

Who does What to Whom:

Introduction of Referents in Children's Storytelling from Pictures

Phyllis Schneider and Denyse Hayward

University of Alberta

Running head: Introduction of referents in children's storytelling

This paper has been published. If citing directly from the paper, please see the journal for the correct page numbers to cite. The reference for this paper is:

Schneider, P., & Hayward, D. (2010). Who does what to whom: Introduction of referents in children's storytelling from pictures. *Language, Speech, and Hearing Services in Schools, 41*, 459-473.

Contact information:

Phyllis Schneider, Ph.D.
Dept. of Communication Sciences and Disorders
University of Alberta
2-70 Corbett Hall
Edmonton, AB T6G 0G2
CANADA
Email: phyllis.schneider@ualberta.ca

Abstract

Purpose: This paper describes the development of a measure, First Mentions, that can be used to evaluate the referring expressions that children use to introduce characters and objects when telling a story. *Method:* Participants were 377 children aged 4-9 (300 typically developing, 77 with language impairment) who told stories while viewing six picture sets. Their first mentions of eight characters and six objects were scored as fully adequate, partially adequate, inadequate, or not mentioned. *Results:* There were significant differences for age and language group, as well as a significant interaction. Within each age group except age 9, typically developing children attained higher scores than children with language impairment. *Conclusions:* These results suggest that the First Mentions measure is a useful tool for identifying problems with referential cohesion.

Story tasks have become a common feature of clinical assessment and intervention in the field of speech-language pathology, for many good reasons. Stories are a part of everyday communication, at home and in school. They can provide a more holistic context of language than most tests of language, which assess the use of words and sentences in isolation, because stories require children to combine words and sentences for a particular purpose (Schneider, Hayward, & Dubé, 2006). Thus they provide information about how well children can use their discrete language skills to communicate. Oral stories are considered to be a form of literate language and to serve as a bridge between oral and written language styles (Westby, 1999). Support for this claim comes from several studies finding that, unlike conversation, children's stories had linguistic complexity characteristic of written language (MacLachlan & Chapman, 1988; Wagner, Nettelbladt, Sahlén, & Nilholm, 2000; Westerveld, Gillon, & Miller, 2004).

To be clinically useful, tasks and materials used for story assessment need to have normative information associated with them. Without normative information, it is impossible to determine with certainty whether a particular child is telling stories as we should expect for the child's age.

Our purpose in this paper is to present a scoring system for evaluating the introduction of referents in stories. The stories were elicited from children using the Edmonton Narrative Norms Instrument (ENNI; Schneider, Dubé, & Hayward, 2009). We will first review the notion of referential cohesion and how it has been measured in previous studies. We will then describe our scoring system for evaluating referent

introduction, First Mentions, and present results using the system on data from the ENNI normative sample.

Cohesion

The term *cohesion* refers to the use of various linguistic means to link utterances together into a unitary text (Hickmann & Schneider, 2000). Two aspects of cohesion that have been extensively studied in recent years are *conjunctive* or *connective cohesion*, which is the use of connectives to tie clauses and sentences together within a text (e.g., The girl went into the cave. *But then* she ran out.), and *referential cohesion*, in which text is connected by linked references within the text (e.g., *the girl...she*). Cohesion is pertinent to any type of discourse or text, including conversation, stories, and expository text. In this paper, we will focus on cohesion in the context of stories.

Referring expressions are linguistic forms used to refer to *referents* such as animate beings (*an elephant, Ella, she*), objects (*a toy, it*), and other entities such as places (*the park, there*) and concepts (*an idea*). Appropriate use of referring expressions in discourse contributes to referential cohesion. When a story is told, referring expressions must be used to mention referents for the first time in such a way that the listener understands that they are new to the story. Thereafter, referring expressions are used to continue to refer to the referents in a way that allows the listener to recognize them as the same referents introduced earlier. Referring expressions can be considered adequate if they are appropriate for the listener's knowledge, shared physical context, and the preceding linguistic context. For example, an indefinite noun phrase such as *an elephant* is adequate for introducing a new character in a story in the absence of a shared physical context, because indefinite articles signal that the referent is not known

to the listener (Givón, 1992). In contrast, *the elephant* or *she* would only be adequate for mentioning the character later on in the story, or if the referent (or a representation of it) were in the context shared by the speaker and listener, because definite forms signal that referents are expected to be known to the listener. The following is an example of adequate first and second mention of a character and an object in the absence of a shared context:

An elephant was bouncing a ball. She accidentally dropped it.

The establishment of referents in discourse when there is no shared context is termed *endophora*, which is reference accomplished through language, as shown in the above example. The alternative way to refer is by using expressions that 'point' directly to referents in the extralinguistic context, which is termed *exophora*. The following example illustrates the difference in adequate reference when the extralinguistic context is shared:

(Context: adult says to child as they look at a picture in a book of an elephant bouncing a ball) Look, *she's* bouncing *it!*

In this example of exophoric reference, the adult is able to point directly to the shared extralinguistic context with the referring expressions and thus can use the pronouns *she* and *it*. The references are understood because the child can refer to the context and relate the referring expressions to the pronouns. In contrast, endophoric reference essentially creates the referents linguistically for the listener (e.g., *There is an elephant bouncing a ball*). Older children (age 11 and up) and adults tend to use endophoric reference even when the context is shared, for example, when pictures are visible to the listener (Kail & Hickmann, 1992).

Reference introduction is an important aspect of all types of connected discourse, from casual conversation to more planned and tightly structured forms such as narratives and expository text. Regardless of the discourse type, it is necessary to introduce referents in a way that is comprehensible to the audience. The current paper focuses on reference introduction in stories told by children.

Development of referent introduction

Children's ability to introduce and maintain referents in narratives effectively develops gradually throughout the early school years (e.g., Hickmann, 1991, 1997, 2003; Kail & Hickmann, 1992; Karmiloff-Smith, 1987; Peterson, 1993; Schneider & Dubé, 1997; Villaume, 1988; Warden, 1981; Wigglesworth, 1990). Young children frequently introduce referents in a confusing way, tending to use referring expressions that are exophoric, even when the listener does not have access to the extralinguistic context and thus cannot understand exophoric referents (Kail & Hickmann, 1992). In referent introductions, young children often use some pronouns and definite noun phrases, forms not generally adequate for first mentions of referents, which suggests that they are using them to refer directly to the extralinguistic context (i.e., exophorically) rather than endophorically (Hickmann, 1995; Karmiloff-Smith, 1987). Thus, the young child might introduce referents in an utterance such as "She's bouncing it" in the absence of a shared context. Note that the issue here is not mastery of particular forms such as pronouns and noun phrases; young children can use these forms adequately in less complex contexts (Hickmann, 1995; Wittek & Tomasello, 2005). Rather, the young child struggles with the appropriate choice of these forms in the context of extended

discourse, in which the current knowledge state of the listener must be constantly monitored (Kail & Hickmann, 1992).

Kindergarten children introduce referents adequately more often when retelling a fictional story that they had just heard than when formulating the story themselves from pictures for a listener who cannot see the pictures (Schneider & Dubé, 1997), suggesting that when they must choose referents themselves based on listener knowledge, they tend to use less adequate means. Pratt and MacKenzie-Keating (1985) obtained similar results with videotaped versus orally presented fictional stories – children in grades 1 and 3 made higher proportions of referential errors in their retellings after viewing videotapes of puppets presenting the story through dialogue than they had after listening to a story. Even when telling stories about personal experiences, which is believed to be an easier task for children than fictional stories (McCabe, 1996), Peterson & Dodsworth (1991) found that one in five referents was not adequately introduced by young preschool-aged children. Younger school-aged children (age 7) continue to introduce referents inadequately, using definite forms to a greater extent than older children; after age 9 or so, children are introducing characters in simple stories in ways similar to adults (Vion & Colas, 1998). Studies have found similar results in a number of languages (e.g., English, French, German, Mandarin), with minor variations in development due to language-specific means of referring (Hickmann & Hendriks, 1999; Hickmann, Hendriks, Roland, & Liang, 1996; Wong & Johnston, 2002).

