



Long versus short language samples: A clinical procedure for French language assessment



Comparaison des échantillons de langage longs et courts : une procédure clinique pour l'évaluation du langage en français

KEY WORDS
LANGUAGE
SAMPLE ANALYSIS
SAMPLE LENGTH
FRENCH
MLU
MORPHOLOGY
LEXICAL DIVERSITY

Elin Thordardottir

Abstract

Introduction: Language sample analysis (LSA) is a main method of evaluation of children’s language production in both research and clinical contexts, providing unique insights that differ from those of formal tests. In spite of available procedures for LSA in French, their clinical use is low in Quebec.

Purpose: With a view to making LSA in French a more realistic clinical procedure, this study examined the effect of sample length on French LSA measures of both children with typical development (TD) and children with language impairment (LI). Effects of length were examined on global measures, such as Mean Length of Utterance (MLU) and detailed measures of morphological diversity.

Method: Conversational language samples collected within several previous studies using the same method were pooled, including samples from 124 children with TD and 25 children with LI, divided into 5 age groups from 2 to 6 years. All children were monolingual speakers of Quebec French. Results of sample lengths of 100, 50, 25, and 12 utterances were compared.

Results: Remarkable stability was found for all measures across sample lengths of 100, 50, 25, and (to a lesser degree) 12 utterances. MLU in words and morphemes were nearly perfectly correlated in both the TD and the LI samples. Greater morphological diversity and a greater number of word types and tokens were seen in longer samples, but differences between sample lengths were systematic. Based on high correlations for all LSA measures between sample lengths, a clinical shortcut procedure was proposed, involving the use of Mean Length of Utterance in words (MLUw) derived from a carefully collected sample of 25 utterances to estimate the more complex language use reported in accumulated descriptive data for 100 utterance samples. The study provides data that can serve as a clinical reference for LSA in Quebec French-speaking children with TD and with LI.

Elin Thordardottir,
 McGill University
 Montréal, QC
 CANADA

 Centre de recherche
 interdisciplinaire en réadaptation
 du Montréal métropolitain (CRIR)
 Montréal, QC
 CANADA

Abrégé

Introduction : L'analyse des échantillons de langage (AÉL) est l'une des principales méthodes d'évaluation de la production de langage d'un enfant, tant en recherche qu'en clinique. Cette méthode d'évaluation fournit des informations uniques qui diffèrent des informations recueillies avec les tests formels. Quoique des procédures d'analyse pour les échantillons de langage soient disponibles en français, elles sont peu utilisées en clinique au Québec.

Objectif : Ayant pour objectif de fournir une procédure clinique plus réaliste avec l'AÉL, cette étude examine l'effet de la longueur des échantillons sur les mesures de l'AÉL chez des enfants au développement typique (DT) et des enfants avec un trouble du langage (TL). L'effet de la longueur sur les mesures globales (telles que la longueur moyenne de l'énoncé, LMÉ) et sur les mesures plus détaillées de la diversité morphologique a été examiné.

Méthodologie : Les échantillons de langage conversationnel utilisés dans cette étude ont été recueillis dans le cadre de plusieurs études antérieures qui utilisaient la même méthode d'analyse. Les échantillons de 124 enfants au DT et de 25 enfants TL ont été inclus. Les enfants ont été divisés selon 5 groupes d'âge, de 2 à 6 ans. Tous les enfants étaient des locuteurs franco-québécois unilingues. Une comparaison d'échantillons de différentes longueurs, composés de 100, 50, 25 et 12 énoncés, a été effectuée.

Résultats : Les résultats montrent une stabilité considérable des mesures entre les échantillons composés de 100, 50, 25 et (à un moindre degré) 12 énoncés. Les LMÉs en mots et en morphèmes sont presque parfaitement corrélées, tant dans les échantillons des enfants au DT que dans ceux des enfants avec un TL. Une plus grande diversité morphologique, un plus grand nombre total de mots et un plus grand nombre de mots différents ont été observés dans les échantillons plus longs, mais les différences étaient systématiques entre échantillons de longueur différente. En s'appuyant sur les corrélations élevées retrouvées entre les mesures de l'AÉL pour les diverses longueurs des échantillons, une procédure clinique plus courte a été proposée. Cette procédure utilise la longueur moyenne des énoncés en mots, dérivée à partir d'un échantillon de 25 énoncés, pour estimer l'utilisation plus complexe du langage rapportée dans les données descriptives accumulées à partir d'échantillons de 100 énoncés. L'étude fournit des données pouvant servir de référence clinique pour l'AÉL chez des enfants franco-québécois au DT et avec un TL.

Language sample analysis (LSA) has a long-standing history as a main method of assessment of children's language development. In a review of the principal clinical uses of LSA with English-speaking children, Heilmann, Miller, and Nockerts (2010) point out that it has been used systematically for over 50 years, and has served as the basis for much of our current knowledge on children's typical language production, as well as of the way language production breaks down in children presenting with, or at risk for, language impairment. This article focuses on a commonly used LSA procedure, which, following the work of Brown (1973), uses the child's average utterance length as a yardstick of global language level, based on the observation that increased language skill leads to longer utterances. This has been shown with conversational samples for preschool children, and with conversational and, more clearly, with narrative samples for older children (Brown, 1973; Leadholm & Miller, 1992; Scarborough, Wyckoff & Davidson, 1986).

Mean length of utterance can be computed as the average number of words per utterance (Mean Length of Utterance in words, MLUw). Alternatively, by coding certain grammatical morphemes and including them in the length count, the yardstick (Mean Length of Utterance in morphemes, MLUm) reflects not only children's ability to string together an increasing number of words, but also their growing ability to use grammatical morphology. This coding also allows analysis of children's morphological development. In Brown's (1973) analysis of English, the morphological coding involves a set of 14 grammatical morphemes known collectively as Brown's morphemes; these are associated with Brown's five stages of morphological development. Standard LSA measures also include lexical measures and measures of verbal fluency. Lexical measures include total number of word tokens (TW), total number of different word types (TDW), used to estimate vocabulary size as well as verbal fluency, and the ratio of these (type-token ratio, TTR, or TDW/TW). The number and size of mazes is another measure of verbal (dys)fluency. Mazes are defined as material that does not contribute to conveying the message of the utterance, such as fillers, repetitions, and reformulations. The Systematic Analysis of Language Transcripts (SALT) software (Miller, Andriacchi & Nockerts, 2011) was developed expressly to yield these measures in English and contains normative data to help interpretation; however, the analysis can also be performed by hand. The analysis procedure used in this study was developed as a French adaptation of the SALT procedure; the French coding procedures can be found in Elin Thordardottir (2005) and in the online version of the SALT manual (Miller et al., 2011).

LSA in French

Language sample analyses are found in many cross-linguistic studies of language development and language impairment, often for the purpose of analyzing particular linguistic elements or error types. Detailed morphological coding procedures have also been developed in a number of languages, permitting the computation of MLUm (e.g. Arlman-Rupp, Van Niekerk de Haan & Van de Sandt-Koenderman, 1976; Elin Thordardottir & Ellis Weismer, 1998; Hickey, 1991; Miller et al., 2011); however, languages vary in the extent to which LSA procedures have been developed for or are commonly found in clinical use.

In French, the language of interest in the present study, longitudinal language sample corpus data on individual children have been collected and analyzed by several researchers, providing crucial information on the sequence of development of various grammatical structures in young children and of the development of various word classes (e.g. Bassano, 2000; Bassano, Maillolchon, Klampfer & Dressler, 2001; Morgenstern & Parisse, 2007). An LSA method for French was presented by Rondal (2003), with data for a single child spanning ages 2 to 3 years. Rondal's procedure included a detailed analysis of inflectional grammatical morphology that was similar, although not identical, to the procedure used in this study. Le Normand, Parisse, and Cohen (2008) reported normative data on 316 Parisian French children in nine age groups, ranging from 24 to 48 months. Measures included MLUw and vocabulary diversity (TW, TDW, and TTR) derived from 20-minute samples collected with a familiar adult. In addition, an automatic tagger was used to identify certain grammatical word classes. French language sample data have also been used to develop an adaptation of Language Assessment, Remediation, and Screen Procedure (LARSP) analysis (Maillart, Parisse & Tommerdahl, 2012), which focuses on phrase structure and morphology. All of these analysis methods have been shown to be sensitive to language development in French-speaking children, supporting their relevance for clinical application.

The analysis method used in this study was developed over a decade ago as a French adaptation of SALT coding procedures (Elin Thordardottir, 2005); however, SALT conventions were not applied directly, but rather, a parallel procedure was developed based on similar principles, taking the structural characteristics of French into account. The first data set analyzed with the French SALT procedure showed that young Quebec French-speaking children (18 to 47 months) had higher MLUs than English-speaking age mates and used more complex morphology with fewer

errors; however, their vocabularies were smaller than those of the anglophones, whether measured by LSA or by parent report (Quebec French version of the McArthur-Bates CDI, Trudeau, Frank, & Poulin-Dubois, 1999). Over the last decade, the procedure has been used in various studies of children with typical development (TD) and language impairment (LI) conducted within the same research lab (Elin Thordardottir, 2015; Elin Thordardottir, Kehayia, Lessard, Sutton & Trudeau, 2010; Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007). A unique aspect of this analysis system compared to other existing LSA procedures for French is that it provides a detailed focus on productive morphology and its relationship to MLU, in a sense replicating the tradition of Brown's morphemes, although in a manner that reflects the complexity of French morphology (see Elin Thordardottir, 2005, 2016a, and for bilingual children, Elin Thordardottir, 2014, 2015). The procedure has, therefore, been well documented in both monolingual and bilingual speakers of French, both with and without LI. In French, however, unlike in English, there has not been a strong tradition of use of such morphological data from spontaneous samples. We discuss the clinical utility of such measures in the next paragraph.

