary to reading comprehension independent of the influence of mediator was still significant. In other words, expressive vocabulary measured as a composite VOCD, still had a small independent contribution to reading comprehension skills. Unexpectedly, the results suggest that the relationship between expressive vocabulary and reading

comprehension was completely mediated by receptive grammar, since direct effects from composite VOCD to reading comprehension independent of the mediator were no longer significant.

Finally, there was also evidence for expressive grammar's influence on reading comprehension to be mediated by both receptive vocabulary and receptive grammar. Remarkably, expressive grammar measured as composite MLT-w was only partially mediated by either receptive vocabulary or receptive grammar, i.e. there were still significant direct effects on reading comprehension independent of either receptive skill.

Therefore, even when regression analyses had indicated that expressive narratives variables could not account for further unique variance, these more detailed mediation analyses showed that these were partially or fully mediated by either receptive vocabulary or grammar. In particular, expressive grammar at least, was not fully mediated, indicating the possibility for this specific variable to explain unique variance in a larger sample.

5.3.6 Moderation analyses of gender and SES differences

To examine gender and SES differences a moderator analysis was carried out, to identify if and when the effects of each language sub-skill on reading comprehension occur depending on whether a child belonged to a specific gender or SES.

The relationship between each sub-skill and reading comprehension was plotted, and regression lines were obtained separately for each SES group and each gender. The effects of receptive vocabulary, receptive grammar, discourse comprehension, expressive vocabulary and expressive grammar on reading comprehension were all examined visually first (see Appendix E for the full group of scatterplots). After inspecting scatterplots of each sub-skill with reading comprehension with corresponding R^2 for each group, those where groups differed by 10% or more of variance explained were identified as candidates to examine the significance of their interactions. Those cases where the lines differed in their intercept, but had virtually the same slope, were not examined.

Different slopes for SES groups were identified in this way for discourse comprehension and expressive vocabulary. In preparation for moderation analysis, SES

and gender, as well as the two linguistic variables identified were centred (mean was subtracted from each value), and interaction terms were computed (Holmbeck, 1997).

Moderation analysis showed that only the interaction for expressive vocabulary and school-wide SES was significant for predicting reading comprehension (p <.05), indicating that at least for composite VOCD, SES plays a role on whether it is predictive of reading comprehension: disadvantaged children's vocabulary in narratives was more predictive of their reading than the middle-class cohort's vocabulary. The scatterplot of the actual regression lines is shown in Figure 5.9, and it illustrates that while the regression line is somewhat predictive for the disadvantaged children, for the middle-class children the prediction is virtually non-existent.
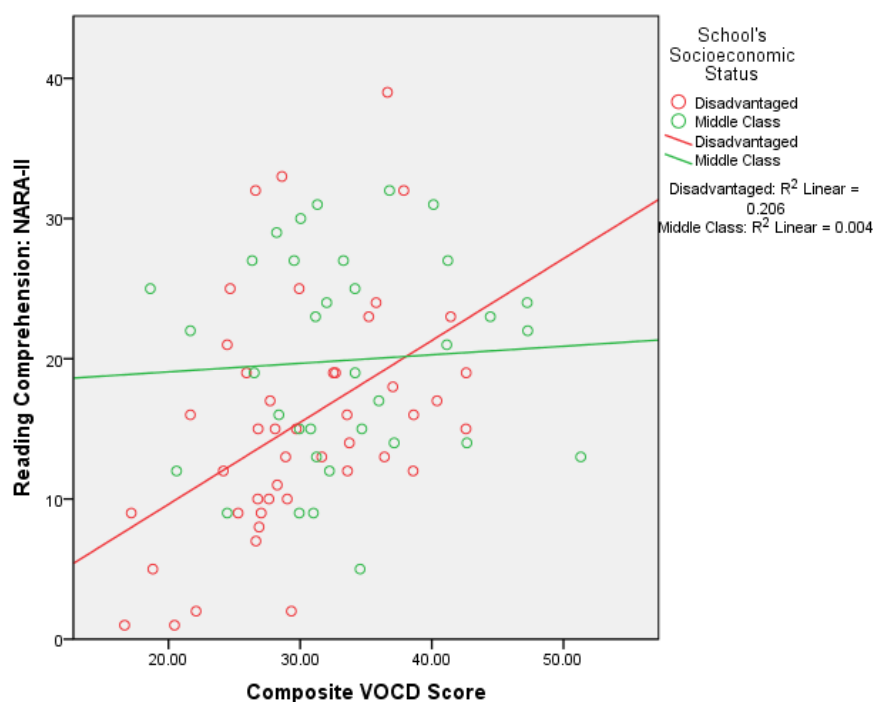
Figure 5.9 Interaction between school-wide SES and expressive vocabulary when predicting reading comprehension.

In the same manner, scatterplots of the relationships between the five linguistic indices and reading comprehension varying by gender were examined. Different slopes for boys and girls were visually identified for receptive vocabulary and expressive grammar, when predicting reading comprehension. Variables were centred and interaction terms computed. However, moderator analyses revealed these differences were not significant (p >.05).

5.4 Discussion

This discussion addresses the hypotheses proposed at the beginning of this chapter. However, before addressing these hypotheses, the findings about the relationships between the expressive narrative measures and their receptive standardised counterparts are discussed.

Expressive narrative indices and standardised receptive assessments.

The analyses linking receptive and expressive measures intended to explore how the associations varied by each single story and by using a composite made up of the three stories for two lexical indices and one syntactic index. Since the discourse measures were applied to different stories, no composite was made, and the relationships were only examined for the single stories where the discourse measures were applied, the Frog and Beach stories.

The previous chapter addressed the question of how these measures behaved across stories between ages 5 and 10. With these additional analyses presented in this chapter, albeit in a very narrow developmental sample at 7 and 8 years of age, finding some degree of association gave some external validity to these expressive measures, while a very high degree of association between the two modalities would have suggested these measures were redundant, indicating that once receptive measures are available expressive measures are in fact not needed.

Associations between the experimental expressive indices and their receptive counterparts were moderate, but significant in the 7- and 8-year-olds. Expressive vocabulary in the form of composite VOCD was related to receptive vocabulary at $r = .27$, $p < .05$, a coefficient similar to the reported correlation for the same expressive index in 11-year-old second-language learners written texts of $r = .29$ (Jarvis, 2002, p. 79). The composite VOCD had a strong correlation with receptive vocabulary scores that was slightly lower than the highest correlation shown by one of the VOCD scores in a single story (the Frog Story, $r=.29$). Meanwhile, the low non-significant correlations of types to the receptive vocabulary assessment suggesting that the index types is measuring a rather different construct.

Expressive grammar in composite MLT-w was also significantly related to receptive grammar at $r = .36$. More interestingly, in the case of the syntactic index, the composite had a higher correlation than any of the measures obtained from individual stories, suggesting that the composite was both more reliable and valid. These findings add further evidence for the validity of this particular syntactic index and they also support the methodological choice of obtaining a composite over an index extracted from a single story. However, a surprising finding was that of all the correlations to single stories, the Frog story, which in the previous analysis had identified the lowest scores for MLT-w in a wider age range, had also the highest measure for association to receptive grammar in the 7- and 8-year-olds. This discrepancy could be explored in future investigations.

For the associations between our discourse-level narrative indices, the INC and Information Content from ERRNI, with the standardised assessment of discourse comprehension, they both failed to reach conventional significance levels, but showed similar magnitudes of association. As this analysis was exploratory, no *a priori* hypothesis was tested. Nonetheless, these results were still unexpected, particularly for the INC, which is built around the story elements that have been well identified in research (Labov, 1972; Peterson & McCabe, 1983; Stein & Glenn, 1979).

In sum, these findings provide additional evidence for the external validity of VOCD and MLT-w, and suggest that the use of composites might be optimal over indices derived from a single story. The lack of significant associations with discourse measures, on the other hand, points to the need for further research to clarify either alternatives to our two narrative indices, or alternatives to the standardised assessment used here.

Research hypotheses

For the first hypothesis, the evidence from the first regression analysis suggests that, as expected, expressive measures from picture-book narratives would predict reading comprehension. Results indicated that on their own, expressive skills were able to predict over a fifth of the variability in reading comprehension at 7 and 8 years of age, when most weak receptive skills might go unnoticed. Although this contribution might not be as great in magnitude to be of clinical significance from the relatively small

sample in this study, it still highlights how a combination of simple narrative tasks might provide a window into children's reading comprehension skills.

Results also indicated that this contribution was mainly driven by expressive grammar: children who were able to formulate longer and more complex sentences by subordination were indeed better at understanding the texts they read, at this developmental point in reading comprehension skills, after the first couple of years of formal instruction. Although smaller, the contribution of expressive vocabulary to reading comprehension was still significant, indicating that those children who used more diversified vocabulary in the narrative tasks were also better at comprehending the texts they read.

Regarding the second hypothesis, that expressive measures from elicited picture-book narratives would contribute unique variance after controlling for both receptive language and non-verbal ability, the hierarchical regression analyses indicated that the contribution of expressive skills beyond variability explained by non-verbal ability was rather small yet still significant. However, once considering receptive skills, expressive skills as measured here were unable to explain further unique variance in the present sample. These results also imply that non-verbal intelligence does not contribute to our prediction of reading comprehension skills once receptive language skills have been taken into account. Whatever skill the assessment with non-verbal matrices is capturing in children at this age, it seems to overlap with those measured by the battery of receptive tests: the ability to distinguish, from an array of possible meanings, the one representing the correct information in a sentence, coupled with vocabulary knowledge and listening comprehension.

Contrary to the expectations set out in the third hypothesis of finding a primacy of vocabulary, observed the pilot study results, this hypothesis was not supported in the present study. Instead, grammar in both modalities had the strongest contribution in our regression analyses in 7- and 8-year-olds. The preponderance of grammar in both its receptive and expressive modalities at this stage, when for many children decoding skills are starting to be mastered in a way that they no longer place a limit to their reading comprehension skills, could be interpreted as being the natural intermediate step between understanding written words and processing written discourse. In other words, with some mastery of decoding skills, their already developed knowledge and production of grammar could be facilitating the comprehension of written sentences.

Overall, the results from the pilot were not replicated in the same way. In fact, there seemed to be a big discrepancy between the results of the pilot and the results from the main study for vocabulary and grammar in both modalities, While the bivariate correlation with reading comprehension increased for receptive grammar from $r = .50$ to .64 in the present study, the other three variables' correlations actually decreased: for receptive vocabulary, from $r = .69$ to .51; for expressive vocabulary, from $r = .61$ to .32; and for expressive grammar from $r = .53$ to .41. Comparing Tables 3.7 and 5.1, listing the means for all variables for each study, children in the pilot displayed unusually strong vocabulary scores in both modalities, and slightly weaker scores in receptive grammar, compared to children in the main study. The expressive grammar scores measured by MLT-w, it should be noted, were nearly identical.

Looking at the possible reasons behind the discrepancies between the pilot and the main study, there were slight differences in the elicitation, transcription and retrieval of the expressive indices between the two studies. For example, the present study included a visual barrier to maximise the opportunity for the production of decontextualized language. Children were also specifically instructed to tell a story "like a story you would read in a book". Transcription differences included the exclusion of onomatopoeia for the vocabulary analysis and exclusion of fragments for the syntactic analysis. However, the main difference was how the sample was conformed, in other words, how participants were recruited.

First, while the pilot study sample was composed of middle-class children, for the present study a deliberate effort was made to include schools in disadvantaged areas, which might account for the differences found in vocabulary: children in the pilot had a higher mean vocabulary score than children in the present study (both receptive and expressive). This finding was expected from the literature as less privileged children often have smaller vocabularies (Farkas & Beron, 2004; Hoff, 2003). With lower vocabularies, its relationship with reading comprehension could change, and as seen with the SES analysis, that was the case.

In addition, the request of parental consent only took place for one of the schools in the pilot, but for all the schools in the present study. At the school where written consent was only obtained from the headmaster, a somewhat wider range of abilities appeared, not so large as to make the reading skills in both schools significantly different, but included more children with low scores. At the pilot study's school where parental

invites and consent were sent at the request of the headmaster, there appears to be a restriction in the range of abilities by comparison.

Likewise, in the current study, which required parental consent, an even more marked effect was observed where volunteer rates were generally higher for high-achieving children, even when the language in the invitation letter was simplified and a decision was taken to exclude a questionnaire about parental education to avoid alienating disadvantaged parents. Although the mean reading comprehension score for the study's sample was only slightly over the norms, a closer look at the sample enables us to see that more competent students agreed to participate: from the total sample of 80 children in the present study, only 14 pupils had standard reading comprehension scores lower than 90 (standard mean=100).

Following the results from the pilot study, the most unexpected result was the lack of an association between the two measures of expressive discourse, the Index of Narrative Complexity and Information Content from ERRNI, with reading comprehension. Evidently, these discourse measures were completely different from the one used in the pilot study, where 20 narratives were manually scored for three implicit elements and three explicit elements. As we wanted to rely on discourse measures already developed and tested, the INC and Information Content were adopted in the main study. Furthermore, as a developmental trend had been observed using Information Content in the narrative study reported in the previous chapter, there was an expectation that there would be a significant relationship between this measure and reading comprehension skills. This non-significant result might be related to either the methodology of the task, the stage of development or both, and that should be explored further.

Regarding the nature of the task used to produce the Information Content score, it must be emphasised that this story was always elicited last and fatigue effects could have occurred. In addition, the directions of giving support to the elicitation of the beach story by pointing, as indicated in the ERRNI Manual were not followed, in order to avoid the elimination of the distance created in the first two stories and also to get a more authentic measure of what the child could perceive on their own. Nonetheless, comparing the scores from our sample to the norms in the manual shown in Table 5.2 suggests that these considerations might not have played such a strong role as to affect these scores.

Considering the developmental stage of Information Content score, results in Section 5.5 suggest that in Year 3, the ability to express sufficient information as measured by Information Score is not yet fully developed for all children, but some already have this greater awareness of the elements of the story. Since there is wide variability in this measure for expressive discourse skills, the possibility exists that this variability could explain *later* reading comprehension, but not concurrently at the stage examined here, particularly if grammar is so predominantly playing a role in comprehension.

Regarding receptive discourse, results in the present study did replicate a previous finding of a relatively small association between discourse comprehension and reading comprehension at this stage. In terms of strict age-matched comparisons, the magnitude of the correlations found here are consistent with the patterns presented in a Simple View meta-analysis (see Table 2.2 in Chapter 2; Gough et al., 1996): the meta-analysis grouped together children from 6 to 8 years of age, the correlations between listening comprehension and reading comprehension were $r = .41$, while the correlation found here in a group of 7- to 8-year-olds was $r = .37$. However, formal schooling begins a year earlier in the UK. We would have expected UK Year 3 children's listening comprehension to be more similar to older US Grade 3 children as both have already received two full years of reading instruction. Finding that even if formal schooling starts earlier in the UK, listening comprehension, at least in this sample, still follows its own pace, could suggest that reading instruction as it is currently carried out seems unrelated to the rate of development for this particular skill.

Regarding the fourth and last hypothesis that pupils with a combination of expressive and receptive weaknesses will have the poorest reading outcomes, it was considered that since expressive skills did not provide additional unique variance beyond receptive skills, at least in this sample, with concurrent measurements, then this in itself eliminated the possibility that combinations of weaknesses in both modalities could predict the lowest reading outcomes. That is the reason why mediation analyses, which were not originally planned, were carried out.

Mediation effects

Mediation analyses were conducted to examine whether expressive skills' influence on reading comprehension was mediated by either of the two main contributors to the

prediction of reading comprehension skills, receptive vocabulary and receptive grammar. Mediation analyses were also carried out to examine whether these two proposed mediators also played a role for discourse comprehension and non-verbal ability.

Non-verbal intelligence was found to be completely mediated by receptive grammar, and partially mediated by receptive vocabulary. Therefore, it seems that the specific role of non-verbal intelligence in receptive skills seems to be particularly important via the ability of a child to identify the correct picture corresponding to an orally heard sentence.

Meanwhile, discourse comprehension's influence on reading comprehension skills was found to be completely mediated by either receptive vocabulary or grammar. At this stage in development at least, the role of discourse comprehension is not yet clearly seen once considering the role of any of these two receptive skills, vocabulary or grammar.

The primary question of these mediation analyses was however whether the influence of the two expressive skills proposed here on reading comprehension was mediated by any of these two receptive skills, and the evidence supported several mediation paths for both expressive vocabulary and expressive grammar. These findings help in explaining why expressive skills were unable to find a unique contribution to reading comprehension once accounting for the influence of receptive skills in the previous regression analyses.

Nonetheless, the mediation paths were not completely as expected. First, expressive vocabulary's influence on reading comprehension was partially mediated through receptive vocabulary, while the mediation through receptive grammar was found to be complete. In other words, expressive vocabulary measured as a composite VOCD, still had a small independent contribution to reading comprehension skills when considering its influence through receptive vocabulary, but none when considering its influence through receptive grammar, which is a counterintuitive finding. Still it must be noted that the significance for the direct effect of expressive vocabulary on reading comprehension when considering receptive vocabulary was barely under conventional levels at $p=.049$, so more studies with larger samples are needed to determine whether this mediation is indeed partial.