Young children also use some inadequate referring expressions when referring to characters after the first mentions. However, preschool and young school-aged children have more difficulty with first mentions of referents than with subsequent mentions (Pratt

& MacKenzie-Keating, 1985; Schneider & Dubé, 1997), at least with simple stories (i.e., relatively short stories with few characters). Difficulty in subsequent mentions is likely to vary with story complexity. When telling short, simple stories, it may not be necessary to refer to characters more than once or twice after the referent introduction. The subsequent reference will vary in difficulty not only in story length but in complexity of the character set; for example, multiple male characters can be introduced as "another boy" but then may need to be differentiated as "the boy who came second" or "the guy with the hat" in subsequent mentions, while in a simple story with one character of each gender, characters could be referred to subsequently as "he" and "she". An additional factor is the way the speaker chooses to tell the story; one speaker may limit subsequent references to characters and objects and make them adequate while another tries to tell a more complete story and thus has more difficulty keeping track of referents when referring to them. With first mentions, we can expect speakers to refer to each character and object once; with subsequent mentions, speakers will vary in whether and how often they refer to each established referent.

Studies of reference and cohesion by children with language impairment

Studies of the use of cohesion markers have used a number of approaches to quantify differences between children with and without language impairment. Some studies have looked at types of cohesive markers whereas others have considered adequacy of markers, that is, whether or not the markers contributed to cohesion. These differences in approaches are related to whether the studies found differences between groups, and also result in differences in what can be inferred about children's skills in this

area. Thus we will review these studies with a focus on the way cohesion was measured in each.

Liles (1985a, 1985b) used some of Halliday and Hasan's (1976) categories of cohesion (Personal Reference, Demonstrative Reference, Conjunction, and Lexical) and compared their relative occurrence in older children with and without language disorders. She found that children with typical language used more Personal Reference ties, whereas the children with language disorders used more Demonstrative and Lexical ties. However, it is not clear that choice of type of cohesion is related to quality or comprehensibility of narrative discourse; it is likely that different combinations of cohesion types could be used to tell stories that would be judged as equally "good," cohesive stories.

Other studies have counted the number and/or rate of all types of cohesive ties used, including referring expressions and conjunctions (such as *and*, *then*, *because*, *since*, etc.) (Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001; Liles, 1985a, 1985b; Strong & Shaver, 1991; Vallance, Im, & Cohen, 1999), and found differences between children with and without language impairment. However, as pointed out by Scott (1988), frequencies and relative distributions of cohesive ties have often failed to differentiate children with and without language impairment. Again, it is not clear that number or rate of all cohesive ties is related to quality of story. Additionally, combining different types of cohesive ties into a single count obscures important differences among them. For example, it is impossible to tell a comprehensible story without referring to people and/or objects, but it is perfectly possible to tell an understandable story with few or no conjunctions.

Studies have also examined number and/or rate of 'incomplete/erroneous' cohesive ties (Girolametto et al., 2001; Liles, 1985a, 1985b; Paul, Hernandez, Taylor, & Johnson, 1996; Paul & Smith, 1993; Schneider, 1996). An incomplete reference tie is one that refers back to a referent that was never introduced, while an erroneous reference tie is one that points to the wrong referent (e.g., *he* instead of *she*) or that has several possible antecedents (*he* when there were several males mentioned previously). An erroneous conjunction is one that expresses a relation between sentences that is not appropriate for the context (e.g., "He liked it. *But* he was happy"). The studies that compared children with and without language impairment have found greater numbers of incomplete and erroneous ties used by children with language impairment. In a factor analysis of data from several previous studies, Liles, Duffy, Merritt, and Purcell (1995) reported that their measure of complete cohesive ties contributed to a factor that predicted group membership (typical development or language impairment). Once again, however, counting different types of cohesive ties together obscures some important differences. Use of inadequate (i.e., incomplete or erroneous) reference ties can make a story very confusing, and is typically a major reason for difficulty in understanding a young child's story unless the listener already knows the story. Misleading conjunction tie use does not appear to happen as frequently as misleading reference, and in fact it is common for young children to simply leave out conjunctive ties or to use simple additive ones (e.g., *and*, *then*), which continue to be used in similar frequencies at least through age 9 (Peterson & McCabe, 1987). It is not incorrect to leave out a conjunction between clauses, and speakers have a range of choices regarding when and how to connect clauses with conjunctions. Thus, the use of

referential and conjunctive ties will have different impacts on perceived story quality. In addition, for purposes of treatment planning, because intervention for referential cohesion would be conducted in a different manner than intervention for conjunctive cohesion, it would be helpful to know more precisely in which area a child was having difficulty. Thus a measure focusing on referential ties separately would be potentially useful for clinical purposes.

On the other hand, it is possible to restrict the range of cohesive devices too narrowly by focusing on particular linguistic forms, thereby missing the overall picture of referential cohesion. Finestack, Fey, & Catts (2006) analyzed pronoun use, including percentage of complete (i.e., adequate) pronoun reference, by children in Grades 2 and 4 and found no differences between children with and without language impairment. In a study of children aged 8-11 with fetal alcohol spectrum disorder (FASD), Thorne, Coggins, Olson, & Astley (2007) analyzed stories for pronominal and nominal reference separately and found that rate of ambiguous nominal reference discriminated between FASD and age-matched children. These studies focused on the use of one or two particular linguistic forms that can be used to refer in discourse. However, referential cohesion is best considered in terms of function rather than particular linguistic forms. For example, a speaker might choose to refer to a previously introduced character with a pronoun, a definite noun phrase, or a proper name; all three expressions could be equally adequate in some contexts, while not all would be adequate in others. To capture the referential ability of an individual, the focus needs to be on whether a form adequate for a particular point in a narrative is selected from among the forms available to the speaker. Mastery of referential cohesion is the ability to introduce and maintain

referents in a comprehensible way in a discourse context, rather than the mastery of individual linguistic forms.

Other research studies have looked at referential cohesion separately but comprehensively, focusing on percentage of referential cohesive ties that are adequate (i.e., not incomplete or erroneous) (Boudreau & Chapman, 2000; Klecan-Aker, 1985; Norbury & Bishop, 2003; Strong, 1998). These studies found differences on referential adequacy scores between children with and without language impairment (and Down syndrome in the case of Boudreau & Chapman). This method of analysing referential cohesion is commonly used in research on typically developing children (Hickmann, 1991; Hickmann, Hendriks, Roland, & Jiang, 1996; Schneider & Dubé, 1997; Tsai & Chang, 2008). Using this method, any referring expression that is not fully adequate for its occurrence in a story is considered inadequate, and the percentage of referential adequacy is calculated as number of adequate referring expressions divided by total number of referring expressions. Frequently, adequacy of referring expressions used to introduce characters is assessed separately from adequacy of those used for subsequent mentions (as in Schneider & Dubé, 1997). The use of percentage of adequate reference has been recommended as a part of narrative assessment (Hughes, McGillivray, & Schmidek, 1997; Strong, 1998).

Referential adequacy measures focus on the function of referring expressions in context and thus have advantages over other measures that either combine referential and other cohesion types or focus on a single form type. However, in grouping all expressions that are not fully adequate as "inadequate," referential adequacy measures fail to make some important distinctions in degree of less-than-adequate reference. In

measures of adequacy that involve percentages, introductions with definite noun phrases (e.g., *the elephant*) are typically included with inadequate expressions such as pronouns because both wrongly signal to the listener that the referent should already be known. However, the nouns in definite noun phrases at least allow the listener to understand what is being referred to, even if the determiner may confuse the listener as to whether the referent is new or not. Thus to investigate children's growing competence with referent introduction, it would be useful to distinguish between partially adequate and fully inadequate expressions. In the current study, we decided to develop a scoring system that would incorporate such distinctions.

Another limitation of the referential adequacy approach is that it is typically calculated based on the number of referents a child chooses to mention. The result is that some children will receive a referential adequacy score based on a smaller number of referents than another child. The child with a larger number of attempted referents could have more problems with adequacy simply because the child is attempting to tell a more complete story, whereas the child who attempts only a small number of referents may achieve a high referential adequacy score and will thus appear more competent than the first child. A measure that controlled for number of referents would be likely to better distinguish between these children.

