The Thesis Committee for Erin Adams Dowd
Certifies that this is the approved version of the following thesis:

Evaluating morphosyntactic differences in narrative re-tell tasks between bilingual children with and without language impairment using computational methods

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Lisa M. Bedore

Elizabeth D. Peña

Evaluating morphosyntactic differences in narrative re-tell tasks between bilingual children with and without language impairment using computational methods

by

Erin Adams Dowd, B.A.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2017

Dedication

To my mother, who has given me a love and encyclopedic knowledge of children’s books that I hope to put to good use for future therapy planning, my father, who has shown me how to take a hit, take a stand, and get things done, and my sister, for listening when I needed to talk and talking when I needed to listen.

Acknowledgements

Many thanks are due to my supervisor, Dr. Lisa Bedore, for her time, her guidance, and for every allowance made by her patience and tolerance over the past two years. I have been the needful recipient of many gentle prods in new directions and towards the constraints of reality and reason, all prefaced by “Well…”.

To Dr. Liz Pena, my reader, for her patience, her exuberant explanations of statistics and diagnostic measures.

To Dr. Mirza Lugo-Neris, for her patience, laughter, sanity, and advice on data analysis, all things narrative, and keeping an even keel.

To Dr. Jessica Franco, for taking the time to teach me how to design therapy around individual clients.

To Nahar Albudoor, Jissel Anaya, Alissa Baron, Jennifer Briseno, Scott Marcus, Rachel Tessmer, Ying Hao, Alissa Baron, Dr. Prarthana Shivabasappa, and many more at the HABLA lab and the University of Texas at Austin for commiserating and keeping me on task.

To my externship supervisors, Viki Schmidt and Kathryn Howell, for taking the time to listen to me talk about my thesis much more than was justified.

To Dr. Natasha Cabrera, Dr. Jenessa Malin, and Dr. Liz Karberg-Fein at the FAMILIA lab at the University of Maryland, for introducing me to CLAN and the CHILDES database, and giving me my first introduction to working on bilingual research.

To Mel Chua, Eric VanWyk, Mike Willems, Ginneh Cornelius, Kelcy Adamec, Eric Hughes, Jaime Ravenet, Susan Wheatley, Katie Dodd Syk, and everyone at French House for their inspiration and encouragement.

Abstract

Evaluating morphosyntactic differences in narrative re-tell tasks between bilingual children with and without language impairment using computational methods

Erin Adams Dowd, M.A.

The University of Texas at Austin, 2017

Supervisor: Lisa Bedore

The diversity of linguistic backgrounds and second-language competencies of Spanish-English bilingual school-age children present challenges for accurately diagnosing and treating language impairment. Narrative re-tell samples from Peña, Gillam, & Bedore (2014) were analyzed in two groups of 21 matched language-impaired and typically developing children, aged 4-7 years old attending school in central Texas. Transcribed methods included a custom extension of the Natural Language Processing Toolkit in Python and the IPSYN analytical function in CLAN (MacWhinney, 2000). From these analyses, the complexity and linguistic diversity of nested –ing verb phrases and IPSYN scores were compared across groups. Language-impaired children made significantly more errors in auxiliary verb use, had less diverse vocabulary, and had lower syntactic complexity scores than their typically developing counterparts.

Table of Contents

Chapter I: Introduction: *Using syntactic differences as diagnostic markers*

Speech Language Pathologists are tasked with identifying and treating children with language impairment through early intervention services and school special education services. While a delicate and nuanced task for all populations, diagnosing language impairment is especially challenging in children with unique patterns of language exposure, including bilinguals, defined in this study as children who are regularly exposed to more than one language by the age of 5 in social contexts. Symptoms of language impairment manifest differently in bilinguals than in monolinguals, and the most salient symptoms of language impairment vary between languages and individuals (Kohnert, 2010). As 21% of the United States population over the age of 5 speaks a language other than English in their home (Ryan, 2011), a large portion of the population falls into this category. Furthermore, across demographic categories, about 7.4% of children, on average, may have language impairment (Tomblin, Smith, & Zhang, 1997). While a typically developing second-language learner might have difficulties with speech production in their second language, or make grammatical errors, they will not have similar difficulties in their first language, and these errors will resolve as they master their second language. Children with language impairment, on the other hand, will have problems in every language they speak.