Then, expressive grammar's influence on reading comprehension was found to be partially mediated by both receptive vocabulary and receptive grammar. Predictably, receptive vocabulary was only partially mediating the composite MLT-w influence on reading comprehension skills. However, an unexpected finding was that receptive grammar's role as a mediator was also partial: the influence of the composite MLT-w on reading comprehension skills independent of its effect through receptive grammar was still significant. The findings of partial mediation pathways for expressive grammar seem to suggest that receptive grammar does not tell the whole story and MLT-w could, in future investigations with larger samples potentially provide that possible unique contribution to reading comprehension skills.

Finding that expressive measures from narratives are actually mediated through receptive skills is not the result that was predicted in the hypotheses, but it is useful information and it could have practical implications. The driving force behind the search for the more visible narrative indices was that not all children are going to be tested using the one-on-one individual standardised assessments that were used here, and are routinely used to uncover severe language deficits. Then, by using an arguably more visible form of language assessment, that was closer to what practitioners could observe or elicit in their classrooms, it is possible to suggest that these expressive narrative skills might be a partial window by which carers can glimpse the potentially hidden deficits of their children or pupils.

Socio-economic Status moderation effects

Of all the socioeconomic and gender differences possible, only one of the reading comprehension predictors seemed to be significantly moderated by SES at the conventional *p* level of .05, and that was expressive vocabulary. This fits into the body of knowledge that vocabulary differences are well established by socioeconomic differences (Farkas & Beron, 2004), and that language skills are the underlying cause of achievement gaps by socioeconomic status (Durham et al., 2007). Interestingly, this moderation analysis was only significant for the *expressive* vocabulary index composite VOCD, which would suggest that spoken vocabulary, in narratives at least, could be a visible sign of vocabulary weaknesses in children of disadvantaged backgrounds, although more work is needed with SES indicators at the individual level. That vocabulary deficits are visible in children from lower SES in narrative assessment is

particularly relevant given the findings that children with language deficits from disadvantaged backgrounds tend to be less likely to be referred for specialist language services (Bishop & McDonald, 2009). Since early expressive vocabulary is greatly influenced by language input (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991), this has potential implications for the kinds of vocabulary that children are exposed throughout their preschool and school years.

Given that, as discussed above, this study's sample was formed mostly from competent readers, and that some trends observed failed to reach statistical significance, it is possible that with a larger sample with a wider range of abilities, and more precise SES indicators at the child level, a clearer picture of how these factors influence the development of language and reading comprehension could be obtained.

In general, even when failing to reach statistical significance, a trend was noted for either disadvantaged children's language to be more predictive of their reading abilities, or to have a similar level of prediction of reading (slope) but slightly weaker measures for most of the linguistic variables, confirming expectations of greater language influences in disadvantaged samples. Gender differences, on the other hand, were relatively weaker in this sample composed of more girls than boys, and of mostly skilled readers.

Summary

The modest but significant association of the expressive vocabulary and expressive grammar elicited by experimental measures with receptive standardised scores provides some support to the notion that these might be valid indicators of these linguistic aspects, yet measure linguistic facets not yet tapped by standardised tests.

Our main findings present new evidence that carefully chosen expressive skills, measured from narrative samples are able to partially predict reading comprehension, even in the context of a sample with a very likely self-selection bias. Of the three language sub-skills measured here, expressive grammar and expressive vocabulary, in that order, contribute to reading comprehension at this stage of development. These contributions were reduced, but remained significant, when also considering non-verbal

intelligence. However, these expressive contributions were virtually eliminated once receptive skills were considered.

Mediation analyses revealed, however, that both expressive skills, composite VOCD and composite MLT-w, were partially or completely mediated by receptive vocabulary or receptive grammar.

Our additional exploratory analysis also converge with previous studies in finding that socioeconomic status can have a moderating influence in whether a relationship between broader aspects of language and reading comprehension can be found or not, and this was particularly the case for expressive vocabulary.

The next and final chapter interprets these findings in the context of the literature and of methodological considerations, while it also suggests future lines of research.

Chapter 6

General Discussion and Conclusions

The present study was undertaken to make a particular contribution to the relatively limited literature linking expressive school-age language and reading comprehension. Looking into the diverse ways for measuring expressive skills, narratives were chosen as an ideal way to collect decontextualized language that could be engaging for pupils and be used as a tool for intervention. It also intended to add to the growing body of research on the links between specific broad-language skills and reading comprehension beyond the initial decoding years.

Oral language can be investigated on two different modalities: reception and expression. As described in the literature review in Chapter 2, when compared to the vast research investigating receptive language skills, the expressive skills of school-aged children have been largely overlooked, despite evidence documenting higher prevalence rates of reading disorders amongst those with language disorders (Catts, Adlof et al., 2005; Snowling et al., 2000), and the longitudinal evidence connecting early expressive language, mostly from language samples, with later reading outcomes (Bishop & Adams, 1990; Harlaar et al., 2008; NICHD Early Child Research Network, 2005; Scarborough, 1990b; Shapiro et al., 1990).

Concurrent associations need to be investigated because after the first couple of years of reading instruction focused on mastering the alphabetic principle and decoding words (Adams, 1990; Hoover & Gough, 1990), children face progressively longer continuous texts, and it is here where the impact of broader language skills starts to be seen (Gough et al., 1996). Moreover, at this stage many of the broad language deficits tend to go unnoticed, particularly if children are good 'decoders' (Nation et al., 2004). Furthermore, as very few children are likely to be formally tested using single-skill receptive tests of language, expressive skills measured with language samples would provide a window into the language aspects related to reading that could potentially be perceived, albeit in a lay manner, by those closest to the child.

Even when a few exceptions examining concurrent associations between expressive narrative language and reading comprehension do exist (Cain & Oakhill, 1996; Chen-Wilson, 2005; Klecan-Aker & Caraway, 1997; Snyder & Downey, 1991), their findings are difficult to compare and integrate given the heterogeneity of methods and populations. Indeed, the two main methodological issues in reading research, detailed in our literature review in Chapter 2, highlighted that developmental patterns and methodological choices in reading studies have a critical impact on results.

These two issues seem of particular importance regarding the contribution of expressive measures given reported evidence that earlier measures of language are more predictive of reading skills than later language measures (Scarborough, 2005). However, it has been argued here that this finding can be open to interpretation given that the impact of broad language skills is most clearly seen after the initial decoding years, and it is acknowledged that no consensus exists about the best way to measure expressive language and particularly expressive language from speech samples (Paul, 2007; Scott & Stokes, 1995).

The awareness of the critical role of the methods employed, coupled with the lack of consensus for measuring expressive measures led to a considerable amount of work dedicated to the selection of the optimal expressive measures in narratives. In addition, to avoid the issue of different predictors affecting reading at different rates depending on the stage of development (Scarborough, 2010), a narrow focus on a single developmental cohort aimed to give this study results that would be more useful for deriving practical implications. A very specific developmental window was chosen after the first two years of formal reading instruction at primary Year 3 in the UK, or 7 to 8 years of age, when basic word reading is generally, though not yet fully mastered. This specific age group is also the last developmental frame where reciprocal causal relationships from reading exposure feeding into language have not yet been documented, but are about to start soon after (Verhoeven & Van Leeuwe, 2008).

6.1 Findings in context: Comparisons and unique contributions

The three main studies reported in this thesis fall into two categories: indices derived from narrative language, and the reading-expressive language associations. The following two sections address each category respectively.

6.1.1 Indices from narrative language

While the selection of the optimal receptive measures was relatively straightforward by choosing the standardised tests representing the simplest measure of vocabulary, syntax and discourse, the selection of expressive measures presented specific challenges. First, a decision was made to elicit language samples instead of using expressive standardised tests to avoid the higher processing demands of a test (Scott & Stokes, 1995), and because they would be closer to the competence displayed by children in their natural environments (Shipley & McAffe, 2004), and therefore more similar to what those closer to the child, i.e. parents or teachers, perceive on a daily basis. As described at length in Chapter 3, picture-elicited narratives seemed to strike the best balance between the need for productivity and efficiency, and previous work documented a relationship between indices from these kinds of language samples and reading comprehension (Chen-Wilson, 2005).

The search for the optimal measures within the context of picture-book elicited narratives led this research to examine developmental patterns and reliability of automated linguistic indices in a database and in primary data collected for this purpose, covering a wide age range up to 9 and 10 years of age, or UK Year 5. Optimal vocabulary and syntactic indices were identified that were developmentally sensitive up to 7/8 years of age, and which were relatively reliable across tasks and samples of differing size in terms of tokens.

The contribution of the two studies investigating the reliability and developmental sensitivity of the indices derived from narratives in school-aged children was that the methodological choices for expressive measures were further supported. Previous studies existed that examined developmental patterns and/or reliability on a single linguistic index, such as vocabulary (Richards, 1987; Stromqvist et al., 2002; Vermeer, 2000) or syntax (Kemper et al., 1995; Verhoeven et al., 2002). However, it is important to examine diverse sub-skills together because of the potential trade-offs between different linguistic levels (Crystal, 1987).

Furthermore, our second study using primary data examining two types of discourse-level measures found evidence of developmental variability for one of them,

Information Content, between the two groups of older children, that is between ages 7/8 and 9/10.

A few previous studies had investigated developmental patterns in a comprehensive set of indices and in a wide range of primary school ages (Justice et al., 2006; Scott & Windsor, 2000). However the elicitation procedures in these comprehensive studies had led to either very short language samples in the first study, or very large time-consuming samples in the second one. In addition, the second study (Scott & Windsor, 2000) had mainly focused on those indices that best distinguished normal from impaired language development, while the focus of both of our narrative indices studies was on typical language development. Moreover, neither of these two studies had used the vocabulary index VOCD that was identified in the literature as showing developmental variability in school-aged children's writing and/or other languages (Malvern et al., 2004; McCarthy & Jarvis, 2007; Stromqvist et al., 2002).

A particular contribution from these linguistic indices studies is that they presented original findings about lexical developmental patterns from speech in English, using the more reliable index VOCD. Although the computation of VOCD is not difficult, it is only available in a very specific set of software for language analysis, CLAN (MacWhinney, 2000), which is not the most common software used in the speech-language literature reporting on developmental language patterns (the most commonly used software in the speech-language literature is SALT; Miller, 2008). Specifically, significant differences were documented for the comparatively reliable index VOCD between ages 5/6 and ages 7/8. The primary data study reported in Chapter 4 also documented that while types was able to be developmentally sensitive up to ages 9/10, its loss in reliability was high. To our knowledge, these are novel findings that provide a unique contribution to the literature about the developmental variability and reliability across tasks of lexical indices obtained from narrative samples.

Regarding developmental patterns for syntactic indices, even when the data about the syntactic index MLT-w was mostly replicating previous findings in different elicitation contexts and different populations, the linguistic studies presented here corroborated how reliable this index is across elicitation procedures and populations. The main contribution regarding MLT-w could be however, in our view, the creation of a brief yet fairly detailed manual addressing the most common issues in transcription and segmentation of school-age language, which can be found in Appendix D. Even when

the CLAN Manual (MacWhinney, 2000) contains the necessary transcription conventions to start working in school-age language, it vastly elaborates on the specific transcription issues of early language. More useful is the ERRNI Manual (Bishop, 2004), which provides some basic pointers for a few transcription and segmentation issues. However, both manuals offer little detail into the particular challenges of transcribing and segmenting school-age language, and the manual presented here was created specifically to deal with expressive language from 5 to 10 years of age. Of course, this manual is neither definitive nor exhaustive, but it could provide a very useful starting point for other non-specialist researchers looking into narrative indices in the future, before they embark on the daunting task of transcribing dozens of narratives. It should be noted that the reliability of T-Unit segmentation was greatly improved from the database study to the primary data study, suggesting the guidelines developed for the latter were useful for setting clearer segmentation criteria, at least for these narratives.

Further to the empirical validation of these indices, the data presented in both narrative language studies were able to show that a set of controlled stimuli provided a good opportunity to obtain fairly reliable measures of both vocabulary and syntax across different tasks.

Moreover, two discourse-level measures were evaluated for developmental patterns in two different stories. Although these two measures could not be directly compared as they were not applied simultaneously to the same story, and with the caveat that they were applied in a fixed order, results of the primary data study suggested that while the Index of Narrative Complexity was only able to differentiate between the youngest and oldest groups, more developmental variability was observed with the Information Content measure from ERRNI. Although both measures had poorer inter-rater reliability agreement than the linguistic indices, Information Content performed better in this regard.

In addition, data from the two investigations into narrative indices allowed for the comparison of developmental patterns for different linguistic indices in a wide range of ages. This is relevant in its own right, but more so in the context of identifying language predictors of reading since the peaks and valleys of language have an effect on their predictive power (Scarborough, 2010). While lexical and syntactic indices appeared to follow different (cross-sectional) patterns in the database study, the patterns obtained

from primary data revealed a more similar pattern between these two linguistic indices but a remarkably different one from the patterns followed by Information Score, which served as a proxy for expressive discourse (ERRNI; Bishop, 2004). Ranging from UK Years 1 through 5, or ages 5/6 through 9/10, the developmental patterns from three different expressive skills were illustrated with the empirically-supported indices, and provide a unique picture into the development of expressive narrative language in school-aged children, even when no generalisations can be made about developmental trajectories as these are not longitudinal data. It is recognised, however, that the specific methodology used for the elicitation of narratives with long picture books is possibly best suited to obtain variability in linguistic indices, and discourse-level measures might not offer as great variability as with verbal prompts (Cain, 2003; Cain & Oakhill, 1996).

The main study findings relating expressive language from picture-book elicited narratives are discussed next in the context of these and other methodological considerations and they are compared with similar studies from the existing literature.

6.1.2 The expressive language-reading connection

The main study investigating the concurrent relationship between expressive narrative language and reading comprehension intended to answer the question of how indices from picture-book elicited narratives were related to reading comprehension. In other words, we wanted to see whether some carefully chosen narrative indices were able to predict reading comprehension skills concurrently. We had focused on expressive narrative skills because these could provide visible signs to broad language weaknesses that tend to go unnoticed (Nation et al., 2004). Also, compared to the fraction of children diagnosed for specialist language services, there is a potentially greater proportion of the school population whose mild broad language deficits might be identified and supported with a better understanding of how expressive skills and reading comprehension are related.

Although much literature has linked early expressive language with reading development, the literature connecting school-aged language and reading in general, and school-aged narrative language and reading comprehension in particular, was much scarcer. Still, some previous studies had already documented some associations between reading comprehension and expressive narrative language sub-skills in school-aged

typically-developing children. Evidence exists for its relation with expressive syntax (Klecan-Aker & Caraway, 1997), a manual score for expressive discourse (Cain, 2003; Cain & Oakhill, 1996; Klecan-Aker & Caraway, 1997), and proportion of stories retold (Snyder & Downey, 1991). In the most recent study, which made full use of automated analysis (Chen-Wilson, 2005) MLU in clauses from a personal narrative was the only expressive index to be significantly related to reading comprehension, while MLU in clauses from the Sweets Story was not; in addition, Type-Token Ratio was not related to reading skills in this study. As this last study included children from 6;01 to 8;11, reading comprehension could have been confounded with decoding ability in the youngest children.

Most of these studies however, used a single or at most two stories to elicit narrative scores. With the exception of the Cain studies (Cain, 2003; Cain & Oakhill, 1996) which had multiple stories and a reading-matched design, no other report in this topic focused on a single developmental window, so changing patterns in the prediction of reading development (Scarborough, 2010) had an impact on the results. In addition, none of the studies evaluated, or if examined found, an association with expressive vocabulary.

It was also important to use tasks and indices which had some support of developmental variability, given that using indices such as Type-Token Ratio (Pearson, 2002) or tokens (Cain & Oakhill, 1996; Chen-Wilson, 2005), have been described as lacking both validity and reliability (Chapter 3), and have found inconsistent results. This lack of reliability in the variables' measurements has an effect on whether the results can be replicated. Furthermore, the empirical support for the relative advantages of these indices was necessary in light of the findings that earlier expressive measures are more predictive of reading skills than later measures.

The results presented here have taken full advantage of automated language analysis, in a medium-sized socially diverse sample, eliciting three different narratives from which a more reliable composite could be obtained, using procedures that emphasised the elicitation of decontextualized language. These findings contribute to the literature linking broader expressive and receptive aspects of language and reading comprehension in the following ways.

First of all, we have in our view, appropriately addressed the issue of having a sound methodology for measuring expressive school-age language, *before* addressing the

question of how expressive language and literacy are linked, in light of the findings that earlier expressive language at 3 and 4 years was more predictive of reading outcomes than expressive language at 5 years of age (Scarborough, 2005).

