

Running Head: Narrative assessment for Cantonese-speaking children

Narrative assessment for Cantonese-speaking children

Carol Kit-Sum To¹, Stephanie F. Stokes², Hin-Tat Cheung³ and Benjamin T'sou⁴

¹ University of Hong Kong

² Curtin University of Technology, Perth

³ National Taiwan University

⁴ City University of Hong Kong

Address for Correspondence: Carol To, Division of Speech & Hearing Sciences, University of Hong Kong, 5/F Prince Philip Dental Hospital, 34 Hospital Road, Sai Ying Pun, Hong Kong

Email: tokitsum@hku.hk

Abstract

Background. This study examined the narrative skills of Cantonese-speaking school-age children to fill a need for a normative language test for school-age children.

Aims. To provide a benchmark of the narrative skills of Cantonese-speaking children; to identify which of the microstructure components was the best predictor of age; and to determine the diagnostic accuracy of the test components.

Methods & Procedures. Data were collected from 1120 Cantonese-speaking children aged between 4;10 and 12;01, using a story-retell of a 24-frame picture series. Four narrative components (syntactic complexity, semantic score, referencing and connective use) were measured.

Outcomes & Results. Each measure reflected significant age-related differences in narrative ability. Regression analyses revealed that vocabulary and syntactic complexity were the best predictors of Grade. All measures showed high sensitivity (86% - 94%) but relatively low specificity (60% to 90%), and modest LR^+ (2.15 to 9.42) and LR^- (.07 to .34) values.

Conclusion & Implications. Narrative assessment can be standardized to be a reliable and valid instrument to assist in the identification of children with language impairment. Syntactic complexity is not only a strong predictor of Grade but was also particularly vulnerable in Cantonese-speaking children with SLI. Further diagnostic research using narrative analysis is warranted.

A large body of work on narrative development has emerged from the fields of speech language pathology, developmental psychology, sociology, and linguistics over the past two decades. This may be because narrative has always been noted as reflecting not only children's linguistic and pragmatic knowledge, but also their cognitive development and social awareness. From the clinical perspective then, there are many reasons for the inclusion of a narrative test in a comprehensive language test battery for school-age children (Botting, 2002; Miller, 2004). For example, narratives have ecological validity in that they are frequently encountered by children in their daily environment.

Language develops rapidly in the preschool years. As children grow older, their task shifts from learning language to using language for learning. Language becomes a medium of acquiring knowledge (Westby, 1998). Narrative plays an important role in the classroom discourse between children and teachers, among peers at school and between children and books. For example, children may recount their personal experience and report problematic events to their teachers, as well as share their experience with peers. In the later school years, children's performances at school are often evaluated by their ability to summarize or paraphrase what they have learned in a lesson (Crais & Lorch, 1994). Teachers also estimate how much knowledge children have grasped in the subjects of science and general studies by asking them for personal experience or expression of opinion. Children who use literate styles of communication in oral language match teacher expectations more often than those who do not (Wallach & Bulter, 1994). In other words, narrative development in school-age children is an early step to, and develops hand-in-hand with, academic and literacy skills.

Besides the importance of narrative at school, a narrative task is a versatile and rigorous test of language. Language skills at single word, sentence and text levels are deployed and integrated during the production of a narrative (Berman, 2004). When translating a story into

words, children have to execute several tasks simultaneously or consecutively: organizing the story information based on the narrative schema, retrieving appropriate vocabulary to represent the intended ideas, deploying various syntactic structures to package the story and selecting appropriate referential expressions and connectives to tighten the story as a coherent whole. Therefore, assessment of narrative is rewarding in terms of the amount and diversity of information elicited.

Narratives of children with language impairment

More practically, narratives are reported to be strong predictors of later language outcomes (e.g. Bishop & Edmundson, 1987; Paul & Smith, 1993). Children showing weak narrative skills in the preschool years often exhibited persistent language problems during the school years. It is not difficult to understand that children with better lexical and syntactic resources are more ready to acquire narrative discourse. The narrative schema established then provides children with frames to learn new lexical and sentence forms as well as new text types from classroom discourse during the school years. These two dimensions of narrative schema and language forms act synergistically during the school years. This explains why early narrative skills are good predictors of later syntactic as well as lexical skills (Johnston, 2006), even in children with specific language impairment. For example, Botting, Faragher, Smikin, Knox and Conti-Ramsden (2001) followed a group of children with specific language impairment (SLI) from seven to 11 years of age. A battery of standardized language assessments including tests of lexical comprehension and word association, a narrative test, an articulation test and a measure of nonverbal intelligence was administered to the children at the first assessment. Regression analysis showed that narrative ability was the best predictor of subsequent performances in vocabulary, syntactic comprehension and word association five years later. Children with poor narrative skills in the early school years often have poor language outcomes in the later years, while those with

better narrative skills may go through transient language problems. In this way, children's narrative performances may help speech-language pathologists prioritize candidates for intervention when the demand for speech and language therapy outstrips provision of services (Johnston, 2006). Therefore, when devising a language assessment for school-age children, a narrative assessment is an inevitable component.

Despite the rich source of information obtained from narratives, studies of narratives in school-age children in languages other than Indo-European languages are relatively scarce. Recent research has produced a standardized test of narrative abilities in USA English-speaking school-age children (Gillam & Pearson, 2004). This test can be used as a benchmark of language abilities and a strong supplement to the diagnostic process for the identification of language impairment in USA English-speaking school-age children. Because of the significant impact of language and culture in narrative production (see further below), comparable tools in other languages would be of great benefit to non English-speaking communities, not least those of great size, such as Chinese populations. Such tools would also facilitate studies of bilingual language development and disorders, and second language learning in the school-age years. The current research sought to develop a clinical tool that assessed narrative ability, including norms for Cantonese-speaking children, which could be a useful reference to speech-language pathologist working with Cantonese-speaking children. This instrument should be developmentally robust, showing good differentiation of children by school grade or age on a range of story components. In developing a narrative test for the detection of language impairment in a non-English-speaking population, several parameters must be considered, including approaches to narrative analysis, indicators of language impairment in the target language, and cultural/linguistic differences in narratives. We address these in turn.

Approaches to narrative analysis

The preschool years are a period of rapid growth in vocabulary size and syntactic complexity. In the school years, when basic vocabulary and sentence-level grammar are generally established, narrative skills undergo a period of rapid growth (Miller, 2004; Gillam & Pearson, 2004). In order to capture the milestones of typical narrative development and differences between typical and atypical narrative development, researchers have developed a wide range of approaches and measures.

Macrostructure. Under the cognitive approach, researchers have usually described narratives with reference to their macrostructure (e.g. Johnston, 1982; Stein & Glenn, 1979; Thorndyke, 1977). These scholars investigated the overall organization of narratives in terms of the causal and temporal relationships within fictional stories. The most widely used approach for analyzing the macrostructure of a narrative is the story grammar framework. Many acquisition studies in various languages have shown that as children mature, the number of complete episodes which consist of initiating events, attempts and consequences increases with age (e.g. Berman & Slobin, 1994). Berman and Slobin reported that at about three years of age, most children produce very primitive stories in which clear story grammar components were not yet observed. These children generally failed to conceive of the story as a coherent text and simply treated it as a description task. At five years of age, children were generally able to construct a simple episode including an initiating event and an attempt, but less often with an outcome. The results of many recent studies of narration indicate that most children create stories with initiating events, attempts, and consequences by 6 years of age.

Children with language problems tend to produce stories with fewer total story grammar elements than their age-matched peers, particularly those propositions related to plans, action and reactions, which involve higher cognitive functioning (e.g. Johnston, 1982; Merritt & Liles, 1987). While it is clear that narrative macrostructure poses problems for children with

language impairment, Liles, Duffy, Merritt and Purcell (1995) reported that it was microstructure features that best discriminated between children with SLI and typically developing children.

Microstructure. Microstructure concerns the local structures of words, clauses, and sentences that contribute to the cohesion of a text. Halliday and Hasan's (1976) seminal work served as an impetus for many subsequent studies on the development of textual cohesion. Cohesion is formally defined as "a semantic relation between an element in the text and some other element that is crucial to the interpretation of it" (Halliday & Hasan, 1976, p.8). This relation is marked by linguistic devices, called cohesive ties, which include conjunction, reference, ellipsis, substitution and lexical ties. Following Halliday and Hasan's model, Liles (1985) investigated the usage of cohesive ties in English-speaking typically developing (TD) children and children with language impairment (LI) with and without comprehension problems. Liles (1985) found that the two LI groups used significantly less adequate cohesion than the TD children. Therefore a measure of cohesion in terms of connective use was adopted in this study.

Indices of syntactic complexity have also been widely used to capture developmental maturity in narratives produced by school-age children (e.g. Bishop, 2004; Gillam & Johnston, 1992; Scott, 1988; Scott & Windsor, 2000). However, findings for children with and without LI are less clear and vary depending on which aspect of microanalysis is investigated. Some studies reported insignificant differences between children with LI and TD children in clausal connectivity when compared with other measures (Gillam & Johnston, 1992; Roth & Spekman, 1989) whereas some showed that children with LI performed significantly worse than their TD peers on this aspect (Bishop, 2004). Given the significant weight of a measure of syntactic complexity identified by Justice, Bowles, Kaderavek, Ukrainetz, Eisenberg and Gillam (2006), we included a composite measure of syntactic complexity in the present study.

One aspect of microstructure not explored by Justice et al is that of referencing. Referencing has been described as “a potentially interesting aspect of narrative production” (Norbury & Bishop, 2003). This is because not only does referencing reflect a child’s ability to conjoin sentences into a text, but it also indicates the child’s ability to assess a listener’s knowledge needs during different points of the story. Karmiloff-Smith (1985) suggested a three-stage model to describe overall narrative development. The first stage is characterized by the deictic use of pronouns. No recognizable overall narrative organization could be observed in this stage. In the second stage, at the ages of six and seven, referencing skill was typified by the use of the so-called ‘thematic-subject strategy’. Children reserved the subject slot to denote a character which was the highest in topicality (i.e. the main character in a story) and their story showed more recognizable overall organization. In the final stage, at about eight and nine years of age, referencing becomes flexible while overall narrative structure became more well-organized and detailed. A similar gradual developmental trend in the types of referencing strategies used by story tellers was observed by Wigglesworth (1997). A flexible anaphoric strategy was predominantly observed in adult subjects who could use pronouns to maintain just-referred-to referents and an explicit NP to switch references. A thematic subject strategy as proposed by Karmiloff-Smith (1981, 1985) was observed in older children at the ages of six and seven whereas children at about four years of age tended to overuse pronominal forms without establishing a clear antecedent in the previous discourse.