Clinical Uses of LSA

Language sample analysis, particularly in English, owes its widespread and long-standing use in research and clinical settings in large part to its ability to function as a measure of overall language level, and to the strong association that exists between global measures such as MLU_w and syntactic and morphosyntactic development (Brown, 1973). Clinically, a child's MLU does a better job than the child's age at predicting which grammatical structures the child has mastered given his or her overall language development (Brown, 1973). This is also a principal reason why MLU is frequently used as a matching variable in research, representing overall language level. The strong association between MLU and morphological development was demonstrated clearly in French for children aged 18 to 47 months, where the productive use of different grammatical morphemes was shown to be far more systematically predicted by MLU level than by age group (Elin Thordardottir, 2005). LSA also allows the documentation of the types of grammatical morphemes used by children. Data from children with TD at various MLU levels provides a crucial roadmap of the typical sequence of acquisition of grammatical morphology, which is an important guide to goal setting in intervention. In addition, it permits an assessment of whether an individual child's length of utterance actually results from the expected advances in morphosyntactic skills. Given that languages vary in their

structural characteristics and developmental sequence, it is important to use procedures and reference datasets that adequately reflect the language being assessed.

Another principal advantage of LSA in relation to formal tests is its high ecological validity and the preservation of a true communicative intent. In this respect, LSA provides a different and complementary type of information about a child's language abilities than do standardized tests. Correlational analyses between various French language measures, including standardized tests, measures of verbal memory, and language samples, indeed indicated that MLU contributes unique information on language abilities (Elin Thordardottir et al., 2010). It has been suggested that language sample data are more in line with clinicians' perceptions of language difficulties in children than standardized assessment results (cf. Heilmann, Miller, et al., 2010). This validity issue takes on an even greater importance in languages in which relatively few appropriate language tests are available that are adequately supported by research on these languages. Few standardized language tests are available in Quebec French, and most of the available tests have been translated and adapted from English (such as the EVIP (Dunn, Thériault-Whalen & Dunn, 1993) and the CELF-Canadian (Wiig, Secord, Semel, Boulianne & Labelle, 2009). Even though these particular tests have been renormed on French speaking children, they were not initially constructed to represent the characteristics of the French language or the typical sequence of acquisition of linguistic structures in French. Descriptive data from spontaneous language samples offer a measure that reflects spontaneous production data obtained from native speakers rather than responses to a predetermined set of test items based on another language.

Diagnostic accuracy of LSA measures.

In terms of diagnostic accuracy, research on global LSA measures in English, such as mean length of utterance (MLU) and vocabulary size and diversity, indicate sensitivity only in the fair range (79%) with specificity in the acceptable range (84%) (Heilmann, Miller, et al., 2010). A study on Quebec French comparing various measures for the identification of LI at age 5 years (Elin Thordardottir et al., 2011), including the MLU_w and MLU_m, found that the two MLU measures had similar and very low sensitivity (46%) but better specificity (80%) for the identification of LI. This means that low MLUs, whether computed in words or morphemes, suggested the presence of LI fairly strongly; however, because many children with LI obtained normal-range scores, the presence of LI was often missed. Overall, then, the conclusion for both English and French

is that clinical identification of language impairment is not a main benefit of global LSA measures, although they can contribute to such assessment. That said, LSA measures are superior to most other measures in providing a detailed picture of the child's current language abilities in real life settings. The more authentic portrait of the child's communicative abilities allows the clinician to determine which specific skills should be targeted in intervention, based on which needed skills have not been mastered and on which prerequisite supporting skills are in place.

Language sample length and complexity of analysis

In spite of the availability of various language sample measures for both European and Quebec French (Elin Thordardottir, 2005; 2015; 2016a; Le Normand et al., 2008; Rondal, 2003), systematic LSA is not in widespread use in clinical work with French-speaking children in Quebec. Anecdotal evidence suggests that clinicians may rely on spontaneous utterances glossed over the course of a clinical session to draw conclusions about the mastery of different structures. The clinical interpretation of such a set of utterances is problematic, in part due to selection bias – the clinician will tend to gloss the sentences that he or she finds most interesting or noteworthy. However, the careful collection of a sample of continuous utterances in a specified context and their detailed analysis appear to be rare. A principal reason is likely to be the time-consuming nature of LSA – this reason is reported as a main hindrance to clinical LSA, even in English (Heilmann, Miller, et al., 2010). For a full analysis, a language sample needs to be recorded, transcribed, and coded, a process that can take in excess of an hour. Another reason may be the low diagnostic precision of LSA in identifying LI, and the lack of a strong tradition of other uses of systematic LSA for more in-depth assessments in French.

Previous efforts to make English LSA more feasible within clinical settings have looked at the extent to which language samples can be shortened and still be reliable and informative. There is no general consensus on the necessary or ideal length of language samples; however, normative databases generally use samples of at least 100 utterances, and many research studies have used considerably longer samples. Tilstra and McMaster (2007), looking to develop a brief measure to assess gains in clinical intervention, showed that short narrative samples elicited from a single picture produced reliable results across three such pictures. Across children in K, 1st, and 3rd grade, measures of verbal fluency (such as number of words and C-units per minute) were reliable across the three short samples in all grades. In contrast, measures

targeting productivity (absolute number of words or clauses) were reliable only for the oldest children and grammatical accuracy was reliable only for the youngest children. It was suggested that overlap in age between the grade levels might explain the lack of stability of the productivity measures and, further, that brief samples might not give young children sufficient opportunity to show their productive abilities. As for grammatical errors, the fact that they are more common at the younger than older ages was thought to possibly contribute to the higher reliability of grammatical accuracy in K than in 1st and 3rd grade. Similarly, Heilmann, DeBrock, and Riley-Tillman (2013), examined language samples of kindergarteners at risk for LI, collected using a structured set of questions. The results showed high test-retest reliability across topics and sample lengths; sampling context and length had significantly less impact on the language sample measures than did child factors.

In yet another study, Heilmann, Nockerts, and Miller (2010) demonstrated that global measures of lexical diversity (words per minute and number of different words per minute), number of utterances, and utterance length were highly consistent across samples of 1, 3, and 7 minutes obtained from two age groups of children, 2;8 to 5;11 and 6;0 to 13;3, in both conversational and narrative contexts. These authors chose to focus on global measures rather than more fine-grained analyses such as grammatical morphology because they considered the latter to be less appropriate for short language samples in that they target some low-frequency elements of language. It is important to note that in this study, LSA measures included ones based on ratios (e.g. words per minute) as well as absolute counts (e.g. number of different words). However, the absolute count measures were, at least in some of the analyses, converted into ratios (such as number of different words per minute). An important body of literature has demonstrated that lexical diversity counts are sensitive to the number of words in the sample being considered (Duran, Malvern, Richards, & Chipere, 2004; Richards & Malvern, 1997). As a result, counts such as Total Number of Different Words (TDW) are likely to be higher in longer samples and also in samples with a higher MLU (more words per utterance). One way around this issue has been the use of Type Token Ratio (TTR: total number of different words/total number of words; however, this metric has also been shown to be biased, and other less biased metrics have been proposed (Duran et al., 2004). These results indicate that absolute counts should not be expected to stay constant across sample lengths.

Together, these studies are encouraging in that they indicate that the collection of lengthy samples is not necessary for all LSA purposes; however, they are limited in that they have focused solely on English and in that there has been little focus on grammatical morphology across sample lengths, with one study finding morphological errors not to be reliable across lengths, and another study assuming that morphology would not be a good candidate measure for shorter samples. The effect of sample length on various LSA measures may vary across languages. Notably, the development of grammatical morphology varies greatly across languages. Consequently, morphological findings may contribute in different ways to clinical conclusions across languages. In languages that are moderately or very highly inflected, including Icelandic, Dutch, and Irish, a very high correlation has been found between MLUm and MLUw in samples of TD children (Arıman-Rupp, et al., 1976; Elin Thordardottir & Ellis Weismer, 1998; Hickey, 1991). Furthermore, a near-perfect correlation was found in samples of Icelandic-speaking children aged 4 to 14 years old with and without Specific Language Impairment (SLI) (Elin Thordardottir, 2016b). This suggests that a detailed coding of grammatical morphology may be overkill in some languages if the main purpose is to derive a global measure of utterance length. However, information on grammatical morphology is, in its own right, a major clinical benefit of LSA, in particular for the assessment of language level, selection of intervention goals, and monitoring of treatment gains. Short samples are likely to give a good representation of a core set of high frequency words and structures, whereas low frequency words and structures, including some grammatical morphemes, are less likely to be seen in shorter samples. This may reduce the clinical advantages of short samples. However, given that the frequency of grammatical morphemes varies across languages, negative effects of short samples on morphology may be felt less in relatively more highly inflected languages. Given that languages also vary in which structures are most vulnerable in LI, and at which points in time (Elin Thordardottir, 2016b), languages may vary in whether a high correlation between MLUm and MLUw is found in samples of children with LI. Clearly, more research is needed to better understand the effect of sample length across languages and across language domains.

Purpose of Study

The main purpose of this study was to examine the effect of sample length on French LSA measures, specifically global utterance length, lexical counts, number of mazes, and detailed morphological production. The interest is on one hand theoretical and on the other hand

has the goal of developing a simplified yet informative clinical procedure. To date, little is known about the effects of sample length on clinical LSA measures in languages other than English. However, there are reasons to believe that the effects might not be uniform across measures across languages. Novel aspects of this study include not only the focus on French, but also a focus on the effect of sample length on grammatical morphology counts, both in children with TD and with LI. Language samples from several previous studies were pooled together for the analyses performed in this study. Therefore, an additional outcome of the study is the presentation of a French LSA database for a relatively large group of children with TD, and a smaller group of children with LI. Whereas parts of these data have been published previously for subgroups of the children in the study, this study presents, for the first time, data on vocabulary diversity and on mazes. Further, previous publications of the data have not examined effects of sample length.