Next, to the best of our knowledge, we have presented original findings that, when taken on their own, expressive skills from picture-book elicited narratives, using developmentally-sensitive reliable indices, are able to explain over a fifth of the variance in typically developing readers of 7 and 8 years of age, in the UK primary Year 3, right after the initial decoding years.

More specifically, even when a multitude of longitudinal findings had linked expressive vocabulary and later reading skills (Griffin et al., 2004; Harlaar et al., 2008; Shapiro et al., 1990; Tabors, Snow et al., 2001; Walker et al., 1994), to the best of our knowledge, concurrent associations had not been reported. We have been able to show that expressive vocabulary as measured by VOCD is significantly associated with reading, and contributes to the prediction of reading comprehension in a regression model.

Findings that expressive grammar measures by the Mean Length of T-Units in words could concurrently predict reading comprehension skills, are not novel, but converge with previous findings which have also found this link with this particular index in older children (Klecan-Aker & Caraway, 1997) or with a less ideal variant, Mean Length of Clauses in morphemes, in younger children (Chen-Wilson, 2005). However, having examined a uniquely comprehensive set of expressive language skills, we have been able to define the relative importance of grammar versus vocabulary at this stage of reading comprehension development: comparing different sub-skills, expressive grammar as measured by MLT-w was shown to have a primary role in reading comprehension over expressive vocabulary, at this stage of development.

Even when the initial analysis reported that the contribution of expressive skills becomes redundant, once the receptive single-skill tests are included in the regression model, the mediation analyses revealed that these receptive skills served a mediating role for the expressive indices proposed here. Moreover, finding significant indirect effects of these two expressive indices through the influence of receptive vocabulary and receptive grammar, gives support to the notion that these more visible expressive skills might be a window into the hidden deficits that can go unnoticed in children at this stage of development, particularly if they are decoding well (Nation et al., 2004).

Similar work, but examining longitudinal relations between narrative indices in UK Year 1 and later reading comprehension is being carried out (Silva & Cain, 2010). As the results of this study have yet to be reported at the time of writing this thesis, no comparisons can be made.

Results from the main study were modest to be of clinical importance, particularly when compared to the results of our own pilot described in Chapter 3. Nonetheless, in the context of being a sample of mostly competent readers, it is possible that results could have differed if more struggling readers had been included. In a similar fashion, the magnitude of the prediction had the potential to become larger if reading measures had been collected at a later point where comprehension and accuracy were more dissociated (Oakhill et al., 2003). Given the high correlations between accuracy and comprehension in our particular sample, decoding could have had an influence in our results. Also, the study could have arrived at different results in a longitudinal design given that broad language skills in general have a greater effect as more time passes (Tabors, Snow et al., 2001; Walker et al., 1994).

6.2 Theoretical implications

We started this research trying to evaluate whether a semi-structured naturalistic approach of language assessment – with its promise of greater ecological validity, which does not penalise culture-specific forms of language, and its potential to lend itself to training – would provide a glimpse of the children's sub-skills useful for reading comprehension. Our findings indicate that they do, but to a limited extent.

An additional aim was to explore whether these carefully chosen expressive measures would explain more of the reading comprehension phenomenon that it can already be explained by standardised single-skill receptive tests. In the context of our sample, composed of mostly competent readers, they do not, at least not concurrently. However, further analyses found that receptive language skills mediated the relationship between expressive narrative skills and reading comprehension. Moreover, it remains to be seen how this range of skills would play out in a longitudinal study, as correlations between broad language skills and reading comprehension tend to increase in magnitude with

time, both for receptive and expressive language skills (Table 2.1, Chapter 2; Walker et al., 1994).

Even with the possibility of having a more ecologically valid expressive language measure, which would be more similar to what practitioners can hear from a pupil, the potential of this methodology was not realised. A possibility might be that greater ecological validity is not needed because, after all, reading comprehension is still a culturally bound task; in other words, even if we are able to distinguish between linguistic ability (i.e. great linguistic ability in a non-standard dialect of English) and cultural exposure, with a novel narrative assessment such as the one presented in this thesis, reading comprehension would still require both linguistic skills and cultural exposure to be mastered. Some literature provides evidence that familiarity with Standard English can be predictive of reading achievement in disadvantaged populations (Charity et al., 2004), suggesting that actually cultural exposure to Standard English could be equally important in mastering reading comprehension. Nonetheless, actual usage or expression of Standard English versus just familiarity has not been measured. Although this was never the aim of this study, from a theoretical stance this skill/culture interaction could be explored further, and elicited narratives could provide an appropriate tool for such explorations.

It was also proposed at the beginning of this work that expressive narrative language could prove particularly useful when disentangling socioeconomic influences. Even when our study only used a SES school-wide measure, it was able to find some general trends for disadvantaged children to exhibit weaker receptive and expressive skills, a phenomenon already documented in the literature (Farkas & Beron, 2004; Hoff, 2003). More interestingly, a significant result for the moderation analysis indicated that at least for expressive vocabulary, some level of prediction of reading comprehension skill could only be found for disadvantaged children, but not for middle-class pupils. In the literature of longitudinal studies, it is the family's socioeconomic status which has an effect in language, which in turn has an effect on reading skills (Durham et al., 2007). In the context of the snapshot provided by our data, the SES variable indicated that only for disadvantaged children did the diversity of their vocabulary have a relationship with their reading comprehension skills.

One of the two prerequisites for language development is having access to a language model (Hoff, 2006). The role of input, i.e. speech, in children's language has been

heavily documented. Early language learning has been shown to be related to the rate and quality of parental speech. For example, early children's vocabulary has been shown to be related to the amount of words that parents speak to their children (Huttenlocher et al., 1991), while the syntax of 8-year-olds has been shown to be related to the syntactic quality of parental speech (Huttenlocher et al., 2002). Parental talk also facilitates children to use language for expressing their experience in literate ways (Snow et al., 1998). However, for some children their language model is characteristically different, particularly in communities from disadvantaged backgrounds (Heath, 1983). It still remains to be seen however, if a language model provided by teachers, i.e. teacher's daily instructional speech, can also influence the rate of literate language development.

In addition to these issues to support the use of narratives, evidence from a small-sample study suggest that they are useful tools not only for assessment, but also for intervention. This study found that narrative training in children with poor oral narration abilities had a direct impact on learning a specific set of vocabulary words (Cable, 2007) at precisely the 7-8 years of age developmental window studied here. The data are still scarce, however, so the evidence is only suggestive.

Independently of the possible pathways in which literate language can be enhanced and whether narrative training can have an effect on reading comprehension, our findings that carefully chosen developmentally sensitive expressive narrative measures can be, even if modestly, related on their own to reading comprehension skills at this stage of development, are a novel contribution to the reading literature in general, and to the research linking expressive school-age language and reading, in particular. Moreover, our findings that receptive skills were found to mediate the relationships between these expressive measures and reading comprehension skills highlight that even if partial, these narrative indices could provide a window into a child's linguistic repertoire.

6.3 Practical implications

One of the main reasons why expressive skills measured with elicited narratives are considered relevant in general (Paul, 2007), and were considered even more appropriate for this particular study, was their great potential for application (Cable, 2007). Although the very first markers of future language difficulties are receptive and can be

identified by measuring how newborns perceive sounds in a laboratory (Guttorm et al., 2005), the majority of language difficulties will be identified when parents or care workers first notice a child is delayed in his/her speech (Bates et al., 1995). In the same manner, we wanted to know whether there were noticeable school-aged narrative language markers that could be observed in a lay manner by those closer to the children, parents and practitioners, and in a systematic manner by researchers, to further understand how some expressive features could potentially affect reading comprehension. The aim was to uncover these markers as a way of targeting the often ignored broad language skills that ultimately contribute specifically to comprehension, in a way that would allow diverse linguistic forms to be examined.

Although the concurrent evidence presented here was too small to be of clinical significance, elicited narratives can still offer some insight into the linguistic repertoire of children. Given that pupils will not be routinely assessed with the battery of single-skill receptive language tests used here, our partial findings open up the possibility of using a simpler form of the narrative protocol reported in these studies to get a coarse measure of linguistic ability in their diverse sub-skills which, according to our results, could be particularly useful for observing disadvantaged children's vocabulary.

One of the main ideas guiding the selection of expressive measures from language samples, particularly narrative language samples, was that teachers, who are the ones closest to the pupils, could in some way perceive the first language weakness that in turn might potentially affect reading comprehension. Although the literature is scarce and indirect, there is some evidence that teachers in Australia's Kindergarten, or age 5 (Jessup, Ward, Cahill, & Keating, 2008) and Year 1, or age 6 (Williams, 2006), are able to reliably judge children who would require specialist language services. The analysis presented in this study however, entails more training and resources than those normally available to teachers, but it nonetheless has underscored the importance of identifying language strengths and weaknesses in a format similar to what they could experience in a classroom environment.

The findings that expressive grammar is the main expressive language contributor to reading skills at this developmental window, would suggest for teachers to pay more attention to the length of T-units pupils use when presenting some form of discourse, be it a narrative or an expository presentation, as potential markers of current or future reading comprehension weaknesses. Further studies could validate whether informal

observations such as these correlate concurrently or longitudinally with standardized or institutional reading comprehension assessments.

Findings that expressive vocabulary is only related to reading comprehension skills in disadvantaged populations could also be further explored and examined. It could be the case that a teacher-led focus on developing expressive vocabulary as part of discourse could be developed and promoted in disadvantaged schools.

Another promising feature of the narratives as elicited in this study was the opportunity of creating an expressive assessment that was not only valuable, but also interesting for pupils, a characteristic which would have made it helpful as a tool for intervention. Observations from data collection suggest that this is indeed a stimulating task that most children at this age are eager to engage in and enjoy. Moreover, for some children the idea of an adult not seeing the pictures, made them particularly motivated in terms of having some autonomy in an otherwise ordinary school activity. Recognising that it is the provision of communicative opportunities that drive language development (Hoff, 2006), there may be a place for elicited narratives not only as instruments of assessment, but also as pedagogical tools, although this application was outside the scope of this study. When parents are taught interactive strategies to challenge and involve their young children in story telling during reading time, these children's language presented significant differences from the control group whose parents continued reading in a traditional way (Arnold, Lonigan, Whitehurst, & Epstein, 1994; Whitehurst et al., 1988).

It is reasonable to expect that more systematic exposure to broader language skills such as vocabulary, grammar and discourse abilities should facilitate acquisition of a more literate style of oral language, which in turn should have an effect on reading comprehension abilities.

There is considerable attention already given to the development of decoding abilities in the UK national literacy curriculum (Rose, 2006). And although the need to develop speaking and listening skills has been stated (Department for Education and Skills, 2003) the specific rate of development, and its subsequent impact on reading skills, is not yet fully researched and what is known is not always widely disseminated amongst practitioners. Our results highlight, like other language-reading studies, the need to emphasise broader language skills in the classroom.

6.4 Limitations

From the outset, it was acknowledged that by choosing a cross-sectional design, both the narrative and the reading studies would not able to produce findings that could generate either developmental trajectories for language in the first case, or causal statements from language to reading in the second case. The causality had been conceptualised as coming, not from the study itself, but from the literature reported in Chapter 2 about longitudinal data where early language differences are in direct relationship with later reading differences.

Still, a longitudinal design would have allowed for a time-2 measurement of reading comprehension at a later date than the one allowed by a concurrent design, so that it would not have been as closely correlated to our proxy of decoding, reading accuracy, as it was in our sample. A longitudinal design is also likely to have improved the significance of these results given that the evidence reviewed in Chapter 2 suggests that the more time passes between time 1 of the language assessment and time 2 of the reading assessment, these relationships tend to become stronger for both receptive (Chapter 2) and for expressive measures (Snow et al., 1995; Tabors, Snow et al., 2001; Walker et al., 1994).

Therefore, even when basing the direction of the influence to go from language variables to reading skills, it is still possible that concurrent associations like the ones presented in this work could also represent the other direction of reciprocal relations, going from print to language.

Nonetheless, within the framework and resources provided, it was considered that resources would be better spent finding additional developmental and intra-child validity for the proposed narrative measures before examining the relationship between reading comprehension and expressive narrative language. Future investigations will have the opportunity to take the findings from this study, and apply them longitudinally and possibly even in intervention studies.

Therefore, our findings can only be considered as to provide a partial answer to the research question into identifying the relative contributions between narrative language

and reading comprehension. Once more, future lines of research will be able to provide more data to complete the picture across the developmental spectrum.

In the context of choosing a cross-sectional design, a decision could have been made to select a later developmental window, where indeed reading comprehension and decoding could have been more differentiated. However, in the context of well-documented reciprocal relationships in the literature between print exposure and language (Cunningham & Stanovich, 2001), a decision was made to specifically choose the developmental window where the existing reading and language data suggest that causality still goes from language to reading, and reciprocal relations are yet to begin (Verhoeven & Van Leeuwe, 2008).

Another limitation of our study is that the specific procedures used here could have influenced pupils' communicative performance. As mentioned in the previous chapter, the three stories were collected, for most children, in a single session, and this could have had an impact in the youngest and oldest pupils. Even when the results for tokens in the narrative study were fairly similar to those from the database study (with the exception of some very talkative 5-year-olds in the database cohort) for the Sweets story, results for Beach story, which was always administered last, suggest that there might have been fatigue effects for the youngest and oldest pupils, that is children in Years 1 and 5. Collecting the data in two sessions, in retrospect, could have been advisable.

Furthermore, it should be noted that even when three different picture-books were examined, different results could have been obtained with different methodological choices, as illustrated in the Cain studies (Cain, 2003; Cain & Oakhill, 1996). Specifically, the most notable differences in those studies were seen when comparing narratives elicited from verbal prompts versus the ones elicited from a picture sequence. Less-skilled readers' performance was better supported, and therefore, better overall when using a picture sequence than when only hearing a verbal prompt to elicit their narratives. This could mean that some of the less-skilled readers in our sample could have displayed greater variability in terms of discourse-level structure, if their narratives had been elicited with verbal prompt. Our choice was to control the stimulus with the picture-books to provide a level playing field in terms of linguistic output, but it must be considered that the methodology used here might not have been optimal for the measurement of discourse-level coherence.

Sampling issues could be, however, the greatest limitations in this study. First, in statistical terms this was a convenience sample, even when this term would hardly describe the difficulties in gaining access to schools and children. With volunteers and an informally observed self-selecting bias, the sample was composed mostly of competent participants, which makes the sample a truncated one (Tabachnick & Fidell, 2007), an issue that has been described at greater length in the previous chapter. Having mostly competent readers did not allow for enough variation to examine how the language and reading skills of less competent readers were related. While the pilot study results suggested that having less-skilled readers increased the strength of the associations between language and reading, the socioeconomic analysis in the previous chapter suggested that including children from disadvantaged areas was likely to strengthen this relationship as well. From the data gathered in the pilot and the reading study described here it would seem that having few struggling readers had more of an impact on the results than the inclusion of competent readers from disadvantaged areas.

From communication with teachers a possibility emerged as how this could happen: less competent children might have been less likely to remember to give the invitation to their parents and/or bring the form back to the teacher, or even less likely to want to participate in the study themselves (although the less skilled children who did participate were happy to do so). In addition, one of the teachers commented on how parents of less able children might have been wary of their child being in a study, regardless of the nature of the research.

Moreover, this volunteer bias effect seems to have also taken place for schools: one of the disadvantaged schools agreeing to participate had recently been commended for their outstanding performance. In any case, it can be argued that a self-selecting bias was in place, where brighter pupils, and their parents and high-achieving schools were happy to volunteer, but less skilled pupils, their parents and lower-achieving schools were not.

The comparison between the results from the pilot and from the present study highlights the dangers of having a truncated sample (Tabachnick & Fidell, 2007), composed in this case of mostly competent readers. Vellutino and colleagues (1996) had already noted that, in their intervention study, many middle-class children did not receive such a great benefit from these educational interventions, as they were not needed. This, in fact,

guided the effort to include a more socioeconomically diverse sample, but this in itself might have been insufficient to include enough numbers of less able readers. Unfortunately, this self-selecting bias works precisely against those children who would potentially perceive a greater benefit from this and any other study in the field of reading research.

Of course, requiring parental consent was part of the ethical considerations for this and any other study requiring the participation of children. The challenge for future studies looking at expressive indices derived from narrative language would be to reach that difficult balance where all ethical considerations are taken into account, while at the same time the disadvantaged and less able children, who are less likely to volunteer, are nevertheless well represented in these kinds of studies.

To avoid the dangers of obtaining a truncated sample, the gold standard is still a random or even a stratified sample. Of course, random sampling is rare in education, though some well-funded studies compare interventions with the gold-standard of a randomised-controlled-trial (Kaiser et al., 2011; Sylva, Scott, Totsika, Ereky-Stevens, & Crook, 2008). Even when more random sampling should be done in education, whenever ethical, financial and logistical issues can be resolved, it can be argued that there is a place for small fine-grained studies such as the one presented here. The distinctively comprehensive set of language skills sampled here has the potential to, with more access and consequently better sampling, provide a much richer and detailed picture of how specific sub-skills relate to reading, something that becomes difficult in the large-sampled broadly-defined studies.