To date there is no normed narrative instrument that includes a measure of referential cohesion. The likely reason is the fact that it is difficult to specify the rules for determining adequacy of subsequent mentions. Adequacy of subsequent mentions depends on the length of a story and the number and order of referents mentioned. However, the rules for adequate first mentions are more straightforward than for

subsequent mentions. In addition, if subsequent mentions are included, the number of references to each character and object will vary from child to child, as well as the number of characters and objects referred to at all. If analysis is restricted to first mentions, the analysis can include the same set of referents for all children. For this reason, and because first mentions appeared to discriminate well among age and language groups (Schneider & Dubé, 1997; Schneider & Hayward, 2004), we decided to develop a scoring for first mentions as our measure of referential cohesion.

Referential adequacy can vary greatly depending on the complexity of the narrative stimuli used to elicit stories. For example, using three of Mercer Mayer's Frog stories as stimuli with typically developing 8-10 year olds, Strong (1998) obtained mean percentages of referential cohesion errors varying from 6% to 27%. The variation appears to be due to story differences, with the story having the most characters (*Frog Goes to Dinner*; Mayer, 1974) having the highest mean percentage of problem reference. Thus we felt it was very important to use stories that were controlled for number and type of referents.

The current study investigated whether a First Mentions score would be a useful measure of the development of cohesion in storytelling from ages 4 to 9, using stories controlled for number and type of referents. We also wanted to determine whether and to what degree First Mention scores would differentiate between groups of children with and without previously identified language impairment. Since the goal of this paper was to investigate the usefulness of the First Mention analysis for assessment purposes, the research questions focused on examining developmental trends and group differences. The research questions were:

1. Are there significant effects for age and language group in the First Mentions scores?
2. Are there differences in First Mention scores between children with and without language impairment within each age group?
3. How do results of the First Mention analysis compare to results using a referential adequacy measure?

Methods

Participants

Participants were 377 Edmonton children aged 4 through 9 whose stories form the database of the Edmonton Narrative Norms Instrument (ENNI). Within each one-year interval there were 50 typically developing children (25 boys and 25 girls) and a smaller sample of children with language impairment. Prevalence of specific language impairment (SLI) has been estimated at 7.4% of the population (Tomblin, Records, Buckwalter, Zhang, Smith, & O'Brien, 1998). The population of children with language impairment is a very heterogeneous group (Conti-Ramsden, Crutchley, & Botting, 1997). If children with language impairment had been added to the normative sample for the ENNI, there would be only 4 children per age group with language impairment, and they would be unlikely to be sufficiently representative of the range of impairment that can be found in the population. For the purpose of calculating norms for the ENNI, group of children with language impairment was oversampled (i.e., sampled at more than 7.4% of the total), with subsequent weighting of subsample data so that data from the children with language impairment made up 7.4% of the total sample. Oversampling and weighting were done to assure as representative a subsample as possible without overrepresenting children with language impairment. For the analyses reported here,

the unweighted data were used. The target sample for children with language impairment was 15 per age group; the obtained sample varies from 10 to 17 children per age group. Gender was left to vary in this group. As expected, there were more boys than girls (48 of 77 – 62%) in the group with language impairment. Sample information is summarized in Table 1. All children in the ENNI sample were included in the current study.

Place Table 1 about here

Children in the school-age range were chosen from children attending Kindergarten through Grade 4 in Edmonton public and separate schools. The younger children were chosen from those attending preschools, daycare centres, and Kindergarten programs in Edmonton. Schools were randomly selected from areas across Edmonton to assure a cross-section of socioeconomic groups. In all, 34 schools and 13 daycares, preschools and independent Kindergarten programs were visited to collect the data. All participants spoke English as a first language at home; information about other languages spoken in the home was not collected.

To identify potential typically developing children for the study, teachers were asked to refer two children in the upper level of achievement, two children from the middle level, and two children in the lower level (one boy and one girl at each level). In all cases, the children who were referred for the typically developing sample were not to have speech or language difficulties or any other diagnostic label such as attention deficit disorder, learning disability, or autism. Information and parental consent forms

were sent to the homes of children referred by the teachers; children whose parents returned the forms were included in the study.

The subsample of children with language impairment was obtained with the cooperation of 3 sites: a public school serving children with language/learning disabilities; a rehabilitation hospital that has several programs for children with language impairment; and Capital Health Authority, which served preschool and school-aged children throughout the city. Children could be referred if they had fine or gross motor delays, attention deficit disorder with or without hyperactivity (ADD/ADHD) with medication, a diagnosed learning disability, or mild or moderate speech disorder, in addition to language impairment. (Information regarding concomitant conditions of children referred to the study was not collected.) Sites were asked not to refer children who had received diagnoses of mental retardation, ADD/ADHD without medication, autism, hearing impairment, severe visual impairment that would result in inability to see pictures even with correction, or severe speech impairment that would preclude accurate orthographic transcription of their stories. IQ test information was not collected; the speech-language pathologists referring children for the study were asked to refer children for whom they had no concerns regarding cognitive abilities. Because we do not have IQ scores to confirm that our participants were in the normal range of cognitive ability, we will refer to our participants in this group as having language impairment rather than specific language impairment.

Demographic information was collected on the families of participating children to permit description of socioeconomic status and ethnic composition of the sample. The purpose of collecting demographic information was to ensure a sample representative of

the Edmonton population. Socioeconomic status was estimated from parents' occupations using Blishen, Carroll and Moore's (1987) Socioeconomic Index for Occupations in Canada. Based on Canadian census information, this index reflects equally weighted components of education and income level by occupation. Scores of the index range from 17.81 (newspaper carriers and vendors) to 101.74 (dentists) with a mean for Canada of 42.74 (SD = 13.28). Socioeconomic index (SES) information is reported for each age and language group in Table 1. Blishen SES scores were compared across age and language groups using an analysis of variance. Results revealed no significant differences in SES for age group, $F(5, 358) = .99, p < .43$, or language group, $F(1, 358) = 1.84, p < .18$.

Ethnic information was collected on a checklist based on Statistics Canada categories of visible minorities; the ethnic composition of the sample corresponded closely to the range of ethnic diversity in the city of Edmonton according to Statistics Canada data (Statistics Canada, no date). Ethnic information for the sample, the city of Edmonton, and Canada is displayed in Table 2. Data collection was conducted throughout the school year, with care taken to collect data from the full age range throughout the year so that no one age group was sampled at a different point in the school year than another age group.

Insert Table 2 here

All children were tested on two subtests of the Clinical Test of Language Fundamentals (CELF), using either the CELF-Preschool (Wiig, Secord, & Semel, 1992)

for children under 6 years of age or the CELF-III (Semel, Wiig, & Secord, 1995) for children age aged 6 and over. Subtests from the CELF-P were Linguistic Concepts and Recalling Sentences in Context; these subtests were chosen because they are recommended for use in screening by the test authors (Wiig, Secord, & Semel, 1992). Subtests from the CELF-III were Concepts and Directions and Recalling Sentences, which are analogous to the CELF-P subtests used. Means for the typically developing and specific language impairment groups are reported in Table 3. The purpose of administering the CELF subtests was to obtain language information for the children with typical development. In addition, 160 participants in the typical development group (29%) and all of the participants in the language impairment group were given the full CELF-R or CELF-III. CELF data are reported in Table 4. Collection of full CELF data permitted calculation of correlations as a measure of concurrent validity and yielded comparable information across the group of children with language impairment, who presumably would have been tested on a variety of language instruments for their initial diagnosis. In addition to the 77 children in the language impairment group, there were 19 children referred to the study as having language impairment who attained a score of 85 or above on both the receptive and expressive language total score of the CELF-P or CELF-III; these children were excluded from the sample. The decision was made to exclude these children because we did not have other information to back up the referral information indicating language impairment. Children in the group with typical development were not excluded on the basis of CELF score, because it is possible for a child with functional language skills to score below the normal range on a standardized test. Additionally, only the two subtest scores were available for most of the children in

the TD group, which would not be adequate for identifying language impairment without supporting information. Finally, eliminating the children from the TD group who had the lowest CELF scores would potentially bias the sample in the direction of greater differences between the groups on the ENNI. Consequently, some of the children in the typical language group had subtest scores below one standard deviation on the CELF-P or CELF-III. Specifically, 19 children (from 1 to 6 per age group) had standard scores below 7 on the Linguistic Concepts (CELF-P) or Concepts and Directions (CELF-III) subtest, 11 children (1-3 per age group) had scores below 7 on Recalling Sentences in Context (CELF-P) or Recalling Sentences (CELF-III), and 6 children (0-2 per age group) had scores lower than 7 on both subtests. Of the 160 children in the typical language group who were given the full CELF, 5 (from 0-3 per age group) had a Receptive Language composite score below 85, 3 (0-2 per group) had an Expressive Language composite score below 85, and 2 (both 7-year-olds) had a Total Language score below 85.