Specific research questions are the following: 1) Are the various LSA measures in French sensitive to development in children with TD and children with LI? 2) How do children with TD and LI compare on the various LSA measures in French? 3) How stable are global language sample measures (utterance length, lexical, and maze counts) across sample length? 4) How stable are more fine-grained measures of morphological diversity across sample length? 5) What is the shortest sample length that can provide reliable and clinically useful information?

Methods

Participants

Participants included a total of 149 monolingual French-speaking children: 124 children with typical development (TD, age range 21 to 71 months) and 25 children with primary (specific) language impairment (LI, age range 37 to 77 months). These children were participants in previous studies conducted in the same research lab using the same language sampling and analysis procedures and collection of background information (Elin Thordardottir, 2005; Elin Thordardottir, 2015; Elin Thordardottir et al., 2010; Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007; Elin Thordardottir, Rothenberg, Rivard, & Naves, 2006). Of a total of 163 language samples gathered from these studies, 14 were excluded because they did not contain a full set of 100 utterances, leaving 149 samples. Diagnostic status as TD or LI was determined within each of the previous studies. Children with TD had no history of delayed development, major illnesses or hospitalizations, or pre- or perinatal complications as per parent report. They were

given a number of language measures, which are reported within each of the respective studies. Children with LI were recruited through clinical referral; their diagnostic status was verified as part of the studies in which they participated (Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007). As the data from these various studies were pooled, age groups were formed: 2 year olds (24 months \pm 6 months, or 20 to 29 months inclusive), 3-year-olds (36 months \pm 6 months, or 30 to 41 months), 4-year-olds (48 months \pm 6 months, or 42 to 53 months), 5-year-olds (60 months \pm 6 months, or 54 to 65 months), 6-year-olds (72 months \pm 6 months, or 66 to 77 months). Background characteristics as well as the distribution of children into these groups is displayed in Table 1. Background characteristics included gender, maternal education as a proxy for socio-economic status (SES), and nonverbal cognition (brief IQ scale of the Leiter International Performance Scale-Revised, Roid & Miller, 1997). The Leiter was not administered in one of the studies targeting young children with TD.

Procedures

Language samples were collected as part of a larger assessment protocol that varied across studies. The language sampling and analysis method was the same across all studies: samples were collected in a conversational play context, using a standard set of toys (for the younger children a house with people and furniture, household and food items; for the older children Playmobil and Polly Pocket toys). The children interacted with a trained examiner who was a native speaker of Quebec French. The examiner was instructed not to put pressure on the child, to give the child time to speak, and to refrain from asking many questions, particularly ones that would elicit a yes/no response. The examiner was instructed to show interest in the child's utterances and to respond to them. If children did not spontaneously engage in talk, the examiner was instructed to engage in self-talk and parallel talk in order to engage the child in conversation by modelling conversational behaviors.

Table 1. Background Characteristics for TD and LI Groups by Age Group.

TD Group:						
	Age in months	n	girls	boys	Mat.Ed.	Leiter Brief IQ
2 years	24.14 (2.03)	7	4	3	15.0 (4.8)	not available
3 years	35.33 (3.17)	28	11	17	15.6 (2.6)	109.8 (17.2)
4 years	48.00 (3.78)	19	12	7	16.6 (3.0)	112.3 (17.2)
5 years	59.10 (3.74)	58	26	32	16.4 (2.9)	99.9 (19.2)
6 years	68.17 (1.85)	12	5	7	17.4 (2.5)	104.6 (20.2)
LI Group:						
	Age in months	n	girls	boys	Mat.Ed.	Leiter Brief IQ
2 years	no participants					
3 years	38.33 (1.53)	3	1	2	19.0 (1.4)	100.0 (18.4)
4 years	47.00 (3.80)	10	1	9	14.8 (2.5)	102.9 (17.7)
5 years	56.80 (3.27)	5	1	4	14.5 (2.6)	111.3 (10.9)
6 years	68.57 (4.11)	7	7	0	13.7 (4.2)	94.9 (22.5)

Language samples were transcribed orthographically using SALT software (Systematic Analysis of Language; Miller et al., 2011). Grammatical morphology was coded following the French adaptation of SALT conventions. For a full description, see Elin Thordardottir (2005) or the online SALT manual (Miller et al., 2011). Transcription and coding reliability was verified and reported within each of the studies in which the samples were originally collected. For each child, a 100-utterance sample was obtained, excluding utterances that were exact repetitions of a previous utterance, but including utterances that contained unintelligible words. This procedure was used because unintelligible segments frequently make up a very small proportion of an otherwise grammatical and intelligible utterance. Given trade-off effects in language use, there is also a danger that unintelligible segments may tend to occur with higher frequency in longer and more complex utterances: excluding them might, therefore, bias the sample.

In order to examine effects of sample length, shorter sample cuts of 50, 25, and 12 utterances were obtained. The shorter samples were each taken from the middle of the original 100-utterance sample. Each shorter sample is, therefore, a subsample of the longer samples. Measures derived from each sample, using SALT, included mean length of utterance in words (MLUw), mean Length of utterance in morphemes (MLUm), total number of words (TW), total number of different words (TDW), and morphological diversity (MD). Morphological diversity refers to the number of different types of grammatical morphemes found in the sample. The set of grammatical morphemes documented in all the samples included these 16: verb person marking, compound past tense (*passé composé*), imperfect past tense (*imparfait*), pluperfect past tense (*plus-que-parfait*), periphrastic future tense (*futur proche*), simple future tense (*futur simple*), simple past tense (*passé simple*), imperative verb mood, subjunctive verb mood, conditional verb mood, past participle when not part of a compound tense, gender marking of adjectives, gender marking of pronouns, plural marking of adjectives, plural marking of pronouns, and plural marking of nouns. Other verb tenses exist in French that did not occur in the samples in this age range, but which would have been coded had they occurred. Therefore, the 16 morphemes represent the maximal morphological complexity found in this age range in a 100-utterance sample.

Results

100-Utterance Samples

The first research question asked whether the different

LSA measures are developmentally sensitive for children with TD and children with LI. We first report results for the 100-utterance samples – a sample length frequently used in normative reference databases, including our previous reports on French language samples from both monolingual and bilingual speakers of Quebec French (Elin Thordardottir, 2005; 2015; 2016a). Data are reported in Table 2, displaying MLUw, MLUm, TW, TDW, MD, and number of mazes, for age groups of 2-, 3-, 4-, 5-, and 6-year-old children with TD. The table also gives results for 3-, 4-, and 5-year-old children with LI. The results for the children with LI need to be interpreted with caution because of the small size of some of the age groups and also because of potential differences in severity levels that likely contribute to variability within each group. These data should not be seen as a reflection of the expected performance of children of the corresponding ages who have LI, as children with LI are a more heterogeneous group than children with TD. However, these data do offer descriptive information about these children's developmental trajectory, including the relationship between MD and MLU, as well as the sequence of acquisition of grammatical morphemes. For both the TD and LI groups, the 100-utterance measures of utterance length, vocabulary diversity, and morphological diversity increased systematically with age. For the TD group, MLUw increased from 2.17 to 4.83, and MLUm from 2.63 to 6.61 between ages 2 and 6 years. For the LI group, MLUw increased from 2.21 to 3.40 and MLUm from 2.66 to 4.30 between ages 3 and 6 years. Significant positive correlations were found between each measure and age for the TD group, with the exception of number of mazes: MLUw: $r = .506, p < .01$; MLUm: $r = .498, p < .05$; TW: $r = .506, p < .01$; TDW: $r = .599, p < .01$; MD: $r = .550, p < .01$; number of mazes: $r = .099, p = .639$. For the children with LI, all the measures were significantly correlated with age at $p < .01$: MLUw: $r = .594$; MLUm: $r = .630$; TW: $r = .630$; TDW, $r = .692$; MD: $r = .499$, and number of mazes: $r = .348$.

ANOVA analyses were performed for the TD group to examine age group effects, revealing a significant effect of age group for each measure (MLUw: $F(2, 123) = 17.566, p < .001, \eta^2 = .37$; MLUm: $F(4, 123) = 19.733, p < .001, \eta^2 = .40$; TW: $F(4, 123) = 28.448, p < .001, \eta^2 = .49$; TDW: $F(4, 122) = 20.131, p < .001, \eta^2 = .41$; MD: $F(4, 122) = 11.521, p < .001, \eta^2 = .28$; number of mazes: $F(4, 122) = 4.620, p < .001, \eta^2 = .28$). Post Hoc Tukey tests (family-wise alpha set at $p < .05$) revealed a similar pattern for MLUw and MLUm, TW, TDW, and MD: 2-year-olds and 3-year-olds did not differ significantly from each other, but each differed from the 4, 5, and 6-year-olds. The 5- and 6-year-olds did not differ significantly from each other, but differed from the 2-year-olds and 3-year-olds for all five measures. For MLUm, TW, and TDW, 4-year-olds differed

Table 2. Means and (Standard Deviations) of Language Sample Measures in 100-Utterance and 25-Utterance Samples for Children with TD and LI