Another important sampling issue concerns the relatively small sample size. With six language predictors, and intelligence as a control variable (at the outset there were two controls, but age turned out to be non-significant for such a narrow window at this stage), it would have been desirable to have a much larger sample size of $50+8k$, where k is the number of predictors, or 106 participants just for testing the overall fit of the model (Field, 2009). A sample of $104+k$, or 111 participants would have allowed for more confidence when evaluating individual predictors.

An even larger sample size would also have allowed the use of multilevel modelling, a much more powerful statistical technique because this study, as well as a majority of studies in educational research, is made of a sample of data in sub-groups. Multilevel modelling is a very appropriate technique in this context, because it is able to deal

precisely with correlated errors in clustered data (Field, 2009). Moreover, as our exploratory SES analysis showed, errors can occur when heterogeneous populations are analysed together (Stride, 2008).

Nonetheless, it should be noted that considering the labour-intensive nature of the speech sample data used here, in terms of transcription, and accuracy/segmentation reliability checks, a sample of 80 is by no means inconsequential.

6.5 Directions for future research

The previous sections have clearly emphasised the need for reading comprehension to be measured at a later time. For all the reasons described here, a longitudinal design would have been the ideal next step for this reading study, if we assume that strength of correlations would have increased as decoding influences would have waned. With more access, and true random sampling, the methodology proposed here could still be used in a longitudinal design to look for early predictors of late-emerging reading comprehension deficits.

Given that this methodology allows observing a glimpse of the child's linguistic repertoire and discourse abilities, the procedures and measures used here could be used to examine links with and devise interventions for writing skills, specifically since these tools would provide a means to observe the child's own linguistic repertoire.

An additional facet that should be examined in relation to the elicited narratives methodology is how much the temperament of a child moderates the relationship with language. Extraversion and negative affect have been shown to moderate conversational language samples in young children (DeThorne, Deater-Deckard, Mahuring-Smith, Coletto, & Petrill, 2011). Even when care was taken to minimise confounding variability by providing children with a structured stimulus, this issue still warrants further investigation as a possible confounding variable.

Regarding the indices from narrative language, the instrument comprising procedures and measures used here could be further validated using longitudinal designs, which would minimise the possibility of cohort effects. This methodology could also potentially be used for the analysis of first language oral skills in Bilingual children

wherever there are homogeneous minorities, which might provide an alternative to standardised tests of expressive skills.

The measures used here, could also be further validated in the context of expository discourse, which might produce more rich measures of grammar. The potential of expository discourse lies in that it can be used to measure language in older children (Nippold et al., 2008).

Finally, further research could possibly solve the apparent contradiction where the Frog Story appeared somewhat more valid, in that it was more strongly correlated to standardised tests, than the other stories used here, while at the same time, it appeared the least reliable in that it was the story which threw significantly different results to the other two stories. Further addressing the validity of these measures in a wider developmental window could potentially clarify this issue.

While this issue is being addressed, it is possibly best to use a composite measure, as the one used here, derived from several elicitations.

6.6 Conclusion

Can we state that better communicators are indeed better readers? Based on our data, they are, but only to a limited extent in a sample consisting of mostly competent readers. In particular, vocabulary narrative indices seem as likely candidates to display language deficits in disadvantaged children, as the relationship between expressive language and reading comprehension was moderated by SES.

Vocabulary and grammatical narrative indices did not seem to provide additional information once accounting for receptive skills. However, further analyses revealed that receptive vocabulary and receptive grammar played a mediating role between expressive narrative indices and reading comprehension. Therefore, even if expressive skills measured using a more ecologically valid form of assessment, were not able to account for additional variance in reading comprehension, the children's diversity of vocabulary and complexity of their sentences by subordination can still be viewed as more visible indices of potentially hidden language deficits. Moreover, these narrative markers could become more predictive of later reading comprehension, given that

concurrent reading comprehension was still strongly associated with decoding skills at this particular developmental window.

We expect that more systematic exposure to broader language skills such as vocabulary, grammar and discourse abilities, should facilitate acquisition or a more literate style of oral language, which in turn should have an effect on reading comprehension abilities.

Even when giving adequate consideration to developmental differences in the reading acquisition process, a crucial element of carrying out research in education involves sampling, and this was the greatest limitation of the present study. Having adequate access to schools, though not enough to guarantee better sampling, is a pre-requisite for producing generalizable results which can later become useful at the practitioner level, and benefit children of all levels of skills.

References:

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adlof, S. M., Catts, H., & Lee, J. (2010). Kindergarten predictors of Second versus Eight grade reading comprehension impairments. *Journal of Learning Disabilities, 43*(4), 332-345.

Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies* Oxford: Routledge.

Allen, M. S., Kertoy, M. K., Sherblom, J. C., & Petit, J. M. (1994). Children's narrative productions: A comparison of personal event and fictional stories. *Applied Psycholinguistics, 15*, 149-176.

American Speech-Language-Hearing Association. (2010). Retrieved 7 September, 2011, from http://www.asha.org/public/speech/disorders/

Anderson, R. C., & Nagy, W. (1993). *The vocabulary conundrum*: Center for the Study of Reading, University of Illinois at Urbana-Champaign.

Applebee, A. N. (1978). *The Child's Concept of Story: Ages Two to Seventeen*. Chicago: The University of Chicago Press.

Arnold, D. H., Lonigan, C. J., Whitehurst, G. J., & Epstein, J. (1994). Accelerating language development through picture book reading: replication and extension to a video tape training format. *Journal of Educational Psychology, 86*(2), 235-239.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *Handbook of Child Language* (pp. 96-151). Malden, MA, USA: Blackwell Publishing.

Bates, E., & Goodman, J. (1999). The emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum.

Beals, D. E. (2001). Eating and reading. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning Literacy with Language: Young Children at Home and School* (pp. 75-92). Baltimore, MD: Brookes Publishing.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: a crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berninger, V., & Abbott, R. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: related yet unique language systems in grades 1, 3, 5, and 7. *Journal of Educational Psychology, 102*(3), 635-651.

Beron, K. J., & Farkas, G. (2004). Oral language and reading success: a structural equation modeling approach. *Structural Equation Modeling, 11*(1), 110-131.

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research, 38*(2), 446-462.

Bishop, D. V. M. (1991). Developmental reading disabilities: The role of phonological processing has been overemphasized. *Mind and Language, 6*, 97-101.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove: Psychology Press.

Bishop, D. V. M. (2003). *Test for the Reception of Grammar- Version 2*. Manchester, UK: Chapel Press.

Bishop, D. V. M. (2004). *Expression, Reception and Recall of Narrative Instrument*. London: Harcourt Assessment.

Bishop, D. V. M., & Adams, C. (1990). A prospective study of the relationship between Specific Language Impairment, phonological disorders and reading retardation. [empirical]. *Journal of Child Psychology and Psychiatry, 31*(7), 1027-1050.

Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders, 44*(5), 600-615.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: same or different? *Psychological Bulletin, 130*(6), 858-886.

Bliss, L. (2002). *Discourse impairments: Assessment and intervention applications*. Boston: Allyn & Bacon.

Booth, J. N., Boyle, J. M. E., & Kelly, S. W. (2010). Do tasks make a difference? Accounting for heterogeneity of performance of children with reading difficulties on tasks of executive function: Findings from a meta-analysis. *British Journal of Developmental Psychology, 28*(1), 133-176.

Bowey, J. A. (1986). Syntactic awareness and verbal performance from preschool to fifth grade. *Journal of Psycholinguistic Research, 41*, 282-299.

Bradley, L., & Bryant, P. (1983). Categorising sounds and learning to read: a causal connection. *Nature, 301*, 419-421.

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

Bryant, P., MacLean, M., & Bradley, L. (1990). Rhyme, language and children's reading. [empirical]. *Applied Psycholinguistics, 11*, 237-252.

Cable, A. L. (2007). *An oral narrative intervention for second-graders with poor oral narrative ability.* University of Texas at Austin, Austin, TX.

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Developmental Psychology, 21*(3), 335-351.

Cain, K. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics, 28*, 679-694.

Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology, 14*, 187-201.

Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*, 697-708.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96*(1), 31-42.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meaning from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*, 671-681.

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36*, 948-958.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are Specific Language Impairment and Dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*, 1378-1396.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: a case for the Simple View of Reading. *Journal of Speech, Language, and Hearing Research, 49*(2), 278-293.

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading Achievement Growth in Children With Language Impairments. *Journal of Speech, Language, and Hearing Research, 51*, 1569–1579.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142-1157.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. [empirical]. *Scientific Studies of Reading, 3*(4), 331-361.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilites* (pp. 25-40). New York: Lawrence Erlbaum Associates.

Charity, A. H., Scarborough, H. S., & Griffin, D. M. (2004). Familiarity with school English in African American children and its relation to early reading achievement. *Child Development, 75*(5), 1340-1356.

Chen-Wilson, C. (1997). The sweets story. Unpublished Illustrated book.

Chen-Wilson, C. (2003). *Temporality in English- and Mandarin-Speaking Children's Narratives.* Unpublished PhD., University of Manchester, Manchester.

Chen-Wilson, C. (2005). Are better communicators better readers? *Proceedings of the British Psychological Society, 13*(1), 33.

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of Discourse Understanding* (pp. 10-63). Cambridge: Cambridge University Press.

Clifton, C., & Duffy, S. A. (2001). Sentence and text comprehension: Roles of linguistic structure. *Annual Review of Psychology, 52*, 167-196.

Cooper, D. H., Roth, F. P., & Speece, D. L. (2002). The contribution of oral language skill to the development of phonological awareness. *Applied Psycholinguistics, 23*, 399-416.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), 1-9.

Cragg, L., & Nation, K. (2006). Exploring Written Narrative in Children with Poor Reading Comprehension. *Educational Psychology, 26*(1), 55-72.

Crais, E. R., & Lorch, N. (1994). Oral narratives in school-age children. *Topics in Language Disorders, 14*(3), 13-28.

Crystal, D. (1987). Towards a 'bucket' theory of language disability: taking account of interaction between linguistic levels. *Clinical Linguistics & Phonetics, 1*(1), 7-22.

Crystal, D. (2008). *A dictionary of Linguistics and Phonetics* (6th ed.). Oxford: Blackwell Publishing.

CTBS. (1987). *The comprehensive test of basic skills*. Monterey, CA, USA: Macmillan/McGraw-Hill.

Cunningham, A. E., & Stanovich, K. E. (2001). What reading does for the mind. *Journal of Direct Instruction, 1*(2), 137-149.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: relative contributions of word recognition, language proficiency and other cognitive skills can depend on how comprehension is measured. [empirical]. *Scientific Studies of Reading, 10*(3), 277-299.

de Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension in the development of reading. *Scientific Studies of Reading, 6*(1), 51-77.

de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research, 2*, 267-278.

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25*, 223-238.

Demont, E., & Gombert, J. E. (1996). Phonological awareness as a predictor of recoding skills and syntactic awareness as a predictor of comprehension skills. *British Journal of Developmental Psychology, 66*(3), 315-332.

Department for Education and Skills. (2003). *Primary National Strategy. Speaking, Listening, Learning: working with children in Key Stages 1 and 2*. Retrieved. from.

DeThorne, L. S., Deater-Deckard, K., Mahuring-Smith, J., Coletto, M., & Petrill, S. A. (2011). Volubility as a mediator in the associations between conversational language measures and child temperament. *International Journal of Language & Communication Disorders, 46*(6), 700-713.

DeThorne, L. S., Hart, S. A., Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., et al. (2006). Children's history of speech-language difficulties: Genetic influences and associations with reading-related measures. *Journal of Speech, Language, and Hearing Research, 49*(6), 1280-1293.

Duncan, G. J., Yeung, W. J., Brooks-Gunn, J., & Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *American Sociological Review, 63*, 406-423.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale II*. Windson, England: Nfer-Nelson.

Durham, R. E., Farkas, G., Hammer, C. S., Tomblin, J. B., & Catts, H. W. (2007). Kindergarten oral language skills: a key variable in the intergenerational transmission of socioeconomic status. *Research in Social Stratification and Mobility, 25*, 294-305.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. [review]. *Reading Research Quarterly, 36*(3), 250-287.

Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 95-113.

Entwisle, D. R., Alexander, K. R., & Olson, L. S. (2005). First grade and educational attainment by age 22: A new story. *American Journal of Sociology, 110*(5), 1458-1502.

Farkas, G., & Beron, K. J. (2004). The detailed age trajectory of oral vocabulary knowledge: Differences by class and race. *Social Sciencie Research, 33*, 464-497.

Farmer, M. (1996). *Associations and disassociations in specific language impairment. Issues in normal and disordered child language: from phonology to narrative.* Paper presented at the Child Language Seminar, University of Reading, Reading, UK.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: Sage Publications.

Fry, R. (2010). *Understanding household poverty at small area level*. Retrieved. from.

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*, 659-671.

Gavin, W. J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of pre-school children: A review. *Journal of Speech and Hearing Research, 39*, 1258-1262.

Gazella, J. (2003). Children's story retelling under different modality and task conditions: Implications for standardizing language sampling procedures. *American Journal of Speech-Language Pathology, 12*, 61-72.

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes, 23*(3), 265-303.

Gibbons, P. (1998). Classroom talk and the learning of new registers in a second language. *Language and Education, 12*(2), 99-118.

Gillam, R. B., & Pearson, N. (2004). *Test of Narrative Language*. Austin, TX: Pro-Ed.

Gillam, R. B., Pena, E. D., & Miller, L. (1999). Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders, 20*(1), 33-47.

Girolameto, L., Wiigs, M., Smyth, R., Weitzman, E., & Pearce, P. S. (2001). Children with a history of expressive vocabulary delay: Outcomes at 5 years of age. *American Journal of Speech-Language Pathology, 10*(4), 358-369.

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing, 18*, 583-616.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hove: Lawrence Erlbaum.

Gottardo, A., Stanovich, K. E., & Siegel, L. S. (1996). The relationships between phonological sensitivity, syntactic processing, and verbal working memory in the reading performance of third-grade children. [empirical]. *Journal of Experimental Child Psychology, 63*, 563-582.

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

Griffin, T. M., Hemphill, L., Camp, L., & Wolf, D. P. (2004). Oral discourse in the preschool years and later literacy skills. *First Language, 24*(2), 123-147.

Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading, 3*(3), 231-256.

Guttorm, T. K., Leppanen, P. H. T., Poikkeus, A., Eklund, K. M., Lyytinen, P., & Lyytinen, H. (2005). Brain Event-Related Potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia *Cortex, 41*(3), 291-303.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hansson, K., Nettelbladt, U., & Nilholm, C. (2000). Contextual Influence On The Language Production Of Children With Speech/Language Impairment. *International Journal of Language & Communication Disorders, 35*(1), 31-47.

Harlaar, N., Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2008). Why do preschool language abilities correlate with later reading? A twin study. *Journal of Speech, Language, and Hearing Research, 51*(3), 688-705.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation and Conditional Process Analysis: A Regression-Based Approach*. New York: Guildford Press.

Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. Cambridge, UK: Cambridge University Press.

Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. D. (2008). Narrative transcription accuracy and reliability in two langauges. *Topics in Language Disorders, 28*(2), 178-188.

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the Narrative Scoring Scheme. *American Journal of Speech-Language Pathology, 19*, 154-166.

Hess, C. W., & Haug, H. T. (1989). The reliability of Type-Token Ratios for the oral language of school age children. *Journal of Speech and Hearing Research, 32*, 536-540.

Hickmann, M. (2004). Coherence, cohesion and context in narrative development. In S. Stromqvist & L. Verhoeven (Eds.), *Relating Events in Narrative, Volume 2: Typological and Contextual Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development, 74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55-88.

Holmbeck, G. N. (1997). Toward terminological, conceptual and statistical clarity in the study of mediators and moderators: examples from the child-clinical and pediatric psychology literatures. *Journal of Consulting and Clinical Psychology, 65*(4), 599-610.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to Narrative Language: Procedures for Assessment*. Eau Claire, WI, USA: Thinking Publications.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels (Research Report No. 3)*. Urbana, IL: National Council of Teachers of English.

Hunt, K. W. (1970). Recent measures in syntactic development. In M. Lester (Ed.), *Readings in Applied Transformational Grammar*. New York: Holt, Rinehart & Winston.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relations to language input and gender. *Developmental Psychology, 27*(2), 236-248.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1992). Early vocabulary growth: Relations to language input and gender. *Developmental Psychology, 27*(2), 236-248.

Huttenlocher, J., Vasilyeva, M., Cymerman, E., & Levine, S. (2002). Language input and child syntax. *Cognitive Psychology, 45*, 337-374.

Jarmulowicz, L., Hay, S. E., Taran, V. L., & Ethington, C. A. (2007). Fitting derivational morphology into a developmental model of reading. *Reading and Writing, 21*(275-297).

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Jessup, B., Ward, E., Cahill, L., & Keating, D. (2008). Teacher identification of speech and language impairment in kindergarten students using the Kindergarten Development Check. *International Journal of Speech-Language Pathology, 10*(6), 449-459.