In a study on the referential ability of children with SLI van der Lely (1997) found that a relatively homogenous group of children with SLI, so called “grammatical SLI” used a mature strategy of anaphoric reference that was not different from an older language matched group, suggesting relatively intact pragmatic development in these “grammatical SLI” children. Norbury and Bishop (2003) compared narrative referencing skills in four groups of children, those with SLI, autism spectrum disorder (ASD), pragmatic language impairment

(PLI) and age-matched controls. While there were no significant differences among the clinical groups, all three groups used significantly more ambiguous referencing than their age-matched controls. Similar problems with referencing were reported by Wong (2001) who compared the use of referential expressions by a group of typical Cantonese-speaking children with those of a group of children with SLI from a functional perspective. Judgments about referential clarity were made by a panel of listeners. Listeners found it more difficult to resolve the referential expressions produced by the children with SLI than their TD counterparts. The discrepancy between van der Lely's (1997) findings and those of Norbury and Bishop (2003) and Wong (2001) may be explained by the composition of the participant groups included in these studies. The children with SLI in van der Lely's study showed specific grammatical deficit while the children with SLI in the latter two studies were generally impaired on a range of language measures. Given the potential value of referencing in the identification of SLI, we included a measure of referencing in this study.

Another aspect of interest in measuring narrative skill is the measure of the amount of relevant vocabulary provided in a narrative. The idea of measuring the amount of information successfully conveyed in discourse has been widely adopted in the aphasiology literature to evaluate a patient's functional communication ability (Nicholas & Brookshire, 1993). This measure, named the Correct Information Unit, was defined as "words that are intelligible in context, accurate in relation to the picture(s) or topic, and relevant to and informative about the context of the picture(s) or the topic" (p. 348). It has proved reliable across different raters and stable over time (e.g. Brookshire & Nicholas, 1994; Nicholas & Brookshire, 1993). In the developmental literature, similar measures have been adopted and used widely, for example, the information score of the Bus Story (Renfrew, 1969) wherein the score refers to the number of propositions or main concepts expressed in the story. Although the original Bus Story test did not provide strong evidence of the reliability and validity of the information score, its

psychometric properties have been tested in subsequent studies. Paul and Smith (1993) compared the information scores of three groups of children, those with a history of a SLI, those with a current SLI, and those with normal language (NL). Their results suggested that the information score could distinguish the NL group from the other two groups. Consistent findings were also reported in recent studies (e.g. Bishop, 2004; Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001, Pankratz, Plante, Vance and Insalaco, 2007; Reed, Patchell, Coggin, & Hand, 2007). On the other hand, Norbury and Bishop (2003) examined children's ability to convey relevant story information in terms of the number of relevant propositions and found no significant differences among three clinical groups and their age-matched controls. Likewise Botting (2002) reported information scores that were within the normal range for children with PLI and children with SLI in the later school years. Given this controversy, we aimed to investigate the usefulness of a measure of vocabulary use within narratives in distinguishing between Cantonese-speaking children with SLI and their TD peers.

In summary, research indicates that measures of cohesion, syntactic complexity, referencing and vocabulary should be included in a narrative diagnostic instrument.

SLI in Cantonese

The nature of SLI in Cantonese-speaking children has received much attention over the last two decades with researcher exploring a range of language structures that have been found to be impaired in English-speaking children with SLI, including grammatical morphemes (Stokes & Fletcher, 2003; Wong, Stokes, & Fletcher, 2003; Fletcher, Leonard, Wong, & Stokes, 2005), vocabulary deployment (Stokes & Fletcher, 2000; Klee, Stokes, Wong, Fletcher, & Gavin, 2004), utterance length (Klee et al., 2004; Wong et al., 2004), narrative referencing (Wong, 2001), syntactic structures (Leonard, Deevy, Wong, Stokes, & Fletcher, 2007; Wong, Leonard, Fletcher, & Stokes, 2004; Leonard, Wong, Deevy, Stokes, &

Fletcher, 2006) and phrase markers (Stokes & So, 1997). Note that the narrative referencing abilities of children with CSLI (Wong, 2001) are discussed above.

Since grammatical inflections were found to be a hallmark of SLI in English-speaking children (Leonard, 1998), researchers also examined if the same holds in Cantonese-speaking children with SLI (CSLI). Cantonese does not possess syntactically obligatory tense markers as occurs in English, but it does possess aspect marking that is deployed on some pragmatic and syntactic grounds. Stokes and Fletcher (2000, 2003) and Wong et al., (2003) found that CSLI used significantly fewer aspect markers, and attached aspect markers to fewer verb types in naturalistic speech than their TD age-matched peers. An experimental study by Fletcher, et al., (2004) reported consistent findings of poor aspect markers use in CSLI.

Question and passive sentence forms, and modal marking of main verbs have also been studied for CSLI. Wong et al., (2004) studied question formation by children with CSLI and reported difficulties with *who*-object questions but not *who*- subject questions relative to their AM and LM peers. Leonard et al., (2006) reported that children with CSLI were less able at passive sentence construction than their AM peers but were as able as their LM peers. In addition, children with CSLI had a mean utterance length similar to their LM peers, but a significantly shorter length in comparison with TD peers (Klee et al., 2004).

In terms of lexical abilities, children with CSLI differed significantly from their TD, but not LM, peers on overall lexical diversity (Klee, et al., 2004), but Stokes and Fletcher (2000) also reported reduced diversity of noun forms in comparison with their LM peers. Finally, a study of noun phrase structures (use of classifiers) reported that children with CSLI did not differ from their AM peers on the number of classifiers used, but did differ in the complexity of classifiers used (i.e., a result similar to aspect marker use). Overall then, in a language where some syntactic markers are not obligatory, there are qualitative rather than quantitative differences between children with CSLI and their AM or LM peers, but on the whole, children

with CSLI show vulnerabilities that are similar to their English-speaking counterparts. For a summary of findings on CSLI, see Fletcher, Leonard, Stokes, and Wong (2008) and Fletcher, Stokes and Wong (2005).

One further report is of note. Wong, Au, & Stokes (2004) reported age-sensitive differences for cohesive devices in a story re-telling task with Cantonese-speaking children. The section on cohesion above indicates that this is a viable candidate variable for distinguishing between TD children and children with CSLI. Given this body of literature on CSLI, a narrative analysis should include components of syntactic, lexical, and referencing abilities. In addition, given the age sensitivity of cohesive devices in Cantonese, a category of cohesion should also be included.

Language and Cultural Specific Features in Narratives

Communication style and use of rhetorical devices in narratives are susceptible to language and cultural differences (Westby, 1994). During comprehension of stories, people make use of schemas which they develop with their exposure to daily routines and stories of one's culture. A schema can be considered an organized cognitive structure; that is, an organized mental representation of knowledge about some entity. It can be an object, a scene, or an entire event (Mandler, 1978). These schemas serve as structures of expectation (Chafe, 1990) guiding text comprehension and production. A mismatch between text schema and personal schema results in difficulty in understanding the sense of the text and/or text misinterpretation (Westby, 1994). In a story retelling experiment, Pritchard (1990) found that both American and Palauan participants recalled significantly more idea units and produced more elaborations and fewer distortions for culturally familiar, relative to culturally unfamiliar, texts. Cultural and language differences are also observed in in telling the stories and these differences are mainly at grammar level. A cross-linguistic study by Hickmann and Hendriks (1999) compared the referential skills in English, Mandarin Chinese, German and

French-speaking children in a story telling task with picture support. Given the similarity between the grammars of Cantonese to Mandarin Chinese, the findings from this study can be largely generalizable to Cantonese. Hickmann and Hendriks (1999) found that zero forms and pronouns are particularly frequent in Chinese children and adults as shown in the following excerpt from Hickmann et al. produced by a five year old child (1999, p. 443):

jiu4 [o] xia4-lai2, [o] xia4-lai2 yi3hou4 ne, [o] ba3 ta1 gei3 xia4 pao3

Then down-come, down com after particle preposition 3rd person pronoun give
frighten-run

(‘Then [cat] came down, after [cat] came down, [dog] scared him away.’)

Another cross-linguistic study conducted by Chafe (1980) comparing Chinese and other language found that although Chinese does not possess verb tense marking, speakers can still tell temporally very precise stories (Erbaugh, 1990). This may be because Chinese and Cantonese speakers generally make use of lexical devices such as temporal adverbials as the rhetorical strategy to express the time concept, rather than using morphological markings. Finally, a further typological feature that may call for our attention in interpreting our data is that the temporal and causal links play a role in textual cohesion in Chinese narratives (Su, 2000). However, clauses are often pulled together by juxtaposition without explicit use of connectives in conversation as illustrated in the example (Matthews & Yip, 1994). Listeners are expected to infer implicitly encoded relations between the clauses.

keoi5 m4 sik2, dou1 jiu4 heoi3

3rd not-can also need go

(‘(Even if) he can’t, he still has to go.’)

These features highlight the importance of cultural and language-specific rhetoric forms which should be considered when testing children’s narrative skills to ensure unbiased appraisal (McCabe & Bliss, 2003).

Integration of the prior literature on the narrative abilities of English-speaking children with SLI, the linguistic characteristics of children with CSLI and cultural/linguistic differences in narrative construction, we propose that an assessment of narrative performance for Cantonese-speaking school-age children should include syntactic, cohesive, lexical and referencing components.

The Present Study

As noted above, the main aim of the study was to develop a narrative assessment instrument that would be useful in the identification of language impairment in Cantonese-speaking children. An implied objective is to determine the strength of the microstructure components as predictors of language development between 5 and 11 years of age. By doing so, we expected to provide further support of narrative ability as an integrated skill building on lexical and sentence level ability as well as textual comprehension ability (e.g. Justice et al., 2006). More specially, the present study addressed the following research questions:

- (1) Are there significant differences across grade levels in vocabulary, syntactic complexity, reference, connective use and total narrative score?
- (2) Which variables(s) best predict(s) age? and
- (3) Can the measures proposed differentiate Cantonese-speaking children with SLI and typically developing (TD) children?

Method

The present study stemmed from a large scale project, the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS, T'sou et al., 2006)¹, which aimed to develop a standardized language assessment tool for testing school-age Cantonese-speaking children in

¹ HKCOLAS was published in 2006 and made available from the Child Assessment Service of the Hong Kong SAR Government to all qualified speech therapists.

Hong Kong. Hong Kong is located on China's south coast on the Pearl River Delta with a population of 6.99 million (Hong Kong Statistics – Population and Vital Events, 2006). More than 90% of the Hong Kong population is of Chinese descent and the majority speaks Cantonese as their first language. In addition to Cantonese, English is also the de facto official language, spoken by more than 38% of the population 2006 (Hong Kong Statistical and Census Department, 2006).