Insert Table 3 here

Insert Table 4 here

Materials

Stimuli for the current study were the two story picture sets of the ENNI, originally developed by Dubé (2000). The stories were created according to story grammar principles (Glenn & Stein, 1979) and contain information considered to be essential to good stories. Each story set has two main animal characters (different species), a young

male and a young female, introduced in the first story in the set (5 pages long, single basic episode); these characters appear in all of the stories in their set. The second story of 8 pages (two episodes) introduces a third character who is an adult animal (the same type of animal as one of the main characters), and the third story (13 pages, 3 episodes) introduces a fourth character in addition to the previous three (another adult of the same type of animal as the third character, opposite gender). Thus the stories increase in referential difficulty; the first two animals can be distinguished in a number of simple ways (e.g., gender, type of animal), while the later characters are more difficult to differentiate when referring to them.

Each story page was put into its own plastic page protector, and each story was put into its own binder, permitting each page to be presented separately. The pictures in which the targeted characters and objects first appeared are provided in Appendix A. The full picture sets may be viewed on and downloaded from the ENNI website (Schneider, Dubé, & Hayward, 2009).

As the ENNI was designed as a storytelling task rather than a retell task, no verbal story is presented to children; rather, children are asked to tell the story from the pictures to the examiner.

Procedure

Data were collected by three female research assistants. Children were seen in their school or preschool settings. The child was first given a training story consisting of a single episode story in 5 pictures with a main character (a boy) and a minor character (a man). The purpose of the training story was to familiarize the child with the procedure and to allow the examiner to give more explicit prompts if the child was having difficulty

with the task, such as providing the story beginning (e.g., “Once upon a time ... there was a ...”) or encouraging the child to go beyond labelling (“You’ve told me who is in the pictures – now can you tell me a STORY about the pictures?”). After the training story was administered, the two story sets were given. Administration of the story sets was counterbalanced, with half of the children telling stories from Set A first and the other half telling stories from Set B first. For the sets A and B stories, the examiner was restricted to less explicit assistance than in the training story such as general encouragement, repetition of the child’s previous utterance, or if the child did not say anything, a request to tell what was happening in the story.

Each child was presented with each story, one page at a time, before telling the story. Then the child was again presented with the stories page by page and was asked to tell the story to the examiner, who held the story binder in such a way that she could not see the pictures. The child was reminded before each story that the examiner would not be able to see the pictures. Inability of the examiner to see the pictures was established so that pointing or other exophoric reference would not be adequate for indicating referents and children would need to refer endophorically for adequate introduction of referents. The procedure was repeated for each of the six stories, with a break between the two story sets. Stories were audiorecorded using JVC digital minidisk recorders.

Children's story retellings were transcribed in full using the CHAT transcription system from the CHILDES database (MacWhinney, 2000). The CHILDES database is a collection of transcripts from many researchers of primarily children’s language samples in a number of languages. CHILDES also provides a system for analysing transcripts

using the CLAN program, which was used for the analyses of storytelling described below. All transcripts were checked against the recordings by the first author before being analysed. A research assistant transcribed 5% of the checked stories for transcription reliability purposes; word-by-word reliability was calculated to be 96.5%.

Development of the First Mentions scoring

To develop the First Mentions scoring, the introduction to all 8 characters and 6 story objects was examined. Preliminary analyses of stories indicated that some objects such as *picnic basket* were mentioned less consistently than other objects, and in fact could be left out without making the story deficient. For example, a child could mention that the characters in the second story of set B were "having a picnic", without mentioning a picnic basket, while in the complex story of set B, it would not be possible to tell the story as depicted without mentioned the balloon that one of the characters lets loose. To ensure that the target referents were those that were likely to be mentioned, we chose referents that were mentioned by the majority of the oldest participant groups. The 6 objects selected for the analysis (3 from each story set) were each mentioned by 98-100% of the 8- and 9-year-olds in the typical development group, as were all 8 story characters.

A scoring system was developed in which a score of 3 indicated a fully adequate referring expression for its context (e.g., indefinite determiner plus noun, as in 'There was *an elephant* bouncing *a ball*'; name, as in '*Ella the elephant*'; possessive pronoun plus noun when the possessor has been introduced, as in '*her ball*'). A score of 2 indicated a less than adequate expression that was still partially informative (e.g., definite determiner plus noun, as in '*The elephant* was bouncing *the ball*'). A score of 1 indicated

an inadequate referring expression (e.g., personal or demonstrative pronoun, as in '*she* was bouncing *that*'; use of definite determiner with a noun that had been used for a previously-introduced character, as in "*the elephant*" for the third character). Referents that were omitted altogether received a score of 0. The point system was developed to provide a more graduated scoring than systems using an adequate/inadequate dichotomy, such as the referential adequacy scoring described earlier. Scoring was not dependent on the use of a particular term; for example, the giraffe character could be referred to as *horse*, *zebra*, or *cow*, and the ball could be referred to as *balloon* or *egg*. Scoring was dependent on the appropriateness of the linguistic form for first mention (indefinite or definite determiner, pronoun, etc.). The complete FM scoring system is available on the ENNI website (Schneider, Dubé, & Hayward, 2009), including examples of scoring for 1, 2, and 3 points for each of the 14 target referents. Examples of scoring criteria for three referents are provided in Appendix B.

Each child's First Mention score was the total number of points awarded for the 14 targeted referents. For example, if a child received scores of 3 for 6 referents, scores of 2 for 3 referents, and scores of 1 for 3 referents, omitting the last 2 referents, the child would receive a total First Mentions score of 27. Maximum possible score was 42.

For purposes of comparison with scoring used in previous studies, we also calculated a referential adequacy score for the 14 referents used in the FM scoring, as follows: The number of adequate referring expressions (i.e., those scored as 3 in the FM scoring) was divided by the total number of the 14 referents mentioned by each child. The number of referents mentioned by each child was used rather than the total number used for the FM score (14) because referential adequacy scores are typically calculated

on referents actually mentioned by each individual rather than a predetermined number of potential referents. Note that, to keep the FM and referential adequacy scoring comparable, we included only the 14 referents in the referential adequacy scoring, rather than scoring all referents mentioned by the child, as is common in previous research. For example, the child in the previous example referred to 12 of the 14 referents and used an adequate referring expression for 6 of them, and thus the child's referential adequacy score would be 6/12 or 50%. Note that this would result in the same referential adequacy score for children who mention 7 of 14 referents adequately as those who mention 2 of 4 adequately. This is the procedure typically used in studies of referential adequacy (and in fact is one of the limitations of the measure, as discussed earlier).

Data Scoring and Reliability

The first author scored the transcripts using the First Mentions scoring system. To check scoring reliability, the second author scored 20% of the transcripts (entire story sets from 20% of the children, randomly chosen). Cohen's kappa was computed; this statistic takes into account differences between scorers on each item as well as the probability of agreement by chance on individual items, and thus is considered a more rigorous way to calculate reliability than point-to-point reliability for multi-item scoring systems (Bakeman & Gottmann, 1986). A Cohen's kappa of .85 was obtained, indicating excellent reliability (Landis & Koch, 1977).

