8 TalkBank Resources for Psycholinguistic Analysis and Clinical Practice

Nan Bernstein Ratner and Brian MacWhinney

Introduction

The formation of a network of linked linguistic open data (LLOD) can contribute in many important ways to the advancement of the study of language structure, usage, processing, and acquisition. The chapters in this book present a comprehensive overview of various efforts to build this new structure. The current chapter will show how the TalkBank system has already succeeded in realizing many of these goals and could eventually support still others. TalkBank had its origins in 1985 with the Child Language Data Exchange System (CHILDES), founded by Brian MacWhinney and Catherine Snow; both the first and second author of this chapter continue to work to enlarge and maintain its growing resources.

TalkBank (https://talkbank.org) is now the largest open repository of data on spoken language. Initially, these data were represented primarily in transcript form. However, new TalkBank corpora now include linkages of transcripts to media (audio and video) on the utterance level, as well as extensive annotations for morphology, syntax, phonology, gesture, and other features of spoken language.

An important principle underlying the TalkBank approach is that all its data are transcribed in a single consistent format. This is the CHAT format (talkbank.org/manuals/chat.pdf), which is compatible with the CLAN programs (talkbank.org/manuals/clan.pdf). This format has been developed over the years to accommodate the needs of a wide range of research communities and disciplinary perspectives. Using conversion programs available inside CLAN, the CHAT format can be automatically converted both to and from the formats required for Praat (praat.org), Phon (phonbank.talkbank.org), ELAN (tla.mpi.nl/tools/elan), CoNLL (universaldependencies.org/format.html), ANVIL (anvil-software.org), EXMARaLDA (exmaralda.org), LIPP (ihsys.com), SALT (saltsoftware.com), LENA (lenafoundation.org), Transcriber (trans.sourceforge.net), and ANNIS (corpus-tools.org/ANNIS). For each of these conversions, the CHAT format recognizes a superset of information types (dates, speaker roles, intonational patterns, retrace markings, and so on). This means that, when data are converted into the other formats, there must always be a method for protecting data types not recognized in those programs against loss. This is done in two ways. First, users can often hide CHAT data in special comment fields that are not processed by the program but that will be available for export. Second, when employing the other programs, users must be careful not to alter codes in CHAT format that mark aspects that cannot be recognized by the other programs. There are no cases in which information created in the other programs cannot be represented in CHAT, because CHAT is a superset of the information represented in these other programs.

TalkBank is composed of a series of specialized language banks, all using the same transcription format and standards. These include CHILDES (https://childes.talkbank.org) for child language acquisition, AphasiaBank (https://talkbank.org/Aphasiabank) for aphasia, PhonBank (https://phonbank.talkbank.org) for the study of phonological development, TBIBank (https://tbi.talkbank.org) for language in traumatic brain injury, DementiaBank (https://dementia.talkbank.org) for language in dementia, FluencyBank (https://fluency.talkbank.org) for the study of childhood fluency development, HomeBank (https://homebank.talkbank.org) for daylong recordings in the home, CABank (https://ca.talkbank.org) for Conversation Analysis, SLABank (https://slabank.talkbank.org) for second language acquisition, ClassBank (https://class.talkbank.org) for studies of language in the classroom, BilingBank (https://biling.talkbank.org_ for the study of bilingualism and code-switching, LangBank for the study and learning of classical languages, SamtaleBank (https://samtale.talkbank.org) for Danish conversations, the SCOTUS corpus in CABank with 50 years of oral arguments linked to transcripts at the Supreme Court of the United States, and the spoken portion of the British National Corpus, also in CABank. We and our collaborators are continually adding corpora to each of these collections. The current size of the text database is 1.4TB and there are an additional 5TB of media data. All the data in TalkBank are freely open to downloading and analysis, with the exception of the data in AphasiaBank, HomeBank, and research data in FluencyBank, which are password protected. The CLAN program and the related morphosyntactic taggers are all free and open-sourced through GitHub (http://github.com).

These databases and programs have been used widely in the research literature. CHILDES, the oldest and most widely recognized of these databases, has been used in over 6,500 published articles. PhonBank has been used in 480 articles and AphasiaBank has been used in 212 publications. In general, the longer a database has been available to researchers, the more that its use has become integrated into the basic research methodology and publication history of the field.

Metadata for the transcripts and media in these various TalkBank databases have been entered into the two major systems for accessing linguistic data: OLAC (see Simons and Bird in this volume) and CMDI/TLA (see Trippel and Zinn, also in this volume). Each transcript and media file has been assigned a PID (permanent ID) using the Handle System (www.handle.net). In addition, each corpus has received a DOI (digital object identifier) code. The metadata available through these systems, along with the data in the individual files, implements each of the requirements of the DTA system (Blume et al., this volume). The PID numbers are encoded in the header lines of each transcript file and the DOI numbers are entered into HTML web pages that include extensive documentation for each corpus, photos and contact information for the contributors, and articles to be cited when using the data. All these resources are periodically synchronized using a set of programs that rely on the fact that there is a completely isomorphic hierarchical structure for the CHAT data, the XML versions of the CHAT data, the HTML web pages, and also the media files. If information is missing for any item within this parallel set of structures, the updating program reports the error and it is fixed. All this information is then published using an OAI-PMH (www.openarchives.org/pmh) compatible method for harvesting through systems such as the Virtual Linguistic Observatory at https://vlo.clarin.eu (VLO) developed through the CLARIN initiative (https://clarin.eu).

For 10 of the languages in the database, we provide automatic morphosyntactic analysis using the MOR, POST, and MEGRASP programs built into CLAN. These languages are Cantonese, Chinese, Dutch, English, French, German, Hebrew, Japanese, Italian, and Spanish. Tagging is done by MOR, disambiguation by POST, and dependency analysis by MEGRASP. Details regarding the operation of the taggers, disambiguators, and dependency analyzers for these languages can be found in MacWhinney (2008). Processing in each of these languages involves differing computational challenges. The complexity and linguistic detail required for analysis of Hebrew forms is perhaps the most extensive. In German, special methods are used for achieving tight analysis of the elements of the noun phrase. In French, it is important to mark various patterns of suppletion in the verb. Japanese requires quite different codes for parts of speech and dependency relations. Eventually, the codes produced by these programs will be harmonized with the GOLD ontology (Langendoen in this volume). In addition, we compute a dependency grammar analysis for each of these 10 languages, which we will harmonize with the Universal Dependency tagset (https://universaldependencies.org).

Because these morphosyntactic analyzers all use a parallel technology and output format, CLAN commands can be applied to each of these 10 languages for uniform computation of indices such as MLU (mean length of utterance), vocd (vocabulary diversity), pause duration, and various measures of disfluency. In addition, we have automated language-specific measures such as DSS or Developmental Sentence Score (for English and Japanese) and IPSyn. Following the method of Lubetich and Sagae (2014), we are now developing language-general measures based on classifier analysis that can be applied to all 10 languages using the codes in the morphological and grammatical dependency analyses. However, there are many other languages in the database for which we do not yet have morphosyntactic taggers. This means that it is a priority to construct MOR systems for languages with large amounts of CHILDES and TalkBank data, such as Catalan, Dutch, Indonesian, Polish, Portuguese, and Thai.