The HKCOLAS instrument was comprised of seven subtests (textual comprehension test, word definition test, lexical-semantic relations test, test of Hong Kong Cantonese grammar, expressive nominal vocabulary tests, nonword repetition test, and the narrative test) which were constructed to represent the language components of later language development and processing ability. HKCOLAS was designed to provide a holistic picture about an individual's strengths and weaknesses across a wide range of language ability. Data reported in the present study were from the Narrative Test.

Participants

Participants were 1120 children aged between 4;10 and 12;01 and they were sourced from primary schools, nurseries and kindergartens. Stratified random sampling by gender, grade level, and three major geographic districts in Hong Kong was used. Because of the rapid rate of language development during the early preschool years, preschool children's development was examined more closely in two smaller age ranges than the primary children. Preschool children were grouped into two six-month age bands (i.e. 5;00 and 5;06) while school-age children were divided into six grade levels from Primary 1 (P1) to P6. The sample size was calculated on the basis of 8 grade levels x 2 sexes x 3 districts x 20 participants, for a total of 960 participants. With the addition of a further 10% to act as replacements for any possible drop-outs, the final number of subjects to be recruited was estimated to be 1056. Children were sampled randomly from classes and with reference to their grade level.

Approximately equal numbers of boys and girls were included. The parents of the child participants received an invitation letter which explained the purpose of the study. They signed the consent letter and returned it with a questionnaire asking for demographic information such as the child's native language, history of receiving speech therapy services and total family income. In the final sample, there were 1,120 participants. Because the children were sampled in grade groups, there was a small degree of overlap in the ages in the four of the grade levels (affecting four participants), for example the oldest child in the Primary 3 group was 8;7 years old, while the youngest child in the Primary 4 group was 7;9 years old. As narrative develops with literacy, we accepted these grade groupings as reflective of each age group. Subject information is summarized in Table 1. Initially, no exclusionary criteria were applied, and children with speech-language-impairment or history of this were not excluded from the sample so that the sample better represented the mainstream population in the region (Mcfadden, 1996). Thus all children were entered into the first set of analyses that examined variable scores as a function of grade. However, in order to examine the diagnostic accuracy of the narrative test, two exclusion criteria were imposed. These were a history of attending speech and language therapy services and suspected presence of SLI based on the clinical judgment of the assessing speech-language pathologist. These children were subsequently removed from this population sample (N = 40) leaving only children with no history of intervention (N = 1080).

Table 1 about here

Based on the salary scale reported in the bi-census data, three categories of the total family income were devised to represent three levels of the social economic status (SES) of the child's family. The SES distribution is shown in Table 2. An adult group comprised of 15 undergraduate university students (aged 21 years) was included and their responses were used to aid with the development of the model story and scoring protocols.

Children with SLI. A group of children with SLI was included to establish the validity for the whole assessment tool. Fifty-six Cantonese-speaking children aged 5;06 to 12;00 who were diagnosed as language impaired by experienced speech therapists were recruited from the Child Assessment Service of the Department of Health in Hong Kong (see Table 3). The diagnosis was made by experienced speech-language pathologists using the service's normal procedure of completion of an informal assessment checklist that observed aspects of semantics, morpho-syntax and pragmatics. This was because there was no standardized reference test in Hong Kong at the time of this study (hence the development of the HKCOLAS). All children achieved a performance IQ at or above 85 on the Hong Kong Wechsler Intelligence Scale for Children (1981). Inclusionary criteria for SLI were no sensory impairments, no psychiatric problems and no frank neurological deficits. Six children who were diagnosed as suffering from Attention Deficit and Hyperactivity Disorders (ADHD) were excluded and this left a total of 50 children with SLI.

Tables 2 and 3 about here

Materials and Testing Procedures

Narrative samples were collected using a story retell task in order to better control the vocabulary, syntactic structure, story structure and story length compared with a story generation task or personal narrative (Hadley, 1998). A 24-frame picture series depicting a story with four interconnected episodes was carefully constructed based on the framework of story grammar analysis (please see Appendix A). The story consisted of a series of problem-resolution events in which six protagonists make attempts to rescue a cat. A script incorporating different syntactic structures and connectives coding various semantic relations was written as the model story. The story was read by a male voice. The narration and page-flipping sounds that signaled picture boundaries were recorded onto a mini-disk (MD).

Each child listened to the model story through ear phones with the support of the pictures.

The story book was put in an upright position so that the investigator was not able to see the picture. The child was instructed to flip the pictures themselves on hearing the page-flipping sounds. The investigator would ensure the child was looking at the correct picture by tracking the track number displayed on the MD player and the page number printed on the back of each picture. After listening to the model story, the child was asked to retell the story, using the picture series, to the “naïve” investigator. A trial story was used to familiarize the children with these procedures before the test story. Using the above set-up and procedures, mutual knowledge between the investigator and the child about the story was kept to a minimum.

Prior to data collection proper, a pilot study was conducted that included 196 children aged between four and 12 years, and an adult group of 15 university students (who were different from the main study). This allowed us to trial the proposed methods of analysis and to define the desirable end-state (adult-like) performance for the chosen story.

Utterance segmentation. Samples were first transcribed verbatim and segmented into utterances. Segmentation relying on syntactic structures alone lead to considerable inter-rater variation. Instead, segmentation relied mainly on intonation patterns and semantic meaning and to a lesser extent on syntactic structures. The decision to use intonation patterns instead of syntactic structures, as in the T-unit (Hunt, 1965) and the C-unit (Loban, 1976) calculations, was due to the fact that prosodic elements play the most important role in determining the boundaries of a sentence in Chinese (Chao, 1968). Sentences in Chinese can be conjoined by juxtaposition without an explicit connective and multi-predicate utterances incorporating serial verb constructions are productivity structures in Chinese. Eight narrative samples from the pilot data were used to assess the inter-rater reliability of utterance segmentation. The method relying on intonation for segmentation yielded the best inter-rater reliability at 85%. Therefore, utterance segmentation was determined by intonation patterns.

Narrative Measures

Macrostructure (story grammar). The story for the pilot study was comprised of 46 pictures. The first measurement explored was the number of story grammar components recalled by the children. Ceiling effects were observed in even the three youngest groups of children. This may have been due to the elicitation method of story retelling with heavy picture support, because the main idea of each story grammar component was depicted explicitly in each picture (Norbury & Bishop, 2003). Even very young children who failed to convey the textual nature of the narrative gained credit for their description of the pictures. While older and younger children did not differ on the number of story grammar components recalled, it was the older children's ability to use sophisticated vocabulary, syntactically complex utterances and greater referential cohesion that set them apart from the younger children. Therefore, analysis focused on microstructure components (see below) rather than a macrostructure analysis (i.e., story grammar). This was fitting given that Liles et al (1995) reported better discriminatory ability of microstructure analysis. In light of these observations, the final measures adopted in the main study were a semantic score, a syntactic complexity score, a connective score, and a referencing score. Finally, the number of pictures was reduced to 24 to reduce picture dependence in story re-telling.

Semantic score. Points were awarded for the use of sophisticated vocabulary. Adults' productions were taken as indices of what could be considered 'sophisticated vocabulary'. To illustrate, consider an example from the story. Picture 20 of the story depicts children rescuing a cat up a tree. In the pilot study, most adults used the term 'piggyback' (see sentence (a) below). Older children used the form in (b) below, and the youngest children used the simple form in (c) below. The most immature forms were often characterized by omission of specific vocabulary (c).

(a) "*The kids piggyback and rescued the cat.*" (7 words)

(b) "*The kids were piling up the tree to get the cat.*" (11 words)

(c) “*They get it down.*”

To ensure reliable and valid scoring of these structures, a list of words was selected based on adults’ productions. Children’s narratives were scored against this pre-selected set of ideas in terms of gist, as for previous standardized assessment (e.g. Bishop, 2004). Extra credit (i.e. two marks) was given to the use of more sophisticated vocabulary or more precise ideas. In the above examples, (a) would receive a score of 2 while (b) would receive a score of 1 (see Appendix B). Operationally, sophisticated items here refers to those that belong to a higher register of vocabulary that occurs more often on formal occasions or that requires collocation of single morphemes, as is common in Cantonese. In Cantonese, multi-syllabic words are derived from monosyllables through compounding and derivation, or phrases with words that are collocated to denote specific meanings. For example, in the model story, the vocabulary item *baaul zaat3* “wrap-tie up” not only refers to “wrap” but also entails “to bandage the wound” which is comprised of collocated compounds and so this item is regarded as a higher register counterpart of the word *baaul* “wrap”. The use of collocated or compounded forms in Cantonese spoken language is very obvious and the judgment of basic and sophisticated items was straight-forward. The identification of sophisticated items was quite clear-cut as demonstrated in the high inter-rater reliability of 98% for scoring of vocabulary. Words used by the adults were indicative only, and if a child produced a sophisticated word which was semantically appropriate to the context and referred to the target ideas, but was not listed in the model story or even the pre-selected list, he/she will also be given credit. The maximum possible semantic score was 92.

Syntactic complexity. We examined the use of the measure ‘mean length of utterance in words’ (MLU_{word} per clause) as a proxy for syntactic complexity in Chinese. However, since pauses play a very important role in utterance segmentation in Chinese, significant subjectivity was observed in inter-rating coding (To, Cheung & T’sou, 2008). Five percentage

of all the normative child samples were re-analysed (segmentation and coding) by one of the RAs and an independent speech therapist. Intra-rater and inter-rater agreement was calculated using a point-to-point percentage of agreement procedure. This yielded 86% and 74% accuracy of intra- and inter-rater reliability respectively (To, Cheung & T'sou, 2008). Given that inter-rater reliability is an important concern in establishing a standardized test, other measures that better capture children's syntactic ability precisely and reliably were employed.

In Chinese, there are seven basic structures to construct sentences. They include (1) subject-predicate *zyu2wai6*, (2) subordination *pin1zing3*, (3) verb-object *wai6ban1*, (4) verb complement *wai6bou2*, (5) serial verb *lin4wai6*, (6) pivotal structure *gim1jyu5*, and (7) coordination *bing6lit6* (Cheung, H-N, 2007). All of the seven structures can be used to form more complex structures. Among these seven types, the subject-predicate structure was not chosen as a measurable structure because of its limited value in contributing to complex syntax during the school years. Coordination, which refers to a series of items in the same word class (e.g. noun phrases, verb phrases, adjectives) conjoined by juxtaposition without any overt verbal and nonverbal (e.g. pause) marking, was also not chosen. Because of the occurrence of zero marking in Cantonese, it poses many problems in accurate identification. We selected a total of seven structures to be examined in the pilot study, based on their contribution to complex constructions, and they are summarized as follows (see Appendix C):

- | | |
|---|------------------------------------|
| (i) subordination <i>pin1zing3</i> : | (1) Relative clause (RC) |
| (ii) verb-object <i>wai6ban1</i> : | (2) Preverbal manner modifier (MM) |
| (iii) serial verbs <i>lin4wai6</i> : | (3) Clausal complement (CC) |
| | (4) Prepositional phrase (PP) |
| | (5) Serial verb construction |
| (iv) verb complement <i>wai6bou2</i> : | (6) Verbal complement |
| (v) pivotal constructions <i>gim1jyu5</i> : | (7) Pivotal construction |

The first four types of syntactic structures were found to be significantly different across grade levels in the pilot study. The sum of these four most sensitive structures was then used as an index of children's syntactic complexity. The total score for syntactic complexity in the model story was 38.