Results

An analysis of variance was used to investigate the first research question, *Are there significant effects for age and language group in the First Mentions scores?*

Means and standard deviations for the First Mentions data are displayed in Figure 1 by age and language group. There was a significant main effect for age group, $F(5, 376) = 54.48, p < .001, \text{partial } \eta^2 = .43$. Because Levene's test of equality of variances yielded a significant difference, Games-Howell post hoc tests were used to look at differences between age groups. The post hoc tests revealed that the 4-year-old age differed from all the other age groups, as did the 5 and 6 year olds; 7, 8 and 9 year olds differed from the younger age groups but did not differ from one another. Thus it appears that the First Mentions measure used with the ENNI stories reveals development in adequacy of referent introduction between the ages of 4 and 7. The effect size indicates that age accounts for about 43% of the variance in scores.

Figure 1 goes about here

The main effect for language group was also significant, $F(1, 376) = 110.91, p < .001, \text{partial } \eta^2 = .25$. There was a significant interaction between age group and language group, $F(5, 376) = 3.09, p = .01, \text{partial } \eta^2 = .04$. Inspection of Figure 1 suggests that the interaction was due to greater differences between the language groups at the younger ages, with the language groups being relatively close by age 9.

The significant interaction between Age and Group suggests that the difference between groups may decrease with age and may no longer exist by age 9. To investigate our research question 2 regarding differences between typically developing children and children with language impairment within each age group, post hoc comparisons were made of the two groups' First Mentions scores within each age group

using Least Squares Difference (LSD) tests with Bonferroni correction for multiple comparisons and corrections for unequal variances where appropriate. The two groups were significantly different within each age group at $p < .0001$, with effect sizes (partial η^2) ranging from .20 to .43, with the exception of the 9-year-old group, in which the two language groups did not differ, $p = .11$, partial $\eta^2 = .12$. These results indicate that the First Mentions measure yields significant differences between children with and without language impairment in the age range of 4 to 8.¹

For purposes of comparison, the same analyses were conducted using the referential adequacy measure (percentage of first mentions that were fully adequate). Means and standard deviations for referential adequacy are illustrated in Figure 2. Results for the ANOVA were similar to those for First Mentions in that there was a main effect for Age Group, $F(5, 376) = 28.34$, $p < .001$, partial $\eta^2 = .28$, and for Language Group, $F(1, 376) = 71.13$, $p < .001$, partial $\eta^2 = .16$. However, there was no significant interaction between Age Group and Language Group, $F(5, 376) = .96$, $p = .44$, partial $\eta^2 = .01$. Post hoc tests with Games-Howell were similar to those for First Mentions except that there was no difference between 5 and 6 year old age groups. As with the First Mentions measure, children in the TD group had higher scores within each age group than the children with LI except in the 9-year-old group. Effect sizes were smaller than in the First Mentions analyses, indicating that less variance was accounted for in the referential adequacy analyses. Thus, while the overall pattern of results was similar for the two measures, the First Mention measure provided a bit more information than the Referential Adequacy percentage measure, revealing an interaction between age group and language status and differences among the younger age groups.²

Insert Figure 2 about here

Discussion

The purpose of this study was to develop a measure of referential cohesion using first mentions of characters and objects in stories told from the ENNI. The measure was applied to data collected for the local normative sample of the ENNI. Results of the scoring were examined to see whether there were differences among age groups. Analysis of variance revealed that scores did increase between the ages of 4 and 7, but did not appear to change from ages 7 to 9. Thus it appears that in these fairly simple stories, children's ability to use adequate referring expressions to introduce characters and objects gradually improves until age 7, when it appears to be mastered by the majority of children. However, the measure captures differences in children of different language abilities beyond age 7. Within-age comparisons of scores from children with and without language impairment revealed that the two groups' First Mentions scores were significantly different in each age group except age 9. In our data, the means for ages 7-9 are very close, with the standard deviation getting smaller over this range. This suggests that the typically developing children had reached a plateau on this measure and in these simple stories, allowing the children with language impairment to catch up by age 9. It is possible that children with language impairment catch up in their referring abilities by age 9, at least when telling stories comparable in complexity to the ENNI stories. However, it must be noted that the 9-year-old sample contained only 10 children with language impairment, and thus had limited power to show a difference in this age

group. Further studies would be needed to address the question of whether and when differences are no longer found using the ENNI FM measure with children with language impairment.

Preliminary data from a related study (Schneider, 2008) compared the First Mentions scores of the ENNI sample of 9-year-old children with typical development to a group of 10 adults aged 25-33. The goal was to explore whether the First Mentions measure would show differences between adults and children in referent introduction. Despite the lack of change between ages 7 and 9 found in the current study, the Schneider (2008) study found a significant difference between the 9-year-olds and adults, with a large effect size (partial eta squared) of .85. These results suggest that development of referent introduction is not completed by age 7 but continues to develop for some time after age 9.

Similar to previous research on cohesion in stories reviewed earlier (e.g., Hickmann, 1991, 1997, 2003; Kail & Hickmann, 1992; Karmiloff-Smith, 1987; Liles, 1985a, 1985b; Peterson, 1993; Schneider & Dubé, 1997), we found that older children achieved higher scores on referent introduction than younger ones, and children with typical development had higher scores than children of the same age with language impairment. The advantages that the First Mentions score has over previous scoring are: it is limited to one type of cohesion, namely, referential; it focuses on adequacy of expressions for cohesion within stories rather than on the mastery of particular linguistic forms; and it facilitates the qualitative analysis of error types for intervention planning.

Examples from our normative data illustrate the difference between FM and referential adequacy measures and make it clearer why FM might be a better measure of

the skill. One 6-year-old child with typical language development introduced 7 of the 14 referents with definite noun phrases (e.g., *the elephant*, *the airplane*) and introduced one with a pronoun; he scored just within one standard deviation of his age group mean with the FM measure (based on 33 of 42 or 79% possible raw score points) but 1.5 standard deviations below the mean with the referential adequacy measure (with a raw referential adequacy score of 43%). His story began, "The elephant and the giraffe play with the ball." The three referents were scored as inadequate (0) in the referential adequacy scoring but were awarded 2 points each in the First mentions scoring. In contrast, consider a 6-year-old child with language impairment who scored higher on referential adequacy than on first mentions. He obtained a score in the normal range for referential adequacy (57%) but his FM score (31/42 or 74%) was 1.24 standard deviations below the mean. He introduced five referents with expressions that were scored 1 (mainly pronouns) and one with a definite determiner that was scored 2. In the referential adequacy scoring, these expressions were scored 0. Note that this child's RA score would be no different than a child who used 6 definite determiners and no pronouns. A story in which referents are introduced with definite noun phrases is easier for listeners to follow than a story with pronouns, and thus the FM scoring appears to reflect the quality of referent introduction better than the referential adequacy scoring.

Thus we feel that the FM measure is a potentially useful one for identifying problems in establishing characters and objects in stories. By evaluating reference to a pre-established set of target characters and objects, the measure avoids a major limitation of measures that are calculated on all referents introduced by children, such as the referential adequacy measure. With a consistent set of referents, it is not possible

for two children to achieve the same score based on different numbers of attempts – children attempting fewer referents will obtain a lower score. In addition, the FM measure takes into account different degrees of referential inadequacy, by awarding more points for referring expressions that provide some information (e.g., definite determiner + noun) than to those providing minimal information (e.g., pronouns).

Clinical implications

The results of this study suggest that the First Mentions measure is a useful tool for assessing an important aspect of narrative ability, namely, referential cohesion. The First Mentions measure could be utilized to determine whether a child is having difficulty with this aspect of storytelling. In conjunction with other measures, such as story content, the clinician could determine the type and overall severity of a child's discourse problems. Working on referring expressions in intervention could be very helpful in making a child's storytelling more comprehensible. The First Mentions scoring system provides information about types of errors that can be used in planning intervention. If a child uses pronouns to introduce referents, one could encourage the child to provide nouns; if the child uses definite noun phrases, the child could be taught to use the indefinite article, and to emphasize that a new referent is being mentioned, the child could learn to use it within introductory phrases such as "Once there was a..." and "then along came a...".