Using these data and methods, researchers have been able to evaluate the use of different approaches to comparable data. Such comparisons have been particularly fruitful in studies of the acquisition of morphology and syntax. For example, the debate between connectionist models of learning and dual-route models focused on data regarding the learning of the English past tense (Marcus et al. 1992; Pinker and Prince 1988; MacWhinney and Leinbach 1991) and later on data from German plural formation (Clahsen and Rothweiler 1992). In syntax, emergentists (Pine and Lieven 1997) have used CHILDES data to elaborate an item-based theory of learning of the determiner category, whereas generativists (Valian, Solt, and Stewart 2009) have used the same data to argue for innate categories. Similarly, CHILDES data in support of the Optional Infinitive Hypothesis (Wexler 1998) have been analyzed in contrasting ways using the MOSAIC system (Freudenthal, Pine, and Gobet 2010) to demonstrate constraint-based inductive learning. In these debates, and many others, the availability of a shared open database has been crucial in the development of analysis and theory.

Through these various methods of transcript format conversion, metadata publication, grammatical analysis, and data sharing, TalkBank has already fulfilled many of the goals of the LLOD Project. As a result of these efforts, TalkBank has been recognized as a Center in the CLARIN network (clarin.eu) and has received the Core Trust Seal (https://coretrustseal.org). TalkBank data have also been included in the SketchEngine corpus tool (http://sketchengine.co.uk).

However, there are other goals of the LLOD Project that seem to be currently out of the reach of spoken language corpora like TalkBank. The type of linkage proposed by Chiarcos and colleagues (this volume) and perhaps even the LAPPS system (Ide, this volume) would require a major effort to cross-index the individual lexical or morphological items in the many TalkBank databases. Such linkage makes sense for lexical databases or coding systems, because these involve linkages that can directly yield secondary analyses. For example, linkages between WordNet systems (http://wordnet.princeton.edu) in various languages or grammatical coding features (Langendoen, this volume) can directly facilitate a variety of NLP (natural language processing) tasks, such as translation, tagging, metaphor analysis, and information extraction. However, the value of linkages between entities for spoken language corpora has yet to be demonstrated. For these corpora, the role of individual lexical items depends entirely on the overall syntactic and discourse context, and it is not clear how these relations can be evaluated through simple links on the lexical or featural level. For these resources, the most important analytic tools involve corpus-based searches, such as those available in the TalkBankDB system at https://talkbank.org/DB.

An additional problem facing the task of linkages across spoken language data arises from the fact that many data centers do not make their data publicly available. For example, the majority of the materials in The Language Archive (tla.mpi.nl) cannot be directly accessed, and many are not available for access at all. The materials collected by the Linguistic Data Consortium (ldc.org) are only available to subscribers, thereby making them off limits for linked open access. Of the major databases for spoken language data, only TalkBank provides completely open access to records in a consistent XML format. Thus, TalkBank would seem to be a good target for integration into the LLOD project, once methods for dealing with spoken language corpora have been developed.

Rather than focusing on LLOD linkages across spoken language corpora, TalkBank has developed other methods for between-corpus linkage. Two of these methods have already been discussed. The first method involves the construction of programs that can convert between CHAT format and formats used by other analytic programs. That work has largely been completed. The second method is the construction and publication of metadata to the VLO system for indexing corpora, transcripts, and media. This work, too, has mostly been completed.

We are now actively engaged in the development of a third approach to between-corpus linkage. This method permits automatic quantitative comparisons between corpora or subsections of a given corpus. The goal here is to be able to compare data from speakers at different ages, speaking different languages, in different tasks and situations, at different stages of learning, and with different clinical profiles. In the balance of this chapter, we will outline the development of one of these methods, called KIDEVAL, for comparing child language data. A parallel system, called EVAL, has also been developed for making comparisons across samples of speech from persons with aphasia (PWAs). The EVAL system makes use of the fact that the data in AphasiaBank were all collected with a single consistent protocol. Based on these protocol data, we can extract group means for individual aphasia types (Broca's, Wernicke's, anomia, global, transcortical motor, and transcortical sensory), which we can then use as comparisons for the results from individual PWAs. For child language data, we have identified a subset of the database that can be used in a similar way to make comparisons within age groups. Comparisons of this type are fundamental to the process of clinical assessment, as well as to the study of basic developmental processes.

Child Language Sample Analysis

For the assessment of child language abilities, language sample analysis (LSA) provides a very high degree of ecological validity and "authenticity," as mandated by current educational policies (Overton and Wren 2014). It supplements standardized assessment by providing a snapshot, as it were, of a given child's language "in action." More critically, it provides baseline insights into the child's strengths and weaknesses across the range of language skills necessary for age-appropriate communication, from vocabulary to syntax to pragmatics. These skills can be tracked in natural contexts over time (Price, Hendricks, and Cook 2010). LSA provides clinicians with tangible goals for therapy unlikely to emerge from results of standardized testing but that can be prioritized for intervention (Overton and Wren 2014). In the absence of norm-referenced assessments for children speaking non-mainstream dialects or English as a Second Language, LSA also can provide less biased and more informative information about a child's expressive language skills and needs (Caesar and Kohler 2007; Gorman 2010).

80151-10990_ch01_1P.indd 135 7/19/19 6:17 PM

However, there are a number of practical issues in using LSA for clinical purposes that tend to diminish the frequency (and depth) of its use in actual clinical practice (Gorman 2010). While the self-reported use of LSA has been steadily climbing in reports from 1993 to 2000 (Hux 1993; Eisenberg, Fersko, and Lundgren 2001; Kemp and Klee 1997), most SLPs (Speech-Language Pathologists) report compiling relatively short samples in real-time notation and using them primarily to compute Mean Length of Utterance (MLU; Price, Hendricks, and Cook 2010; Finestack and Satterlund 2018), despite the fact that MLU is not a good stand-alone measure for identifying language impairment (Eisenberg, Fersko, and Lundgren 2001). In addition, Lee and Canter (1971) found that less than one-third of respondents computed an additional measure, the most popular being DSS. Very recently, Finestack and Satterlund (2018) found that only about 30% of American SLPs compute “informal” language sample measures. Of these, from 86 to 94% (depending upon age of child) used MLU. Type-token ratio (TTR) was used by only about 25–32% of respondents. Use of DSS had fallen to roughly 15% of SLPs, and other measures were used by fewer than 10% of SLPs who conducted LSA.

It is well acknowledged that good LSA can be quite time-consuming (Overton and Wren 2014). Some studies have estimated that it takes up to 8 hours of training and from 45 minutes to one hour of work after a transcript has been generated to compute DSS (Long and Channell 2001; Cochran and Masterson 1995). One study (Gorman 2010) estimated that it takes more than 30 minutes per sample following transcription to compute the Index of Productive Syntax (IPSYN; Scarborough 1990). Hand computation of most LSA measures, even the time-honored MLU, is quite prone to error. It is difficult to use the same worksheet to compute multiple linguistic measures, and it is a waste of time to transfer handwritten scribbles of what the child said to most scoring protocols. Thus, even by self-report, LSA is not used by many clinicians, and is not intensively exploited by most to inform child language assessment. Those who do LSA often use a sample that is much too short to meet the intended sample size for the measures that are computed (Westerveld and Claessen 2014), sometimes 50–75% fewer utterances than recommended.