Referencing. This measure determines the accuracy of a referential form used by a child when introducing and switching reference. A previous study of the acquisition of referential forms in Cantonese-speaking children showed that young children were prone to use attenuated forms such as zero marking and pronouns to refer to characters rather than being too explicit or redundant to their listeners (To, 2006). In this way, anaphoric pronominal forms may not be easily differentiated from deictic use for the function of referent maintenance in narrative contexts (To, 2006). Only referent introduction and switching which called for an explicit noun phrase form were taken into account in representing children's referencing ability in this study. The model story contained 4 male and 2 female characters and 1 animal character which presented the participants with the task of providing clear references for their listener's referent identification. Fourteen points of referent introduction and switching where all adult participants used consistent accurate noun phrase forms (as opposed to pronouns and zero forms) were identified for scoring, for a maximum possible score of 26. Scoring criteria included the appropriateness of the linguistic form, and the distinctiveness of the adjectives used to contrast the characters (see Appendix D).

Connectives. Connectives in Mandarin Chinese play a role in conjoining utterances and formulating complex sentences in conversation and narrative contexts (Su, 2000). Although no comparable findings have been reported for Cantonese, connectives also contribute to textual cohesion in Cantonese in a similar manner. All connectives were first identified with the computer program PowerGREP (Goyvaerts, 2006) which highlighted all the possible

connectives in the transcriptions. The acceptability of the connectives was judged manually with reference to the logical relation they encoded. These include causality, concession, hypotheticality, coordination/addition and temporality (see Appendix E). Only those connectives that encoded correct logical relations were given credit. Discourse markers (e.g. *gan1zyu6* ‘then’, *jin4hau6* ‘and then’ occurring at the utterance initial position) without coding a specific semantic relation of two clauses were not included. The total number of connectives used correctly was then computed. The total score for connectives in the model was 28.

Reliability

The narratives were transcribed by five trained research assistants. This involved orthographic transcription and utterance segmentation. Five percent of the samples were subjected to measures of inter-rater reliability. Syllable-by-syllable inter-rater agreement was computed by the percentage of correctly transcribed syllables to the total number of syllable transcribed. Inconsistency occurring in mazes was not included in the calculation of agreement. The level of agreement for syllable transcription was very high (99%). Three research assistants who transcribed the narratives were trained in scoring. Five percent of the samples scored by each of the research assistants were counter-scored manually by the first author. The inter-rater scoring agreement was calculated by dividing the number of items matched by the total number of matched and mismatched items. The agreement levels of the four measures were high, with 98% for semantic score, 94% for syntactic complexity, 96% for connectivity and 97% for referencing. Reliability of each subtest was also examined by computing the reliability coefficients of all the items in this narrative test. This correlation coefficient measures the degree of consistency of a proposed measure (Reinard, 2001) and the result of 0.89 for the whole sample was regarded as good.

Measurement of diagnostic accuracy of the narrative test

Validity of a test refers to its ability to differentiate individuals with and without a disease. Validity has two components: sensitivity and specificity. The principles and purposes of using diagnostic accuracy measures are outlined by Dollaghan (2007) and Klee (2008) and will not be repeated here, but brief descriptors are included, based on Stokes and Klee (in press). Sensitivity is a measure of the value of the test in correctly identifying language impaired (LI) children as LI and specificity is a measure of the value of the test in correctly identifying typically developing (TD) children as TD. Establishing sensitivity and specificity for the present study was in fact challenging as there is no previously available standardized language assessment in Hong Kong to assess school age children hence the original motivation of developing this project. Due to the lack of standardized language assessment, clinical decisions were mainly based on pieces of available research findings on Cantonese and English as well as speech therapists' clinical experience. The reference standard for comparison of the validity of the test in the present study is the clinical judgment of experienced speech therapists and the clinical history of speech therapy of a child.

The positive likelihood ratio (LR +) [$\text{sensitivity}/(1-\text{specificity})$] shows how likely it is that a fail score on the narrative test comes from a child with LI rather than a TD child. A negative likelihood ratio (LR -) [$(1-\text{sensitivity})/\text{specificity}$] shows how likely it is that a pass score on the narrative test comes from a child with LI. In using LRs to diagnose individuals, McAlister, Straus and Sackett (1999) suggest that values of $\text{LR}^+ \geq 10.0$ and a $\text{LR}^- \leq 0.10$ are minimal levels for acceptable index tests (i.e. the narrative test in this case). If an index test achieved these desirable LR^+ and LR^- values (and had acceptable confidence intervals for these measures), we would conclude that a LI child would be at least 10 times as likely as a TD child to fail the index test, and that a LI child was only 0.10 times as likely as a TD child

to pass the index test. The last measure to be used here is the diagnostic odds ratio (DOR). A DOR is “the ratio of the odds of positivity in disease relative to the odds of positivity in the nondiseased” (Glas, Lijmer, Prins, Bossel & Bossuyt, 2003, p. 1130), and is computed as LR^+/LR^- such that as LR^+ increases and LR^- decreases, DOR increases with values between zero and infinity. A DOR value of 1 suggests that the index test has no value as a discriminator between LI and TD children. In this study we generated DORs for each of the components of the narrative test to determine which of the variables was the most useful in detecting SLI in this sample of children.

Results

Normative Sample

Table 4 summarizes the descriptive statistics of the four measures in the narrative test as a function of age. The first research question (are there significant differences across grade levels in the quantitative narrative measures derived in this study?) was addressed by univariate analyses of variance (ANCOVA) with alpha levels set at 0.01 (0.05/4) for a Bonferroni correction for multiple comparisons. In order to control for the length of the story produced, the measures of semantic score, connectives and syntactic complexity were adjusted for the measure of number of words by putting this as a covariate in the analysis of variances. There was a significant main effect for the semantic score ($F_{(7, 1101)} = 142.86, p < 0.001$, partial $\eta^2 = 0.48$); syntactic complexity ($F_{(7, 1101)} = 330.07, p < 0.001$, partial $\eta^2 = 0.52$); reference ($F_{(7, 1102)} = 58,56, p < 0.001$, partial $\eta^2 = 0.27$) and connectives ($F_{(7, 1101)} = 10.83, p < 0.001$, partial $\eta^2 = 0.64$).

Post-hoc analyses were then used to isolate significant differences among the grade levels. For semantic score, all groups were significantly different from each other, except for K2 and K3 groups, and P4 and P5 groups. For syntactic complexity, the groups that were not

statistically significant were (i) K2 and K3; (ii) P2 and P3; (iii) P3 and P4, P5; (iv) P4 and P3, P5; and (v) P5 and P4. For referencing, the 8 groups can generally be differentiated into 3 significantly different groups, K2 to P1 as one group, P2 alone as the other group, and P3 to P6 as another group. Finally, for connectives, the younger group from K3 to P1 is significantly different from the older group from P3 to P6.

Difference scores from the post-hoc analyses were also used to detect the biggest growth periods for each variable. Increase in the semantic score was seen right up through 11 years and the biggest gain was between K3 and P1. Children at P3 showed a two-fold increase in semantic score over the youngest group. Younger children did not recall as much relevant information and used less precise vocabulary than the older children but their story content was generally relevant to the original story. Syntactic complexity was also a robust developmental measure. Children demonstrated a steady increase in the number of complex structures across the grade levels. The score increases most remarkably from K3 (7.80) to P1 (14.51). Younger school-age children generally used fewer syntactic structures such as clausal complementation of perception and mental verbs, relative clauses, and prepositional phrases than their older peers in packaging their stories. Some degree of sophistication in the handling of clausal linkages (connectives) and anaphoric dependencies (referencing) in framing the narratives was observed during the early school years. The biggest leap in both connective scores and referencing scores occurred between P1 and P2. At P2, children used more than double the number of connectives than the youngest group at K2. At about P3 and P4, the performance on connectives reached a plateau. The standard deviations shown in Table 4 reveal that children in all grade levels demonstrated greater variability in using connectives than any of the other measures. Great improvement in the use of explicit reference occurs before P2 with a very protracted improvement from P3. By P6, children used appropriate

nominal forms to introduce discourse referents and for marking switched referents which is close to adult performance.

Table 4 about here

We then subjected the four measures to a principal components factor analysis. The results showed that the four measures were highly and significantly correlated (Table 5) and that the semantic score accounted for 78% of unique variance in the model, followed by syntactic complexity with 12%, referencing with 6% and connectives with 4% unique variance. Only one factor was extracted and the subtests were highly inter-correlated, therefore, we generated a composite score by summing across the four measures for a total narrative score which represents an index of the general narrative skill.

Table 5 about here

The second research question (which variables(s) best predict(s) grade?) was addressed via a multiple regression. The four variables (semantic score, syntactic complexity, reference and connectives) were entered into a multiple regression analysis with grade as the dependent variable. Forward elimination was used. The final significant model was obtained with the variables of semantic score and syntactic complexity ($F_{(1,1104)} = 1840.42, p < 0.001$). Semantic score accounted for 63% of the unique variance in grade, and syntactic complexity accounted for an addition 5% of unique variance. Together they accounted for 69% of the variance in grade. The other two variables (reference and connectives) were eliminated from the model. Next we tested the usefulness of the composite score as a predictor of grade, wherein the composite score alone accounted for 65% of the variance in Grade ($F_{(1,1104)} = 2070.36, p < 0.001$). Given the high inter-correlation among the variables, referencing and connectives could be excluded in an assessment. However, given the strength of the narrative composite score, it is recommended that all measures be retained.