The First Mentions scoring as described here is specific to the ENNI stories. Scoring was based on the particular referents that were pictured in ENNI stimuli. Evaluation of the results is made with reference to the ENNI normative database. It would not be advisable to use in its current form with other story sets unless normative

data were available for those stories as well. However, First Mentions scoring could be easily adapted to other stories by describing criteria for selected referents in the stories and collecting normative data for the appropriate age range. It would also be possible to make the task more appropriate for older children by choice of target referents. For example, characters could be incorporated into stories who were similar in many ways, making it more difficult to distinguish them in referring expressions, as in the film used in studies by Liles (1985a) and in the Pear Stories study (Chafe, 1980), both of which depicted a large number of young male characters. Of course, to make the stories and scoring clinically useful, normative data on the specific stories would need to be collected. Future research could focus on FM scores that can be expected using more referentially complex stories at different ages.

Although the FM scoring as presented here is tied to the ENNI stories, we believe that the principle of attending to type of referential error can be applied more broadly when working with children on storytelling. It is important to note not only the frequency but also the types of errors made by children when introducing characters and objects. As noted earlier, type of error can guide the way that clinicians help children, and may also help in selecting appropriate stories for children (e.g., number of potential confusable characters in a story).

The current study did not investigate whether subtypes of language impairments would be related to adequacy of referent introduction. It is possible that children with receptive language problems might score differently than children with expressive language problems only. Additionally, it would be interesting to investigate how children with other developmental disorders, such as autism spectrum disorder, would score on

the FM measure. Future studies could include samples of children with different patterns of communication impairment.

Future research should be conducted to explore referential ability in the context of other narrative measures. It is expected that measures of a range of narrative ability, including story information, cohesion, and semantic and syntactic measures, would provide a complete picture of a child's competence in narrative contexts. As studies of adults' judgments of stories have shown (McCabe & Peterson, 1984; Schneider & Winship, 2002), different narrative measures contribute to adults' overall assessment of the quality of a story, and no one measure appears to capture everything that contributes to perceptions of story quality. In order to be able to provide effective narrative intervention, it is necessary to have measures that will pinpoint the exact nature of a particular child's difficulty with storytelling. It is important to look at stories from a number of different perspectives in order to characterize a child's narrative ability and to design an intervention that targets areas of difficulty.

References

- Bakeman, R., & Gottman, J. M. (1986). *Observing interaction: An introduction to sequential analysis*. Cambridge, UK: Cambridge University Press.
- Blishen, B. R., Carroll, W. K., & Moore, C. (1987). The 1981 socioeconomic index for occupations in Canada. *Canadian Review of Sociology and Anthropology*, 24, 465-488.
- Boudreau, D. M., & Chapman, R.S. (2000). The relationship between event representation and linguistic skill in narratives of children and adolescents with Down syndrome. *Journal of Speech, Language & Hearing Research*, 43, 1146-1159.
- Brace, N., Kemp, R., & Snelgar, R. (2003). *SPSS for psychologists: A guide to data analysis using SPSS for Windows (2nd Ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chafe, W. L. (1980). *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with SLI. *Journal of Speech, Language & Hearing Research*, 40, 765-777.
- Dubé, R. V. (2000). An instrument for language assessment of Deaf children: American Sign Language and English. Unpublished dissertation, Faculty of Rehabilitation Medicine, University of Alberta.
- Finestack, L. H., Fey, M. E., & Catts, H. W. (2006). Pronominal reference skills of second and fourth grade children with language impairment. *Journal of Communication Disorders*, 39, 232-248.

- 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.
- 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: Advances in discourse processing* (pp. 53-120). Norwood, NJ: Ablex.
- Givón, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics, 30*, 5-56.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longmans Group.
- Hickmann, M. (1991). The development of discourse cohesion: Some functional and cross-linguistic issues. In G. Piéraut-le Bonniec and M. Dolitsky (Eds.), *Language bases...discourse bases: Some aspects of contemporary French-language psycholinguistics research* (pp. 157-185). Amsterdam: John Benjamins.
- Hickmann, M. (1995). Discourse organization and the development of reference to person, space and time. In P. Fletcher and B. MacWhinney (Eds.), *Handbook of child language* (pp. 194-218). Oxford: Blackwell.
- Hickmann, M. (1997). Information status and grounding in children's narratives: A crosslinguistic perspective. In J. Costermans & M. Fayol (Eds.), *Processing interclausal relationships: Studies in the production and comprehension of text* (pp. 221-243). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hickmann, M. (2003). *Children's discourse: Person, space and time across languages*. Cambridge, UK: Cambridge University Press.

- 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.
- Hickmann, M., Hendriks, H., Roland, F., & Liang, J. (1996). The marking of new information in children's narratives: A comparison of English, French, German, and Mandarin Chinese. *Journal of Child Language, 23*, 591-619.
- Hickmann, M., & Schneider, P. (2000). Cohesion and coherence anomalies and their effects on children's referent introduction in narrative retell. In M. Perkins & S. Howard (Eds.), *New directions in language development and disorders* (pp. 251-260). New York, NY: Plenum Publishers.
- Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.
- Kail, M., & Hickmann, M. (1992). French children's ability to introduce referents in narratives as a function of mutual knowledge. *First Language, 12*, 73-94.
- Karmiloff-Smith, A. (1987). Function and process in comparing language and cognition. In M. Hickmann (Ed.), *Social and functional approaches to language and thought* (pp. 185-202). Orlando: Academic Press.
- Klecan-Aker, J.S. (1985). Syntactic abilities in normal and language deficient middle school children. *Topics in Language Disorders, 5*, 46-54.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.
- Liles, B. Z. (1985a). Cohesion in the narratives of normal and language-disordered children. *Journal of Speech and Hearing Research, 28*, 123-133.

- Liles, B. Z. (1985b). Production and comprehension of narrative discourse in normal and language disordered children. *Journal of Communication Disorders, 18*, 409-427.
- 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.
- MacLachlan, B. G., & Chapman, R. S. (1988). Communication breakdowns in normal and language learning-disabled children's conversation and narration. *Journal of Speech and Hearing Disorders, 53*, 2-7.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (Third Ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayer, M. (1974). *Frog goes to dinner*. New York: Puffin Pied Piper.
- McCabe, A. (1996). Evaluating narrative discourse skills. In S. F. Warren & J. Reichle (Series Eds.), K. N. Cole, P. S. Dale, & D. J. Thal (Vol. Eds.), *Communication and language intervention series: Vol. 6. Assessment of communication and language* (pp. 121-141). Baltimore: Paul H. Brookes.
- McCabe, A., & Peterson, C. (1984). What makes a good story? *Journal of Psycholinguistic Research, 13*, 457-480.
- Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language and Communication Disorders, 38*, 287-313.
- Paul, R., Hernandez, R., Taylor, L., & Johnson, K. (1996). Narrative development in late talkers: Early school age. *Journal of Speech and Hearing Research, 39*, 1295-1303.