CHALLENGES OF EVALUATING BILINGUALS

Complicating this diagnostic challenge, many clinicians evaluating bilingual children do not speak both of the child's languages (Jordaan, 2008), and many children are initially evaluated in the language of the majority culture, by a monolingual clinician. Standardized tests that are not normed on bilingual populations that match a particular child's linguistic history may not be valid for that child (Goldstein, 2012), and clinicians may not be able to effectively evaluate a bilingual child's ability from existing standardized tests or informal samples alone. Additionally, when learning two different languages at home and at school, even typically developing children will make errors that would be cause for concern in a monolingual child; diagnostic tools must be able to differentiate between linguistic patterns common in second-language learners and children with language impairment. In short, in order to effectively deliver speech and language interventions to bilingual children with language impairment, patterns of second-language production unique to bilingual children with language impairment must be identified, and diagnostic methods that can be used by both monolingual and bilingual clinicians must be developed.

Differentiating between symptoms of language impairment and second language acquisition

As error patterns in children with language impairment often look similar to errors made by typically developing bilingual children (Crago & Paradis, 2003); conventional assessment tools that look only at whether errors exist may overidentify bilinguals as having language impairment. However, the errors common to typically developing second language learners can be distinguished from errors indicative of language impairment with careful analysis (Paradis, Rice, Crago, & Marquis, 2008) and comparison to developmental norms for typically developing children from similar language backgrounds. While bilingual Spanish-English children with language impairment produce more noun-phrase (i.e., clitic and gender) errors than typically developing bilingual peers, the type of errors may vary depending on language input, dialect spoken, and environment (Morgan, Restrepo, & Auza, 2013). In order to reliably use morphosyntactic features to distinguish between bilingual children with and without language impairment, it may be necessary to analyze multiple features; however, unlike standardized tests, this type of analysis is time consuming to do by hand, and requires a deep understanding of all the languages spoken by the child. As many clinicians are expected to evaluate children from a variety of language backgrounds that the clinician may not be familiar with (Jordaan, 2008), analytical tools and processes that can be designed by a person familiar with a given population but implemented by anyone may be diagnostically useful.

Appropriate diagnostic criteria

In cases where a child's linguistic history makes finding an appropriate standardized measure difficult, detailed analysis of language samples can provide the data necessary to distinguish between these groups (Bedore, Peña, Gilliam, & Ho, 2010). Many of the features: phrase complexity, number of words used, measures of grammaticality- analyzed in a language sample are similar in content across languages, if not directly comparable between one language and another. For instance, while two different languages may use very different words, a child's vocabulary diversity can still be easily evaluated from language samples taken in both languages. For instance, a child might be able to name a wide variety of foods and toys in the language she speaks at home, and a large number of animals, shapes, etc. in the language she is learning at school, while being unable to name the same items in the opposite language. Similarly, while phonetic and morphological features vary from language to language, the complexity of structures used by a given child may be compared between two languages. Theoretically, if the complexity of features and structures used by a child could be analyzed automatically, even a monolingual clinician might be able to perform a language sample analysis from transcribed samples in order to diagnose children with language impairment. One way to simplify this process might be to automate it; natural language processing tools exist in multiple languages, and theoretically the same tools that allow Facebook to curate your newsfeed or your spam filter to throw out unwanted mail could be altered to analyze syntactic differences in the language produced by children with and without language impairment.

GENERALIZABILITY OF ERROR TRENDS ACROSS LANGUAGES

While the specific features that children with SLI struggle with vary from language to language, there are general trends that are present across language families. For instance, Swedish monolinguals with SLI also produce predictable errors in verb phrases (Hedenius et al., 2011), and Farsi and Azeri bilingual children with SLI also showed predictable verb phrase differences (Ahadi, Nilipour, Rovshan, Ashayeri, & Jalaie, 2014). While different parsing functions and norms would need to be made for different languages, if a corpus analysis approach can be shown to be effective, these tools could be optimized for use in as many languages as necessary. In reference to Spanish-English bilinguals, trends common to both Romance and Germanic languages should be considered.

Language-specific symptoms of language impairment

In order to perform this type of analysis, comparable features that differ between children with and without language impairment must first be identified. Children with specific language impairment (SLI) produce more errors in functional morphological inflection than typically developing children (Bishop, 1994).

Symptoms of language impairment in English

In general, English-speaking children with SLI have particular trouble with verb usage, producing fewer unique verbs and overusing bare stems more often than typically developing children (Conti-Ramsden & Jones, 1997, Rice, Wexler, & Cleave, 1995). Even when children with language impairment master the morphology of more complex English verb forms, they often struggle to use them in appropriate contexts (Leonard et al., 2007).