Diagnostic accuracy

Finally, we examined the diagnostic accuracy of these measures. Sensitivity, specificity, positive likelihood ratios (LR+), negative likelihood ratios (LR-) and diagnostic odds ratio (DOR) were calculated for each of the four variables. In this calculation, children in the normative study with a previous history of speech therapy as reported by the parents in the questionnaire and confirmed diagnosis of language impairment from the interviewing speech therapist were excluded leaving the 1080 assumed TD children. A total of 50 children with SLI recruited from the Child Assessment Service along with these 1080 TD children were used to establish the sensitivity and specificity of the index test. Table 6 summarizes the descriptive statistics of the TD and SLI groups. ANOVA tests with alpha levels set at 0.01 (0.05/4) for a Bonferroni correction for multiple comparisons were conducted. A test of homogeneity using Levene's test confirmed the equal variances assumption between these two uneven groups (n=1080 vs. 50). SLI children scored significantly lower than the TD children for all of the four measures [semantic score ($F_{(1, 1120)} = 198.63, p < 0.001, \text{partial } \eta^2 = 0.151$), syntactic complexity ($F_{(1, 1120)} = 272.68, p < 0.001, \text{partial } \eta^2 = 0.20$), referencing ($F_{(1, 1120)} = 124.52, p < 0.001, \text{partial } \eta^2 = 1.0$) and connectives ($F_{(1, 1120)} = 49.50, p < 0.001, \text{partial } \eta^2 = 0.04$)].

We firstly set a cut point of -1.25 SD below the mean on each variable (i.e. the 10th percentile) for each grade (Table 7 shows the 10th percentile cut points for each variable by each grade). In addition, we generated ROC curves to more fully examine the diagnostic accuracy of these measures from a mathematical perspective to identify the best cut-point. The optimal cut point for diagnostic accuracy for pass/fail rates on each narrative measure was determined by identifying the cut-points that yield the best sensitivity (i.e. the maximum number of true positives, along the y-axis) and specificity (i.e., maximum number of true

negatives). Children scoring at or below the cut-off were defined as impaired and children scoring above it were classified as non-impaired.

Table 6 and 7 about here

Cross-tabulations of each variable for original diagnosis and index test diagnosis were generated for both sets of cut-points (-1.25SD and that found from ROC curves). The results were entered into an online diagnostic test calculator (Centre for Evidence Based Medicine University of Toronto). The results are shown in Table 8 and 9. Using the -1.25SD, sensitivity ranged from 34 to 94%, specificity ranged from 88 to 89%, LR+ ranged from 2.79 to 7.85, LR- ranged from .07 to .75 and DOR ranged from 3.72 to 112. From the ROC curve, the cut points were slightly different and the validity measures were better. Sensitivity ranged from 86 to 94%, specificity ranged from 60 to 90%, LR+ ranged from 2.15 to 9.42, LR- ranged from .07 to .34 and DOR ranged from 9.20 to 140.6. As noted in the methods section, the DOR allows for a direct comparison of the value of each variable as a potential identifier of children with SLI. In this study, regardless of which cut point was used, the best marker was the syntactic complexity variable, followed by the semantic score.

Table 8 and 9 about here

Discussion

Narrative as a sensitive test of language in the school-age years

Standardized narrative tests for school age children have become popular in the last decade (e.g. Bishop, 2004; Gilliam & Pearson, 2004, Strong, 1998). The normative results reported here provide an objective benchmark for documenting a Cantonese-speaking child's level of performance in narrative production with reference to his/her age peers. The findings not only assist in the diagnosis of children with narrative problems but also improve our understanding of Cantonese language development during the school years. Children's

narrative skills improved significantly across the school years with even continual growth up through 10 and 11 years as measured on the parameters of semantic score, syntax, connective use and referencing ability. Previous research had documented the development of the linguistic forms under investigation in this study in preschool children, including the use of particular NP structures for achieving referential cohesion, connectives and various syntactic structures for conjoining clauses. (c.f. Fletcher, Leung, Stokes & Weizman, 2000; Lee & Wong, 1998). These studies showed that these forms emerge early, before five years of age. However, being a proficient native speaker is not only a matter of knowing many words or grammatical rules, but also being able to exploit that knowledge effectively for various communicative purposes or in different genres such as narratives. Each of the quantitative measures explored here exhibited different developmental rates across the school years.

Semantic score was the best growth predictor and accounted for the largest proportion of variance in Grade in the regression model. Overall there was a robust growth pattern among cohorts of different grades. There was considerable variability in scores within grades, as shown in Figure 1. Justice et al (2006) reported similar results for both growth in vocabulary scores across time and variability within age groups, for 250 English-speaking children aged 5-12 years, using the number of different words (NDW) as a measurement of vocabulary deployment in spoken narratives. The usefulness of a measure of vocabulary deployment for studying the language abilities of school aged children was also reported in a study of the written narratives of 120 children in grades 3-6 (Puranik, Lombardino, & Altmann, 2008). Although measured differently (as 'number of ideas expressed'), there was a similar growth pattern for a semantic score across time.

The ability to use a range of complex syntactic markers in a discourse (narrative) genre was another significant predictor of Grade in this sample of 1120 children. Whereas preschool

children tended to use simple syntactic structures such as canonical SVO sentences to retell the story, and relied on connectives for text cohesion, older children were able to use more sophisticated, and a wider set of syntactic structures to complete the task, with less use of connectives. To have an idea of how mature Cantonese speakers exploited these two grammatical devices to package the story, we examined the performance of the adult group from the pilot study. In the model story, connectives could occur to a maximum count of 28, and the number of syntactic markers could occur to a maximum count of 38. Our adult participants used, on average, 12.86 and 34.33 connectives and syntactic markers respectively (46% and 90% of maximum scores respectively). Thus, for adults, using complex syntactic structures is the preferred method for achieving tighter textual cohesion. This is because connectives are often optional in Cantonese. Mature speakers often juxtapose two clauses without using any explicit markings and listeners anticipate the need to infer the logical relation between the two propositions. Embedding clauses through complementation or relativization appears to be a productive way to conjoin clauses in narratives in Cantonese. The interaction of the optionality of connectives and coding cohesion through syntactic complexity makes interpretation of developmental growth patterns less than straightforward.

In early Cantonese narratives, connectives are a dominant linking device until P3 when connectivity of the narrative text is achieved by using complex sentence structures (increasing syntactic complexity). This may explain why there is marked connective growth among the younger groups but this growth reached a plateau quite early, and did not provide a significant contribution to the prediction of Grade. Syntactic complexity continued to show a strong growth pattern up through 11 years. The high correlation between the sentence complexity measure and NDW ($r_{(248)} = 0.81$) in Justice et al (2006) reflects a similar pattern for English-speaking school aged children.

The referencing measure in the present study demonstrated the notion of obligation (providing required reference points for the listener) among the four variables. Referential clarity plays a crucial role in Cantonese narratives because the pro-drop property means sentence subjects and objects are easily omitted. Pre-primary children tended to use unclear referential expressions such as pronominal forms when introducing and switching referents which called for more explicit NPs for clarity. Marked development was observed during the early school years and then gradually leveled off from P3. Recall that we only selected the “obligatory” points that required explicit nominal forms as our scoring criteria. The plateau of reference skills may *at most* reflect that children can gradually make correct assumptions about how much information the listener needs to interpret the story. It is possible that if a reliable measure of gauging children’s referential maintenance skills (not just introduction and reintroduction) can be developed then referencing may make a stronger contribution to Grade prediction.

Differentiation between TD children and children with SLI

In the present study, the children with SLI scored significantly lower than the TD children on all four microstructure measures. While previous research has suggested that sentence-level measures such as MLU are not age sensitive beyond 5 years, and are probably not useful for identifying children with language-impairment from the late preschool years onwards (Blake, Quartaro & Onorati, 1993; Miller, 2004; Johnston, 2006), syntactic complexity measured in terms of the use of particular forms in the narrative genre is illustrated to be a reliable and more valid indicator of language impairment during later language development (Scott, 2004). Though sentence connectives can distinguish TD children from children with SLI, the between group differences of this measure was not as substantial as the other measures. It seems that when children with SLI failed to deploy

complex syntactic structures to achieve cohesion, connectives appear to be an easier strategy for them to link simpler utterances together.

The other aspect that signifies the presence of language impairment in Cantonese is a limited ability to include as many ideas, and deploy precise lexical items to represent ideas, efficiently in narratives. The children with SLI tended to employ general purpose rather than specific lexical items. For example, one child with SLI said *go3 neoi5zai2 gan1zyu6 zing2 zek3 maau1zek3 goek3* “the girl then **did** the leg of the cat” while his age-matched peer said *go3 neoi5zai2 bong1 zek3 maau1baau1zaat3 soeng1hau2* “The girl helped the cat to *bandage* the wound”. Lexical choices in narratives may be a potential area to further explore in Cantonese-speaking SLI children by separating the variable into the number of ideas recalled and the sophistication of vocabulary employed (Klee, Stokes, Wong, Fletcher & Gavin, 2004).

Some stories from SLI children were characterized by a total lack of referential clarity by using the third-person pronoun *keoi5* throughout the story, resulting in fragmented and unclear stories. Consistent findings were also reported in Norbury and Bishop’s (2003) study who found significant referential problems in language disordered children regardless of whether they were children with SLI, PLI or ASD.

Sensitivity, specificity and DOR

The four microstructure measures produced different levels of sensitivity and specificity as reflected in different DOR values. Although the semantic score accounted for most of the variance in Grade in the regression model and was the best predictor of Grade, it was the syntactic complexity variable which had the highest value of DOR and best differentiated children with and without SLI. Cantonese SLI children’s poor narrative production seemed to be ascribed to a large extent to their reduced use of complex syntactic structures when

compared with their peers. As mentioned before, some children showed an adequate grasp in referencing, connectives, and even the semantic score, but their syntactic complexity still lagged behind their TD peers. Building on previous studies of CSLI during the pre-school years, the present findings may suggest that the syntactic difficulties identified in younger children continue to be a problem to CSLI children in the school years, manifested as a restricted deployment of complex syntactic structures at a textual level.

With the significant group difference between children with SLI and TD children and along with the relatively good sensitivity of all of the four measures, we have attested that narrative production is also a vulnerable aspect to Cantonese SLI children as has been found for English-speaking children (e.g., Justice et al., 2006). The relatively low specificity levels (at or below 90%) in this study may simply reflect that in this sample of over a thousand children there were children who scored below the cut point (i.e. were coded as 'impaired' on at least one of the microstructure measurements. Given that our exclusion criteria (for being coded as TD) were a history of previous speech/language intervention, and the clinical judgment of the assessing speech/language pathologist, it is entirely possible that there were some children who were performing in the borderline range of language ability were entered into the TD group. This is not a novel idea. It has been suggested that children with language impairment may remain undiagnosed in the community, or have an illusionary recovery period, where children's submerged problems, that seem to have resolved before preschool, reappear in the school years (Johnston, 2006). From a clinical perspective, these narrative measures cannot yet be used solely for the diagnosis of SLI. Instead, the narrative sub-test should be interpreted in conjunction with other language measures for identifying language disorders in school-age children. As recommended by Leonard (1998) and Paul (2001), the diagnosis of language impairment should be based on at least two different language subtests.