Age	MLUw	MLUw	MLUm	MLUm	TW	TW	TDW	TDW	MD	MD	#Mazes	#Mazes
	100	25	100	25	100	25	100	25	100	25	100	25
Children with TD:												
2y	2.17	2.57	2.63	3.22	217.3	64.3	48.9	25.9	7.9	4.0	8.1	1.0
	(0.91)	(1.07)	(1.06)	(1.58)	(90.4)	(28.9)	(15.5)	(13.5)	(3.2)	(2.6)	(5.0)	(0.8)
3y	3.03	3.03	3.72	3.75	303.3	75.0	81.9	35.7	10.4	6.0	20.6	4.2
	(0.78)	(1.04)	(1.10)	(1.40)	(79.4)	(26.0)	(26.1)	(11.2)	(3.2)	(2.4)	(21.0)	(6.3)
4y	3.89	3.72	5.03	4.62	390.1	91.7	116.4	45.4	12.4	7.1	25.3	5.1
	(0.81)	(1.03)	(1.14)	(1.33)	(80.4)	(24.3)	(21.0)	(8.4)	(2.5)	(2.3)	(21.5)	(7.5)
5y	4.56	4.52	5.83	5.80	453.8	113.1	127.0	53.6	13.6	8.4	38.3	8.7
	(0.97)	(1.23)	(1.36)	(1.72)	(97.3)	(30.8)	(26.9)	(11.9)	(2.6)	(2.7)	(26.7)	(8.5)
6y	4.83	4.98	6.61	6.37	508.5	124.5	142.7	59.8	13.3	8.2	38.5	11.1
	(1.98)	(1.88)	(2.43)	(2.60)	(164.7)	(46.9)	(32.5)	(16.5)	(1.8)	(3.0)	(30.0)	(10.7)
Children with LI:												
3y	2.21	2.20	2.66	2.93	221.3		56.7	25.0	7.3	4.7	14.3	5.0
	(0.85)	(1.0)	(1.05)	(1.31)	(85.0)		(37.6)	(19.9)	(3.2)	(1.2)	17.0)	(4.5)
4y	2.22	2.26	2.68	2.81	221.8		56.4	24.0	7.6	4.0	6.2	1.5
	(0.70)	(0.74)	(0.97)	(0.95)	(69.4)		(20.6)	(10.1)	(2.7)	(3.2)	(5.7)	(1.8)
5y	3.14	3.29	3.95	4.02	314.0		90.4	38.4	11.6	6.4	22.6	5.8
	(0.45)	(0.92)	(0.62)	(1.24)	(44.9)		(15.5)	(11.2)	(1.5)	(3.1)	(9.5)	(4.0)
6y	3.39	3.03	4.30	4.18	339.7		97.7	39.0	11.9	6.4	10.1	1.29
	(1.37)	(1.18)	(1.97)	(1.83)	(137.0)		(29.1)	(15.4)	(4.1)	(1.9)	(11.1)	(2.6)

significantly from 6-year-olds; whereas for MLUw and MD, no difference was found between 4-, 5-, and 6-year-olds. For MD, 4-year-olds differed significantly from 2-year-olds. A different pattern emerged for the number of mazes: the only significant difference between age groups was that the groups of 2- and 3-year-olds each differed significantly from the group of 5-year-olds. ANOVA analysis on age effects in the LI group was not performed because of the small size of the age groups. Visual inspection of the LI data suggests a clear distinction between the two younger groups on one hand, and on the other hand, the two older groups.

Comparison of children with TD and LI

The second research question was how children with TD and LI compare on the various LSA measures. Comparison of the means of the TD and LI groups suggests that the children with LI are roughly 1 to 2 years behind their TD peers in their language sample measures. For a statistical comparison of the two groups, matched subgroups were formed: each of the 25 children with LI were matched with a child from the TD group on age (within 2 or 3 months) and on gender. Exceptions to this matching were that the oldest child in the LI group (77m) could not be matched closer than within 6 months as the oldest child in the TD group was 71 months old. Further, for two children, a gender match and age match could not be obtained. Thus one girl in the LI group was matched with a boy of the same age from the TD group, and one boy was matched with a girl. The resulting groups thus included the 25 children with LI (mean

age 53.96 m (*SD* 11.24) and 25 children with TD (mean age 54.0 months (*SD* 10.62). T-tests were used to compare the two groups. The two groups did not differ significantly in age ($p = .990$), but differed significantly on all the other measures. Children with TD had a significantly higher MLUw, and MLUm, greater number of words and different words as well as greater diversity of grammatical morphemes than did children with LI. Children with TD also produced a significantly greater number of mazes than children with LI. Detailed results are reported in Table 3.

Global measures (MLU, lexical and maze counts) across sample lengths

The third research question concerned the stability of the global LSA measures across sample length. Results on MLUm and TDW at sample lengths of 100, 50, 25, and 12 utterances are graphed in Figures 1 and 2, respectively, for the TD and LI group. These graphs provide a visual illustration of the patterns observed: those measures that are averages (MLUm and MLUw) changed little across sample lengths, whereas those that are based on absolute numbers of items, namely TW and TDW, increased systematically with increasing sample length.

Group means and standard deviations for each LSA measure are reported in Table 2, for sample lengths of 100 and 25 utterances. In comparing MLUw at 100 and 25 utterances lengths, the average difference was 0.49 (*SD* 0.42, range 0 to 1.93). MLUw changed by 0.5 or less for 61% of the children (91/149), by 0.5 to 1.0 for 27% of the

Table 3. Comparison of Matched Subgroups of Children with LI and TD for Language Sample Measures at a Length of 100 Utterances: Means, (Standard Deviations), T-test Results, and Effect Sizes

	LI	TD	t (48)	p	Cohen's d
MLUw	2.73 (1.05)	4.26 (1.64)	-1.52480	.001	-1.11
MLUm	3.38 (1.44)	5.66 (2.16)	-2.28200	.001	-1.24
TW	273.2 (102.5)	437.5 (151.4)	-164.320	.001	-1.27
TDW	74.8 (30.4)	123.6 (33.6)	-48.800	.001	-1.52
MD	9.6 (3.6)	12.8 (2.7)	-3.28000	.001	-1.01
Mazes	11.6 (11.0)	32.5 (26.5)	-20.800	.001	-1.03

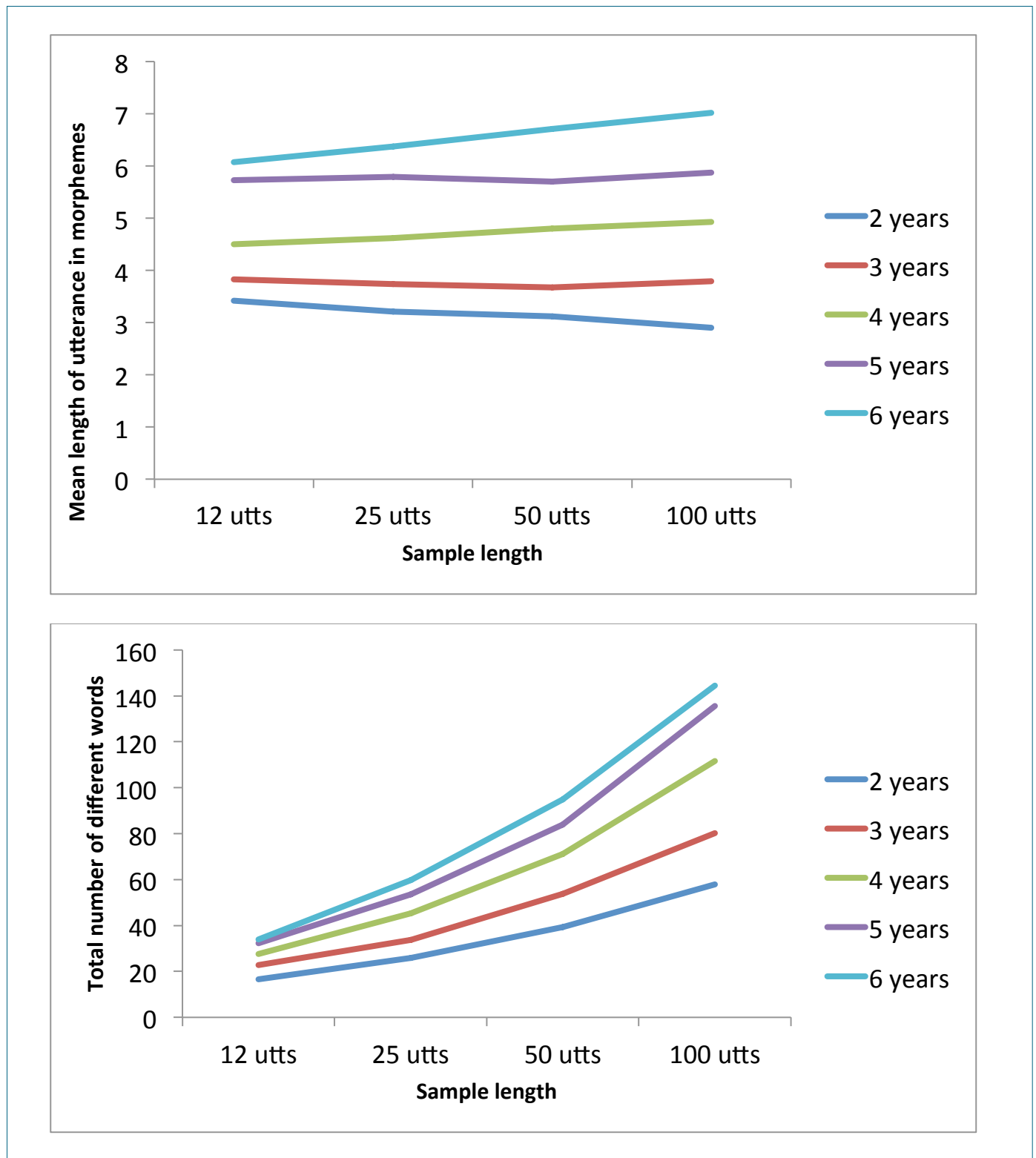


Figure 1. Group means for MLUm (upper panel) and TDW (lower panel) across sample lengths of 12, 25, 50, and 100 utterances, for age groups of children with TD.