In particular, children with SLI appear to overuse the infinitive form of the verb for an extended period of time (Rice et al., 1995); once children with SLI acquire a common infinitive or bare form, they use them at a higher rate than typically developing children (Beverly & Williams, 2004). Children with SLI also acquire morphological agreement features later than typically-developing children and use copula phrases at lower rates than typically-developing children when matched for mean length of utterance (MLU) (Rice et al., 1995), vocabulary diversity (Leonard, Miller & Owen, 2000), to the extent that these features can be used to differentiate between typically-developing (TD) and language impaired (LI) populations by creating a scoring method that weights these different factors together as a composite score (Bedore & Leonard, 1998), including extended verb phrases and copulas using *be* and *do* (Rice, Wexler, & Redmond, 1999).

Possible computational targets in English

One way to address this in a computational method would be to target regular verb inflection patterns, build the necessary syntax tree that word form requires off of that information, and then check the surrounding words in order to see if, say, and auxiliary be is in the correct location in a syntax tree before an –ing verb in English.

Symptoms of language impairment in Spanish

In contrast, Spanish-speaking children with SLI have relatively strong past- and present-tense verb usage, but have more trouble with noun-phrase morphological features, such as clitic usage and agreement in number and gender (Bedore & Leonard, 2005). Unfortunately, these features have fewer variables that would make the speaker's understanding of the underlying function relevant- indeed, the underlying morphological representation of neuter romance clitics is a subject of academic debate (Bonet, 1995; Pescarini, 2010). Furthermore, the salient features distinguishing the grammatical and ungrammatical forms are usually binary in Spanish: even if a child does not yet understand the rules of Spanish gender agreement, for instance, they have at least a 50% accuracy every time they choose between *el* and *la*; this makes clitic and article agreement difficult to use as a diagnostic marker at single-node determiner phrase level. Nested or distributed noun phrases can test a speaker's command of attribute and agreement rules more rigorously, but probing with a longer phrase also increases the difficulty of the task significantly and makes greater demands of short term memory. Fortunately the frequency with which noun phrases are used in speech provides many opportunities for the speaker to use these features, and just as many opportunities for a clinician to assess a speaker's command of these features; while this measure may be of questionable use in a small number of utterances, that utility increases over a longer sample. Additionally, analyzing noun phrases allows for a more varied semantic analysis than a verb analysis alone would allow, as children frequently know a large number of nouns even if they have not yet acquired the rules of attribute agreement.

STRENGTHS OF EVALUATING BILINGUALS USING MULTIPLE FEATURES

As the features that children with specific language impairment struggle with are different in both Spanish and English, this suggests that weaknesses in specific areas of morphology may be language specific. If these features vary across languages, and a clinician may be called upon to evaluate bilingual children from a variety of linguistic backgrounds, a feature-specific test may not be able to evaluate bilingual children effectively, especially if the clinician is forced to use a translated test, or to modify existing materials at his or her disposal. However, a test that evaluates multiple features may be more useful across multiple languages.

Methods of evaluating multiple features

One way to efficiently analyze multiple linguistic features in narrative samples is to use natural language processing tools and statistical analyses to compare how often particular features, structures, and words appear in different configurations and distributions between groups; Solorio et al. (2011) have shown the efficacy of doing this type of analysis on the same dataset analyzed in this study to identify language dominance. While Solorio et al. (2011) looked at words and number-based phrase units, the analyses performed there were done without reference to larger phrase structures; instead of being structure-driven, predictive text strategies parsed sentences based on word order, and was limited by ungrammatical or out-of-context utterances. Building on those findings, this study seeks to do a more focused analysis using natural language processing tools available in Python's Natural Language Processing Toolkit to break down noun and verb phrase usage in narrative samples at the phrase level, and drive relational analyses of those phrases from not only the heads of each phrase but also from the phrase structures themselves. This structure-based analysis will allow language use to be analyzed with more context than the previous analysis had access to, and will allow for a more general picture of each subject's language use patterns to be built.

Computational methods of analyzing morphosyntax

Corpus analysis using the NLTK has been done for literary texts to analyze language use, semantics, content, and for a variety of language-based projects and analytical purposes (Perkins, 2010); however, the default processing methodology of the functions provided in the NLTK depend on text being in grammatically correct English. While the NLTK is not made for this purpose, similar tools have been used to measure children's syntactic complexity (Lu, 2009), and existing markup conventions such as that used in the CHILDES (MacWhinney, 2000) and SALT databases provide a foundation to make new tools for these tasks. In order analyze utterances typical of children in general and language-impaired children in specific, extensions of these functions needed to be written that were capable of analyzing text with and without obligatory elements, and these tools needed to be extended to work in Spanish as well. In theory, once a working model for a given language is developed, and appropriate measures and useful analyses identified, a program set up for any one transcribed language sample should also work on other language samples transcribed in the same format. Similarly, once a program is developed to analyze one sample, the same code can be used to analyze a large set of similar samples without significant additional work, or used to compare one child's narrative to a set of norms or a database collected from children with similar characteristics. Additionally, this type of analysis can be tweaked and repeated for research purposes. In this study, these tools will be used to find new, statistically significant differences between language usage, vocabulary, and types of morphosyntactic errors within and between typically developing and language impaired Spanish-English bilingual children.