In fact, the whole assessment battery for Cantonese school aged children was made up of six language subtests and one nonword repetition test. The narrative subtest is only part of the whole test. Diagnosis of language impairment in Cantonese-speaking children is determined by failure on two or more of the six subtests in the whole assessment battery. Coverage of all of the subtests is beyond the scope of the current report but a subsequent report will describe diagnostic agreement across subtests. The current report on the narrative results highlight the need for continued verification of the work presented here. We hope that this report will serve as a catalyst for continued research using this narrative instrument.

Limitations of the present study and future research

This study has demonstrated the robustness of the narrative components as developmental measures. However, there are several shortcomings in this research. First, before this body of research, there was no standard reference test for the diagnosis of language impairment in Cantonese-speaking children. In the absence of such a test, language status (TD versus SLI) was determined by the criteria of a history of intervention and clinical judgment. A subsequent publication will document the degree of agreement across the seven subtests of the entire HKCOLAS in detecting language impairment. In addition, diagnostic research is generally construed within a four-phase model of investigation (e. g., Sackett & Haynes, 2002). We have yet to explore phase IV (tracking the long term outcome of those who failed the diagnostic test). Our phase III research reported here (does the test distinguish between those with and without the target disorder among those in whom it is clinically reasonable to suspect that the disorder is present) needs to be replicated with a larger sample of children with CSLI. In addition, our preliminary studies on CSLI have suggested that children may exhibit different language profiles. Further investigation should be conducted to explore the different narrative profiles of children with specific language impairment, with

reference to the measure of semantic score, syntactic complexity, reference and use of connectives and the relationship with other language components examined in the HKCOLAS.

Acknowledgments The authors gratefully acknowledge the Child Assessment Service, Government of HKSAR and all the speech therapists participated in this project. We would also like to thank Dr. Anita Wong, for her comments on the early version of this paper, and also the schools and children who helped us with this research.

References

Berman, R. (2004). Between emergence and mastery. In R. Berman (2004). *Language development across childhood and Adolescence* (pp. 9-34). Amsterdam: John Benjamins B. V.

Berman, R., & Slobin, D. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Erlbaum.

Bishop, D. V. M. (2004). *Expressive, Receptive and Recall of Narrative Instrument*. London: Harcourt Assessment.

Bishop, D., & Edmundson, A. (1987). Language impaired 4 year olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders*, 52, 156-173.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, 20, 139-152.

Botting, N. (2002). Narrative as a clinical tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18, 1-22.

Botting, N., Faragher, B., Simkin, Z., Knox, E., & Conti-Ramsden, G. (2001). Predicting pathways of specific language impairment: What differentiates good and poor outcomes? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42, 1013-1020.

Brookshire, R. H., & Nicholas, L. E. (1994). Speech sample size and the test-retest stability of connected speech measures for adults with aphasia. *Journal of Speech and Hearing Research*, 37, 399-407.

Centre for Evidence-Based Medicine. (2008). *Stats calculator, EBM toolbox*. Accessed March 26, 2008, from <http://www.cebm.utoronto.ca>

Chafe, W. (1980). *The Pear stories: cognitive, cultural, and linguistic aspects of narrative production*. Norwood, N.J.: Ablex Pub. Corp.

Chafe, W. (1990). Some things that narratives tell us about the mind. In B. K. Britton & A. D. Pellegrini (Eds.), *Narrative thought and narrative language*. Hillsdale, NJ: Erlbaum.

Chao, (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Crais, E. R., & Lorch, N. (1994). Oral narratives in school-age children. *Topics in Language Disorders, 14*, 13-28.

Cheung, H-N. S. (2007). *Xianggang Yue yu yu fa de yan jiu. (revised ed.)*. Hong Kong: Chinese University of Hong Kong Press.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, Maryland: Brookes.

Erbaugh, M. S. (1990). 'Mandarin oral narratives compared with English: The Pear/Guava Stories'. *Journal of the Chinese Language Teachers Association, 25*, 21-42.

Fletcher, P., Leonard, L. B., Stokes, S. F., & Wong, A. M-Y. (2005). The expression of aspect in Cantonese-speaking children with Specific Language Impairment. *Journal of Speech, Language and Hearing Research, 48*, 621-634.

Fletcher, P., Leonard, L. B., Stokes, S. F., & Wong, A. M-Y. (2008). Morphosyntactic deficits in Cantonese-speaking children with specific language impairment. In S-P. Law, B. S. Weekes, & A. M-Y. Wong (Eds.), *Language disorders in speakers of Chinese*, Bristol: Multilingual Matters, (pp. 75-88).

Fletcher, P., Leung, C.S., Stokes, S., & Weizman, Z. (2000). *Cantonese pre-school language development: A guide*. Hong Kong: Dept. of Speech & Hearing Sciences, University of Hong Kong. (Report of a project funded by the Language Fund, Hong Kong, 1996-1999) (in English and Chinese versions). http://cd.emb.gov.hk/scolar/html/rindex_c02696_tw.htm

Fletcher, P., Stokes, S. F., & Wong, A. M-Y. (2005). Constructions and language development: implications for language impairment. In P. Fletcher & J. Miller (Eds.), *Developmental theory and language disorders*, Amsterdam: John Benjamins Publishing, (pp.

35-51).

Gilliam, R. B., & Johnston, J. (1992). Spoken and written language relationships in language/ learning impaired and normally achieving school-age children. *Journal of Speech and Hearing Research, 35*, 1303-1315.

Gilliam, R. B., & Pearson, (2004). *Test of Narrative Language*. Texas: Pro-ed.

Girolametto, 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*, 358-369.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology, 56*, 1129-1135.

Goyvaerts, J. (2006). PowerGREP (Version 3.3.2) [Computer software]. Thailand: Just Great Software Co. Ltd.

Hadley, P. A. (1998). Language sampling protocols for eliciting text-level discourse. *Language, Speech, and Hearing Services in Schools, 29*, 132-147.

Halliday, M. A. K. & Hansen, R. (1976). *Cohesion in English*. London : Longman.

Hickmann, M., & Hendriks, H. (1999). Cohesion and Anaphora in Children's Narratives: A Comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language, 26*, 419-452.

Hong Kong Statistical and Census Department. [Hong Kong Statistics - Population and Vital Events](#). Government of the Hong Kong SAR. Retrieved on [2008-02-01](#).

Hong Kong Wechsler Intelligence Scale for Children (HKWISC) (1981) (Translated Chinese version). New York: The Psychological Corporation.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Urbana, IL:

National Council of Teachers of English.

Johnston, J. (1982). Narratives: A new look at communication problems in older language-disordered children. *Language, Speech and Hearing Service in the Schools*, 13, 144-155.

Johnston, J. (2006). *Thinking about child language*. Eau Claire, WI: Thinking Publication.

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, 177-191.

Karmiloff-Smith, A. (1981). The grammatical marking of thematic structure in the development of language production. In W. Deutsch (Ed.), *The child's construction of language* (pp. 121-147). London: Academic Press.

Karmiloff-Smith, A. (1985). Language and Cognitive Processes from a Developmental Perspective. *Language and Cognitive Processes*, 1, 61-85.

Kintch, W., & Green, E. (1978). The role of culture-specific schemata in the comprehension and recall of stories. *Discourse Processes*, 1, 1-13.

Klee, T. (2008). Considerations for appraising diagnostic accuracy studies in communication disorders. *Evidence-Based Communication Assessment and Intervention*, 2, 34-45.

Klee, T., Stokes, S.F., Wong, A., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47, 1396-1410.

Lee, T., & Wong, C. (1998). "CANCORP: The Hong Kong Cantonese Child Language Corpus. *Cahiers de Linguistique d'Asie Orientale*, 27, 211-228.

Leonard, L. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. B., Wong, A. M-Y., Deevy, P., Stokes, S. F., & Fletcher, P. (2006). The production of passives by children with specific language impairment acquiring English or Cantonese. *Applied Psycholinguistics*, 27, 267-299.

Leonard, L. B., Deevy, P., Wong, A. M-Y., Stokes, S. F., & Fletcher, P. (2007). Modal verbs with and without tense: a study of English- and Cantonese-speaking children with specific language impairment. *International Journal of Language and Communication Disorders*, 42(2), 209-228.

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., Duffy, R. J., Merritt, D. D., & Purcell, S. L., (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research*, 38, 415-425.

Loban, W. (1976). *Language development: Kindergarten through grade twelve*. Urbana, IL: National Council of Teachers of English.

Mandler, J. M. (1978). A Code in the Node: The Use of a Story Schema in Retrieval. *Discourse Processes*, 1, 14-35.

Matthews, S., & Yip, V. (1994). *Cantonese: A comprehensive grammar*. London: Routledge.

McAlister, F. A., Straus, S. E., Sackett, D. L. (1999). Why we need large, simple studies of the clinical examination: the problem and a proposed solution. *Lancet*, 354, 1721-1724.

McCabe, A., & Bliss, L. (2003). *Patterns of narrative discourse: a multicultural life span approach*. Boston, Mass.: Allyn and Bacon

McFadden, T. U. (1996). Creating language impairments in typically achieving children:

The pitfalls of "normal" normative sampling. *Language, Speech, and Hearing Services in Schools, 27*, 3-9.

Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling, and story comprehension. *Journal of Speech and Hearing Research, 30*, 539-552.

Miller, J. (2004). *New approaches to language sample analysis*. Paper presented at the annual meeting of the British Columbia Association of Speech-Language Pathologists and Audiologists, Kelowna, BC.

Nicholas, L. E., & Brookshire, R. H. (1995). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*, 338-350.

Norbury, C. & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*, 287-313.

Pankratz, M. E., Plante, E., Vance, R., & Insalaccon, D. (2007). The diagnostic and predictive validity of the Renfrew Bus Story. *Language, Speech, and Hearing Services in Schools, 38*, 390-399.

Paul, R., & Smith, R. L. (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research, 36*, 572-598.

Pritchard, R. (1990). The effects of cultural schemata on reading processing strategies. *Reading Research Quarterly, 25*, 273-295.

Puranik, C. S., Lombardino, L. J., & Altman, L. J. P. (2006). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology, 17*, 107-120.

Reed, V. A., Patchell, Frederick C., Coggins, T. E., & Hand, L. S. (2007).

Informativeness of the spoken narratives of younger and older adolescents with specific language impairment and their counterparts with normal language. *Clinical Linguistics & Phonetics*, 21, 953-960.

Renfrew, C. (1969). *The Bus Story: A test of continuous speech*. Oxford, UK: Collins & Company, Ltd.