- 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*, 592-598.
- Peterson, C. (1993). Identifying referents and linking sentences cohesively in narration. *Discourse Processes, 16*, 507-524.
- Peterson, C., & Dodsworth, P. (1991). A longitudinal analysis of young children's cohesion and noun specification in narratives. *Journal of Child Language, 18*, 397-415.
- Peterson, C., & McCabe, A. (1987). The connective 'and': Do older children use it less as they learn other connectives? *Journal of Child Language, 14*, 375-381.
- Pratt, M. W., & MacKenzie-Keating, S. (1985). Organizing stories: Effects of development and task difficulty on referential cohesion in narrative. *Developmental Psychology, 21*, 350-356.
- Schneider, P. (1996). Effects of pictures vs. orally presented stories on story retellings by children with language impairment. *American Journal of Speech-Language Pathology, 5*, 86-96.
- Schneider, P. (2008, June). Referent introduction in stories by children and adults. Poster presented at the triennial meeting of the International Association for the Study of Child Language, Edinburgh, Scotland.
- Schneider, P., & Dubé, R. V. (1997). Effect of pictorial versus oral story presentation on children's use of referring expressions in retell. *First Language, 5(3)*, 283-302.
- Schneider, P., Dubé, R. V., & Hayward, D. (2009). *The Edmonton Narrative Norms Instrument*. <http://www.rehabmed/ualberta.ca/spa/enni>

- Schneider, P., & Hayward, D. (2004, June). Measuring referring expressions in a story context. Poster presented at the Symposium for Research in Child Language Disorders, Madison, Wisconsin.
- Schneider, P., Hayward, D., & Dubé, R. V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. *Journal of Speech-Language Pathology and Audiology, 30*, 224-238.
- Schneider, P., & Winship, S. (2002). Adults' judgments of fictional story quality. *Journal of Speech, Language, and Hearing Research, 45*, 372-383.
- Scott, C. M. (1988). A perspective on the evaluation of school children's narratives. *Language, Speech, and Hearing Services in Schools, 19*, 67-82.
- Semel, E., Wiig, E., & Secord, W. (1995). *Clinical Evaluation of Language Fundamentals III*. San Antonio, TX: The Psychological Corporation.
- Statistics Canada (no date). *Canada dimensions: The people* [on-line]. Retrieved September, 2003 from <http://www.statcan.ca>.
- Strong, C. J. (1998). *The Strong Narrative Assessment Procedure*. Eau Claire, WI: Thinking Publications.
- Strong, C. J., & Shaver, J. P. (1991). Stability of cohesion in the spoken narratives of language-impaired and normally developing school-aged children. *Journal of Speech and Hearing Research, 34*, 95-111.
- Thorne, J. C., Coggins, T. E., Olson, H. C., & Astley, S. J. (2007). Exploring the utility of narrative analysis in diagnostic decision making: Picture-bound reference, elaboration, and fetal alcohol spectrum disorders. *Journal of Speech, Language, and Hearing Research, 50*, 459-474.

- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1998). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*, 1245-1260.
- Tsai, W., & Chang, C. (2008). "But I first...and then he kept picking": Narrative skill in Mandarin-speaking children with language impairment. *Narrative Inquiry, 18*, 349-377.
- Vallance, D. D., Im, N., & Cohen, N. J. (1999). Discourse deficits associated with psychiatric disorders and with language impairments in children. *Journal of Child Psychology and Psychiatry, 40*, 693-704.
- Villaume, S. K. (1988). Creating context within text: An investigation of primary-grade children's character introductions in original stories. *Research in the Teaching of English, 22*, 161-182.
- Vion, M., & Colas, A. (1998). L'introduction des référents dans le discours en français : Contraintes cognitives et développement des compétences narratives [Introducing referents in French: Cognitive constraints and development of narrative skills]. *L'année psychologique, 98*, 37-59.
- Wagner, C. R., Nettelbladt, U., B. Sahlén, B., & Nilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language and Communication Disorders, 35*, 83-93.
- Warden, D.A. (1981). Learning to identify referents. *British Journal of Psychology, 72*, 93-99.
- Westby, C.E. (1999). Assessing and facilitating text comprehension problems. In Catts, H.W., & Kamhi, A.G. (Eds.), *Reading disabilities: A developmental language perspective* (pp. 154-223). Boston: Allyn and Bacon.

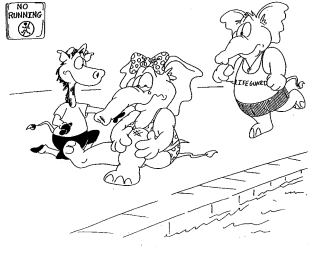
- Westerveld, M. F., Gillon, G. T., & Miller, J. F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech Language Pathology, 6*, 195-208.
- Wigglesworth, G. (1990). Children's narrative acquisition: A study of some aspects of reference and anaphora. *First Language, 10*, 105-125.
- Wiig, E.H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals – Preschool*. San Antonio, TX: The Psychological Corporation.
- Wittek, A., & Tomasello, M. (2005). Young children's sensitivity to listener knowledge and perceptual context in choosing referring expressions. *Applied Psycholinguistics, 26*, 541–558.
- Wong, A. M.-Y., & Johnston, J. R. (2002). The development of discourse referencing in Cantonese-speaking children. *Journal of Child Language, 31*, 633-660.

Appendix A. Pictures in which the target referents are introduced in the ENNI
(Permission was granted by the artist, Terry Willis, to reproduce the pictures in this
article -- © Wooket.)

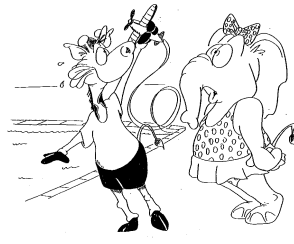
Set A



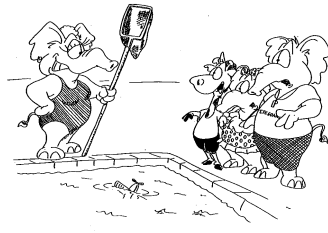
Giraffe, elephant, ball



Second elephant

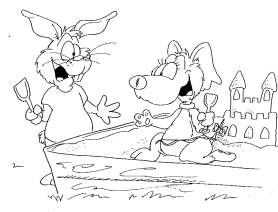


Airplane



Third elephant, net

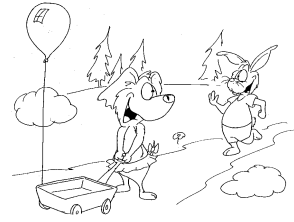
Set B



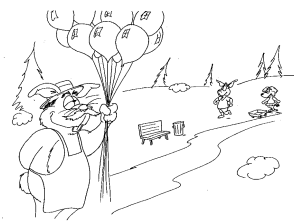
Rabbit, dog, sandcastle



Second rabbit



Balloon



Third rabbit



Balloons (end)

Appendix B. Example of scoring criteria for the First Mentions measure

Scores of 3 indicate expressions that are fully adequate for first mentions; scores of 2 indicate less than adequate expressions that are still partially informative; scores of 1 indicate expressions inadequate for first mentions. Full scoring information is available at <http://www.rehabmed.ualberta.ca/spa/enni>.

Character	Score as 3	Score as 2	Score as 1
Giraffe – story A1	<p>a/this _____ (e.g., <i>a giraffe, this cow</i>)</p> <p>name (e.g., <i>Gerry, Geegee</i>)</p> <p>possessive + noun (e.g., <i>her friend</i> if 'she' already introduced)</p> <p>another animal</p> <p>the other animal (if C mentioned 2 animals and one animal mentioned previously)</p>	<p>the/that _____ (e.g., <i>the giraffe</i>)</p> <p>a [invented word], e.g., <i>a geegee</i></p> <p>someone / somebody</p> <p>possessive + noun (if other character not yet introduced)</p> <p>another/the other _____ (e.g., <i>the other animal</i> if no animal mentioned previously)</p>	<p>pronoun (<i>he, she, it</i>)</p> <p>the [invented word], e.g., <i>the geegee</i> (an invented name would be scored as 3)</p>
Elephant – Story A1	<p>a/this _____ (e.g., <i>a elephant</i>)</p> <p>name (e.g., <i>Ellie</i>)</p> <p>possessive + noun (e.g., <i>her friend</i> if 'she' already introduced)</p> <p>another _____ (e.g., <i>another animal</i> if other character introduced as animal)</p> <p>the other _____ (e.g., <i>the other animal</i> if C mentioned 2 animals and one animal mentioned previously)</p>	<p>the/that _____ (e.g., <i>the elephant</i>)</p> <p>a [invented word]</p> <p>someone / somebody</p> <p>possessive + noun (if other character not yet introduced)</p> <p>another/the other _____ (e.g., <i>the other animal</i> if no animal mentioned previously)</p>	<p>pronoun (<i>he, she, it</i>)</p> <p>the [invented word] (an invented name would be scored as 3)</p>
Ball – Story A1	<p>a/this _____ (e.g., <i>a ball, a balloon, an orange</i>)</p> <p>possessive + noun (e.g., <i>her ball, the elephant's ball</i>)</p> <p>the ball if character is 'playing ball'</p>	<p>the /that _____</p> <p>vague or empty term, e.g., <i>a thingy/something/whatchacallit</i></p> <p>a [invented word]</p>	<p>pronoun (<i>it</i>)</p> <p>the [invented word]</p>