Computer-assisted LSA can solve all the problems listed above (time, accuracy and depth of analysis; Heilmann 2010; Price, Hendricks, and Cook 2010; Evans and Miller 1999; Miller 2001; Hassanali 2014), but is not very frequently used in practice. A recent study estimated that only 12.5% of SLPs in Australia use computer-assisted transcription and analysis (Westerveld and Claessen 2014), and there is little to suggest that their American counterparts use such procedures at a significantly higher rate (Price, Hendricks, and Cook 2010). Finestack and Satterlund (2018) recently found that computer-assisted LSA was used by only 1–5% of American SLPs. As we will suggest, use of computers to aid in sample transcription and analysis, particularly using free utilities such as CLAN that additionally link the sample to an audio- or video-recorded record of the child’s actual speech sample, can greatly improve the speech, accuracy, and informativeness of language sample analysis and, by extension, can also aid in clinical assessment, therapy planning, and measurement of therapeutic progress.

In this chapter, we will illustrate the utility of LSA conducted using CLAN and the KIDEVAL utility that uses two separate datasets. The first is a large cohort of very young children followed as part of a single research study. The second is a review of data obtained from the CHILDES Project Archive that we use to evaluate the potential utility of certain LSA measures at particular ages. Many LSA measures lack robust normative or comparison reference values, therefore the data in CHILDES can greatly augment what we currently know through measures such as MLU, DSS, IPSYN, VOCD, and others.

KIDEVAL in Action

In this section, we summarize how we have used the KIDEVAL utility to assess the dyadic interactions of a large cohort of infants and their mothers ($n = 125$), who were sampled at 7, 10, 11, 18, and 24 months as part of a larger study examining possible predictors of later child language skills (Newman, Rowe, and Ratner 2015). The scope of the project was quite daunting: we had ~125 families and conducted 5 play sessions, with both child's and mother's verbal interaction being a focus of analysis. This produced a total of roughly 1,250 quarter- to half-hour minute transcripts. Given traditional estimates of time required per transcript to compute multiple measures, we estimated a total time commitment of 6,250 hours to finish this part of the project, and the granting agency did not, in fact, predict that we would obtain any findings during the actual grant time window. However, *they were wrong*. This is because CLAN media linkage in Walker Controller, a CLAN program utility for transcription of spoken language, allows single keystroke playback of the segment being transcribed. This cuts down the time required to make an accurate transcript of the child's sample by roughly 75%. Moreover, because the transcriber can easily repeatedly compare the transcription to the original, accuracy is increased.

Next, we used the automated MOR function to assign and disambiguate grammatical descriptions of all the words in these 1,250 transcripts. The command "mor *.cha" will run MOR, POST, and MEGRASP in sequence on all target transcript files. The output has the form of this excerpt:

```
*CHI:mommy this xxx .
%mor:n|mommy pro:dem|this .
*CHI:these shoes on .
%mor:pro:dem|these n|shoe-PL adv|on .
*MOT:okay I can get her shoes on .
%mor:adj|okay pro:sub|I mod|can v|get det:poss|her n|shoe-PL adv|on .
*CHI:+< tiger .
%mor:n|tiger .
```

```
*MOT:is that a tiger ?
%mor:cop|be&3S pro:rel|that det:art|a n|tiger ?
*MOT:or is that a zebra ?
%mor:coord|or cop|be&3S pro:rel|that det:art|a n|zebra ?
*CHI:zebra .
%mor:n|zebra .
```

Following the running of MOR and POST, we then used the KIDEVAL command to generate spreadsheet output of each child's (and parent's) language features on more than two dozen variables. Some of these variables, such as pause length and MLU, are common across languages; others involving specific morphological features are unique and configurable to each language.

What about Norms?

In reviewing the literature on clinical use of language samples, LSA appears to be used most often when standardized test data cannot be obtained or are difficult to interpret. It seems to be particularly favored for assessment of very young children. However, there are conceptual issues in LSA for children at 24 months of age, which was the outcome measurement period for the toddlers in our study. Many of the normative or reference values are based on relatively few cases at lowest age ranges. For example, for MLU, a relatively recent report (Rispoli, Hadley, and Holt 2008) included 37 children at 24 months. Miller and Chapman (1981), the classic reference for MLU in clinical practice, reported on only 16 children in this age bracket, while the largest recent study to report expected values for MLU (as well as number of different words, NDW) (Rice et al. 2010) had 17 typically developing and 6 late-talking participants in the age bracket from 2;6 to 2;11. These are not extremely large populations on which to generalize impressions of a child's linguistic profile, which is why some researchers have expressed serious concerns about using MLU to identify whether a child is typically developing or impaired (Eisenberg, Fersko, and Lundgren 2001).

For Type-Token Ratio (TTR) or NDW, the situation is similar, since most of the studies referenced above also reported these measures, and few additional studies are available. For DSS and IPSYN, reference cohorts are similarly restricted. DSS reference tables report on only 10 children from 24 to 27 months of age (Lee 1974). In this age range, IPSyn provides data for 15 children (Scarborough 1990).

Our study does not intend to contribute normative data on these measures at this time. However, we can illustrate how the children in our study performed on these measures (all were typically developing, as is often the case in research reports taken from relatively high SES families). In general, data from this sample show values for MLU, DSS, and IPSYN that are consistent with prior, smaller samples (see figures 8.1–8.3).

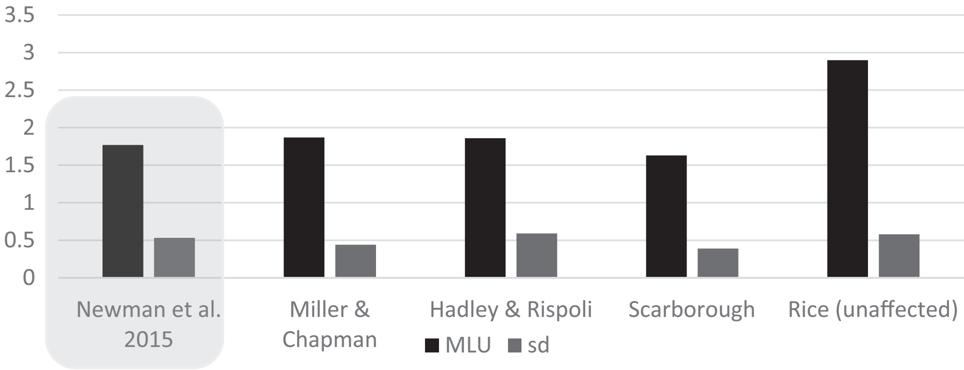

Figure 8.1
MLU values from prior research reports for children at 24 months of age. Note: Current = Newman et al., 2015, $n = 122$; Rice cohort is 2;6–2;11; combined n from other studies = 68.