GOALS OF THIS STUDY

Beyond providing a novel analysis of this dataset and a proof of concept of our extension of the NLTK, this study builds on guidelines developed by other researchers (Bedore & Leonard, 1998; Hadley & Holt, 2006; Gladfelter & Leonard, 2013) to evaluate the most significant factors found in our analysis, and tests those factors' ability to differentiate between language impaired and typically developing children.

Chapter II: Methods: *Implementing computational methods to analyze ungrammatical narrative samples*

DATA COLLECTION

Participants, Spanish-English bilingual children between the ages of 4-7, were recruited from Austin area schools, and given a battery of tests and measures to diagnose them as being typically developing or language impaired during an earlier study (Peña et al., 2014). Participants were then asked to tell a story in English and Spanish from a wordless picture book; English and Spanish narrative samples were obtained at different times.

Transcription

These narratives were transcribed in SALT by student research assistants fluent in Spanish and English, and the files anonymized, and first converted to CLAN files and subsequently to marked-up text files for analysis. Demographic data was used to match typically developing and language impaired children with each other for later comparison purposes. Monolingual comparison data and bilingual training data came from anonymized samples available in the CHILDES database (MacWhinney, 2000; Silva-Corvalan, 1989), and were transcribed using CLAN, and then converted to marked-up text files for analysis. All CHILDES samples were collected in accordance with their home institutions' IRB requirements, and have been released with permission to use them for further research. All candidates with complete, formatted SALT-transcribed transcripts with more than 50 child utterances were selected to be part of the analysis group.

Data cleaning

The resulting text files were first cleaned to remove superfluous text and markup characters, ensuring that all analyzed files began in the same format. Words, phrases, and sentences in these samples were then labeled using a combination of SALT and CLAN markup and an extension of the Natural Language Processing Toolkit in Python to find extended verb and noun phrases in English (i.e., the modal verb marker "mv" and auxiliary markers "auxbe" were used from SALT, the person-indicators from CLAN were used to identify and filter out speakers, and subject and verb markers, along with null markers to indicate empty locations in the syntax tree, were used as additional markers in the NLTK extension); children got credit for appropriate word-level semantic content in either language, but only received credit for verb or agreement morphological features appropriate for the language of the carrying sentence. Omitted obligatory or inappropriate verb phrases were marked, but children were given credit for sentences whether or not they were grammatical. Further analysis in Python was abandoned in Spanish due to inadequate available parsers, and insufficient time to write an adequate one.

Choosing tagging and parsing conventions

In order to perform these analyses, existing SALT transcripts were converted to text files, and existing labeling that conflicted with NLTK tags were removed. Files were formatted and parsed according to content and language use, and then broken down into expanded noun and verb phrases.

Choosing targets to analyze

After these lists of sentence structures and word sequences were generated, post-processing analyses targeting specific morphemes and grammatical structures. An exploratory analysis of frequency table data generated by Solorio et al. (2011) was used to choose targets based on the frequency with which TD and LI children appeared to use specific verb forms. From this analysis and a general exploratory survey using an NLTK extension developed for this purpose, it was decided to focus specifically on

NP+AuxBe+Going+VP/PP (i.e. “He is going running”/”He is going to the store”) syntactic structures, as these provided, ideally, three different opportunities to analyze how children were producing verb phrases, and how children were resolving agreement issues across the syntax tree.

USING CLAN-NATIVE IPSYN AS A COMPARISON MEASURE

In order to evaluate this prototype method against other available methods, it was necessary to find another computational method of scoring syntactic complexity. One method of scoring the syntactic complexity of an utterance is to use the Index of Productive Syntax (IPSYN; Scarborough, 1990). An extension of the CLAN program (MacWhinney, 2000) is now able to perform IPSYN calculations on CHILDES-formatted files (Sagae, Lavie, & MacWhinney, 2010). While the narrative re-tell samples in this dataset frequently had fewer than the 100 utterances needed to calculate IPSYN scores that could be directly compared to normative data, this provided a similar analysis to that attempted by the prototype method. As such, IPSYN scores calculated using CLAN are used here to evaluate the strength of using syntactic analysis to distinguish between language produced by language impaired and typically developing children in general, and to evaluate the strength of the novel technique described here in particular.