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational and quality of life outcomes. *American Journal of Speech-Language Pathology, 19*, 51-65.

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The Index of Narrative Microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology, 15*(2), 177-191.

Justice, L. M., Bowles, R. P., Turnbull, K. L. P., & Skibbe, L. E. (2009). School readiness among children with varying histories of language difficulties. *Developmental Psychology, 45*(2), 460-476.

Kaiser, A., Dickinson, D. K., Roberts, M., Darrow, C., Freiberg, J., & Kerry, H. (2011). *The effects of two language-focused preschool curricula on children's achievement through first grade*. Birmingham, Alabama, USA: Society for Research on Educational Effectiveness.

Katusic, S. K., Colligan, R. C., Barbaresi, W. J., Schaid, D. J., & Jacobsen, S. J. (2001). Incidence of reading disability in a population-based birth cohort, 1976-1982, Rochester, Minn. *Mayo Clinic Proceedings, 76*, 1081-1092.

Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading, 29*(1), 75-91.

Kemper, S., Rice, K., & Chen, Y. (1995). Complexity metrics and growth curves for measuring grammatical development from five to ten. *First Language, 15*, 151-166.

Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765-778.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: a Handbook* (pp. 209-226). Oxford: Blackwell Publishers.

Klecan-Aker, J. S., & Caraway, T. H. (1997). A study of the relationship of storytelling ability and reading comprehension in fourth and sixth grade African-American children. *European Journal of Disorders of Communication, 32*, 109-125.

Klecan-Aker, J. S., & Hedrick, D. L. (1985). A study of the syntactic language skills of normal school-age children. *Language, Speech and Hearing Services in Schools, 16*, 187-198.

Klecan-Aker, J. S., & Lopez, B. (1985). A comparison of T-Units and cohesive ties used by first and third grade children. *Language and Speech, 28*(3), 307-315.

Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language, 12*, 251-269.

Labov, W. (1972). *Language in the inner city*. Philadelphia: University of Pensnsylvania Press.

Labov, W., & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the Verbal and Visual Arts* (pp. 12-44). Seattle: University of Washington Press.

Larrivee, L. S., & Catts, H. (1999). Early reading achievement in children with expressive phonological disorders. *American Journal of Speech-Language Pathology, 8*(2), 118-128.

Law, J., Boyle, J., Harris, F., Harknes, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *International Journal of Language and Communication Disorders, 35*(2), 165-188.

Leadholm, B. J., & Miller, J. F. (1992). *Language Sample Analysis: The Wisconsin Guide*. Madison, Wisconsin: Wisconsin Department of Public Instruction.

Lee, L. (1974). *Developmental Sentence Analysis*. Evanston, IL: Northwestern University Press.

Liles, B. Z. (1985). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28*, 123-133.

Liles, B. Z. (1993). Narrative discourse in children with language disorders and children with normal language: A critical review of the literature. [review]. *Journal of Speech and Hearing Research, 36*, 868-882.

Long, S., & Channell, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology, 10*, 180-188.

Lyytinen, P., & Lyytinen, H. (2004). Growth and predictive relations of vocabulary and inflectional morphology in children with and without familial risk for dyslexia. *Applied Psycholinguistics, 25*, 397-411.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. (2000). *Gates-MacGinitie Reading Tests* (4th ed.). Itasca: Riverside Publishing.

MacWhinney, B. (1996). The CHILDES System. *American Journal of Speech-Language Pathology, 5*, 5-14.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (Third ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Malvern, D., Richards, B., Chipere, N., & Duran, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke, UK: Palgrave Macmillan.

Marzano, R. J. (2004). *Building background knowledge for academic achievement: Reserach on what works in schools*. Alexandria, USA: Association for Supervision and Curriculum Development.

Masterson, J. J., & Kamhi, A. G. (1991). The effects of sampling conditions on sentence production in normal, reading-disabled, and language-learning-disabled children. *Journal of Speech and Hearing Research, 34*, 549-558.

Mayer, M. (1969). *Frog, Where Are You?* New York: Dial Books for Young Readers.

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "specifics" of Specific Reading Disability and Specific Language Impairment. *Journal of Child Psychology and Psychiatry, 41*(7), 869-874.

McCabe, A., Bliss, L., Barra, G., & Bennett, M. (2008). Comparison of personal versus fictional narratives of children with language impairment. *American Journal of Speech-Language Pathology, 17*(2), 194-205.

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459-488.

McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*(3).

Meeks, S., & Murrell, S. A. (2001). Contribution of education to health and life satisfaction in older adults mediated by negative affect. *Journal of Aging and Health, 13*(1), 92-119.

Merritt, D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorders: Story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research, 30*, 539-552.

Merritt, D., & Liles, B. Z. (1989). Narrative analysis: clinical application of story generation and story retelling. *Journal of Speech and Hearing Disorders, 54*, 438-447.

Miller, J. F. (2008). Systematic Analysis of Language Transcripts (SALT). Muscoda, Wisconsin, USA: Salt Software LLC.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research, 24*, 154-161.

Miller, J. F., Long, S., McKinley, N., Thormann, S., Jones, M., & Nockerts, A. (2005). *Language Sample Analysis II: The Winsconsin Guide* (Vol. Bulletin No. 5056). Milwaukee: Wisconsin Department of Public Instruction.

Miranda, E., Camp, L., Hemphill, L., & Wolf, D. (1992). Developmental Changes in children's use of tense in narrative., *Boston University Conference on Language Development*. Boston.

Moats, L. C. (1994). The missing foundation in teacher education: knowledge of the structure of spoken and written language. *Annals of Dyslexia, 44*, 81-102.

Moats, L. C., & Foorman, B. R. (2003). Measuring teachers' content knowledge of language and reading. *Annals of Dyslexia, 53*(1), 23-45.

Munoz, M. L., Gillam, R. B., Pena, E. D., & Gulley-Faehnle, A. (2003). Measures of language development in fictional narratives of Latino children. *Language, Speech and Hearing Services in Schools, 34*, 332-342.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665-681.

Nagy, W., Berninger, V., & Abbott, R. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology, 98*, 134-147.

Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004). The development of early literacy skills among children with speech difficulties: A test of the "critical age hypothesis". *Journal of Speech, Language, and Hearing Research, 47*, 377-391.

Nation, K. (2005). Children's reading comprehension difficulties. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: a Handbook* (pp. 248-265). Oxford: Blackwell Publishers.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: parallels between poor reading comprehension and specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*(1), 199-211.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39*, 85-101.

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics, 21*, 229-241.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342-356.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Report of the subgroups*. Bethesda, MD: National Institute of Child Health and Human Development.

Neale, M. D. (1997). *Neale Analysis of Reading Ability: Second Revised British Edition*. Windsor, England: Nfer-Nelson.

Nelson, N. W. (1988). The nature of literacy. In M. A. Nippold (Ed.), *Later Language Development: Ages 9 through 19* (pp. 11-28). Austin, TX: Pro-Ed.

NICHD Early Child Research Network. (2005). Pathways to reading: the role of oral language in the transition to reading. [empirical]. *Developmental Psychology, 41*(2), 428-442.

Nippold, M. A. (1988). Introduction. In M. A. Nippold (Ed.), *Later Language Development: Ages 9 through 19* (pp. 1-10). Austin, TX: Pro-Ed.

Nippold, M. A., Hesketh, A., Duthie, J. K., & Mansfield, T. C. (2005). Conversational versus expository discourse: A study of syntactic development in children, adolescents and adults. *Journal of Speech, Language, and Hearing Research, 48*(5), 1048-1064.

Nippold, M. A., Mansfield, T. C., & Billow, J. L. (2007). Peer conflict explanations in children, adolescents, and adults: Examining the development of complex syntax. *American Journal of Speech-Language Pathology, 16*(2), 179-188.

Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech-Language Pathology, 17*, 356-366.

O'Donnell, R. C., Griffin, W. J., & Norris, R. C. (1967). A transformational analysis of oral and written grammatical structures in the language of children in grades three, five and seven. *The Journal of Educational Research, 61*(1), 35-39.

Oakhill, J. (1994). Individual differences in children's text comprehension. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 821-848). San Diego, CA: Academic.

Oakhill, J., Cain, K., & Bryant, P. (2003). The dissociation of word reading and text comprehension: evidence from component skills. *Language and Cognitive Processes, 18*(4), 443-468.

OECD, & Statistics Canada. (2000). *Literacy in the information age: Final report of the International Adult Literacy Survey*. Ottawa: OECD.

Office for National Statistics. (2008). Households in Poverty Estimates. Retrieved 29 February 2012, from http://neighbourhood.statistics.gov.uk/HTMLDocs/poverty.html

Ouellette, G. P. (2006). What's meaning got to do with it: the role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554-566.

Pan, B. A., & Snow, C. E. (1999). The development of conversational and discourse skills. In M. Barret (Ed.), *The development of language* (pp. 229-249). Hove: Psychology Press.

Paris, S. G. (2005). Reinterpreting the development of reading skills. [review (theory and methodology)]. *Reading Research Quarterly, 40*(2), 184-202.

Paris, S. G., Carpenter, R. D., Paris, A. H., & Hamilton, E. E. (2005). Spurious and genuine correlates of children's reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's Reading Comprehension and Assessment* (pp. 131-160). Mahwah, N. J.: Lawrence Erlbaum Associates.

Paris, S. G., & Myers, M. (1981). Comprehension monitoring, memory, and study strategies of good and poor readers. *Journal of Reading Behavior, 13*, 5-22.

Paris, S. G., & Paris, A. H. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*(1), 36-76.

Paul, R. (2007). *Language Disorders from Infancy through Adolescence: Assessment and Intervention* (3rd. ed.). St Louis, Missouri, USA: Mosby Elsevier.

Pearson, B. Z. (2002). Narrative competence among Monolingual and Bilingual school children in Miami. In D. K. Oller & R. E. Eilers (Eds.), *Language and Literacy in Bilingual Children* (pp. 135-174). Clevedon, UK: Multilingual Matters Limited.

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among speech, language and reading. *Annual Review of Psychology, 60*, 238-306.

Perfetti, C. A. (1985). *Reading ability*. Oxford: Oxford University Press.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamins.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: a Handbook* (pp. 227-247). Oxford: Blackwell Publishers.

Perna, L. W. (2003). The private benefits of higher education: An examination of the earnings premium. *Research in Higher Education, 44*(4), 451-471.

Petersen, D. B., Gillam, S. L., & Gillam, R. B. (2008). The Index of Narrative Complexity. *Topics in Language Disorders, 28*(2), 115-130.

Peterson, C., & McCabe, A. (1983). *Developmental Psycholinguistics: Three Ways of Looking at a Child's Narrative*. New York: Plenum Press.

Plaut, D. C. (2005). Connectionist Approaches to Reading. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 24-38). Oxford: Blackwell Publishers.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilites. *Psychological Bulletin, 131*(4), 592-617.

Preacher, K. J., & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879-891.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93-115.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1984). *Metropolitan Achievement Tests*. New York: Psychological Corporation.

Rapp, D. N., Van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*(4), 289-312.

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology, 8*, 209-225.

Rego, L. L. B. (1997). The connection between syntactic awareness and reading: evidence from portuguese-speaking children taught by a phonic method. *International Journal of Behavioral Development, 20*, 349-365.

Rego, L. L. B., & Bryant, P. (1993). The connection between phonological, syntactic and semantic skills and children's reading and spelling. *European Journal of Psychology of Education, 8*, 235-246.

Ricard, R. J., & Snow, C. E. (1990). Language use in and out of context. *Journal of Applied Developmental Psychology, 11*, 251-266.

Richards, B. (1987). Type/Token ratios: What do they really tell us? *Journal of Child Language, 14*(2), 201-209.

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading, 11*(3), 235-257.

Rose, J. (2006). *Independent Review of the Teaching of Early Reading*. Nottingham: DfES Publications.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *Journal of Educational Research, 95*, 259-272.

Sagae, K., Lavie, A., & MacWhinney, B. (2005). *Automatic measurement of syntactic development in child language.* Paper presented at the 43rd Meeting of the Association for Computational Linguistics, Ann Arbor.

Sanders, T. J. M., & Gernsbacher, M. A. (2004). Accessibility in text and discourse processing. *Discourse Processes, 37*(2), 79-89.

Scarborough, H. S. (1990a). Index of Productive Syntax. *Applied Psycholinguistics, 11*, 1-22.

Scarborough, H. S. (1990b). Very early language deficits in dyslexic children. *Child Development, 61*(6), 1728-1743.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo & A. J. Capute (Eds.), *Specific Reading Disability: A view of the spectrum* (pp. 75-107). Timonium, MD: York Press.

Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 3-34). New York: Lawrence Erlbaum Associates.

Scarborough, H. S. (2010). *When Matters!* Paper presented at the Seventeenth Annual Conference of the Society for the Scientific Study of Reading.

Scarborough, H. S., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech and Hearing Research, 33*, 70-83.

Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language disordered groups. *Applied Psycholinguistics, 12*, 23-45.

Scott, C. M. (1988). Spoken and written syntax. In M. A. Nippold (Ed.), *Later Language Development: Ages 9 through 19* (pp. 49-96). Austin, TX: Pro-Ed.

Scott, C. M., & Stokes, S. L. (1995). Measures of syntax in school-age children and adolescents. *Language, Speech and Hearing Services in Schools, 26*, 309-319.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324-339.

Seigneuric, A., Ehrlich, M. F., Oakhill, J., & Yuill, N. (2000). Working memory resources and children's reading comprehension. *Reading and Writing, 13*, 81-103.

Semel, E., Wiig, E. H., & Secord, W. (2006). *Clinical Evaluation of Language Fundamentals - 4 UK Edition*. Londong: Harcourt Assessment.

Senechal, M. (2006). Testing the Home Literacy Model: Parent involvement in Kindergarten is differentially related to Grade 4 reading comprehension, fluency, spelling and reading for pleasure. *Scientific Studies of Reading, 10*(1), 59-87.

Shankweiler, D., Crain, S., Brady, S. A., & Macaruso, P. (1992). Identifying the causes of reading disability. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading Acquisition* (pp. 275-305). Hillsdale, NJ: Erlbaum.

Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., Brady, S. A., et al. (1995). Cognitive profiles of reading disabled-children: comparison of language skills in phonology, morphology, and syntax. *Psychological Science, 6*(3), 149-156.

Shapiro, B. K., Palmer, F. B., Antell, S., Bilker, S., Ross, A., & Capute, A. J. (1990). Precursors of reading delay: neurodevelopmental milestones. *Pediatrics, 85*, 416-420.

Share, D. L., & Leikin, M. (2004). Language impairment at school entry and later reading disability: Connections at the lexical versus supralexical levels of reading. [empirical]. *Scientific Studies of Reading, 8*(1), 87-110.

Sharpe, T. (2008). How can teacher talk support learning? *Linguistics and Education, 19*(2), 132-148.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modificiations in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development, 38*(5), 1-38.

Shipley, K. G., & McAffe, J. G. (2004). *Assessment in Speech-Language Pathology: A Resource Manual* (3rd. ed.). New York: Delmar Learning.

Shiro, M. (2003). Genre and evaluation in narrative development. *Journal of Child Language, 30*, 165-195.

Silva, M., & Cain, K. (2010). *Narrative comprehension and production: the role of oral language and home literacy*. Paper presented at the Society for the Scientific Studies of Reading 2010 Conference.

Simkin, Z., & Conti-Ramsden, G. (2006). Evidence of reading difficulty in subgroups of children with specific language impairment. *Child Language Teaching and Therapy, 22*(3), 315-331.

Snow, C. E. (1991). The theoretical basis for relationships between language and literacy in development. *Journal of Research in Childhood Education, 6*(1), 5-10.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing Reading Difficulties in Young Children*. Washington, DC: National Academy Press.

Snow, C. E., Cancino, H., De Temple, J. M., & Schley, S. (1991). Giving formal definitions: a linguistic or metalinguistic skill? In E. Bialystok (Ed.), *Language processing in Bilingual children*. Cambridge: Cambridge University Press.

Snow, C. E., Tabors, P. O., Nicholson, P. A., & Kurkland, B. F. (1995). SHELL: Oral language and literacy skills in kindergarten and first-grade children. [empirical]. *Journal of Research in Childhood Education, 10*(1), 37-48.

Snowling, M. J., Bishop, D. V. M., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry, 41*(5), 587-600.

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: individual differences in the precursors of reading skill. *Child Development, 74*(2), 358-373.

Snowling, M. J., & Hulme, C. (2006). Language skills, learning to read and reading intervention. *London Review of Education, 4*(1), 63-76.

Snyder, L. S., & Downey, D. M. (1991). The language-reading relationship in normal and reading-disabled children. [empirical]. *Journal of Speech and Hearing Research, 34*, 129-140.