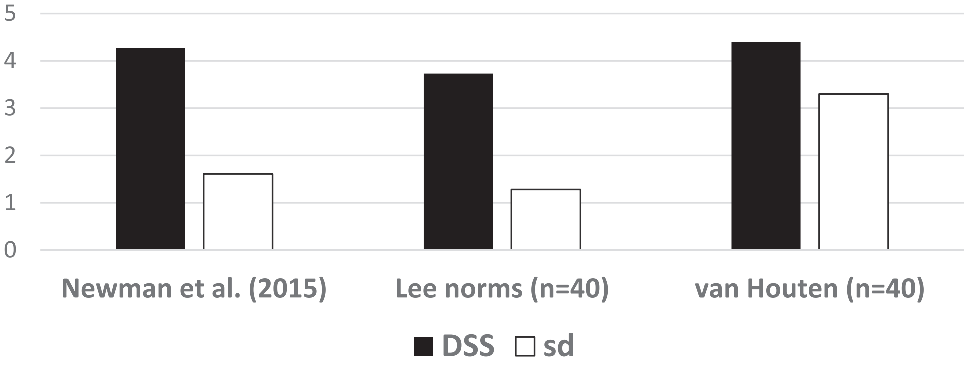

Figure 8.2
Developmental Sentence Score (DSS) values from Newman et al. (2015), reference values reported by Lee (1974), and values derived from the CHILDES van Houten corpus.

These data suggest that KIDEVAL is a useful clinical tool for the assessment of spontaneous language data in 24-month-old children, a group for which few robust measures of LSA performance exist. Our results are comparable, and computed automatically, to data derived from much more time-intensive manual coding. However, we do note that the unaffected sample of Rice et al. did achieve higher MLU values than the other comparison cohorts.

We also computed correlations among LSA values and standardized test outcomes at 24 months of age. We obtained significant but weak correlations that probably justify larger studies of the available measures for toddlers and their construct validity. For instance, we correlated the children's MLU with IPSYN and DSS values; correlations were significant. This should not be surprising, since both IPSYN and DSS award points for various

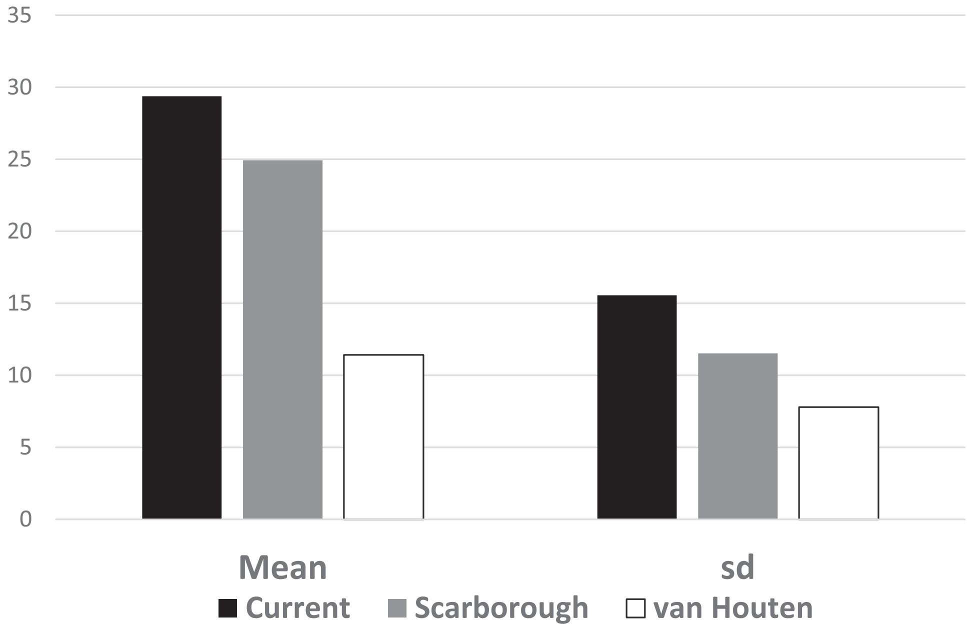

Figure 8.3
IPSYN values for Newman et al. (2015, "Current"), Scarborough (1990), and van Houten corpus (MacWhinney 1991).

syntactic elements, and utterances with longer MLU values have greater opportunity to contain such features. However, it is perhaps surprising that the actual correlations are relatively low, even though they reach significance given our large sample size. (See figures 8.4–8.6.) In particular, DSS correlates more poorly with MLU than does IPSYN, in all likelihood because fewer utterances at 24 months meet DSS eligibility standards and because very early utterances do not achieve DSS sentence points. Likewise, IPSYN and DSS do not correlate well with one another, probably for the same reasons, indicating that they are not interchangeable assessments of a toddler's language sample.

Improving Norms

Our study suggests that, at young ages in English, some potential LSA measures do not appear to be measuring the same constructs. Clearly, a single LSA measure (especially MLU, which has been critiqued extensively; Eisenberg, Fersko, and Lundgren 2001) cannot provide the whole picture, and doing multiple LSAs is much too time consuming, unless more researchers and therapists use computer-assisted analysis to generate data that are more responsive to these concerns. We are, however, encouraged by the fact that the data from our large sample of toddlers do resemble those in smaller reference study

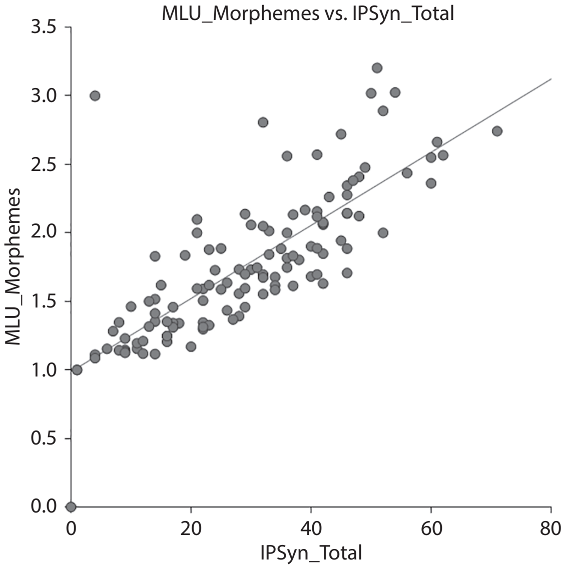

Figure 8.4
Correlation between MLU and IPSyn, r = .78, p = .000.

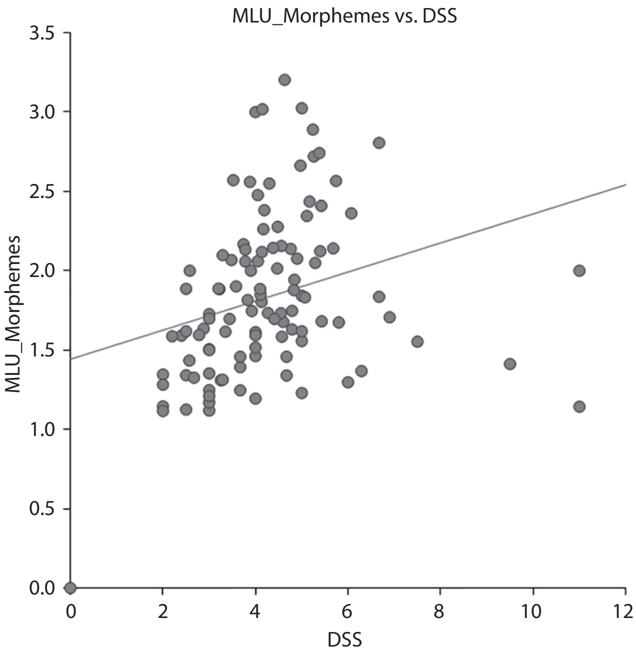

Figure 8.5
Correlation between MLU and DSS, r = .284, p = .003.