Roth, F. & Spekman, N. (1989). The oral syntactic proficiency of learning disabled students: A spontaneous story sampling analysis. *Journal of Speech and Hearing Research*, 32, 67-77.

Sackett, D. L., & Haynes, R. B. (2002). The architecture of diagnostic research. *British Medical Journal*, 324, 539-541.

Scott, C. M. (1988). Spoken and written syntax. In M. Nippold (ed.) *Later language development: Ages 9 through 19*. (pp. 45-95). San Diego, CA: College Hill Press.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with learning disabilities. *Journal of Speech, Language, and Hearing Research*, 43, 324-339.

Scott, C. M. (2004). Syntactic ability in children and adolescents with language and learning disabilities. In R. A. Berman (Ed.) *Language development across childhood and adolescence*. Amsterdam. Philadelphia: John Benjamins Pub.

Stein, N. L., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing* (Vol. 2, pp. 53-120). Norwood, NJ: Ablex.

Stokes, S. F., & Fletcher, P. (2000). Lexical diversity and productivity in Cantonese-speaking children with specific language impairment. *International Journal of Language and Communication Disorders*, 35, 527-541.

Stokes, S. F., & Fletcher, P. (2003). Aspectual forms in Cantonese-speaking children

with specific language impairment. *Linguistics*, 41(2), 381-406.

Stokes, S. F., & Klee, T. (in press). The diagnostic accuracy of a new nonword repetition test for differentiating the late talking and typically developing children. *Journal of Speech, Language, and Hearing Research*.

Stokes, S. F., & So, L. K. H. (1997). Classifier use by language-disordered and age matched Cantonese-speaking children. *Asia Pacific Journal of Speech, Language and Hearing*, 2(1), 83-101.

Strong, C. J. (1998). *The Strong narrative assessment procedure*. Eau Claire, WI: Thinking Publications.

Su, J-C. (2000). The acquisition of Chinese connectives by Mandarin-speaking children. PhD Thesis. Boston: Boston University.

Thorndyke, P. M. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.

To, C. K.-S., (2006). *Use of reference in Cantonese narratives: A developmental study*. Unpublished doctoral dissertation. Hong Kong: University of Hong Kong.

To, C. K.-S., Cheung, H. T., & T'sou, B. (2008). Validity and reliability of three syntactic complexity measures in narratives produced by Cantonese-speaking children. Paper presented in the 12th Congress of the International Clinical Phonetics and Linguistics Association, Turkey, Istanbul.

T'sou, B., Lee, T., Tung, P., Chan, A., Man, Y., To, C. (2006). *Hong Kong Cantonese Oral Language Assessment Scale*. Hong Kong: City University of Hong Kong Press.

Van der Lely, H. K. J. (1997). Narrative discourse in grammatical specific language impaired children: a modular language deficit? *Journal of Child Language*, 24, pp. 221-256.

Wallach, P., & Bulter, K. G. (1994). *Language learning disabilities in school-age*

children and adolescents: some principles and applications. New York: Merrill.

Westby, C. E. (1994). The effects of culture on genre, structure, and style of oral and written texts. In G. P. Wallach, & K. G. Bulter (Eds.), *Language learning disabilities in school-age children and adolescents*. USA: Allyn and Bacon.

Westby, C. (1998). Communicative refinement in school age and adolescence. In W. O. Haynes, & B. B. Shulman (Eds.), *Communication development: foundations, processes, and clinical applications* (pp 311-360). Baltimore: Williams & Wilkins.

Wigglesworth, G. (1997). Children's individual approaches to the organization of narrative. *Journal of Child Language*, 24, 279-309.

Wong, A. M-Y. (2001). *Who are you talking about? The development of reference in Cantonese-speaking children with and without specific language impairment*. Unpublished doctoral dissertation. The University of British Columbia.

Wong, A. M-Y., Stokes, S. F., & Fletcher, P. (2003). Collocational diversity in perfective aspect use in Cantonese children with SLI. *Journal of Multilingual Communication Disorders*, 1, 132-140.

Wong, A. M. Y., Au, C., & Stokes, S. F. (2004). Three language measures for Cantonese-speaking children. *Journal of Speech, Language and Hearing Research*, 47(5), 1164-1178.

Wong, A. M-Y., Leonard, L. B., Fletcher, P., & Stokes, S. F. (2004). Questions without movement: A study of Cantonese-speaking children with and without Specific Language Impairment. *Journal of Speech, Language and Hearing Research*, 47, 1440-1453.

Table 1

The Age and Sex Distribution of the Participants in the Normative Study

Grade	Age	Age	Mean	(SD)	Male	Female	Total
	group	range	age		(n=)	(n=)	(N=)
K2	5 yrs	4.8-5.2	5.04	(.10)	66	66	132
K3	5.5 yrs	5.3-5.9	5.56	(.15)	69	68	137
P1	6 yrs	5.8-6.6	6.27	(.19)	66	72	138
P2	7 yrs	6.8-7.6	7.21	(.19)	69	75	144
P3	8 yrs	7.8-8.7	8.27	(.21)	73	71	144
P4	9 yrs	7.9-9.6	9.22	(.25)	73	74	147
P5	10 yrs	8.9-11.1	10.26	(.31)	71	70	141
P6	11 yrs	9.9-12.5	11.26	(.31)	68	69	137
Total					555	565	1120

Table 2

SES Distribution of the Normative Sample

SES	N	% of child sample
Low	170	15.2%
Mid	556	49.8%
High	329	29.5%
Not classified*	65	5.5%
Total	1120	100%

Note. *Not classified refer to those who either did not return the questionnaire or did not answer the item.

Table 3

Summary of SLI Subjects

Grade	Age group	Age Range	(N=)
K3	5.5 yrs	5.3-5.9	2
P1	6 yrs	5.8-6.6	7
P2	7 yrs	6.8-7.6	12
P3	8 yrs	7.8-8.7	8
P4	9 yrs	7.9-9.6	13
P5	10 yrs	8.9-11.1	5
P6	11 yrs	9.9-12.5	3
Total			50

Table 4

Mean and Standard Deviations of the Four Narrative Measures

	Grade							
	K2	K3	P1	P2	P3	P4	P5	P6
Semantic score								
Mean	25.35	30.48	36.46	44.30	50.55	55.26	57.76	60.81
(SD)	(9.01)	(9.92)	(10.63)	(8.98)	(8.54)	(9.13)	(7.43)	(7.21)
Syntactic Complexity								
Mean	6.45	7.80	14.51	19.52	21.43	22.87	24.58	27.22
(SD)	(3.33)	(3.46)	(5.71)	(5.52)	(5.32)	(5.71)	(4.55)	(5.16)
Referencing								
Mean	8.96	11.34	13.42	17.59	19.63	20.65	21.32	22.16
(SD)	(5.01)	(5.33)	(5.68)	(5.59)	(4.00)	(3.82)	(2.95)	(3.07)
Connective								
Mean	4.02	6.16	7.20	9.47	10.88	11.88	11.97	12.73
(SD)	(3.38)	(5.06)	(4.43)	(4.58)	(4.33)	(4.76)	(4.39)	(4.64)

Table 5

Intercorrelations of the Four Narrative Measures

	Semantic score	Syntactic Complexity	Referencing
Semantic score	-		
Syntactic Complexity	.83 ^a	-	
Referencing	.78	.78	-
Connective	.64	.61	.58

Note. All significant at $p < 0.001$. ^aThese are Pearson r values.

Table 6

Descriptive Statistics of the TD and SLI Children on the Four Narrative Measures.

		Semantic score		Syntactic complexity		Referencing		Connective	
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
TD	(N=1080)	0.05	(0.96)	0.05	(0.97)	0.04	(0.97)	0.02	(0.97)
SLI	(N=50)	-1.92	(1.03)	-2.25	(0.80)	-1.54	(1.18)	-0.97	(0.97)

Table 7
Tenth Percentile Cut Points on the Four Narrative Measures for Each Grade.

Grade	Semantic score	Syntactic complexity	Referencing	Connective
K3	19	4	5	1
P1	24	8	7	2
P2	35	12	10	4
P3	41	15	14	6
P4	43	17	16	6
P5	48	19	18	7
P6	51	21	19	7

Table 8

Sensitivity, Specificity, LR+, and LR- of the Four Narrative Measures based on -1.25SD

	Sensitivity	Specificity	LR+	LR-	DOR
Semantic scores	.76	.89	6.90	0.27	25.56
Syntactic Complexity	.94	.88	7.85	0.07	112.14
Referencing	.66	.88	5.71	0.38	15.03
Connective	.34	.88	2.79	0.75	3.72

Table 9

Sensitivity, Specificity, LR+, and LR- of the Four Narrative Measures based on Cut Points from ROC Curves

	Cut point	Sensitivity	Specificity	LR+	LR-	DOR
Syntactic complexity	-1.40	0.94 (0.84-0.98)	0.90 (0.88-0.92)	9.42 (7.77-11.4)	0.07 (0.22-0.20)	140.60
Semantic score	-1.23	0.86 (0.74-0.93)	0.88 (0.85-0.89)	6.88 (5.67-8.35)	0.16 (0.08-0.32)	43.00
Referencing	-1.16	0.86 (0.74-0.93)	0.70 (0.67-0.73)	2.86 (2.48-3.13)	0.2 (0.10-0.40)	14.30
Connective	-0.39	0.86 (0.74-0.93)	0.60 (0.57-0.63)	2.15 (1.88-2.46)	0.23 (0.12-0.47)	9.20

Appendix A

English Translation of the Model Story

- Pic. 1 One day, a group of boys and girls played on the beach happily.
- Pic. 2 A girl wearing a cap heard some noise.
- Pic. 3 So she used a pair of binoculars to see what was happening. She found out that a cat was about to fall off a tree on the other side of the beach.
- Pic. 4 She asked the children to think of a way to rescue the cat.
- Pic. 5 A very tall boy and a boy with curly hair thought that they were as good at swimming as athletes.
- Pic. 6 So they volunteered to save the cat immediately. The rest of the children who stayed on the beach climbed up on the rails and cheered for them
- Pic. 7 However, since the waves were so strong, even though the boys tried their best, they could not reach the other side.
- Pic. 8 The boys on the rails saw that they were exhausted.
- Pic. 9 Two other boys immediately rescued them. The rest of the children were very worried because the boys looked very pale/exhausted.
- Pic. 10 All of the children were relieved when they saw that the two boys appeared to have recovered.
- Pic. 11 A girl with a ponytail said, "The wave was so fast, even a swimming athlete would not be able to reach the other side."
- Pic. 12 Suddenly, she had an inspiration.
- Pic. 13 She whistled loudly a couple of times since she was sure that if she whistled, it would attract a group of dolphins to come and help them.
- Pic. 14 In a short while, a group of dolphins came and spread out in one line along the shore.
- Pic. 15 Then, the children carefully rode on the dolphins and crossed over to the other side of the beach.
- Pic. 16 After they arrived at the other side, they realized that the cat was trapped tightly between the

branches.