Calculating sensitivity and specificity

Using the prototype NLTK-extension method and the CLAN-native IPSYN calculation method, the diagnostic utility of each was evaluated by choosing a cut-score for “diagnosis” and calculating the sensitivity and specificity of each measure. The most effective scoring method for the IPSYN scores was then used against a set of narratives generated by Spanish-English bilingual children in Los Angeles that is publicly available in the CHILDES database to check the efficacy of the proposed IPSYN scoring method (Silva-Corvalan, 1989).

CHALLENGES OF USING THESE COMPUTATIONAL METHODS

Ungrammaticality

One challenge that arose was that the majority of existing natural language processing tools rely on source data being grammatically correct; this is frequently not the case in narrative samples generated by children, especially children with language impairment. In order to overcome this challenge, some new methods were written to replace the standard code of the NLTK; for instance, if an obligatory word is missing, the new code creates a syntax tree with an item missing, but with the overall tree intact.

Expanding syntax trees with null values

New code was also written to build a model syntax tree around targeted morphemes, i.e. words with –ing endings, so the number of missing elements could be compared while the child still received credit for attempting a more complex structure than they produced. These identified phrases were then sorted by their core word, morphology, and length and output to a new data structure for frequency and accuracy analyses. These initial analyses revealed that "going" verb phrases occurred most frequently with multiple nested verb phrases; subsequent analyses focused on the length, complexity, frequency, and types of errors present in these AuxBe+-ing verb phrase structures.

EVALUATING DIAGNOSTIC UTILITY

The sensitivity and selectivity of each measure was calculated, and the predictive abilities of each verb and noun phrase complexity measure were compared to each other.

Chapter III: Results: Diagnostic utility of different computational methods of morphosyntactic analysis

VOCABULARY DIFFERENCES BETWEEN TD AND LI CHILDREN

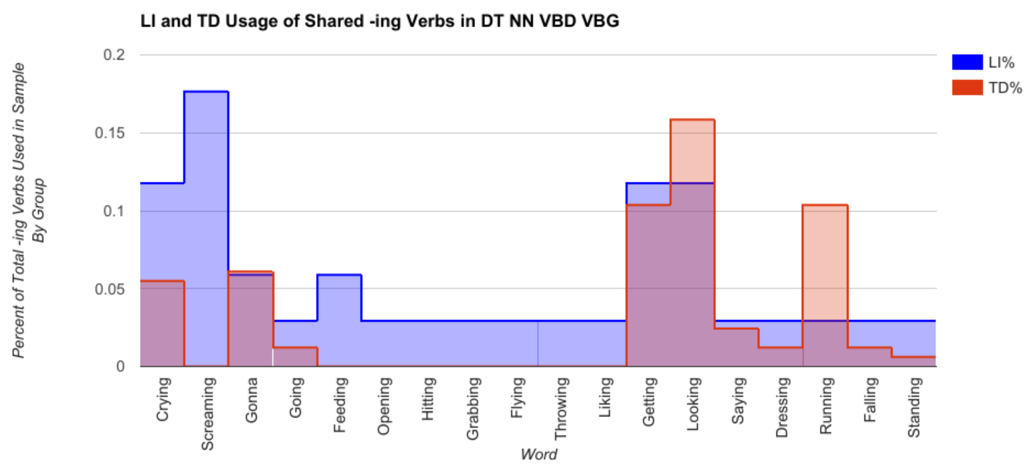

Figure 1: Usage of shared –ing verbs in DT NN VBD VBG.

PRELIMINARY ANALYSIS

An earlier analysis of the sample set compiled a list of associated word combinations, with their corresponding frequency for both TD and LI children (Solorio et al., 2011). An exploratory analysis of these N-grams, i.e. sequences of words of length N, focusing on phrases of structure determiner (DT; i.e. “the”, “a”), noun (NN; i.e. “boy”), past tense verb (VBD, i.e. “started”, “was”), -ing verb (VBG; i.e. “saying”) was conducted.

4-gram nested verb phrases

This phrase structure was chosen as a focal point because forming nested noun and verb phrases requires the speaker to remember and coordinate complex semantic, syntactic, and morphological information across multiple levels in the syntax tree of one phrase, i.e. number ("the boy" versus "the boys", which would make the VBD here, in this case an auxiliary be (AuxBe) verb, match "was" to singular "boy" or "were" to plural "boys"), and nested verbs, i.e. "was" and "saying". Additionally, combining verbs in this way is an alternate way of adding tense markers using syntax rather than verb-internal morphological transformations. For instance, "was saying" and "said" both indicate past tense; the former, however, is a regular and productive syntactic formulation, while the latter is an irregular verb form. Irregular forms are acquired later (Brown, 1973) than regular forms, and present more challenges for children with language impairment than do regular forms.