Steenland, K., Henley, J., Calle, E., & Thun, M. (2003). Individual- and area-level socioeconomic status variables as predictors of mortality in a cohort of 179,838 persons. *American Journal of Epidemiology, 159*(11), 1047-1056.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New Directions in Discourse Processing* (Vol. II, pp. 53-120). Norwood, NJ: Ablex Publishing Corporation.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: the role of language comprehension and working memory skills. *Reading and Writing: An Interdisciplinary Journal, 4*(245-256).

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: a follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*, 407-418.

Stride, C. (2008). Multi-level Modelling using SPSS. Unpublished Manual. Institute of Work Psychology at the University of Sheffield.

Stromqvist, S., Johansson, V., Kriz, S., Ragnarsdottir, H., Aisenman, R., & Ravid, D. (2002). Toward a cross-linguistic comparison of lexical quanta in speech and writing. *Written Language and Literacy, 5*(1), 45-68.

Sylva, K., Scott, S., Totsika, V., Ereky-Stevens, K., & Crook, C. (2008). Training parents to help their children read: a randomized control trial. *British Journal of Educational Psychology, 78*, 435-455.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston: Pearson / Allyn & Bacon.

Tabors, P. O., Beals, D. E., & Weizman, Z. O. (2001). You know what oxygen is? Learning new words at home. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning Literacy with Language: Young Children at Home and School* (pp. 93-110). Baltimore, MD: Brookes Publishing.

Tabors, P. O., Snow, C. E., & Dickinson, D. K. (2001). Homes and schools together: Supporting language and literacy development. In D. K. Dickinson & P. O. Tabors (Eds.), *Beginning Literacy with Language: Young Children at Home and School* (pp. 313-333). Baltimore, MD: Brookes Publishing.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension. *Scientific Studies of Reading, 10*(4), 381-398.

Templin, M. (1957). *Certain language skills in children*. Minneapolis, MN, USA: University of Minnesota Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., & Smith, E. (1997). Prevalence of Specific Language Impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245-1260.

Trumble, W. R. (Ed.) (2007) (6th ed.). Oxford: Oxford University Press.

Tunmer, W. E., Nesdale, A. R., & Wright, A. D. (1987). Syntactic awareness and reading acquisition. *British Journal of Developmental Psychology, 5*, 25-34.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601-638.

Vellutino, F. R., Scanlon, D. M., Small, S. G., & Tanzman, M. S. (1991). The linguistic bases of reading ability: Converting written to oral language. *Text, 11*(1), 99-133.

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonologica, and orthographic coding. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues.* (pp. 279-324). Baltimore: Paul H. Brookes.

Verhoeven, L., Aparici, M., Cahana-Amitay, D., van Hell, J., Kriz, S., & Viguie-Simon. (2002). Clause packaging in writing and speech: A cross-linguistic developmental analysis. *Written Language and Literacy, 5*(2), 135-162.

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*, 407-423.

Verhoven, L., Aparici, M., Cahana-Amitay, D., van Hell, J., Kriz, S., & Viguie-Simon. (2002). Clause packaging in writing and speech: A cross-linguistic developmental analysis. *Written Language and Literacy, 5*(2), 135-162.

Verhoven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*, 407-423.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Wagner, R. K. (2005). *Causal relations between the development of vocabulary and reading comprehension*. Paper presented at the Annual Meeting of the American Educational Research Association.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development, 65*, 606-621.

Wallach, G. P., & Liebergott, J. W. (1984). Who shall be called "learning disabled"?: Some new directions. In G. P. Wallach & K. G. Butler (Eds.), *Language Learning Disabilities in School-Age Children* (pp. 1-14). Baltimore, MD: Williams & Wilkins.

Webster, P. E. (1994). Linguistic factors in reading disability: a model for assessing children who are without over language impairment. *Child Language Teaching and Therapy, 10*(3), 259-281.

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. London: The Psychological Corporation.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of Theory of Mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Wells, G. (1985). *Language development in the preschool years*. Cambridge, MA: Cambridge University Press.

Westby, C. E. (1985). Learning to talk, talking to learn: Oral-literate language differences. In C. Simon (Ed.), *Communication skills and classroom success: Therapy methodologies for language-learning disabled students* (pp. 181-213). San Diego, CA: College Hill.

Wetherell, D., Botting, N., & Conti-Ramsden, G. (2007). Narrative skills in adolescents with a history of SLI in relation to non-verbal IQ scores. *Child Language Teaching and Therapy, 23*(1), 95-113.

Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., et al. (1988). Accelerating language development through picture book reading. *Developmental Psychology, 24*(4), 552-559.

Williams, C. (2006). Teacher judgement of the language skills of children in the early years of schooling. *Child Language Teaching and Therapy, 22*(2), 135-154.

Willows, D. M., & Ryan, E. B. (1986). The development of grammatical sensitivity and its relationship to early reading achievement. *Reading Research Quarterly, 21*(3), 253-266.

Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: how education, income and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health, 82*(6), 816-820.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Hingham, MA: Teaching Resources.

Appendix A: Picture sequence for the Sweets Story
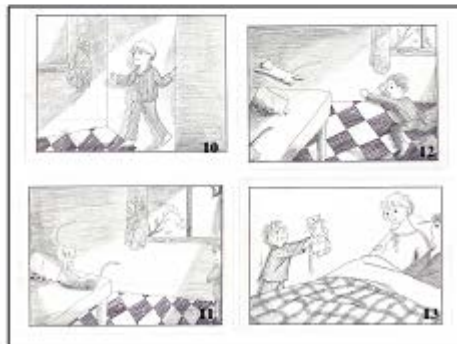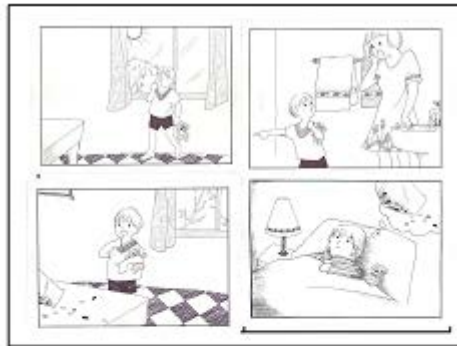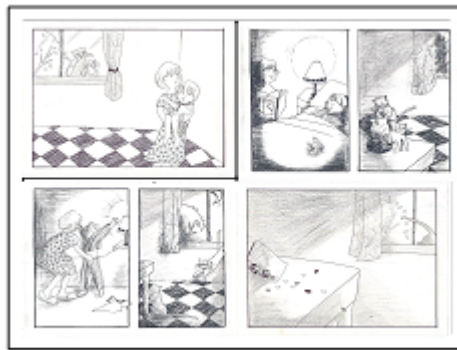
The Sweets Story

Appendix B: Prompting instructions for the Pilot Study

Procedure to elicit a narrative from the picture book 'The Sweets Story'.

- Do you like stories?

- I have a picture story here for you. There are no words in the pictures, so you can make up your own story, ok?

- I am going to record your story so I can listen to it when I go back home, is that all right?

- o Now, this is a story about a little boy and some sweets.

- o First, I want you to look at all the pictures carefully. Pay attention to each picture that you see. Then, you will tell me the story.

- o You can look at the pictures when telling the story, so don't worry about remembering the pictures.

- o You can start to look at the pictures.

- ❖ Did you like it?

- ❖ Now, to tell the story, I want you to imagine a friend from your class. (pause) Which friend have you chosen?

- ❖ You are going to tell him the story but we will not show him the pictures, so you will have to tell the story very carefully so he can understand, ok?

- ❖ Are you ready? You can start now.

Appendix C: Prompting instructions for Narrative Study

C.1 Procedure to elicit a narrative from the picture book 'The Sweets Story'

- Do you like stories? *(wait for response)* I like listening to children's stories, and I am going to ask you to tell me some stories.

- I have a picture book here for you. There are no words in the pictures, so you can make up your own story, ok?

- I am going to record your story so I can listen to it when I go back home, is that all right?

- Now, this is a story about **a little boy and some sweets**.

- First, I want you to look at all the pictures carefully. Pay attention to each picture that you see. Then, you will tell me the story.

- You can look at the pictures when telling the story, so don't worry about remembering the pictures.

- You can start to look at the pictures. *(Allow time for viewing the pictures)*

- Do you like it?

- Now, I want you to tell me a good story, like a story you would read in a book.

- I cannot see the pictures, so you will have to tell the story very carefully so that I can understand, ok?

- Are you ready? You can start now. *(Start recording)*

- That was a very good story. Thank you very much.

C.2 Procedure to elicit a narrative from the picture book 'Frog, Where Are You?'

- Would you like to tell another story?

- Here is another picture story, and now this one is about **a boy, a dog and a frog**.

- Again, I want you to first look at all the pictures carefully. Pay attention to each picture that you see. Then, you will tell me the story and I will record it also, if that is ok with you.

- You can start to look at the pictures. *(Allow time for viewing the pictures)*

- Do you like it?

- Now, I want you to tell me the story, and remember to make it a good one, like a story you would read in a book.

- I cannot see the pictures, so you will have to tell the story very carefully so that I can understand, ok?

- Are you ready? You can start now. *(Start recording)*

- That was also a very good story. Thank you.

Appendix D: Transcription conventions created for Narrative Study.

The T-Unit includes a main clause and any subordinated clauses, with the exception of compound sentences. This is an example illustrating how to define the T-Unit when segmenting children's narratives.

Embedded clauses
The following sentence:

> *"And the thing, which was creeping in, had started to look over the sweets"*

…would be transcribed as a single unit:

```
CHI: and the thing which was creeping in had
     started to look over the sweets.
```

Compound sentences
In the case of compound sentences (joined by a conjunction), the two sentences are treated as separate units, such as in this case:

> *"He had cuts and bruises, but he had something to show for it"*

…would be transcribed as two units:

```
CHI: he had cuts and bruises.
CHI: but he had something to show for it.
```

Two predicates, one subject
However T-Units also class as a single unit any utterance with two predicates where the second subject has been omitted. If the previous utterance had the second *he* omitted, then it would have been classified as a single unit, reflecting a more complex syntactic form:

```
CHI: he had cuts and bruises but had something to
     show for it
```

A list of additional transcription conventions to resolve particular challenges in the Silva corpus was created to ensure consistency. This is by no means an exhaustive description of all transcription issues, but it is focused of the main ones encountered during the transcription of narratives described in Chapter 4.

1. At the word level:
   a. Solve spelling queries by adopting the British standard spelling, found in the Shorter Oxford English Dictionary.

b. Follow ERRNI (Bishop, 2004) guidelines in solving compound word issues also by adopting the standard spelling from the Shorter Oxford English Dictionary.

c. Code onomatopoeic expressions uniformly, when possible, to minimise inflating lexical variations.

d. Code interjections to allow for further analysis.

e. Since the objective of using language sampling is obtaining a more ecologically valid measure of language, the idea is to *not* penalise the use of dialectic variations of British English, particularly in the West Midlands area, where examples abound. Change dialectic variations of English into its equivalent using standard English. For example, the word 'want' is often used to mean 'won't', which will be transcribed 'w(ill) n(o)t'. Transcribe overregularised verbs into the proper spelling in brackets to allow for a choice in the analysis between regular and true spelling. For example 'breaked' will be transcribed 'breaked [: broke]'. Notice the space between the colon and the first letter of the standard word.

f. Identify typing mistakes with freq:

```
Freq +d1 +r6 +u +k *.cha > output.frq
```

This will produce a simple list from which mistakes can be more easily identified.
From the list locate words that are misspelled due to typing errors, having used the American version of a word, extremely inconsistent forms of interjections or onomatopoeia, problems with compound words or contractions.

g. To make changes to a single file use Menu functions to Find & Replace. To make changes across sets of files use CHSTRING and a file called changes.cut. In the file changes.cut list the existing words and the replacement words.

```
chstring +c +f +1 *.cha
```

This will replace/overwrite the file. Since it is definitive, changes should be tried in a single sample first.

h. Eliminate fillers such as hm, hmm, umh, etc.

2. At the sentence level:
  a. Direct speech. Issues with direct speech involve speech running across several sentences with a list characteristic, which are all qualified by a direct-speech verb such as 'said' or 'yelled', etc. The question arises when identifying where to segment the T-Unit. For example, the sentence "The boy said to his dog, 'Leave the frog, lie down, be quiet

and stay awake'" can be segmented into a single T-Unit, allowing all the fragments qualified by the speech verb in the same T-Unit. Instead, it can also be segmented into five separate T-Units, such as 'The boy said to his dog', 'Leave the frog', 'Lie down', 'Be quiet', 'And stay awake'. Each one of these two options runs the risk of either inflating or deflating the MLU count. The original publication by Hunt (1965, 1970) does not address this issue, and considering Hunt's guidelines were created for segmenting written samples, this is not surprising. However, follow ERRNI (Bishop, 2004) guidelines offer a sensible compromise when instructing to add only the first portion of direct speech along with the qualifying verb, which in this example would produce the following four T-Units: 'The boy said to his dog leave the frog', 'Lie down', 'Be quiet', 'And stay awake'.

b. Fragments. Any group of words which do not conform to the subject-verb structure should be coded as fragments to allow for exclusion in syntactic analyses. Use the postcode '`[+ bch]`' at the end of the T-Unit (notice the space between the plus and the bch).

c. How to divide a compound-complex sentence. A compound-complex is a sentence with both coordinating and subordinating sentences. For Hunt (1965, 1970), the compound should be divided at the conjunction level. For example: 'The boy, who could not find the frog, ran outside, and his dog, with a frightened face, followed him' will be transcribed as two T-Units into 'The boy, who could not find the frog, ran outside' and the second as 'and his dog, with a frightened face, followed him'.

d. Omitting the subject. This is a problem with long lists of predicates since they can easily inflate syntactic analyses. A long list of activities, similar to the direct speech example used above can be segmented in a similar fashion, with only the first predicate attached to its subject, and the rest standing each on their own T-Unit as suggested in ERRNI (Bishop, 2004).

e. Include appositions in the same T-Unit. This is when the subject is mentioned twice in two forms for clarification purposes. For example, 'My friend, Sarah, went…' will be transcribed within the same T-Unit.

f. Include nouns of address and interjections in the same T-Unit. For example: 'Frank, wake up' would be a single T-Unit. So would be 'No, I did not break it'.

g. Be careful with the smallest T-Units. Imperative sentences such as 'Kick it' are small, but still constitute a full T-unit. However, 'All right then' is longer, but it is still a fragment.

Appendix E: Language and reading scatter-plots by SES and gender

E.1 Scatter-plots for each linguistic variable and reading comprehension differentiated by socioeconomic status
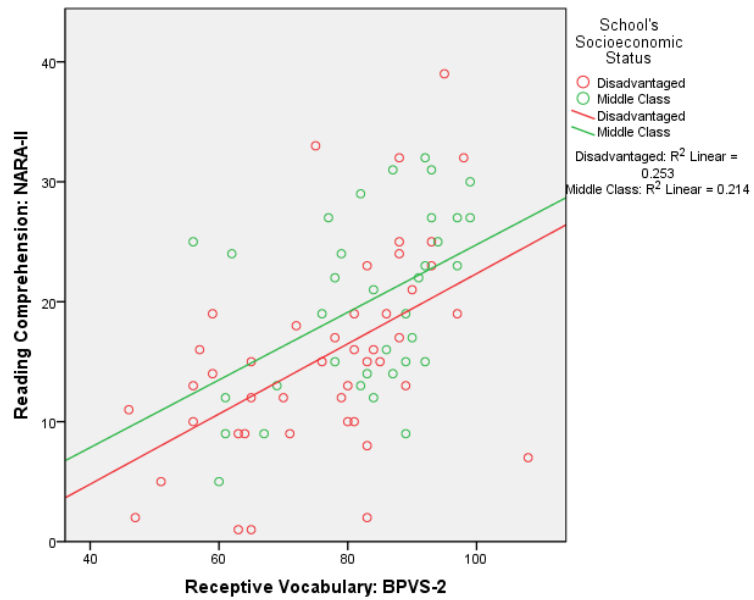

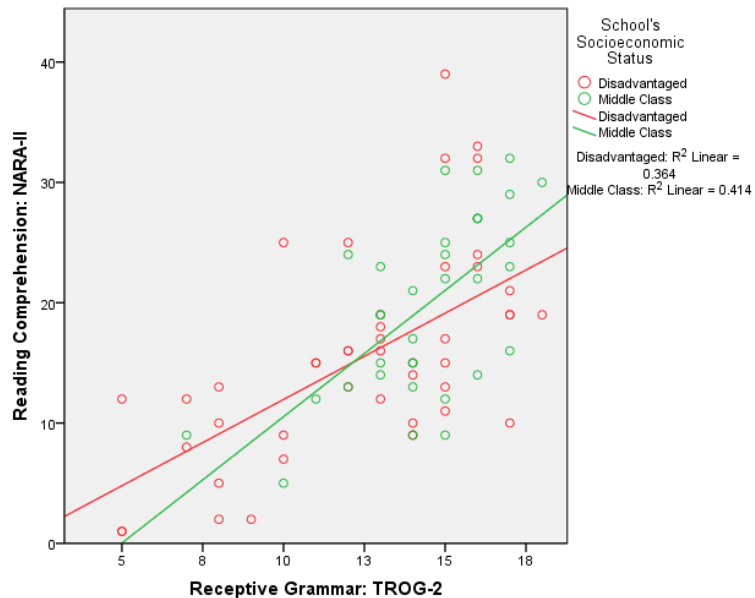Figure E1a Relationship between receptive vocabulary and reading comprehension by school-wise SES