Pic. 17 No matter how hard the cat struggled, it could not get its legs out and shrieked even louder.

Pic. 18 A tall boy assertively said he could climb up the tree to save the cat.

Pic. 19 But because the tree was too tall, he fell off the tree and hurt his buttocks.

Pic. 20 The girl with the cap then suggested that they could piggy-back up to reach the cat.

Pic. 21 First, the fattest boy stayed on the bottom to act as a base. Then the other boys climbed up one by one and carried each other on their shoulders. The girl climbed up to the top.

Pic. 22 When the boys were ready, the girl with the cap saved the cat easily.

Pic. 23 Then the girl with the ponytail helped bandage the cat's wound.

Pic. 24 Finally, all of the children rode the dolphins and left.

Appendix B

Examples for Calculating the Semantic Score

The two examples below illustrate how the semantic score is coded with reference to a preselected list of vocabulary. Each row in the last two columns refers to one piece of information and they differ by the degree of sophistication of the vocabulary.

Example 1:

Picture	Transcription	2 marks	1 mark
Pic. 8	The two older boys <u>found</u> that <u>they cannot swim get there</u>	-	▪ <u>saw</u> , found, knew ▪ exhausted, fatigue, depleted ▪ tired, <u>cannot get</u> <u>there</u>
Pic. 9	Then they <u>rescued them</u> . Other children found that the boys who nearly drowned looked so <u>pale</u> They were very <u>scared/concerned</u>	▪ <u>pulled them to the</u> <u>shore</u> ▪ faint, <u>pale</u> , feeble ▪ worried, concerned, anxious	▪ got them back ▪ did not look good ▪ afraid, <u>scared</u>
Pic. 10	But when they saw that they <u>looked better</u> all the children were then <u>relieved</u>	▪ safe ▪ <u>relieved</u>	▪ <u>better</u> , good ▪ not worried
Pic. 11	A girl who had pigtails then said “The <u>wave</u> was <u>so fast/crashing</u> no one can cross the stream.”	▪ <u>wave</u> , water current, flow, tide ▪ dashing, crashing	▪ stream ▪ <u>fast</u> , quick

Pic. 12	Then she had an <u>idea</u>	▪ inspiration, <u>idea</u>	▪ thought of a way
Pic. 13	She <u>blew a whistle</u> . She thought she can <u>gather (1 mark)</u> a group of <u>dolphins</u> to help them.	▪ <u>blew a whistle</u> ▪ attracted, assembled -	▪ blew ▪ asked, <u>gathered</u> ▪ <u>dolphins</u>

Example 2

Picture	Transcription	2 marks	1 mark
Pic. 20	The girl then <u>said</u> that they could climb up the tree by <u>piggy-pack (2 marks)</u> .	▪ suggested, proposed, thought of an idea ▪ <u>piggy-pack</u>	▪ asked, <u>said</u> ▪ piling up, stacking up, one by one
Pic. 21	First of all, the <u>fat boy stood on the ground</u> to form the base. Then other boys climbed up <u>one at a time</u>	-	▪ <u>the fattest boy</u> <u>acted as the base</u> ▪ <u>one-by-one</u>
Pic. 22	All the boys got ready now. Then the girl with a cap <u>got</u> the cat down.	-	▪ easily, without difficulty, with no trouble ▪ saved, <u>got</u> , rescued, held
Pic. 23	Then the girl with pigtails <u>tied</u> the cat's leg with a cloth.	▪ bandage	▪ <u>tied</u> , wrapped ▪ wound, injury,

sore

Pic. 24 Finally, all the children rode on
the dolphins and went home.

- rode, sat,
 - dolphins,
 - left, went home,
went back
-

Appendix C

Examples of the Seven Syntactic Structures Used in the Pilot Study

- (i) Relative clause (RC) is a form of subordinate clause serving to modify a noun phrase that occurs before the noun phrase.

Daai3mou2 go2go3 neoi5zai2 teng1dou2 di1 seng1

Wear cap that CL(classifier) girl hear vppt(verb particle) CL noise

“The girl who wore a cap heard some noise.”

- (ii) Preverbal manner modifier (MM) refers to the construction or phrases that describe the manner of an action expressed in the verb. It occurs before the verb and is often marked by an adverbial ending *gam2*.

keoi5 hou2 jung6lek6 gam2 ceoi1 zo2 jat1 haa5 hau2saau3.

3rd very use-energy this-way blow ASP one CL whistle

“She blew the whistle hard.”

- (iii) Clausal complement (CC) in VO refers to a clausal structure embedded in the object position of a sentence.

Keoi5 tai5ji5 cyun4bou6 siu2pang4jau5 heoi3 gau3 zek3maau1

3rd suggest all children go rescue CL cat

“She suggested that all of the children go there and rescue the cat”

- (iv) Prepositional phrase (PP) refers to a phrase which is composed of a preposition (which is underlined) and a noun phrase as the complement.

Keoi5dei6 zoeng1 go2 loeng5 go3 siu2pang4jau5 gau3 faan1soeng5 on6

they ZOENG those two CL children rescue vppt up shore

“They saved them (to the shore).”

- (v) Verb complement is a special grammatical device in the Chinese language to indicate the degree of result through an action and it is often marked with a verb particle (vppt) *dak1* or *dou3*.

go2 loeng5go3 naam4zai2 gin3dou2 keoi5dei6 jau4dou3 gan1pei4lik6 zoen6.

those two CL-boys see vppt they swim vppt utterly-exhausted

“The two boys saw that they swam so long that they were already exhausted.”

- (vi) Serial verb construction is also a unique structure in Chinese syntax, in which two or more verbs with the same subject are juxtaposed without any explicit conjunction.

Go3 neoi3zai2 zai6 daai3 maai4zek3 maau1 lei4hoi1

CL girl then bring vppt CL-cat leave

“The girl then held the cat and left.”

- (vii) Pivotal construction is also a special kind of verbal expression in Chinese. In this construction, there is a pivotal noun phrase (the underlined NP in the following example) that serves as the object of the preceding verb and the subject of the following verb simultaneously (Chao, 1968).

Keoi5dei6 zau6 zik1haak1gau3 keoi5dei6 soeng5on6,

they then immediately save them up-land

“They then saved them to the shore at once.”

Appendix D

Examples for Coding Referencing

The table below illustrates how referent introduction (e.g. Pictures 1, 2, 3 and 5) and switching (e.g. Pictures 22 and 23) were scored. The two dimensions for scoring are contractiveness and linguistic form.

Pic. 1	One day, <u>a group of boys and girls</u> played on the beach happily. (Target: 2 marks)
	<ul style="list-style-type: none"> ● “<u>The boy and the girls</u> were playing happily.” (1 mark: using incorrect form which should be indefinite) ● “<u>A group of girls</u> played on the beach.” (1 mark: missing the mentioning of boys) ● “<u>They</u> were playing.” (0 mark: using incorrect form and loss of contrastiveness)
Pic. 2	<u>A girl who wore</u> a hat hears some noise. (Target: 2 marks)
	<ul style="list-style-type: none"> ● “<u>A girl (/she)</u> heard some noise.” (1 mark: missing the contrastive feature) ● “<u>The girls who wears</u> a hat heard some noise.” (0 mark: incorrect form which should be indefinite) ● “<u>The girl (/she)</u> heard some noise.” (0 mark: missing the contrastive feature)
Pic. 3	So, she used a telescope to see what had happened. On the other side of the beach, there was <u>a cat</u> in the tree and it was about to fall onto the ground. (Target: 1 marks)
	<ul style="list-style-type: none"> ● “<u>The cat</u> was in the tree.” (0 mark: incorrect form which should be indefinite)
Pic. 5	<u>One very tall boy and one boy with curly hair</u> thought that they themselves were as smart as swimming athletes. (Target: 4 marks, 2 for each character)
	<ul style="list-style-type: none"> ● “<u>The tall boy</u> and a boy with curly hair say that they have to save the cat.” (1 mark (tall boy): using incorrect form which should be indefinite) ● “<u>The two boys</u> say that they will save the cat.” (0 mark (tall boy): missing

contrastive features and using incorrect form which should be indefinite)

- “A tall boy and a boy with short hair said that they could save the cat.” (1 mark (boy with curly hair): the feature used to describe the boy is not contrastive).

Pic. 22 When the boys were ready, the girl wearing a cap saved the cat easily. (Target: 2 marks)

- “Then the girl got the cat back.” (1 mark: missing the contrastive feature)
- “Then a girl got the cat back.” (0 mark: missing the contrastive feature and using incorrect form which should be definite).

Pic. 23 Then the girl who had pigtails helped bandage the wound of the cat. (Target: 2 marks)

- “Then a girl who had pigtails got the cat back.” (1 mark: using incorrect form which should be definite)
 - “Then she helped the cat.” (0 mark: missing the contrastive feature and using incorrect form which should be explicit and not a pronoun)
-

Appendix E

Coding Conventions and Types of Connectives

The frequency of occurrence of each correctly used connective was counted. The temporal connectives *gan1zyu6*, *jin4hau6*, *jin4zi1hau6* “and, then” occurring at the sentence initial position were excluded. Examples below show the use of connectives which are underlined.

- Example of causality connectives

jan1wai6 seoi2lau4 hou2gap1, so2ji3 keoi5dei6 dim2 dou1jau4m4 dou2 gwo3 heoi3. (2 points)

“Since the wave was too fast, (so) they couldn’t swim and cross (the stream).”

- Example of concessive connectives

seoi1jin3 keoi5dei6 hou2 jung6 lik6 jau4, daan6hai6 dou1 jau4 m4 dou2 gwo3 ho4. (2 points)

“Although they used a lot of energy to swim to cross the river, (but) they still failed to reached the other side.”

- Example of hypothetical connectives

mou4loen4 zek3maau1 dim2 zang1zaat3, dou1mang1m4faan1zek3 goek3 ceot1lai4. (2 points)

“No matter how the cat struggled, it still couldn’t get its leg out.”

- Example of coordinating/ additive connectives

keoi5dei6 m4daan1zi2 seng2 faan1, ji4ce2 min6sik1 zung6 hou2hou2. (2 points)

“They not only awoke from faint, but they looked good.”

- Example of temporal connectives

sau2sin1 keoi5 kei5 hai2 haa6min6, jin4hau6 kei4taa1 jan4 zau6 dap6 soeng5 heoi3, zeoi3 hau6 keoi5 zau6 gau3 dou2 zek3maau1.

“Firstly, they were at the base, then they piled up themselves. Finally, she could get the cat.”