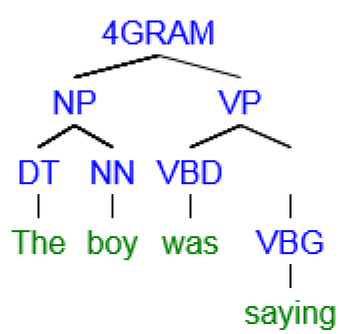

Figure 2: 4Gram: "The boy was saying."

VOCABULARY DIFFERENCES

TD vocabulary

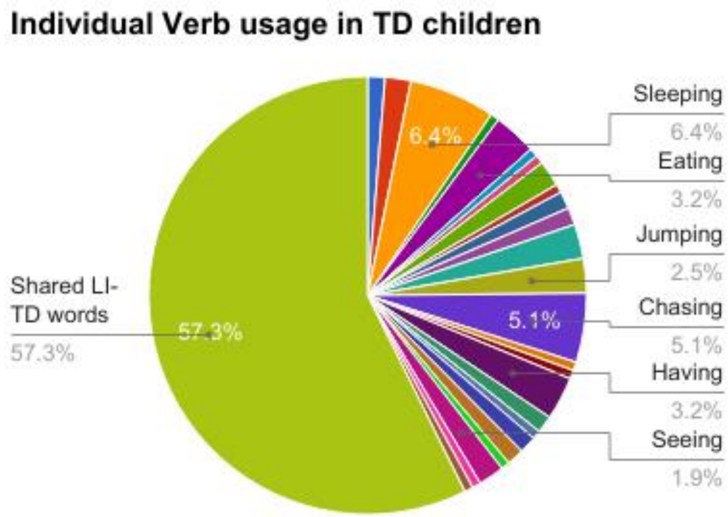

Figure 3: Individual verb usage in TD children.

Within the sample set of utterances that contained this type of 4-gram (DT NN VBD VBG), the entire field of VBG shared between TD and LI children accounted for only 54.9% of the total variety of vocabulary used by TD children. While there were also verbs that were only used by the language impaired children in this sample, there were only 7 such words (feeding, opening, hitting, grabbing, flying, throwing, liking), and other than "feeding", which was used twice, each was only used once in the entire sample set. In contrast, the verbs unique to the TD group included commonly-used words such as "sleeping", used 10 times by TD children in the sample, and "chasing", used 8 times by TD children in the sample.

Shared vocabulary

Words that were shared by both TD and LI groups were used with similar proportional frequency.

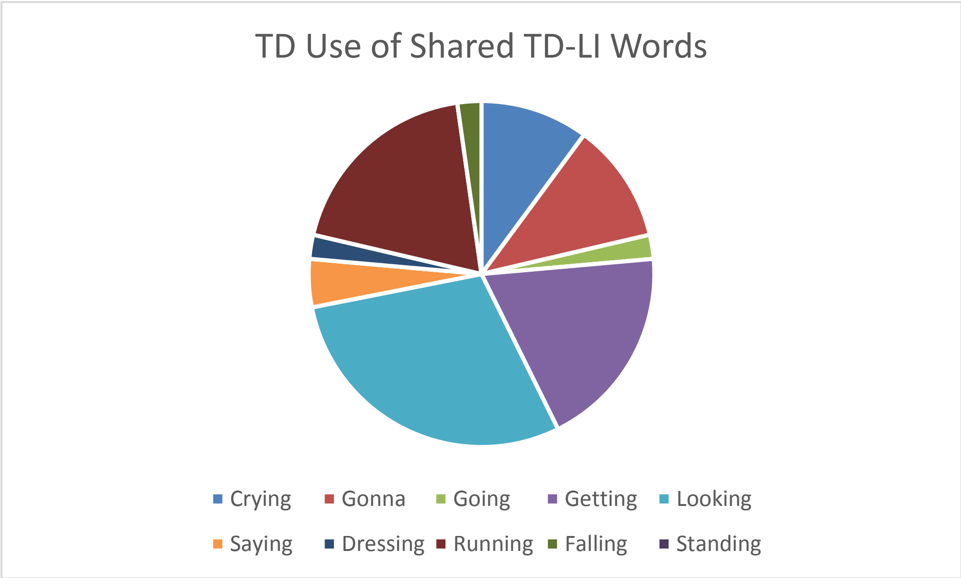

Figure 4: TD use of shared TD-LI words.