Author Note

Author affiliations: Phyllis Schneider, Ph.D., Department of Speech Pathology and Audiology, University of Alberta; Denyse Hayward, Ph.D., Canadian Centre for Research on Literacy, University of Alberta. Funding for this study was provided by the Children's Health Foundation of Northern Alberta. The authors would like to thank Marilyn McAra, Livia Tamblin, and Linda Kaert for their assistance in data collection, and Jess Folk-Farber, Rhonda Kajner, Roxanne Lemire, Marlene May, Michelle Millson, Ignatius Nip, Michelle Trapp, and Kathy Wagner for their assistance with other aspects of the study. Correspondence concerning this article should be addressed to: Phyllis Schneider, Department of Speech Pathology and Audiology, University of Alberta, 2-70 Corbett Hall, Edmonton, AB T6G 2G4. E-mail: phyllis.schneider@ualberta.ca. Web page: <http://www.rehabmed.ualberta.ca/spa/enni>

Footnotes

1. The ENNI website provides a table that can be used to obtain standard scores for each age group based on the raw total FM score. The table is available at <http://www.rehabmed.ualberta.ca/spa/enni/pdf/FM%20norms.pdf>.
2. To check for potential effects of the distribution of the proportion data, further analyses were conducted using transformed referential adequacy scores (arcsine transformation). ANOVA and post-hoc tests showed the same pattern of results using the transformed data.

Table 1. Number, Age, and Socioeconomic Status Information for the Normative Sample

Age Group	Language Group	N	N Boys	Mean Age	Age SD	Age Range	Mean SES	SES SD	SES Range
4	TD	50	25	4.60	.24	4.04-4.97	47.38	13.58	23.70-82.91
	LI	12	9	4.66	.23	4.18-4.97	47.17	10.80	34.45-70.27
5	TD	50	25	5.51	.27	5.01-5.98	46.64	12.12	24.11-73.38
	LI	14	8	5.41	.26	5.07-5.85	46.52	12.00	25.53-63.64
6	TD	50	25	6.56	.29	6.04-6.95	48.31	14.75	25.53-101.53
	LI	11	6	6.64	.26	6.13-6.95	40.26	13.97	26.36-60.73
7	TD	50	25	7.54	.28	7.01-7.98	45.13	13.65	24.11-101.32
	LI	13	10	7.56	.23	7.15-7.92	42.42	13.30	23.70-65.43
8	TD	50	25	8.58	.28	8.01-8.99	45.04	11.55	23.70-75.87
	LI	17	10	8.70	.26	8.11-8.96	42.42	7.40	32.78-60.73
9	TD	50	25	9.49	.28	9.02-9.99	48.79	12.04	25.56-80.32
	LI	10	5	9.50	.21	9.10-9.82	48.71	9.66	27.60-60.73

Table 2. Ethnic composition of the sample

Statistics Canada Category¹	ENNI Sample	Edmonton²	Canada²
Aboriginal	7.36%	4.15%	2.80%
Latin American	2.15%	1.04%	0.62%
Filipino	3.07%	1.64%	0.82%
Chinese	4.29%	6.24%	3.02%
Arab and West Asian	1.23%	1.24%	0.86%
Southeast Asian	1.53%	1.38%	0.61%
Black	2.76%	1.70%	2.01%
Korean	0.31%	0.29%	0.23%
Japanese	0.61%	0.22%	0.24%
Other	76.69%	81.93%	88.79%
Total	100.00%	99.84%	99.99%

¹The categories are those used on the Canadian census for 2001 for visible minorities.

²Data for Edmonton and Canada are from the 2001 Canadian census.

Table 3. CELF subtest scores by Test, Age Group and Language Group

Age Group	Language Group	CELF Subtest 1		CELF Subtest 2	
		M (SD)*	Range	M (SD)**	Range
4	TD	10.82 (3.32)	3-16	9.96 (2.38)	5-18
	LI	4.33 (2.64)	3-11	5.42 (1.17)	4-7
5	TD	10.74 (2.63)	3-15	9.96 (2.79)	3-16
	LI	5.00 (2.88)	3-11	4.43 (1.28)	3-7
6	TD	11.58 (3.03)	6-17	11.76 (3.32)	5-17
	LI	5.72 (1.79)	4-9	5.27 (2.20)	3-10
7	TD	12.24 (3.27)	4-17	11.66 (2.79)	5-17
	LI	6.38 (2.36)	3-11	4.31 (1.50)	3-7
8	TD	12.16 (2.92)	4-17	10.84 (2.74)	4-16
	LI	7.47 (2.38)	4-13	5.00 (1.80)	3-9
9	TD	11.84 (2.80)	6-17	11.14 (2.60)	5-16
	LI	8.10 (2.56)	4-13	5.40 (1.96)	3-8
Total CELF-P	TD	10.78 (2.98)	3-16	9.96 (2.58)	3-18
	LI	4.69 (2.74)	3-11	4.88 (1.31)	3-7
Total CELF-III	TD	11.96 (3.00)	4-17	11.35 (2.88)	4-17
	LI	6.94 (2.40)	3-13	4.96 (1.84)	3-10

* Subtest 1: Ages 4-5, CELF-P Linguistic Concepts; Ages 6-9, CELF-III Concepts and Directions

**Subtest 2: Ages 4-5, CELF-P Recalling Sentences in Context; Ages 6-9, CELF-III Recalling Sentences

Table 4. Composite CELF-P and CELF-3 scores by group

Age Group	Language	N receiving full CELF*	Receptive Language		Expressive Language		Total Language	
	Group		Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
4 (CELF-P)	TD	15	108.07 (14.00)	75-131	106.00 (9.86)	94-130	107.67 (12.13)	90-137
	LI	12	78.33 (15.87)	50-114	77.83 (5.94)	65-85	76.83 (8.62)	68-99
5 (CELF-P)	TD	19	103.37 (9.51)	81-116	104.32 (11.78)	85-133	103.79 (8.51)	88-118
	LI	14	79.86 (15.45)	61-108	74.00 (12.20)	50-92	76.21 (11.35)	58-96
6 (CELF-III)	TD	14	111.57 (12.40)	88-128	110.86 (11.05)	94-128	111.29 (12.03)	91-129
	LI	11	80.79 (11.53)	50-98	79.01 (11.57)	61-100	78.70 (8.34)	63-92
7 (CELF-III)	TD	15	108.53 (24.99)	65-143	112.13 (14.91)	78-139	110.33 (19.43)	70-138
	LI	13	81.62 (12.86)	50-96	69.69 (11.87)	50-86	74.00 (11.05)	51-90
8 (CELF-III)	TD	10	109.70 (10.48)	94-125	105.70 (16.44)	86-131	107.50 (12.94)	90-129
	LI	17	83.24 (16.55)	54-106	70.18 (9.42)	50-82	76.29 (11.94)	55-95
9 (CELF-III)	TD	15	107.87 (14.26)	88-139	97.73 (11.56)	80-118	102.80 (12.22)	86-122
	LI	10	80.00 (13.16)	53-100	70.70 (11.98)	50-90	73.50 (11.27)	55-85
Total CELF-P	TD	34	105.44 (11.75)	75-131	105.06 (10.84)	85-133	105.50 (10.28)	88-137
	LI	26	79.15 (15.35)	50-114	75.77 (9.84)	50-92	76.50 (9.99)	58-99
Total CELF-III	TD	54	109.35 (16.65)	65-143	106.61 (14.34)	78-139	107.96 (14.67)	70-138
	LI	51	81.66 (13.66)	50-106	72.06 (11.34)	50-100	75.68 (10.75)	51-95

*29% of children in the TD group and all of the children in the LI group were given the full CELF appropriate to their age group.

Figure captions

Figure 1. Means and standard deviations for first mention scores by age and language status.

Figure 2. Means and standard deviations for referential adequacy scores (percent of referent introductions that were adequate) by age and language status.