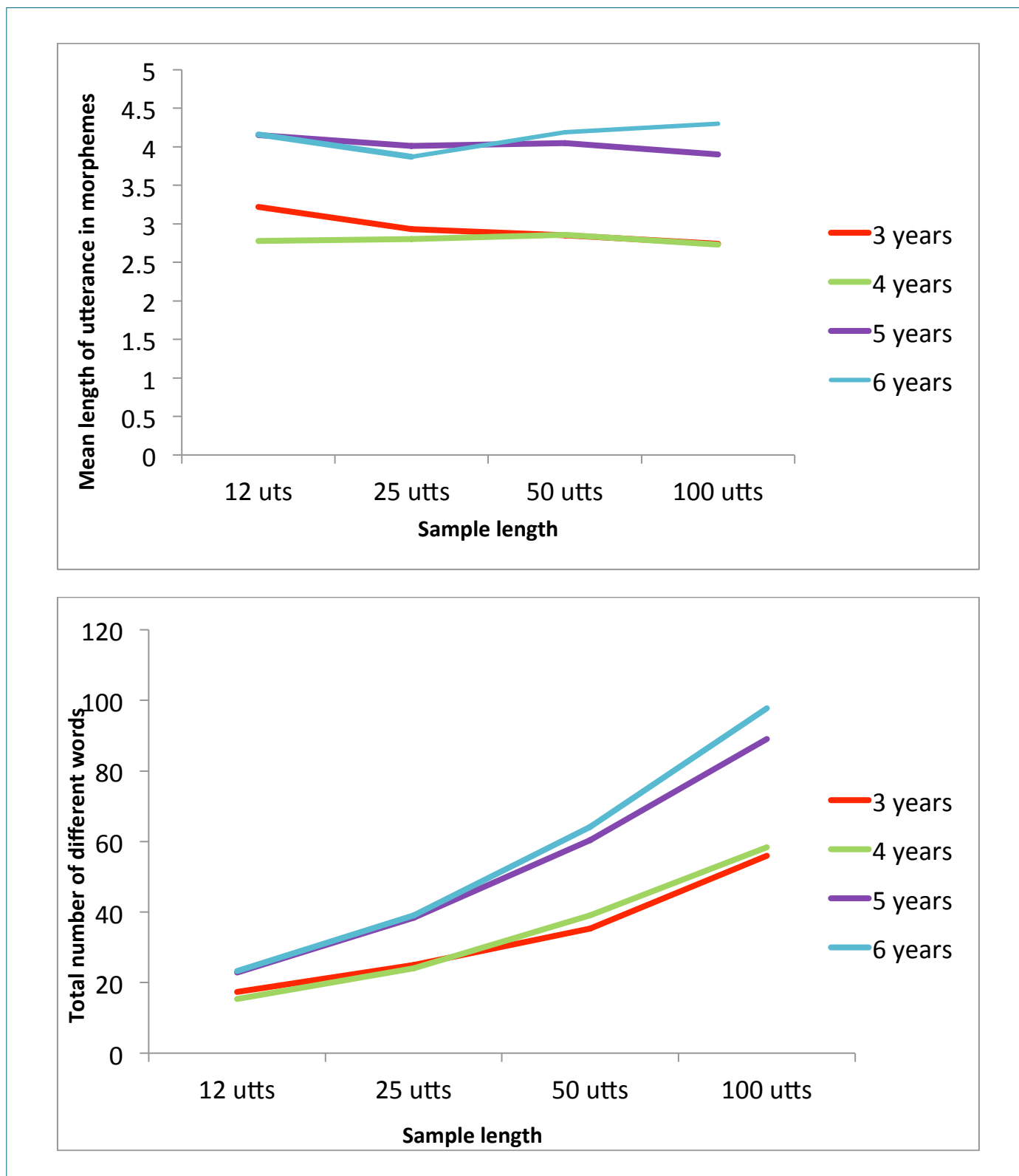


Figure 2. Group means for MLUm (upper panel) and TDW (lower panel) across sample lengths of 12, 25, 50, and 100 utterances, for age groups of children with LI.

children(40/149), and by more than 1.0 for only 12% (18/149) of the children. Table 4 displays results of correlational analysis where results for each measure at each of the shorter sample lengths (50, 25, and 12 utterances) are correlated with the same measure in a 100-utterance sample. The strength of the correlations decreases somewhat with decreasing sample length; however, all the correlations were significant at the $p < .001$ level, and were generally very strong (generally above $r = .800$). Correlations between sample lengths were somewhat stronger for the LI group than the TD group. The weakest correlations were found between samples of 100 and 12 utterances, in particular for the TD group, and for lexical diversity (TDW), also in particular for the TD group. The number of mazes in a sample is another global LSA measure. Table 2 shows that the number of mazes is greater in longer samples and increases with increasing age. The number of mazes in samples of 100 and 25 utterances was highly correlated for children with TD, but not for children with LI (see Table 4).

Morphological diversity across sample lengths

The fourth research question concerned the stability of morphological production across sample lengths. Morphological diversity (MD) involves a more fine-grained analysis than the global measures reported on in the previous section. MD was examined in samples of 100 and 25 utterances, given that samples as short as 25 utterances

retained a high correlation with 100-utterance samples for the global measures. MD is reported for these two sample lengths in Table 2. For each age group, fewer morpheme types are seen in shorter samples. However, a significant and strong correlation was found between MD100 (MD computed from a 100-utterance sample) and MD25 (MD computed from a 25-utterance sample) for children with TD ($r = .707, p < .01$) and for the LI group ($r = .580, p < .01$) (see Table 4). To verify the association between MLU and MD, correlational analysis revealed, for the TD group, that MD100 was significantly correlated with MLUw100 ($r = .665$), MLUm100 ($r = .627$), as well as TW100 ($r = .645$) and TDW100 ($r = .698$), all significant at the $p = .001$ level. For the LI group, MD100 similarly correlated with all global measures: MLUw100 ($r = .884$), MLUm100 ($r = .888$), TW100 ($r = .932$), TDW100 ($r = .932$), all significant at the $p = .001$ level. Given that MD correlated with both MLUm and MLUw, the latter was used for MLU grouping because it is less time consuming to compute.

MLU groups

For a further qualitative analysis, children were divided into groups based on MLUw100 (1.0-2.99, 3.0-4.99, and 5.0-6.99). This procedure follows the tradition of using MLU level to predict morphological level, including in French (Elin Thordardottir, 2005; 2015). The use of one sample length for the MLU grouping ensured that the data for

Table 4. Correlations between LSA Measures Obtained from Samples of 50, 25, and 12 Utterances with the Same Measure Obtained from a 100-Utterance Sample

	50 utterances		25 utterances		12 utterances	
	TD	LI	TD	LI	TD	LI
MLUw	.962**	.977**	.886**	.900**	.777**	.802**
MLUm	.976**	.976**	.889**	.915**	.771**	.829**
TW	.890**	.899**	.823**	.773**	.710**	.733**
TDW	.706**	.957**	.660**	.898**	.555**	.857**
MD			.707**	.580**		
#Mazes			.836**	.370 (NS)		

** $p < .001$

morphological use at 100 utterances and 25 utterances involved exactly the same children. It is worth noting that most, but not all, of the children would have been assigned to the same MLUw group had the grouping been based on 25 utterances (see previous section on global measures across sample lengths). To compare the types of morphemes that children are likely to use in the span of 100 versus 25 utterances, Table 5 shows the percentage of TD children in each MLUw group that were found to use each grammatical morpheme at the two sample lengths. As the table reveals, the different morphemes varied widely in the proportion of children that use them. Morphemes with high use at 100 utterances in all MLUw groups included the gender of adjectives and pronouns, the imperative mood, and person marking. It should be noted that because adjectives and pronouns have no basic gender, gender marking was always coded, whether the gender distinction is audible or not. Morphemes with medium levels of production, even by the lowest MLUw group, included noun plurals, the compound past (*passé composé*), and periphrastic future (*futur proche*) tenses. In general, the percentage of children using each morpheme increased with increasing MLUw group. A particularly noticeable increase with increasing MLUw is seen for the imperfect (*imparfait*) past tense, the simple future, the pluperfect, and the subjunctive and conditional moods. Compared to morphological use in a 100-utterance sample, a lower percentage of children are shown to use each of the morphemes in a 25-utterance sample. However, some morphemes are used by a high percentage of the children even at this short sample length. What is particularly noteworthy is that the relative standing of morphemes as being likely or unlikely to be seen is preserved in the 25-utterance samples compared to the 100-utterance samples.

Probability of use of morphemes by MLU group.

Table 6 shows the percentage of children with LI in each MLUw group using each of the morphemes. Due to the smaller number of children with LI and smaller MLU range, only two MLUw groups could be formed. Morphemes used by a relatively large proportion of children with LI include gender marking of adjectives and pronouns, noun plurals, the imperative mood, and person marking. Morphemes with medium levels of use include the compound past tense (*passé composé*), and periphrastic future tense (*futur proche*). This pattern is parallel to that of the TD children. However, several morphemes were never observed in samples of children with LI: the subjunctive and conditional moods, the simple future (*futur simple*), simple past tense (*passé simple*), and pluperfect (*plus-que-parfait*). Around

20% of children with TD were observed to use these tenses in the MLUw group corresponding to the highest MLU group of the LI children (MLUw 3.00-4.99). As for the TD children, fewer children with LI are observed to use each morpheme in 25- than in 100-utterance samples. However, the pattern of morphemes that are more or less likely to be seen is similar at both sample lengths. Together, Tables 5 and 6 suggest that MD increases with increasing MLUw in both TD and LI groups and that both groups follow a similar sequence of acquisition of these morphemes. The children with LI may need a somewhat higher MLUw to produce at least some of the morphemes.