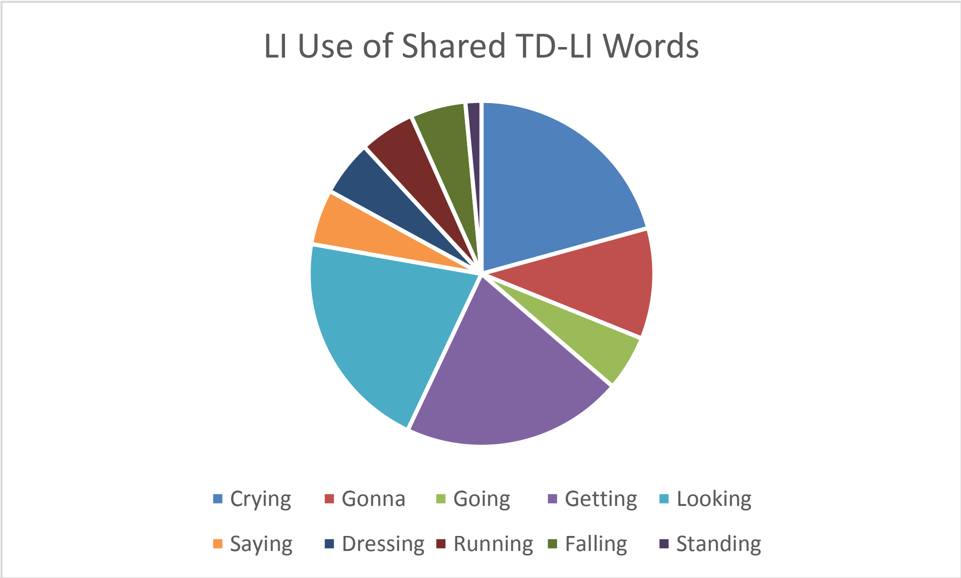

Figure 5: LI use of shared TD-LI words.

LI vocabulary

Children in the LI group depended heavily on a small group of –ing verbs; only 6 of the 17 –ing verbs used by children in the LI group appeared more than once. Of these, two were verbs than can be used to add tense or case meaning to a bare verb, instead of using a verb-internal morpheme, getting and going/gonna. Following analyses of going constructions, additional analyses were conducted on nested verb phrases using looking and getting using the rationale that these were commonly used verbs with a high potential for generating nested verb phrases.

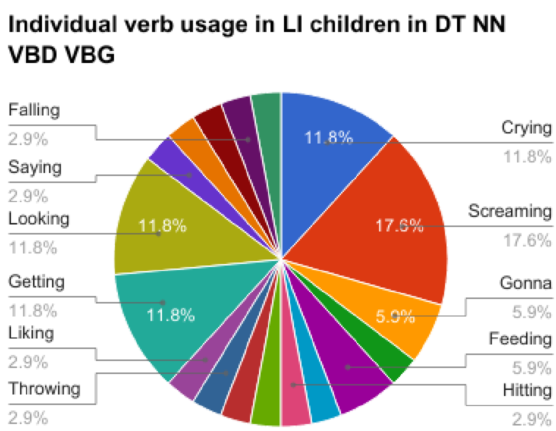

Figure 6: LI use of shared TD-LI words.

OBLIGATORY AUXBE IN NESTED –ING VERB PHRASES

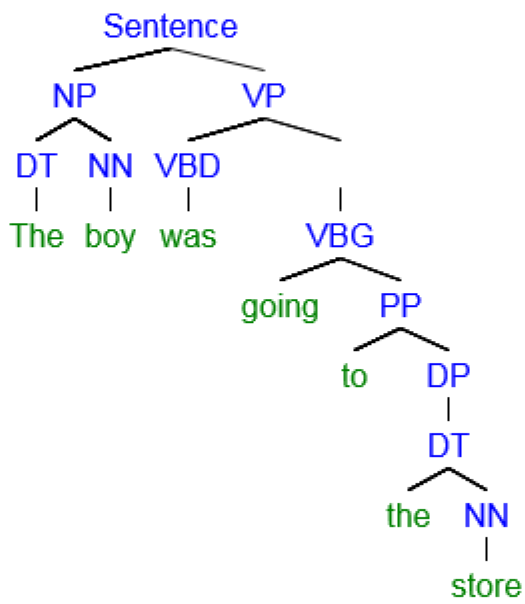

Figure 7: “The boy was going to the store”.