Figure E1b Relationship between receptive grammar and reading comprehension by school-wise SES
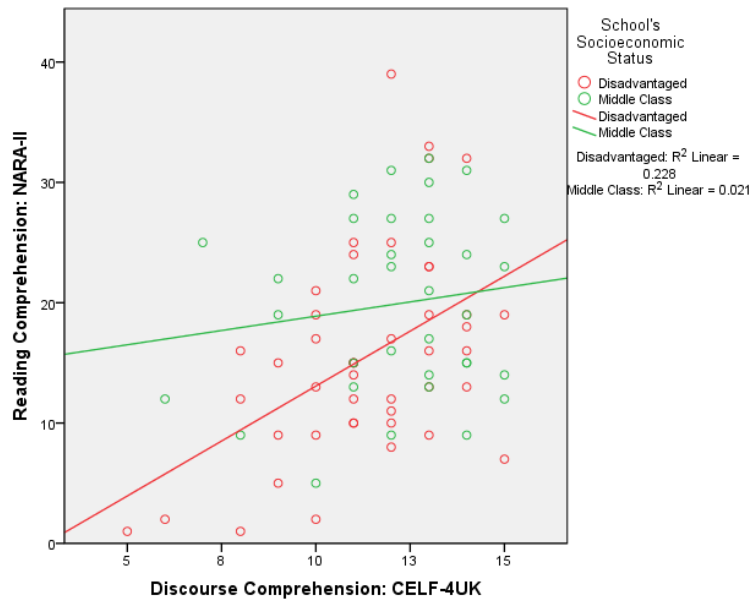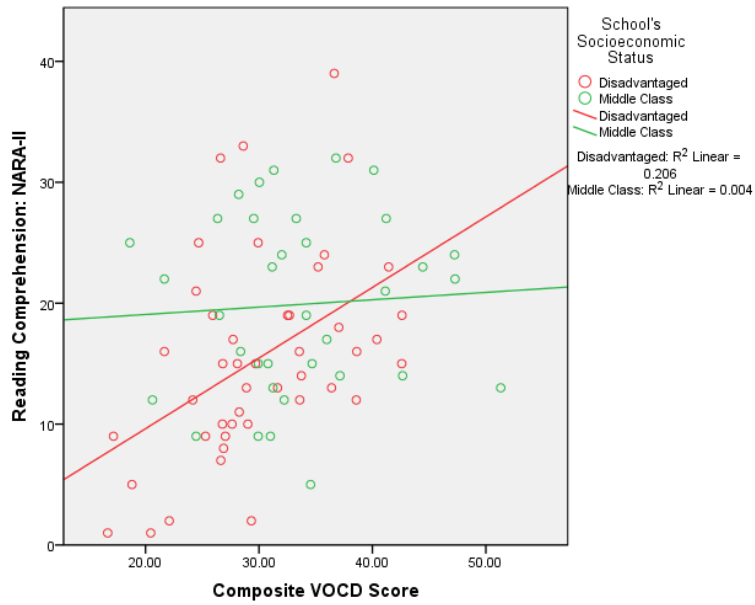
Figure E1c Relationship between discourse comprehension and reading comprehension by school-wise SES

Figure E1d Relationship between expressive vocabulary and reading comprehension by school-wise SES
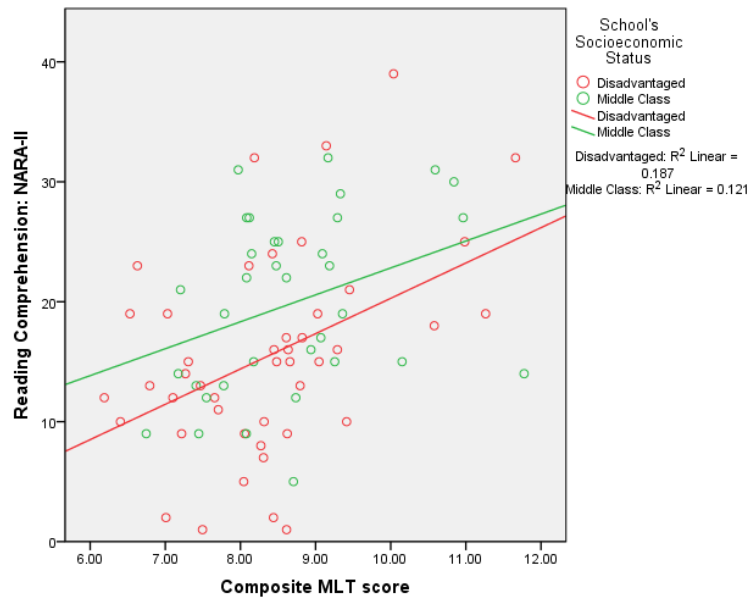
Figure E1e Relationship between expressive grammar and reading comprehension by school-wise SES

E.2 Scatter-plots for each linguistic variable and reading comprehension differentiated by gender
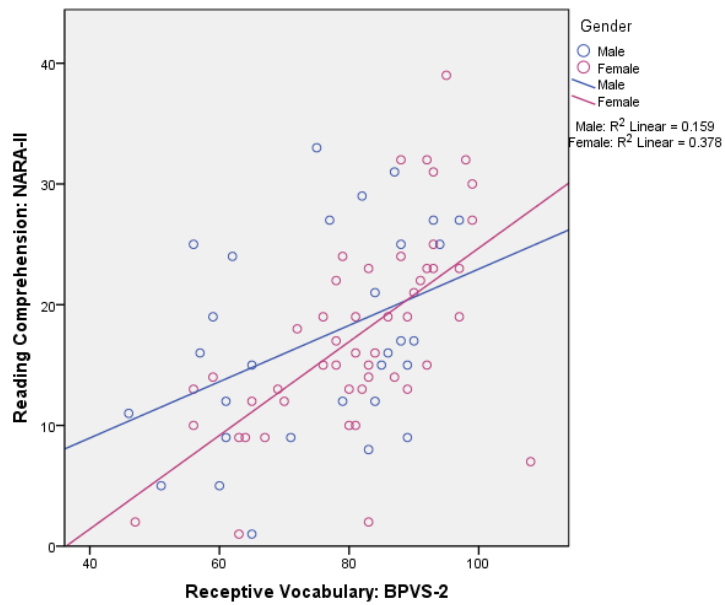

Figure E2a Relationship between receptive vocabulary and reading comprehension by gender
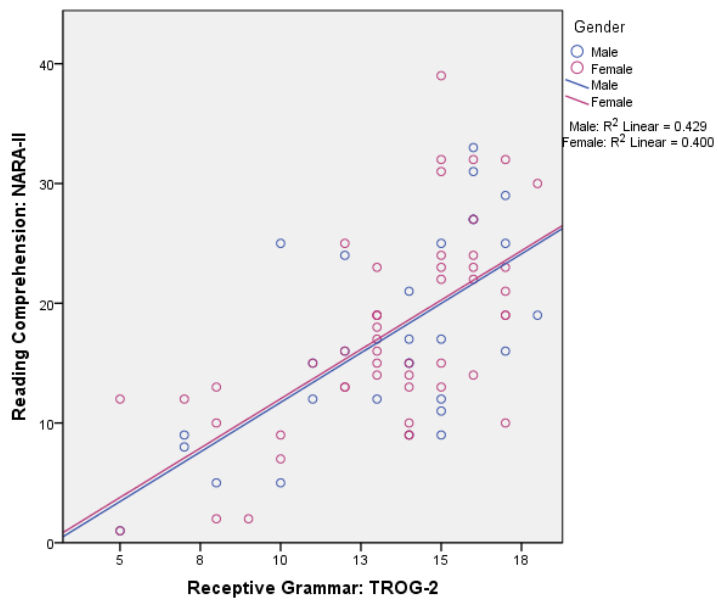

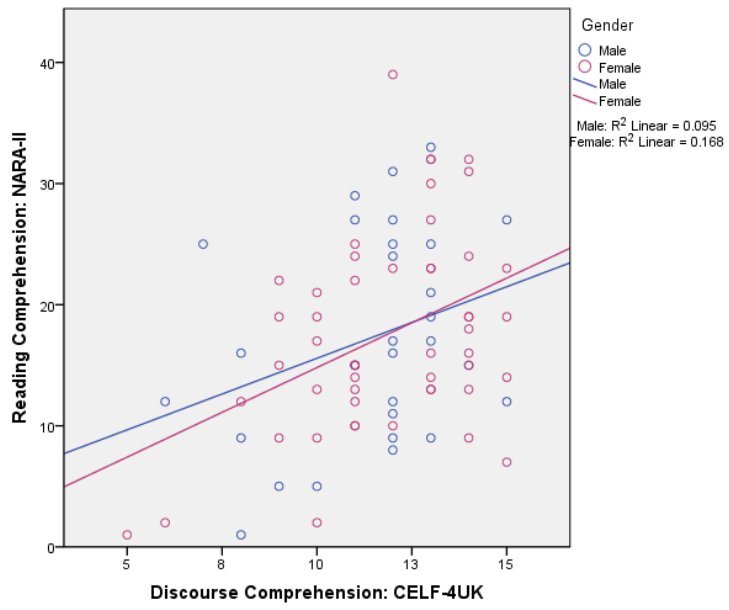Figure E2b Relationship between receptive grammar and reading comprehension by gender

Figure E2c Relationship between discourse comprehension and reading comprehension by gender
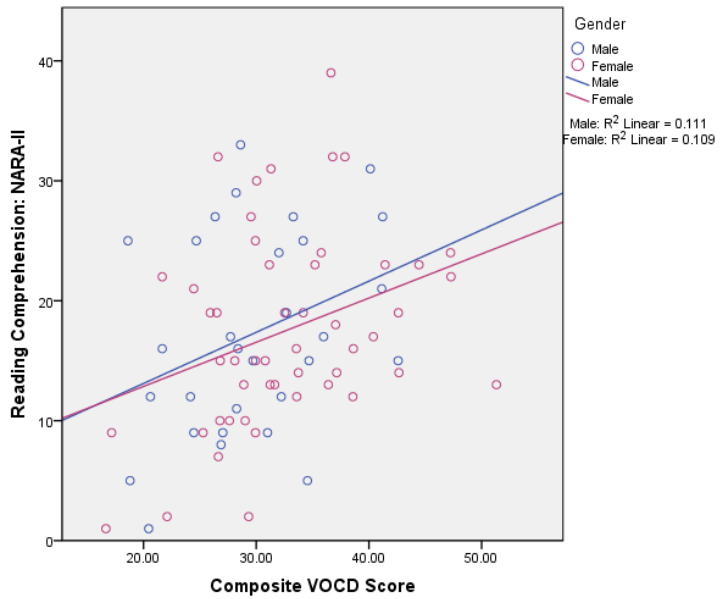

Figure E2d Relationship between expressive vocabulary and reading comprehension by gender
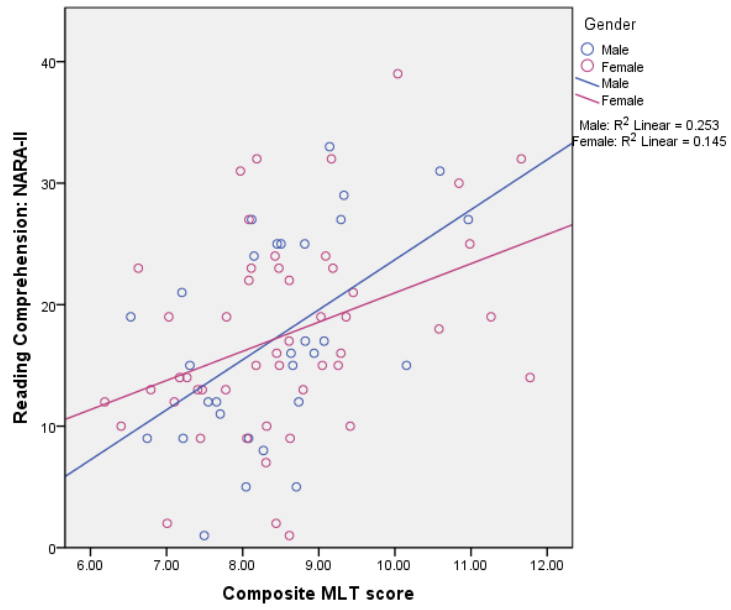
Figure E2e Relationship between expressive grammar and reading comprehension by gender

Appendix F: Examples of children's narratives in all studies

F.1 Examples of children's narratives in Chen-Wilson (2003) corpus (full sample n=60).

F.1a Narrative of the Sweets Story by a 3-year-old.

```
@Begin
@Participants:    CHI    AB      Child, INV    Josephine    Adult
@Birth of CHI:    21-JUL-1993
@Age of CHI:      3;7.12
@Date:      4-MAR-1997
@Sex of CHI:      female
@Situation: This is AB's narration of the Sweets story to her friend.
@Language:  English
@Filename:  s1-3AB.cha
@Tape location:   97E0301 Side A
@Transcriber:     364-405
@Stim:      picture 1
*CHI: one day there are [: is-s] some sweeties [: sweet-s] [=!
      pointing] .
*CHI: <a boy> # [//] <no> [//] a girl and a boy and the sweeties
      [: sweet-s] !
*INV: yeah .
@Stim:      picture 2
*CHI: cat .
*CHI: boy .
*INV: uhhuh .
@Stim:      picture 3
*CHI: and a boy and a girl <and a &ca> [//] and a cat . [+ bch]
*INV: yeah .
@Stim:      picture 4
*CHI: and a cat and a cat .
*CHI: and a boy .
*INV: uhhuh  .
@Stim:      picture 5
*CHI: and the sweeties [: sweet-s] and the sweeties [: sweet-s] and
      the sweeties [: sweet-s] and the sweeties [: sweet-s] ! [+ bch]
%exp: AB took a deep breath after the above utterance .
*INV: uhhuh .
@Stim:      picture 6
*CHI: and # a teddy bear and a:::: boy .
*INV: OK .
@Stim:      picture 7
*CHI: and a teddy bear and a boy .
*INV: yeah .
@Stim:      picture 8
*CHI: and a boy and a teddy bear and a girl . [+ bch]
*INV: uhhuh .
@Stim:      picture 9
*CHI: and a teddy bear <and> [/] and a boy .
*INV: OK .
@Stim:      picture 10
*INV: you missed a page here .
*CHI: a girl and a boy .
*INV: yeah ?
```

```
*INV: what about the boy ?
*CHI: 0*subject 0*be with # no shoes:: [: shoe-s] .
*INV: OK .
@Stim:     picture 11
*CHI: and a cat and a cat .
*INV: OK .
@Stim:     picture 12
*CHI: and a boy and a cat .
*INV: uhhuh .
@Stim:     picture 13
*CHI: and a boy and a cat .
*CHI: and a girl and a boy .
@END
```

_____

F.1b Narrative of the Sweets Story by a 7-year-old.

```
@Begin
@Participants:    CHI    AC    Child,   INV     Josephine        Adult
@Birth of CHI:    10-OCT-1989
@Age of CHI:      7;5.0
@Date:     10-MAR-1997
@Sex of CHI:      female
@Situation: This is AC's narration of the SWEETS story to her friend.
@Language:  English
@Filename:  S1-7AC.cha
@Tape location:   97E0701 Side B
@Transcriber:     045-063
@Stim:     picture 1
*CHI: &th the little boy-'s mom gived [: give-ed] [*] him some sweets
      [: sweet-s] .
%err: gived = gave ;
@Stim:     picture 2
*CHI: it was [: is-ed] bedtime .
@Stim:     picture 3
*CHI: the boy took [: take-ed] his clothes off to put his pyjamas on .
@Stim:     picture 4
*CHI: <he> [//] the little boy laid  [: lie-ed] in bed .
@Stim:     picture 5
*CHI: the sweets [: sweet-s] were [: is-s-ed] all tipped [: tip-ed]
      over.
@Stim:     picture 6
*CHI: the little boy got [: get-ed] up .
@Stim:     picture 7
*CHI: the little boy <saw> [/] saw [: see-ed] the sweets [: sweet-s] .
@Stim:     picture 8
*CHI: <he> [//] the little boy told [: tell-ed] # his mom .
@Stim:     picture 9
*CHI: the little boy went [: go-ed] back in bed .
@Stim:     picture 10
*CHI: he got [: get-ed] up again .
@Stim:     picture 11
*CHI: the cats [: cat-s] ate [: eat-ed] some sweets [: sweet-s] .
@Stim:     picture 12
*CHI: the little boy chased [: chase-ed] after the cats [: cat-s] .
@Stim:     picture 13
*CHI: the little boy got [: get-ed] one of the cats [: cat-s] .
@END
```

_____

F.2 Examples of children's narratives elicited for Pilot Study (full sample n=20).