A clinical short-cut

The final research question asked what the shortest sample length is that can yield clinically reliable and useful information. The results presented thus far have indicated that a 25-utterance sample represents a reasonable compromise for the global measures. Because of the predictable relationship between the more fine-grained MD measures in the 100- and 25-utterance samples, we propose a clinical shortcut that allows complex morphological information based on reference data on 100-utterance samples to be estimated from global analysis of a much shorter clinical sample (see Figure 3 for a summary of the steps).

Rationale for the shortcut procedure.

The ability to predict expected MD from MLU levels is a major clinical benefit of LSA. It is clear that the 100-utterance sample gives a more complete picture of morphological development than the 25-utterance sample. Indeed, the longer samples give more opportunity for use of a variety of morphemes. On average, children in the TD group were found to use 4.92 (*SD* 2.31) fewer morpheme types in the 25- than in the 100-utterance sample. The result was comparable for children in the LI group, who used on average 4.32 (*SD* 3.0) fewer morpheme types in the shorter sample. However, even though absolute numbers of different morphemes differed between the sample lengths, MD100 and MD25 were significantly and strongly correlated, both for children with TD and with LI (see Table 4). Therefore, even though MD is not stable across sample lengths in the sense that the same number of different morphemes is produced, it is stable in the sense that the two sample lengths differ in predictable ways, as detailed above. Given that the information obtained from a 100-utterance sample is more complete and thus more useful, but is time consuming to obtain for individual children, it would be beneficial clinically to be able to predict

Table 5. Percentage of Children with TD in each MLUw100 Group who use Different Types of Grammatical Morphemes in Samples of 100 Utterances and in Samples of 25 Utterances

Sample length	MLUw 1.00-2.99		MLUw 3.00-4.99		MLUw 5.00-6.99	
	100	25	100	25	100	25
Gender of Adj.	88.8	65.2	100	87.1	100	88.5
Gender of Pron.	69.2	7.7	91.4	52.9	95.8	76.9
Plural of Adj.	23.1	7.7	67.1	25.7	92.3	53.8
Plural of Pron.	0	0	50.0	15.7	61.5	34.6
Plural of Noun	50.0	42.3	98.6	72.9	100	92.3
Imperative mood	84.6	46.2	85.7	35.7	92.3	34.6
Subjunctive mood	11.5	0	22.9	4.3	57.7	15.4
Conditional mood	0	0	18.6	7.1	38.5	11.5
Past participle alone	23.1	0	10.0	2.9	3.8	0
Verb Person	100	100	100	100	100	100
Passé composé	57.7	23.1	90.0	48.6	100	57.7
Imparfait	11.5	0	70.0	34.3	84.6	46.2
Futur simple	3.8	0	17.1	4.3	30.8	23.1
Passé simple	3.8	0	1.4	0	4.2	0
Plus-que-parfait	0	0	20.0	1.4	42.3	19.2
Futur proche	46.2	30.8	95.7	52.9	96.1	69.2

Table 6. Percentage of Children with LI in each MLUw100 Group who use Different Types of Grammatical Morphemes in Samples of 100 Utterances and in Samples of 25 Utterances

Sample length	MLUw 1.0-2.99		MLUw 3.0-4.99	
	100	25	100	25
Gender of Adj.	100	100	100	87.1
Gender of Pron.	75.0	0	62.5	50.0
Plural of Adj.	12.5	6.3	18.8	12.5
Plural of Pron.	6.3	0	50.0	12.5
Plural of Noun	62.5	37.5	87.5	37.5
Imperative mood	87.5	56.3	100	87.5
Subjunctive mood	0	0	0	0
Conditional mood	0	0	0	0
Past participle alone	12.5	6.25	0	0
Verb Person	93.8	93.8	100	100
Passé composé	50.0	12.5	100	25.0
Imparfait	6.3	6.3	87.5	50.0
Futur simple	0	0	0	0
Passé simple	0	0	0	0
Plus-que-parfait	0	0	0	0
Futur proche	43.8	6.3	87.5	50.0

MD100 from a simpler LSA measure. It is interesting in this respect that MD100 was shown to be highly correlated with all the global LSA measures, both MLU counts, lexical diversity, and mazes (see previous section). Thus, any of the global measures would be a contender. One aspect that may make MLU more suitable is that, because it is an average, its absolute value changes very little across sample lengths, unlike TW and TDW. Given the near-perfect correlation between MLUm and MLUw, the latter of these two appears to be a better choice because it is much simpler to derive. Finally, because of the strong correlation between all the LSA measures across sample

lengths (Table 4), it may be justifiable to use MLUw25 rather than MLUw100 to predict not only MD25, but also MD100, using the descriptive data presented in Table 6. Further correlational analysis undertaken to evaluate the adequacy of this strategy revealed that MD100 is correlated approximately equally strongly to MLUm100 ($r=.651$), MLUm25 ($r=.640$), and MLUw25 ($r=.615$), all significant at the $p=.001$ level. This suggests that it is indeed justifiable, as a shortcut, to use MLUw25 to predict with reasonable confidence the morphological diversity that would likely have been seen had a 100-utterance sample been collected and analyzed.

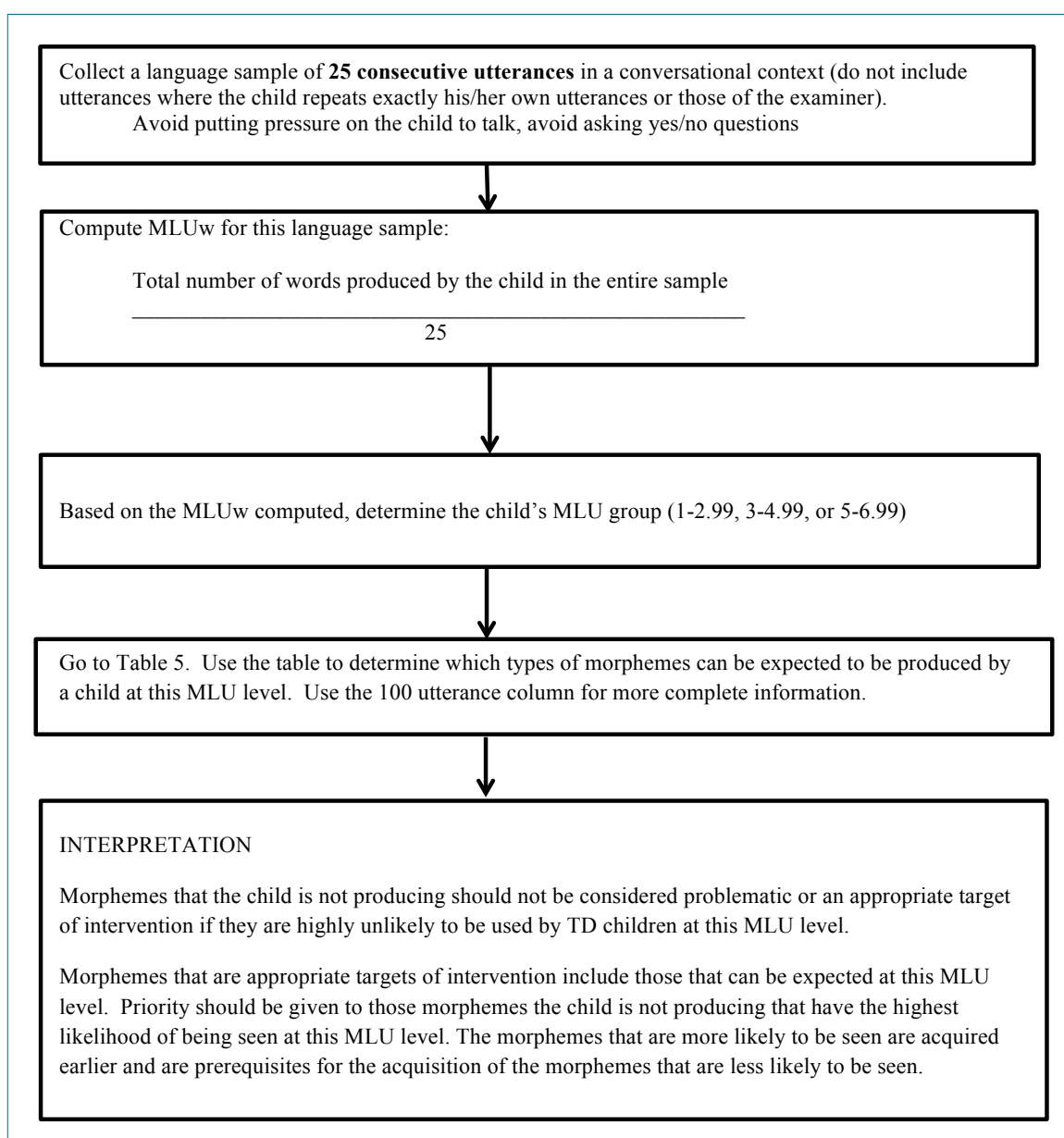


Figure 3. Clinical short cut procedure for the estimation of a child's morphological diversity.