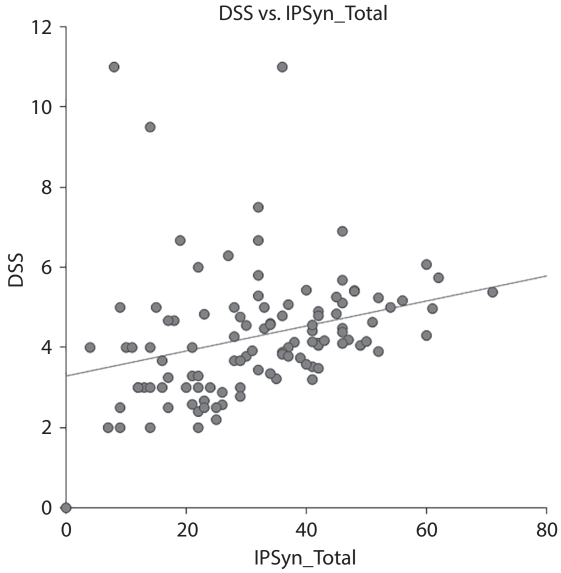

Figure 8.6
Correlation between DSS and IPSyn, r = .283, p = .00.

reports. We also believe that psychometric evaluation of confidence intervals around mean values will be necessary to improve the robustness of measures such as DSS and IPSYN for distinguishing between typical and atypical performance, even though we do have some data to inform this decision-making process.

Fuller Support for SLPs

We are current working to move the CHILDES Project Archive from a repository and resource for researchers to a dynamic source of reference data that can be used to assess and treat children across the world's languages. To this end, the TalkBank project is working to take the following actions that should greatly enhance clinicians' abilities to apply LSA to a broader range of children more easily and insightfully:

1. Increase the number of languages that can be automatically parsed and reported using CLAN utilities. As other contributors to this volume note, the free CLAN utilities now have grammars for a large number of languages; this number is growing yearly. Thus, clinicians working in Spanish, French, German, Dutch, Mandarin, Cantonese and other frequently used languages now have resources to perform accurate LSA of languages other than English.
2. Deploy existing corpora in the CHILDES Archive to improve "norms" for commonly used LSA outcome measures.

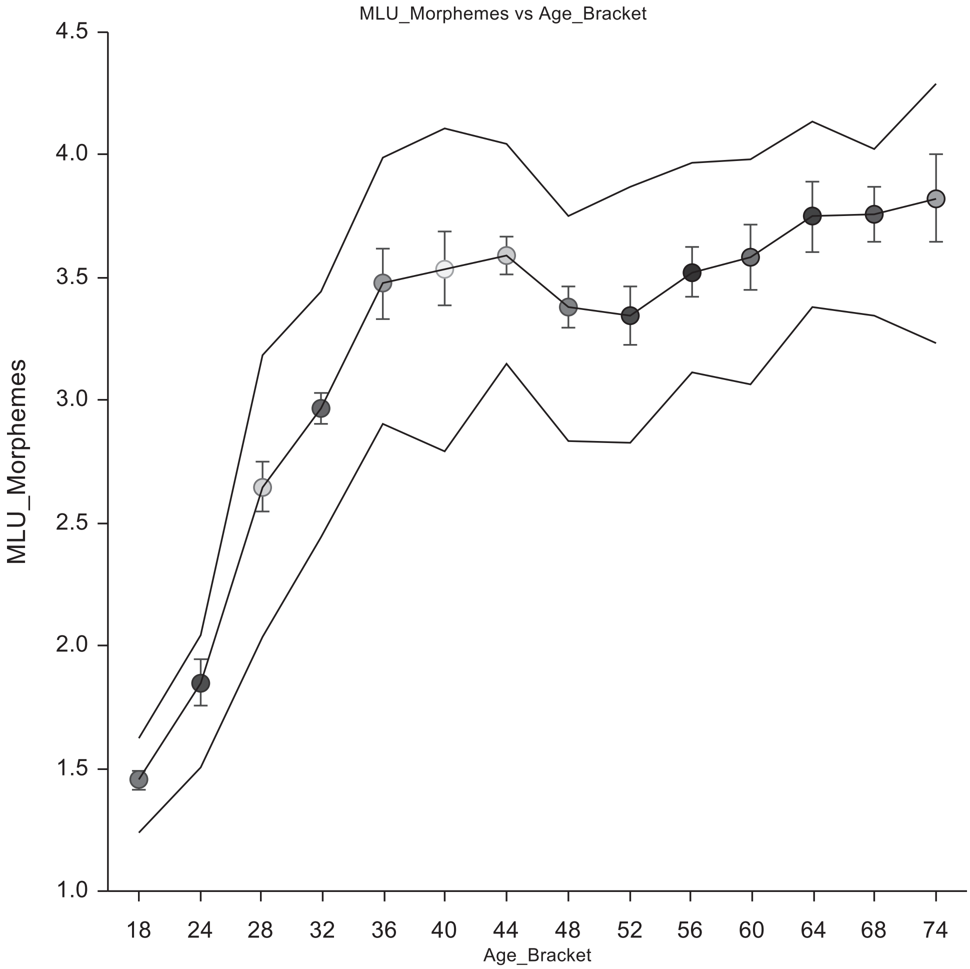

Figure 8.7
MLU values for 630 children in the CHILDES Archive.

We are currently in the process of completing this second ambitious task. Recently, we completed KIDEVAL analysis of a large set of corpora ($n = 630$ children), all of whom spoke North American English, and all of whom were engaged in free play with their parents (a similar context). Results have been fairly interesting, and we provide only a brief taste of our findings here. First, we are happy to note that Roger Brown's (1973) observation that MLU is most useful when the child is fairly young or up until the point that it reaches a value of roughly 4.0 appears to be validated by this large sample, where MLU plateaus for our children past these values and ages (see figure 8.7).

We also note that IPSYN and DSS appear to be differentially sensitive to changes in age, as do two alternative ways of computing lexical (vocabulary) diversity—Type-Token

Ratio (TTR) and vocd (Malvern et al. 2004), a computer algorithm less sensitive to variations in sample size. CLAN reports both in the KIDEVAL utility (see figures 8.8 and 8.9). Similar to our findings reported earlier for the Newman et al. study children, IPSYN and DSS appear to measure different things, particularly across the broader age span covered by the CHILDES data. For example, IPSYN appears more sensitive to growth across very early childhood, whereas DSS appears to be more sensitive at older ages, perhaps as a function of the "sentence point" that provides more credit when a sentence is considered grammatical, an important construct in distinguishing typical from atypical development as children mature.