F.2a Narrative of the Sweets Story by an 8-year-old.

```
@Begin
@Languages: en
@ID:  w1.08swee.0801=CHI
@Date:      17-JUN-2008
@Age of CHI:       8;1.
*CHI: one day # it was just after he had the dinner.
*CHI: <he had eaten> [//] Johnny had eaten all of his dinner.
*CHI: and his mother gave him four sweets.
*CHI: outside there were two cats.
*CHI: they were very hungry because they had n(o)t eaten all day.
*CHI: <they looked they looked> [/] they looked at the sweets and
      thought they looked nice.
*CHI: <after he (ha)d had the sweets> [//] after Johnny had had the
      sweets Mother took him to bed.
*CHI: he put on his pajamas and went to bed while downstairs two
      cats had jumped in from the open window <and so> [//] and tried
      to get up the table.
*CHI: Mother was reading Johnny <a little> [//] a book about a
      bedtime story upstairs.
*CHI: the cats had got into the jar and had strated to eat the
      sweets.
*CHI: Mother <was starting> [//] was coming down.
*CHI: and he could hear the footsteps.
*CHI: <so they> [//] so the cats <spill> [//] tried to get away
      quickly.
*CHI: they accidentally <f kno> [//] knocked over the sweet jar.
*CHI: Johnny yawned and got up.
*CHI: when he got up # he looked at the table.
*CHI: <he saw all the> [//] he saw the sweets and thought +"/.
*CHI: +" oh no that looks like something has been in here.
*CHI: <when he got up> [//] then he rushed back upstairs and
      told his mom exactly what happened.
*CHI: his mommy thought it was him.
*CHI: so he sent him to bed.
*CHI: he was thinking about the sweets.
*CHI: when his mommy said he could finally come down # he went into
      the kitchen and saw the cats eating sweets again.
*CHI: he chased after the cat.
*CHI: <and he> [//] and finally he caught him.
*CHI: he went upstairs with ripped pajamas and a scratch in his face
      to show his parents.
@End
```

F.2b Narrative of the Sweets Story by another 8-year-old.

```
@Begin
@Languages: en
@ID:  w1.17swee.0802=CHI
```

@Date: 30-JUN-2008
@Age of CHI: 8;2.
*CHI: once there (i)s a boy.
*CHI: and he (i)s getting some sweets of his mom.
*CHI: and the cats are looking in the window licking their lips.
*CHI: then the boy started to feel sick.
*CHI: <and> [//] so the mom carried him to his room and put his
      pajamas on.
*CHI: and the cats jumped in the window.
*CHI: the boy was in bed.
*CHI: and the cats were getting the sweets out of the jar.
*CHI: when the cats went out <all the sweets had been kno>
      [//] all the sweets had been knocked over.
*CHI: in the morning # the boy yawned with his teddy in his hands.
*CHI: he saw that the sweets had been knocked over.
*CHI: the boy and their mom were disappointed.
*CHI: when the boy was in bed he was angry.
*CHI: the boy woke up and found the cats eating the sweets.
*CHI: he chased after the cats.
*CHI: and the cats ran away.
*CHI: the boy showed the mom <that it was> [//] and their dad that
      it was the cat that did it.
@End

_____

F.3 Examples of children's narratives elicited for Narrative Study (full sample n=333).

F.3a Narrative of the Sweets Story by a 6-year-old.

```
@Begin
@Languages: en
@ID:   en|silva|CHI||female|primarya||Child||
@Date:        01-MAY-2009
@Age of CHI:        6;1.
@Birth of CHI:    21-MAR-2003
@Situation: Sweets story
*CHI: the boy had sweets.
*CHI: his mum took him to bed.
*CHI: the cats went on the table.
*CHI: the boy was in bed.
*CHI: the cats looked at the sweets.
*CHI: the sweets fell on the table.
*CHI: the boy woke up.
*CHI: he looked at the sweets.
*CHI: he told his mum the sweets fell.
*CHI: he went back to bed.
*CHI: he went to look at the sweets.
*CHI: <the cat> [//] the cats went back on the table.
*CHI: the cat ran away.
*CHI: the boy got the cat.
@End
```

_____

F.3b Narrative of the Frog Story by an 8-year-old.

```
@Begin
@Languages: en
@ID:   en|silva|CHI||female|primarya||Child||
@Date:        28-APR-2009
@Age of CHI:        8;5.
@Birth of CHI:    12-NOV-2000
@Situation: Frog story
*CHI: this is a story with a frog living with his dog and a pet frog.
*CHI: one night a boy <went into> [/] went into his bedroom <and fell>
      [/] and fell asleep.
*CHI: the frog and the dog climbed out.
*CHI: and when the boy woke up the next morning he realised that the
      frog was not there.
*CHI: he looked in his boots.
*CHI: the dog looked in a jar.
*CHI: but there was still no sign of the frog.
*CHI: they looked out the window.
*CHI: and the dog fell down with the jar on his head.
*CHI: <so the dog> [//] so the boy <picked him> [//] picked the dog
      up and looked in the bushes.
*CHI: <he looked around> [/] he looked around to the garden in the
      trees and the tree holes and the holes in the ground.
*CHI: the dog looked up the bee hive.
*CHI: the boy looked down the holes.
```

269

```
*CHI:	and the dog knocked down the bee hive down.
*CHI:	and the bees started following him.
*CHI:	the boy looked in the tree hole.
*CHI:	and a owl came out and pecked him with his nose.
*CHI:	and he found the ground.
*CHI:	<he looked> [/] he looked behind the rock and shouted out his
	name in the air.
*CHI:	then a deer came out and took him to a cliff and threw him and
	his dog in the water.
*CHI:	then <he swam to a> [//] he swam to concrete.
*CHI:	and <then he found> [//] he looked in the log.
*CHI:	and behind the log was the <frog> [//] frogs and his family.
*CHI:	and then the boy hold the little frog in his hand.
@End
```

F.3c Narrative of the Beach Story by a 10-year-old.

```
@Begin
@Languages: en
@ID:   en|silva|CHI||female|primarya||Child||
@Date:        30-APR-2009
@Age of CHI:       10;4.
@Birth of CHI:    06-DEC-1998
@Situation: Beach story
*CHI:	<her mum is> [//] the girl wants her mum to ring her friend to
	go to the beach with her.
*CHI:	she goes upstairs and packs all of her stuff.
*CHI:	she packs a beach volley ball her swimming costume.
*CHI:	and then she goes on her bike and rides to the beach.
*CHI:	she goes past her friend who (i)s fishing.
*CHI:	then she is on the beach.
*CHI:	she sees her friends.
*CHI:	so she parks her bike up.
*CHI:	and she sets all of her stuff out.
*CHI:	and she goes into the beach.
*CHI:	suddenly <she> [//] a crow came and took her watch.
*CHI:	she came back.
*CHI:	and she realised that it was gone.
*CHI:	and she checks all through her bag.
*CHI:	but it was not there.
*CHI:	<she> [/] she rided [: rode] to her friend's.
*CHI:	but she has n(o)t seen her watch.
*CHI:	then she walks around to find her watch.
*CHI:	but she can not see it anywhere.
*CHI:	then they follow the dog.
*CHI:	and the dog leaded them to a bench.
*CHI:	a crow was on there.
*CHI:	and the watch was just lying on there.
*CHI:	and then she got her watch back.
*CHI:	<and they drove> [//] and they went back home.
@End
```

F.3d Narrative of the Frog Story by a 12-year-old.

```
@Begin
@Languages: en
@ID:   en|silva|CHI||male|secondarya||Child||
@Date:      30-SEP-2009
@Age of CHI:       12;11.
@Birth of CHI:    13-OCT-1996
@Situation: Frog story
*CHI: Timmy <and> [/] and Tintin anxiously looked at the frog.
*CHI: Tintin <was sus> [//] was suspecting that the frog was up to
      something.
*CHI: Timmy ignored that <and sca> [//] and continued staring at the
      frog.
*CHI: his mother called up to him and told him to get into his
      pyjamas and ready for bed.
*CHI: it was midnight.
*CHI: and he was still up.
*CHI: and he had school tomorrow.
*CHI: so Timmy and Tintin climbed up onto the big bed and slept.
*CHI: but Tintin had knocked the lid off before he got to bed.
*CHI: and the frog managed to escape.
*CHI: as the light shone through in the morning to wake Timmy and
      Tintin up Tintin noticed that the frog was missing.
*CHI: <and so> [//] and then also did Timmy.
*CHI: they searched everywhere in the bedroom.
*CHI: they turned the whole place upside down.
*CHI: they checked in the slippers under the bed into clothes.
*CHI: and Tintin started <ss ss> [//] to run around in circles in
      case he could sense him.
*CHI: they looked outside the window.
*CHI: but Tintin was starting to flip.
*CHI: and because he was being so stupid he got his head stuck in the
      vase <that> [/] that he kept the frog in.
*CHI: Tintin had fell out the window.
*CHI: and Timmy was anxious now to see what would happen.
*CHI: he found <Tim> [//] Tintin running around like crazy still.
*CHI: <and> [//] but luckily for Tintin Timmy was wearing big boots
      and managed to carry him off the broken glass.
*CHI: they then left the house as he went off on a mystery to find to
      where the frog had escaped.
*CHI: on the way Tintin had <anguish ang> [//] angrily made <a bee>
       [//] a wasp nest very very very anxious.
*CHI: they started then to chase Tintin while Timmy continued
      everywhere looking.
*CHI: he looked inside the mole hole.
*CHI: but he was then actually scratched in the nose because he had
      been too nosy.
*CHI: now that Tintin had knocked the wasp's nest onto the floor it
      was buzzing like crazy.
*CHI: and he was trying to anxiously run off.
*CHI: but Timmy was still looking inside trees and everywhere around
      him <until Timmy fell off> [//] until it all started.
*CHI: Timmy fell off.
*CHI: the owl had got angry for disturbing him while sleeping.
*CHI: the dog Tintin was being chased by bees and wasps and every
      other kind of creature that can fly.
*CHI: as they tried to find shelter a hawk tried to grab Tintin and
      Timmy.
*CHI: but Timmy managed to escape.
```

```
*CHI:  while looking he did n(o)t realise that <an antler an> [//]
       antlers <of a de> [//] of a deer were disguised as tree
       branches.
*CHI:  he leant on them but did not realise this.
*CHI:  he was then stuck on the deer's head.
*CHI:  and the deer had chased and run off into the far corner of a
       cliff which ended in a smoky swamp.
*CHI:  the corner of the cliff started to crumble.
*CHI:  and they fell into the muddy pond.
*CHI:  as Tintin landed on Timmy it was only Timmy that got wet.
*CHI:  but he could hear strange croak@o noises.
*CHI:  he told Tintin to shush <and he ang> [//] as he excitedly swam
       around in circles.
*CHI:  Timmy then also heard a croak@o but <a t a ta> [//] a smaller
       and a tinier one than last time.
*CHI:  he then looked over <the frog the> [//] the log to discover
       that the frog was mating.
*CHI:  and it already had his children.
*CHI:  he waved good bye to the frogs and took away one of the
       smallest of the frogs and named it after the frog Froggie.
```

F.4 Examples of children's narratives elicited for Main Reading Study (full sample n=240).

F.4a Narrative of the Sweets Story by an 8-year-old.

```
@Begin
@Languages: en
@ID:  en|silva|CHI||female|primaryb||Child||
@Date:      10-JUN-2009
@Age of CHI:      8;4.
@Birth of CHI:   25-JAN-2001
@Situation: Sweets story
*CHI: once upon a time there lived a little boy with his mother.
*CHI: his mother got some pebbles.
*CHI: he would like four pebbles.
*CHI: so his mother gave him some pebbles.
*CHI: the next day he was n(o)t feeling very well.
*CHI: his mummy put his (py)jamas on.
*CHI: and he went to bed.
*CHI: there were two naughty cats.
*CHI: and <they decided to> [/] they decided to scratch all of the
      floor boards.
*CHI: his mummy wanted to read him a story.
*CHI: but he did n(o)t like the story.
*CHI: he was too tired.
*CHI: and it was too boring.
*CHI: and all he could think about was these stones.
*CHI: and the cats were trying to steal them.
*CHI: the cats have left all the stones and took them out.
*CHI: and they quickly dashed out of the window.
*CHI: the next morning he got up.
*CHI: and he saw the glass on the table.
*CHI: and all the pebbles fell out.
*CHI: what (ha)s happened he thought to himself.
*CHI: Mum Mum where (i)s my pebbles?
*CHI: they (ha)ve gone missing.
*CHI: last night I fell asleep.
*CHI: and I woke up in the middle of the night thinking that all my
      pebbles have gone.
*CHI: it came true.
*CHI: www.
*CHI: so he came into the bathroom to wash his feet <and he> [//]
      (be)cause it was nearly bed time.
*CHI: then the cats came in and stole more.
*CHI: and then they dashed out of the window again.
*CHI: <as> [//] just as they were jumping out of the window he came
      in.
*CHI: and he said you cats have been stealing my pebbles.
*CHI: <qui> [//] and he quickly dashed to try and get them.
*CHI: <so he> [//] and he got one.
*CHI: he picked it up and took it into his mum's bedroom.
*CHI: why have you done that for?
*CHI: he had a scratch on his cheek.
*CHI: what on earth (i)s the matter?
*CHI: it (i)s the cat.
*CHI: it (ha)s been trying to steal my pebbles.
@End
```

_____

F.4b Narrative of the Frog Story by an 8-year-old.

```
@Begin
@Languages: en
@ID:  en|silva|CHI||male|primaryb||Child||
@Date:      05-JUN-2009
@Age of CHI:      8;1.
@Birth of CHI:   13-APR-2001
@Situation: Frog story
*CHI: a boy and a dog and a frog were in the bedroom.
*CHI: the boy fell asleep and the dog while the frog was hopping out
      of    the jar.
*CHI: the boy was really really scared because he thought the dog ate
      the frog.
*CHI: the boy checked under a shoe.
*CHI: and the dog checked <in the> [/] in the tin.
*CHI: the dog's head was stuck in the tin.
*CHI: and they were shouting.
*CHI: the dog fell out the window.
*CHI: and he broke the jar.
*CHI: the boy was angry at the dog.
*CHI: but the dog still licked him.
*CHI: they shouted again.
*CHI: but it still did n(o)t work.
*CHI: the boy shouted down a rabbit hole.
*CHI: but still no. [+ bch]
*CHI: and the dog was shouting at bees.
*CHI: <a ham> [//] a rabbit came up and banged the boy on the nose.
*CHI: and then all the bees were chasing after the dog <and> [//]
      while the boy was looking in a tree.
*CHI: the boy fell over.
*CHI: and the owl flied [: flew] away <off the tree from the> [//]
      from the tree.
*CHI: and the dog was being chased by bees.
*CHI: the owl was following the boy.
*CHI: but the boy climbed on a rock.
*CHI: and the owl stayed in the tree.
*CHI: there was some branches <on the> [//] behind the rock.
*CHI: <and and> [//] and it was a deer.
*CHI: and the boy fell onto the deer.
*CHI: and then the deer was running.
*CHI: <the deer> [//] the boy and the dog fell off the deer and
      landed in a pool head first for the dog and the boy.
*CHI: but the boy was all right.
*CHI: and the dog landed on the head of the boy.
*CHI: the boy said shush to the doggy because he was barking really
      loud.
*CHI: and the dog and the boy jumped over <the> [/] the log and saw a
      mum and a <dad> [/] dad frog and some babies.
*CHI: they took the frog back home and lived happily ever after.
@End
```

F.4c Narrative of the Beach Story by an 8-year-old.

```
@Begin
@Languages: en
```

```
@ID:    en|silva|CHI||male|primaryb||Child||
@Date:        05-JUN-2009
@Age of CHI:        8;1.
@Birth of CHI:    13-APR-2001
@Situation: Beach story
*CHI: <a boy and a a no> [//] a girl and a dad were in the lounge.
*CHI: the girl was going to go to the beach.
*CHI: so she packed up her stuff.
*CHI: she went on her bike.
*CHI: and her dad said good bye.
*CHI: <they saw> [//] the girl saw a fisherman on the day with a dog.
*CHI: <he at> [//] at the beach she met a boy.
*CHI: and the boy said hello.
*CHI: she went in the sea.
*CHI: and <while> [/] while she was in the sea <the> [/] a bird came
      down and got her necklace.
*CHI: the boy kicked the ball while the bird <was flin> [//] was
      flying away.
*CHI: when they went back to put the ball back she noticed that her
      necklace was gone.
*CHI: they looked in a bag but still could n(o)t find it.
*CHI: they thought that the fisherman would help to find it and his
      dog <when> [/] while the bird was flying away still.
*CHI: they went to a park.
*CHI: but they could n(o)t find it.
*CHI: but they ran up to some flowers.
*CHI: and they saw the bird on the bench and got the necklace.
*CHI: so the girl and the boy went back to the beach and lived
      happily ever after.
@End
```