Tables 5 and 6 provide information on the grammatical morphemes that are likely or unlikely to be found in 100- and 25-utterance samples of specific MLU levels. In order to use these tables, the child's MLU group needs to be established. A crucial issue concerning implementation of the shortcut procedure is whether children stay in the same MLU group whether their group is formed based on 100 or 25 utterances. For the TD children, 99 (79.8%) stayed in the same MLU group whether the group assignment was based on 100 or 25 utterances, whereas 16 (12.9%) went up one group, and 9 (7.2%) went down one group. For the children with LI, 19/25 (76%) stayed in the same group, whereas 2 (8%) moved up one group, and 4 (16%) moved down one group when the assignment was based on 25 utterances rather than 100. This provides further indication that for the large majority of children, estimation of MD100 from MLUw25 is a reasonably safe bet. To give an example of how MD data like the ones reported in Tables 5 and 6 could be used clinically, Table 5 indicates that only 25.7% of children with TD with an MLUw in the 3.0-4.99 range use the *passé composé* (compound past tense) in a 25-utterance sample, whereas 90% of children with this MLUw range will use it in a 100-utterance sample. This indicates that children in this MLUw range are highly likely to have the *passé composé* (compound past tense) in their repertoire even if they do not use it in a short sample. For added certainty of interpretation, the clinician could verify whether the morphemes most likely to appear in the short sample are produced (gender marking of adjective, plural of noun, verb person), which would provide more evidence that morphological use is as expected for MLUw level even in a short sample. This being the case, the data for 100 utterances can then be seen as a better indication of what the child is actually capable of producing.

Discussion

This study examined French LSA measures at four language sample lengths, including the global measures of utterance length, vocabulary diversity, and mazes and the more fine-grained measures of morphological diversity. All the global measures demonstrated remarkable stability of LSA in successively shorter samples, with very high correlations obtained between measures collected at each shorter length compared to the longest sample. Changes in MLU were negligible for the great majority of the children between samples of 100 and 25 utterances, with the great majority of the children being assigned to the same MLU group regardless of which sample length was used. Moreover, differences between age groups were similar at each sample length. Of great importance also, high correlations between samples of different lengths were

found both for children with TD and with LI. Even though samples as short as 12 utterances correlated quite highly with a 100-utterance sample, there was a greater drop in correlation strength between 25 and 12 utterances than between 50 and 25. A 25-utterance sample, therefore, appears to be a reasonable compromise between time investment and information value. It is worth noting here that in this study, shorter samples were a subsample of the longer samples, as the goal was to assess how much a sample can be shortened. This differs from the goal of some previous studies, for example, that of Tilstra and McMasters (2007), which compared short samples collected using three different elicitation pictures. That study addressed the test-retest reliability of short samples, whereas the present study addressed the extent to which a sample collected in a given setting provides more reliable information if it is allowed to be long.

The results on the global LSA measures are in good agreement with previous studies on English that have compared LSA measures at different sample lengths (Heilmann et al., 2013; Tilstra & McMaster, 2007). As expected, and also in agreement with previous studies, it was found that those measures that involve absolute numbers of items or different items differ between sample lengths (such as TW, TDW, and MD), whereas measures that reflect an average (MLU) remain stable. A novel aspect of this study is a more detailed examination of grammatical morpheme diversity across different lengths, revealing that shorter samples do differ from longer samples, but in predictable ways. On average, four to five fewer different morphemes were seen in 25-utterance than in 100-utterance samples. Further, the specific morphemes that were most or least likely to be encountered were the same in long and short samples. Therefore, the pattern of morphological use seen in a short sample, coupled with the descriptive data for both sample lengths (Tables 5 and 6) does give a clinically useful indication of the variety of morphemes that most likely would be seen in a longer sample for the same child. Consequently, in contrast to Tilstra and McMaster (2007) and Heilmann, Nockerts, et al. (2010), who recommended that short samples be used for global measures only, it is proposed here that short samples can be used to estimate the outcomes that would have been found in a longer sample, not only for global measures, but also for morphological diversity. A shortcut estimation procedure was proposed whereby MLUw from a 25-utterance sample is used to predict not only MLU from a longer sample, but also which grammatical morphemes would likely be seen had a 100-utterance sample been collected and analyzed. The justification for this procedure was discussed in an earlier section; it is

based on the high correlations found between measures at 100 and 25 utterance lengths, including the high correlation between MLUw25 and MD100. Estimation procedures are commonplace in clinical practice and are necessary to strike a balance between accuracy of findings and clinical feasibility. Standardized tests of vocabulary and grammar test only a small set of items from which the child's broader language knowledge is estimated. Similarly, language samples of any length, including 100-utterance samples, are but an estimation of the child's countless spontaneous utterances produced throughout a day. Just as a standardized test does not presume to catalogue a child's entire language knowledge, but rather to estimate language level, a language sample provides an estimate of the child's ability to deploy linguistic structures to convey a message in a more spontaneous manner.

Clinical reference data for children with TD and LI

Results of this study provide descriptive data on several global LSA measures in Quebec French for five age groups of children between the ages of 2 and 6 years, including MLUw, MLUm, TD, TDW, MD, and the number of mazes in a sample. Although these data have to be used with some caution because of the small sample size of some of the age groups, the three middle TD age groups (3-, 4-, and 5-year-olds) are of considerable size and their value as a good indication of typical development of Quebec French should not be discounted. The youngest and oldest groups, although smaller, fit into an overall developmental pattern with the middle groups, with measures increasing systematically with age throughout the age range. A systematic and gradual increase in the means with age group as well as a systematic correlation with age show that each of the measures is sensitive to development; post hoc tests on the age groups of the TD children indicate that the increase is not significant between each successive age group, but rather, that a significant shift occurs between the two youngest age groups, on one hand, and the two oldest on the other hand. Although not tested statistically, this pattern is even more evident in the LI group, as seen in Figure 2. The relationship of language measures to age can be expected to be somewhat different for children with LI than children with TD because of variability in severity levels. Nevertheless, these findings raise the possibility that a growth spurt in language development occurs in French in the middle of the preschool years, warranting further research.

A previous study by Le Normand et al. (2008) provided data on French-speaking children in Paris. Their results are reported separately for children of different SES levels.

However, a comparison of their MLUw data for 24, 36, and 48 month-olds reveals a rather close match with the present study for the 36 and 48 month-olds. In contrast, the Quebec French 2-year-olds achieve a considerably higher MLUw than their Parisian-French counterparts (2.17 vs. 1.36). Another difference in the datasets is that whereas there appears to be a slowing in MLU growth after age 4 years in the Paris data, no such slowing occurs in the Quebec data. The samples in the two studies cannot be compared directly because of differences in sampling context – the Paris samples were collected in a 20-minute interaction with a person familiar to the child whereas the present study used an unfamiliar examiner and a standard set of toys. It is nevertheless of interest to observe a fairly close correspondence between the two datasets. Clinically, these comparisons underscore the sensitivity of LSA to the elicitation context and the need to employ the same context as the reference base used to interpret the results; a finding reported previously in numerous studies (Elin Thordardottir, 2008; Hadley, 1998; Leadholm & Miller, 1992;).

Descriptive information on morphological development in this study confirms that of previous single-subject corpus studies (e.g. Bassano, 2000), showing an early preference for compound verb tenses, but also extends this information to higher age ranges, documenting the sequence of acquisition of more complex structures such as the conditional and subjunctive moods. At the age of 6 years, children are still not using a number of verb tenses, such as the *passé antérieur* (past anterior tense) and *futur antérieur* (future perfect tense), or the past tense of the subjunctive. Thus, unlike English, the full acquisition of French grammatical morphology types is not complete at this age.

Previous research had shown that young Quebec French-speaking children with and without LI differ significantly on MLU (Elin Thordardottir et al., 2011; Elin Thordardottir & Namazi, 2007). The present study further shows that they also differ significantly on all the other LSA measures, including vocabulary size and vocabulary diversity (TW, TDW), as well as in morphological diversity and the number of mazes. The group difference for each of these measures is statistically significant, and has a large effect size (Cohen's *d* exceeding 1.0), indicating a significant practical difference as well. This indicates that the language difficulty of the Quebec French-speaking children with LI is not restricted to one area of language, but rather extends across language domains. The effect sizes for the lexical domains are just as large as those for the morphological and syntactic domains. Children with LI produced significantly fewer mazes than children with TD. Although a high number

of mazes has been interpreted clinically as indicating word finding or syntactic formulation difficulties and therefore indicating impairment (cf. Elin Thordardottir & Ellis Weismer, 2002), normative data have also indicated that the number of mazes increases with increased MLU, being greater in the samples of older children and in more complex contexts (Leadholm & Miller, 1992; MacLachlan & Chapman, 1988). Therefore, a high number of mazes is not a clear sign of lower linguistic proficiency, in particular when factors such as length of utterance are not controlled. The lower number of mazes produced by the LI children is likely largely explained by the overall lower MLU and linguistic complexity of their samples. Further, when compared at comparable MLU levels, English-speaking children with LI have been shown to produce greater numbers of only certain types of mazes than children with TD (Elin Thordardottir & Ellis Weismer, 2002). Further research on maze production in French would be of interest.