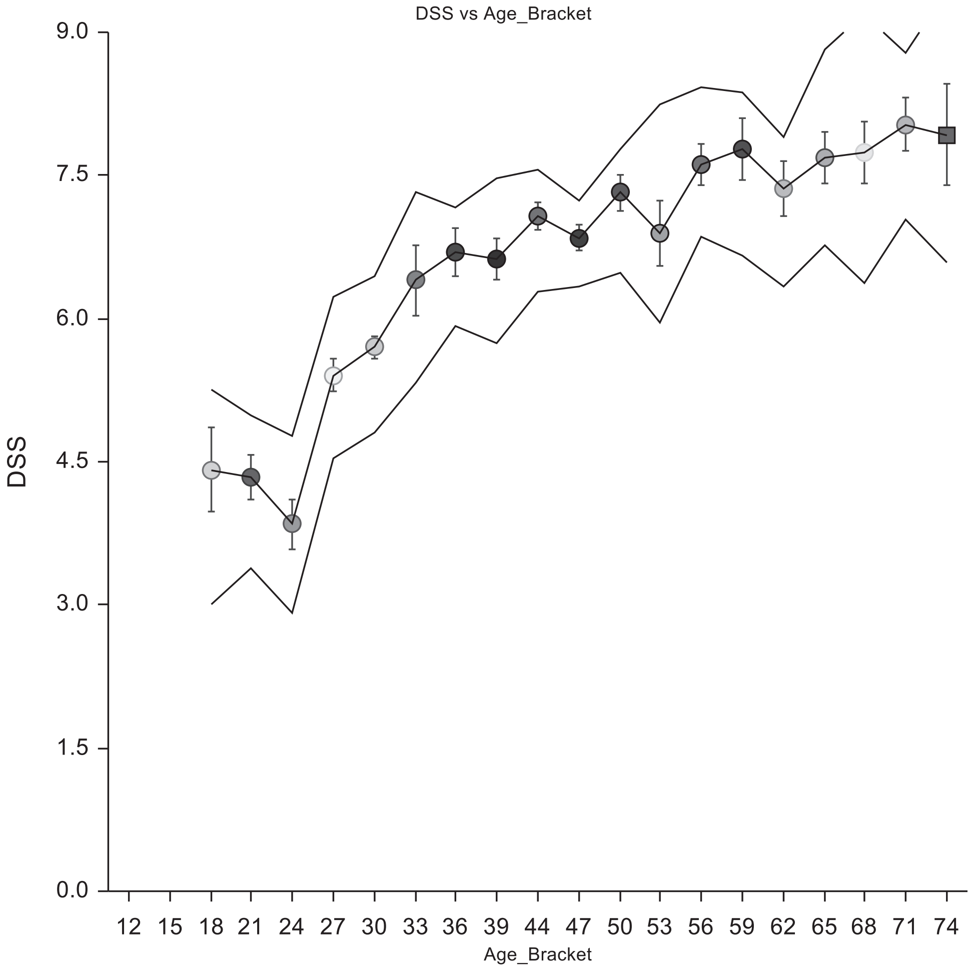

Figure 8.8
DSS scores for 630 children in the CHILDES Archive.

TTR and vocd (see figures 8.10 and 8.11) display a somewhat more difficult profile to evaluate. Vocd appears to track better with age across this sample than does TTR. Currently, vocd is reported in a number of research reports (Pilar 2004; Silverman and Bernstein Ratner 2002; Owen and Leonard 2002; Wong 2010) but has no published norms; we hope to rectify this shortly. TTR has long been known to be vulnerable to a number of issues, particularly sample size; whether Vocd can improve on this to inform clinical assessment remains to be seen. Extending norms and evaluating the utility of various LSA measures is an ongoing initiative of great potential value to SLPs. We also note that there are no robust norms for LSA conducted with bilingual or English Language Learning (ELL)

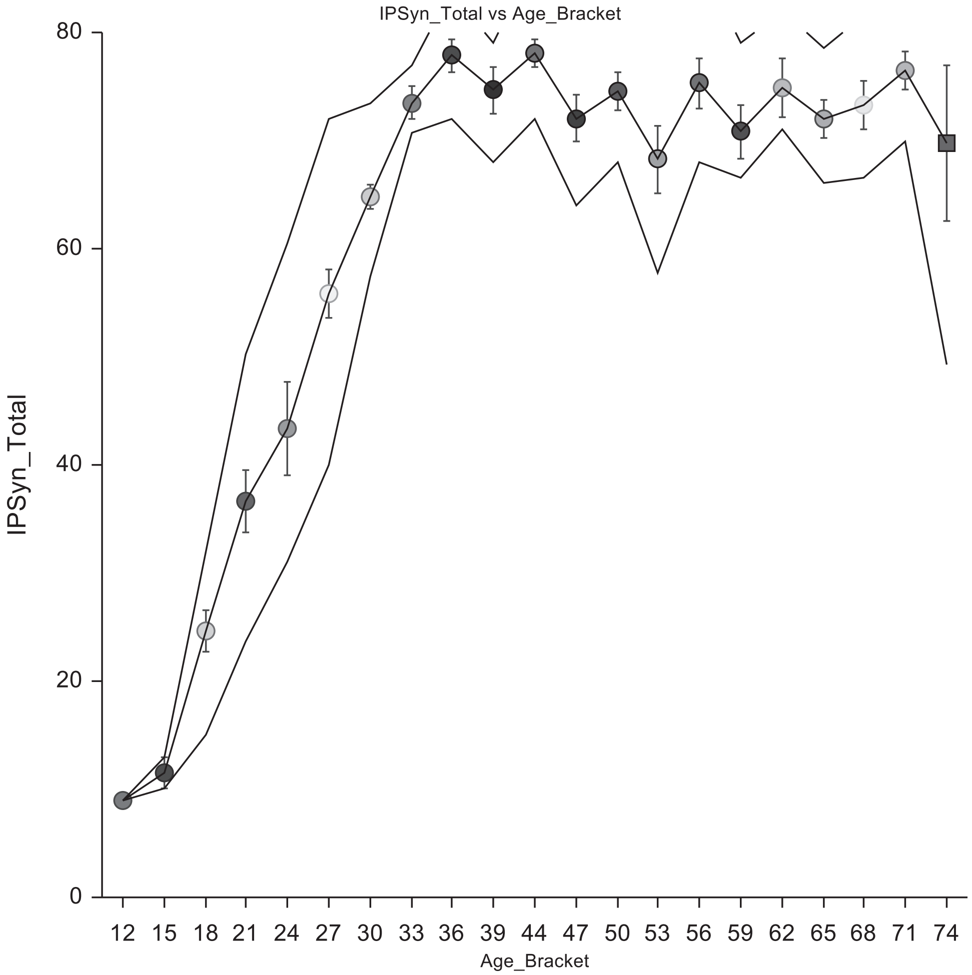

Figure 8.9
IPSyn scores for 630 children in the CHILDES Archive.