The tables on morphological diversity indicate that children with LI proceed through a similar sequence of morphological development as do TD children. The morphemes that are relatively earlier or later developing are overall the same in the two groups. Although the relationship between MLU and the types of morphemes that can be expected to be seen in the sample are generally similar in the two groups, some morphemes are markedly less likely to be seen in the samples of children with LI than in samples of TD children of a similar MLU, or are not seen at all. These include notably some past and future verb tenses. In large part then, it appears that lower morphological diversity in LI samples is a reflection of an overall lower language level which does not provide many obligatory contexts for the missing morphemes - consistent with views of language acquisition that see the development of domains of language as being interconnected and interdependent (e.g. Marchman & Bates, 1994). The high correlations found between MD and both the MLU and lexical measures is consistent with a view of interrelated domains of language. However, some of the paucity of morphemes in the LI samples is unexplained by MLU, as determined by a comparison of what morphology is predicted by MLU group in TD and LI samples. Such a finding could be consistent with views that assume that LI presents a hindrance to the learning of specific morphemes (e.g. Rice & Wexler, 1996; Jakubowicz, Nash, Rigaut, & Gérard, 1999). Alternatively, it could also be that even at a similar overall language level, children with LI need more time or more practice to develop a greater variety of structures - they may need a larger critical mass of examples or may need more input to build the critical mass. Bilingual children who have received relatively little exposure to one of their languages have been

shown to exhibit a pattern characteristic of children with LI in their spontaneous language production in that language, suggesting that LI patterns may be associated with little input or non-efficient use of input (Elin Thordardottir, 2015). It may also be that MLU may be too crude an index of overall language development to adequately address this issue. In future studies, a more in-depth analysis of sentence types and vocabulary use may provide a better understanding of the relationship between language domains in acquisition.

This study has contributed new insights into the effect of language sample length on LSA measures, including both global and more detailed morphological measures in a language that is more highly inflected than English. The findings have important implications for the development of LSA procedures in other languages. Comparatively to English, there appears to be lesser need for the routine morphological coding of French language samples if the goal is mainly to obtain MLU, as MLU_w and MLU_m were almost perfectly correlated. At the same time, detailed morphological information, in relation to age and MLU, does have important clinical uses, which could be exploited much more in clinical work in French than is currently being done. The data presented here suggest that MLU is a useful clinical measure in French, both as a rough estimate of language level in spontaneous production, as well as to set expectations as to which grammatical morphemes should be mastered by the child or should be emerging, and which morphemes are still out of reach, and thus, should not yet be targeted in therapy. A key component of widely used hybrid intervention methods combining aspects of naturalistic and more focused clinician-directed components, such as Focused Stimulation (Ellis Weismer & Robertson, 2006), is the careful selection of therapy targets appropriate to the child's linguistic level. The purpose of this selection is to ensure that the child has the necessary prerequisites to be able to learn the new target. In order to use this method, it is crucial to have a method to document the child's current level and to have information on the normal developmental sequence of the language in question, such as the data and clinical procedure presented here.

References

- Arlman-Rupp, A., van Niekerk de Haan, D., & van de Sandt-Koenderman, M. (1976). Brown's early stages: Some evidence from Dutch. *Journal of Child Language*, 3(2), 267-274.
- Bassano, D. (2000). Early development of nouns and verbs and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language*, 27(3), 521-559.
- Bassano, D., Maillochon, I., Klampfer, S., & Dressler, W. (2001). L'acquisition de la morphologie verbale en français et en allemand autrichien: II. L'Épreuve des faits. *Enfance*, 53(2), 117-148.

- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Dunn, L., Thériault-Whalen, C., & Dunn, L. (1993). *Échelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary Test*. Toronto, ON: PsyCan.
- Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics*, 25(2), 220-242.
- Elin Thordardottir (2005). Early lexical and syntactic development in Quebec French and English: Implications for cross-linguistic and bilingual assessment. *International Journal of Language and Communication Disorders*, 40(3), 243-278.
- Elin Thordardottir (2008). Language specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. *Journal of Speech, Language and Hearing Research*, 51(4), 922-937.
- Elin Thordardottir (2014). Effects of exposure on vocabulary, morphosyntax and language processing in typical and impaired language development. In T. Grüter & J. Paradis (Eds.), *Input and Experience in Bilingual Development* (pp. 141-160). Philadelphia, PA: John Benjamins Publishing Company
- Elin Thordardottir (2015). The relationship between bilingual exposure and morphosyntactic development. *International Journal of Speech Language Pathology*, 17(2), 97-114.
- Elin Thordardottir (2016a). Typical language development and primary language impairment in French-speaking children. In J. Patterson & B. Rodriguez, (Eds.), *Multilingual perspectives on child language disorders* (pp. 176-208). Bristol, UK: Multilingual Matters.
- Elin Thordardottir (2016b). Grammatical morphology is not a sensitive marker of language impairment in Icelandic in children aged 4-14 years. *Journal of Communicative Disorders*, 62, 82-100.
- Elin T. Thordardottir & Ellis Weismer, S. (1998). Mean length of utterance and other language sample measures in early Icelandic. *First Language*, 18(52), 1-32.
- Elin Thordardottir & Ellis Weismer, S. (2002). Content mazes and filled pauses in narrative language samples of children with Specific Language Impairment. *Brain and Cognition*, 48(2-3), 587-592.
- Elin Thordardottir, Kehayia, E., Lessard, N., Sutton, A., & Trudeau, N. (2010). Typical performance on tests of language knowledge and language processing of French-speaking 5-year-olds. *Canadian Journal of Speech Language Pathology and Audiology*, 34(1), 5-16.
- Elin Thordardottir, Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., ... Chilingarian, G. (2011). Sensitivity and specificity of French language measures for the identification of primary language impairment at age 5. *Journal of Speech, Language and Hearing Research*, 54(2), 580-597.
- Elin Thordardottir, & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research*, 50(3), 698-715.
- Elin Thordardottir, Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4(1), 1-21.
- Ellis Weismer, S., & Robertson, S. (2006). Focused stimulation approach to language intervention. In R. J. Cauley and M. Fey (Eds.), *Treatment of language disorders in children* (pp. 175-202). Baltimore, MA: Paul Brookes.
- Hadley, P. (1998). Language sampling protocols for eliciting text-level discourse. *Journal of Speech, Language and Hearing Services at Schools*, 29(3), 132-147.
- Heilmann, J., DeBrock, L., & Riley-Tillmann, T. C. (2013). Stability of measures from children's interviews: The effects of time, sample length, and topic. *American Journal of Speech-Language Pathology*, 22(3), 463-475.
- Heilmann, J., Miller, J., & Nockerts, A. (2010). Using language sample databases. *Language, Speech and Hearing Services in Schools*, 41(1), 84-95.
- Heilmann, J., Nockerts, A., & Miller, J. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech and Hearing Services in Schools*, 41(4), 393-404.
- Hickey, T. (1991). Mean length of utterances and the acquisition of Irish. *Journal of Child Language*, 18(3), 553-569.
- Jakubowicz, C., Nash, L., Rigaut, C., & Gérard, C. L. (1998). Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition*, 7(2-4), 113-160.
- Leadholm, B., & Miller, J. (1992). *Language sample analysis: The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Education.
- Le Normand, M. T., Parisse, C., & Cohen, H. (2008). Lexical diversity and productivity in French preschool development: Developmental, gender and sociocultural focus. *Clinical Linguistics and Phonetics*, 22(1), 47-58.
- MacLachlan, B., & Chapman, R. (1988). Communicative breakdowns in normal and language-learning disabled children's conversation and narration. *Journal of Speech, Language and Hearing Disorders*, 53(1), 2-7.
- Maillart, C., Parisse, C., & Tommerdahl, J. (2012). F-LARSP 1.0: An adaptation of the LARSP language profile for French. *Clinical Linguistics and Phonetics*, 26(2), 188-198.
- Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21(2), 339-366.
- Miller, J., Andriacchi, K., & Nockerts, A. (2011). *Assessing language production using SALT software: A clinician's guide to language sample analysis*. Middleton, WI: SALT Software LLC.
- Morgenstern, A., & Parisse, C. (2007) Codage et interprétation du langage spontané d'enfants de 1 à 3. *Corpus*, 6, 55-78.
- Rice, M., & Wexler, K. (1996). Toward tense as a clinical marker for specific language impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39(6), 1239-1257.
- Richards, B., & Malvern, D. (1997). *Quantifying lexical diversity in the study of language development. The New Bulmershe Papers*. Reading, UK: University of Reading.
- Roid, G., & Miller, L. (1997). *Leiter International Performance Scale-Revised*. Wood Dale, IL: Stoelting.
- Rondal (2003). Langage oral. In J.A. Rondal & X. Seron (Eds.), *Trouble du langage: bases théoriques, diagnostic et rééducation* (pp. 375-411). Sprimont: Mardaga
- Scarborough, H., Wyckoff, J., & Davidson, R., (1986). A reconsideration of the relation between age and mean utterance length. *Journal of Speech and Hearing Research*, 29(3), 394-399.
- Tilstra, J., & McMaster, K. (2007). Productivity, fluency and grammaticality measures from narratives: Potential indicators of language proficiency? *Communicative Disorders Quarterly*, 29(1), 43-53.
- Trudeau, N., Frank, I., & Poulin-Dubois, D. (1999). Une adaptation en français québécois du MacArthur Communicative Development Inventory. *Revue d'orthophonie et d'audiologie*, 23(2), 61-73.
- Wiig, E., Secord, W., Semel, E., Boulianne, L., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales: Version pour francophones du Canada (Clinical Evaluation of Language Fundamentals: French Canadian Version)*. Toronto, Ontario: Pearson Canada Assessment.

Acknowledgements

The data used in this study were pooled from several previous studies, which were funded by grants awarded to the author as sole or Primary Investigator, or as a research team member, by FCAR (Fonds pour la formation de chercheurs et l'aide à la recherche, Québec), MSSS (Ministère de santé et des services sociaux du Québec, and SSHRC (Social Sciences and Humanities Research Council of Canada. Thanks are extended to Katie Torunian for assistance with data entry and management in the data pooling process, and to all the children and parents who participated in the studies.

Authors' Note

Correspondence concerning this article should be addressed to Elin Thordardottir, Ph.D., Professor, School of Communication Sciences and Disorders, McGill University, 2001 McGill College, 8th floor, Montreal, QC H3A 1G1, Canada. Email: elin.thordardottir@mcgill.ca.