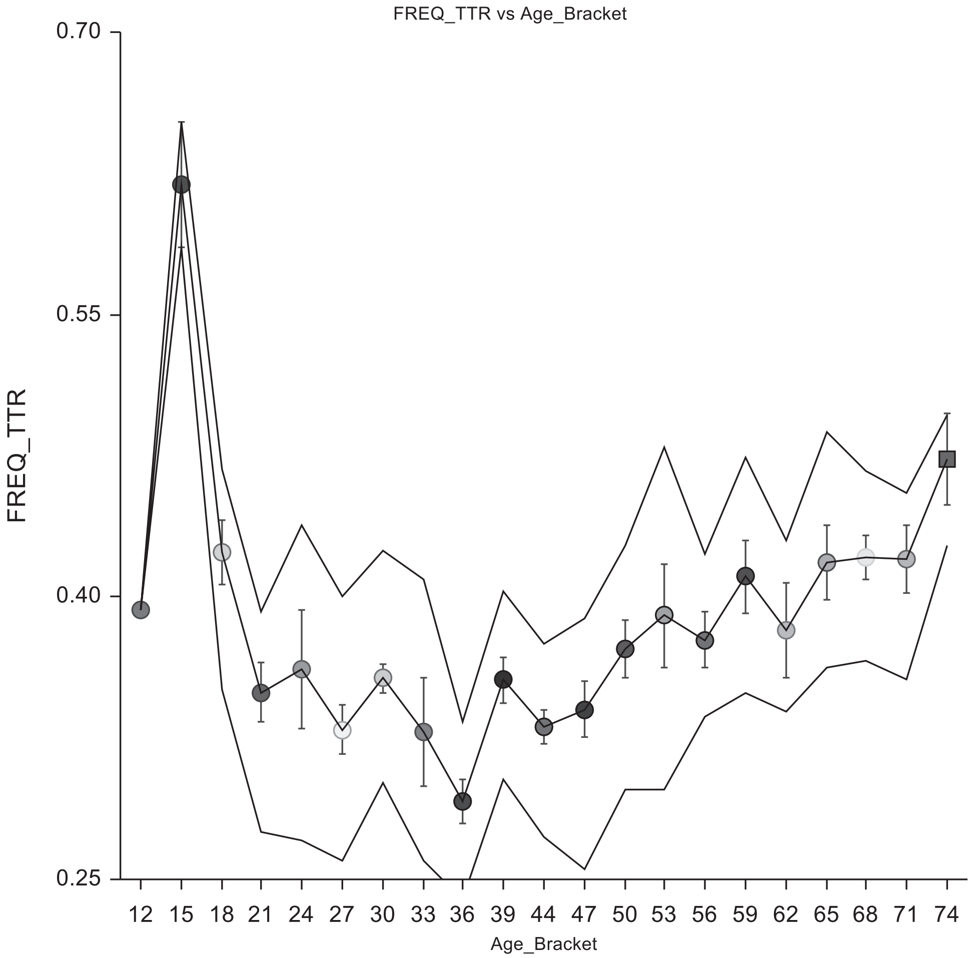

Figure 8.10
TTR values for 630 children in the CHILDES archive.

children, a major clinical cohort where LSA is used, given the parallel lack of standardized assessment norms for this population (Caesar and Kohler 2007).

Take-Away Messages

LSA is an important tool that one can use to appraise and understand child language ability in an ecologically valid way. Having said this, it is underutilized for a number of reasons, primarily because when done "by hand," it is very time-consuming. Because it is time-consuming, we know that clinicians do not fully exploit what can be learned from LSA, transcribing very short samples, and primarily deriving only a few measures such

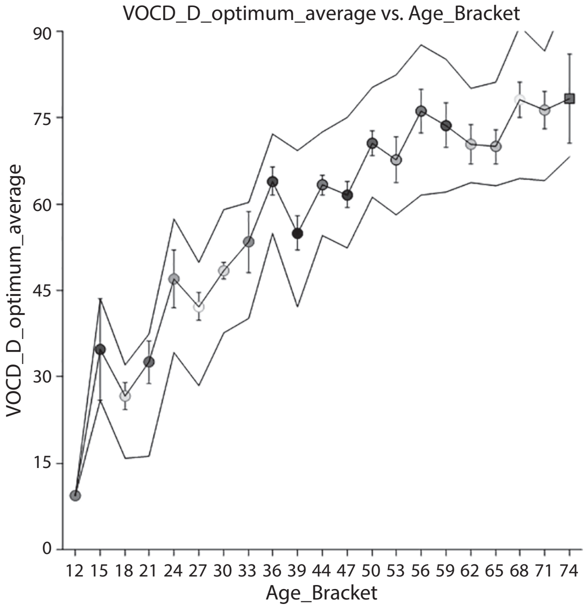

Figure 8.11
vocd values for 630 children in the CHILDES archive.

as MLU, which are not maximally informative for assessment, therapy planning, or outcome measurement. Media-linked transcription, such as is available using the free CLAN utilities available through TalkBank/CHILDES, greatly speeds transcription of a child's language sample. Once completed, this transcript can be used to generate many useful, accurately computed measures of child language performance. These can be used both to augment other assessment measures and to prioritize targets for intervention. Periodic LSA can also judge the child's progress in language growth, using the original LSA as a baseline measure. As clinically focused software evolves, the child's transcript can be paired with other utilities, such as PHON for phonological analysis, or FluCalc for fluency analysis, with little additional effort. CLAN grammatical parsers can also enable clinicians to evaluate bilingual children speaking a variety of languages, a unique benefit when working with a growing and challenging demographic in our profession.

When asked if they would use computer-assisted programs to analyze language samples more quickly and more informatively, the majority of clinicians in a recent survey agreed that they would, if they could identify how to accomplish this (Westerveld and Claessen 2014). We were intrigued to read of a successful pilot program to use SLP assistants or aides to generate transcripts and measures using SALT (Miller 2011), another LSA software program. Thus, we are optimistic that volumes such as this, along with web tutorials and the continued growth of programs available to SLPs, will help clinicians to exploit the potential of LSA more fully. In sum, the CHILDES/TalkBank utilities are an invaluable tool in an SLP's repertoire of clinical resources—free, time-saving, and computationally powerful. So power up your laptop and take computer-assisted LSA for a spin—for we predict that you will become a fast and loyal fan.

Broader Implications

We have examined in depth the ways in which the construction and validation of the KIDEVAL program rely on comparison of a given child language sample with the larger CHILDES database. A similar approach within the EVAL program enables us to compare a transcript from a given person who has aphasia with the fuller AphasiaBank database of 408 PWAs and 254 normal controls. Currently, we have only applied these methods for English and French, but they should work equally well for all 10 languages for which we can automatically compute morphosyntactic analyses.

We plan to build on our ability to automatically compute a wide variety of measures such as MLU, IPSyn, DSS, TTR, and 12 others, by developing norm-referenced clinical profiles such as KIDEVAL (for children) and EVAL (for adults with language impairment). Although a measure such as MLU involves a single construct, measures such as DSS, IPSyn, and QPA (Rochon et al. 2000) involve a complex combination of dozens of decisions about grammatical categories and errors. Using programs to automatically compute variant combinations of these underlying decisions, we will be able to learn which pieces of these larger scoring systems are most predictive of the actual level of language acquisition during development, using age as a proxy for developmental level. Work by Lubetich and Sagae (2014) has already shown that approaches based on data-mining methods such as classifier construction may be able to outperform these older standard measures. By gaining automatic access to large corpora that can be automatically analyzed, we will be able to test out these new and exciting possibilities for clinical diagnosis and developmental evaluation.

Acknowledgments

This work was supported by NSF grants BCS-1626294 and BCS-162-300, NIDCD grant DC008524, and NICHD grant HD082736 to Brian MacWhinney, and NIDCD grants DC015494 and DC017152, to Brian MacWhinney and Nan Bernstein Ratner, respectively.

References

Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge: Harvard.

Caesar, L. G., and P. D. Kohler. 2007. "The State of School-Based Bilingual Assessment: Actual Practice versus Recommended Guidelines." *Language, Speech and Hearing Services in Schools* 38:190–200.

Clahsen, H., and M. Rothweiler. 1992. "Inflectional Rules in Children's Grammars: Evidence from German Participles." In *Yearbook of Morphology*, edited by G. Booij and J. Van Marle. Dordrecht, Netherlands: Kluwer.

Cochran, P. S., and Julie J. Masterson. 1995. "Not Using a Computer in Language Assessment/Intervention in Defense of the Reluctant Clinician." *Language, Speech, and Hearing Services in Schools* 26:213–222.

Eisenberg, S. L., T. M. Fersko, and C. Lundgren. 2001. "The Use of MLU for Identifying Language Impairment in Preschool Children: A Review." *American Journal of Speech-Language Pathology* 10:323.

Evans, J. L., and J. Miller. 1999. "Language Sample Analysis in the 21st Century." *Seminars in Speech and Language* 20:101–198.

Finestack, L. H., and K. Satterlund. 2018. "Evaluation of an Explicit Intervention to Teach Novel Grammatical Forms to Children with Developmental Language Disorder." *Journal of Speech, Language, and Hearing Research* 61:2062–2075.

Freudenthal, D., J. Pine, and F. Gobet. 2010. "Explaining Quantitative Variation in the Rate of Optional Infinitive Errors across Languages: A Comparison of MOSAIC and the Variational Learning Model." *Journal of Child Language* 37:643–669.

Gorman, K. 2010. "Automated Morphological Analysis of Clinical Language Samples." Unpublished manuscript.

Hassanali, K.-N. 2014. "Automatic Generation of the Index of Productive Syntax for Child Language Transcripts." *Behavior Research Methods* 46:254–262.

Heilmann, John. 2010. "Myths and Realities of Language Sample Analysis." *Perspectives on Language Learning and Education* 17:4–8.

Hux, K. 1993. "Language Sampling Practices: A Survey of Nine States." *Language, Speech, and Hearing Services in Schools* 24:84–91.

Kemp, K., and T. Klee. 1997. "Clinical Language Sampling Practices: Results of a Survey of Speech-Language Pathologists in the United States." *Child Language Teaching and Therapy* 13:161–176.

Lee, L. 1974. *Developmental Sentence Analysis*. Evanston, IL: Northwestern University Press.

Lee, Laura, and Susan Canter. 1971. "Developmental Sentence Scoring: A Clinical Procedure for Estimating Syntactic Development in Children's Spontaneous Speech." *Journal of Speech and Hearing Disorders* 36:315–340.

Long, S., and R. Channell. 2001. "Accuracy of Four Language Analysis Procedures Performed Automatically." *American Journal of Speech-Language Pathology* 10:212–225.

Lubetich, Shannon, and Kenji Sagae. 2014. "Data-Driven Measurement of Child Language Development with Simple Syntactic Templates." COLING 2014, Dublin, Ireland.

MacWhinney, Brian. 1991. *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Erlbaum.

MacWhinney, Brian. 2008. "Enriching CHILDES for Morphosyntactic Analysis." In *Trends in Corpus Research: Finding Structure in Data*, edited by H. Behrens, 165–198. Amsterdam: John Benjamins.

MacWhinney, Brian, and J. Leinbach. 1991. "Implementations Are Not Conceptualizations: Revising the Verb Learning Model." *Cognition* 29:121–157.

Malvern, David, Brian Richards, Ngoni Chipere, and Pilar Purán. 2004. *Lexical Diversity and Language Development*. New York: Palgrave Macmillan.

Marcus, Gary F., Steven Pinker, Michael Ullman, Michelle Hollander, T. John Rosen, Fei Xu, and Harald Clahsen. 1992. "Overregularization in Language Acquisition." *Monographs of the Society for Research in Child Development*, i–178.

Miller, J. F. 2001. "Focus on Schools: Having Trouble Monitoring Language Intervention? Language Sample Analysis is the Solution." *ASHA Leader* 6:5.

Miller, J. F. 2011. *Assessing Language Production Using SALT Software*. Middleton, WI: SALT Software.

Miller, Jon F., and Robin S. Chapman. 1981. "The Relation between Age and Mean Length of Utterance in Morphemes." *Journal of Speech, Language, and Hearing Research* 24 (2):154–161.

Newman, Rochelle, Meredith Rowe, and Nan Bernstein Ratner. 2015. "Input and Uptake at 7 Months Predicts Toddler Vocabulary: The Role of Child-Directed Speech and Infant Processing Skills in Language Development." *Journal of Child Language*, 1–16.

Overton, S., and Y. Wren. 2014. "Outcome Measurement Using Naturalistic Language Samples: A Feasibility Pilot Study Using Language Transcription Software and Speech and Language Therapy Assistants." *Child Language Teaching and Therapy* 30:221–229.

Owen, A. J., and L. B. Leonard. 2002. "Lexical Diversity in the Spontaneous Speech of Children with Specific Language Impairment: Application of D." *Journal of Speech, Language and Hearing Research* 45:927–937.

Pilar, D. N. 2004. "Developmental Trends in Lexical Diversity." *Applied Linguistics* 25:220–242.

Pine, J. M., and E. V. M. Lieven. 1997. "Slot and Frame Patterns and the Development of the Determiner Category." *Applied Psycholinguistics* 18:123–138.

Pinker, S., and A. Prince. 1988. "On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition." *Cognition* 29:73–193.

Price, L. H., S. Hendricks, and C. Cook. 2010. "Incorporating Computer-Aided Language Sample Analysis into Clinical Practice." *Language, Speech, and Hearing Services in Schools* 41:206–222.

Rice, Mabel, Filip Smolik, Denise Perpich, Travis Thompson, Nathan Rytting, and Megan Blossom. 2010. "Mean Length of Utterance Levels in 6-Month Intervals for Children 3 to 9 Years with and without Language Impairments." *Journal of Speech, Language, and Hearing Research* 53:333–349.

Rispoli, Matthew, Pamela Hadley, and Janet Holt. 2008. "Stalls and Revisions: A Developmental Perspective on Sentence Production." *Journal of Speech, Language and Hearing Research* 51 (4): 953–966.

Rochon, E., E. Saffran, R. Berndt, and M. Schwartz. 2000. "Quantitative Analysis of Aphasic Sentence Production: Further Development and New Data." *Brain and Language* 72:193–218.

Scarborough, Hollis S. 1990. "Index of Productive Syntax." *Applied Psycholinguistics* 11:1–22. doi: 10.1017/S0142716400008262.

Silverman, Stacy, and Nan Bernstein Ratner. 2002. "Measuring Lexical Diversity in Children Who Stutter: Application of vocd." *Journal of Fluency Disorders* 27 (4): 289–304. doi: 10.1016/s0094–730x(02)00162–6.

Valian, Virginia, Stephanie Solt, and John Stewart. 2009. "Abstract Categories or Limited-Scope Formulae? The Case of Children's Determiners." *Journal of Child Language* 36 (04): 743–778.

Westerveld, M. F., and M. Claessen. 2014. "Clinician Survey of Language Sampling Practices in Australia." *International Journal of Speech-Language Pathology* 16:242–249.

Wexler, K. 1998. "Very Early Parameter Setting and the Unique Checking Constraint: A New Explanation of the Optional Infinitive Stage." *Lingua* 106:23–79.

Wong, Amy. 2010. "Differentiating Cantonese-Speaking Preschool Children with and without SLI Using MLU and Lexical Diversity (D)." *Journal of Speech, Language, and Hearing Research* 53:794–799.