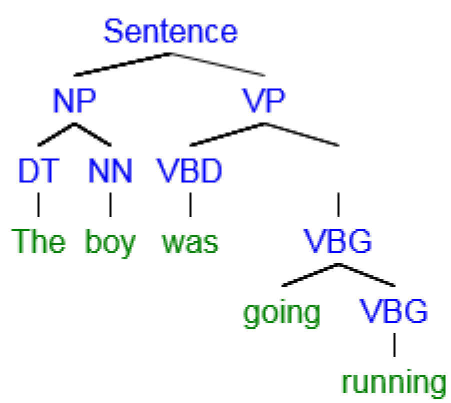

Figure 8: “The boy was going running”.

Nested “-ing” phrases have an obligatory auxiliary “be” verb at their highest branch level, and can take a variety of grammatical units as a daughter tree, including noun phrases, verb phrases, adverb phrases, etc. Additionally, the structure of tensed AuxBe + -ing can allow a speaker who may not have mastered irregular verbs in general to form a regularly structured past-tense ("was going"). From the exploratory data analysis it appeared that children in LI and TD groups attempted to use going and gonna constructions at similar rates; additional frequently used –ing verbs that were also frequently used with nested AuxBe phrases included getting and looking.

COMPUTATIONAL ANALYSIS OF –ING PHRASE USAGE PATTERNS

Researchers used a custom extension of the NLTK to examine how children in the TD and LI groups produced these nested "going" verb phrase structures differently.

Number of nested AuxBe+going phrases

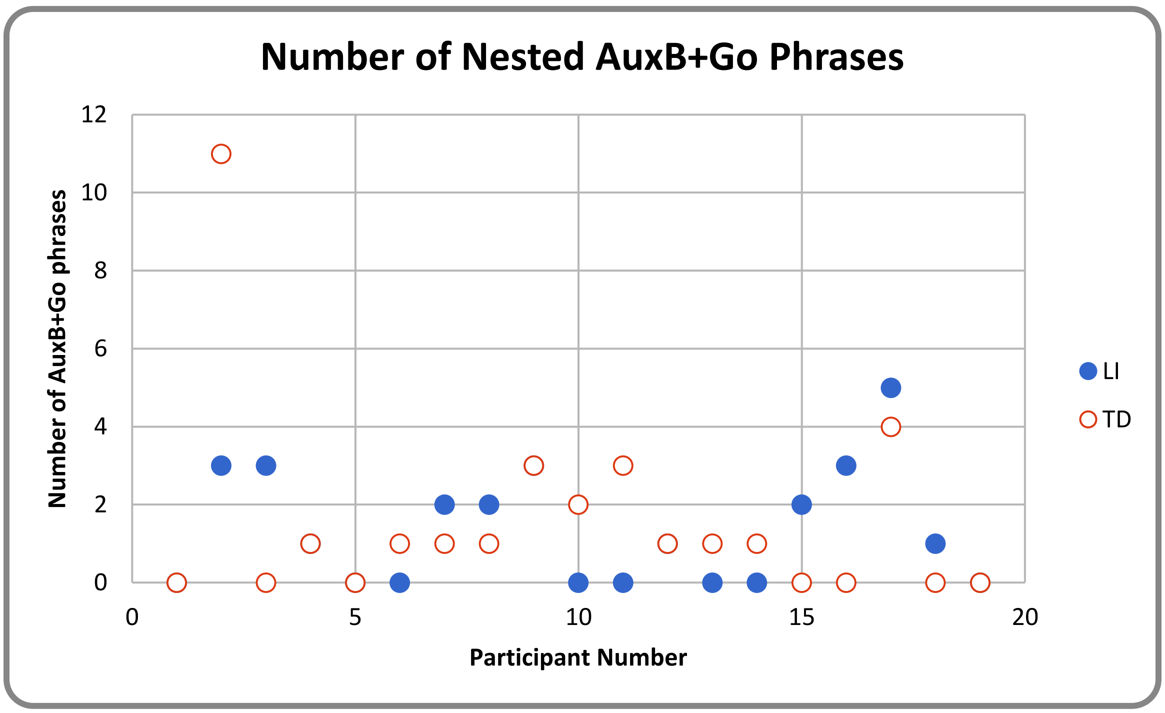

Figure 9: Number of nested AuxBe+ Going phrases.

With some outliers, TD and LI children used AuxBe+Going nested phrases with similar frequencies; LI children used a mean of 1.44 AuxBe+Going nested structures in the dataset, while TD children used a mean of 1.67 AuxBe+Going nested structures. The number of nested verb phrases was similarly comparable across both groups for looking and getting phrases.

MORPHOLOGICAL ACCURACY OF –ING VERB USE

Similarly, all attempted AuxBe+Going structures contained correctly-formed and used "going" verbs, with some idiosyncratic errors (prepositional phrase errors, etc.) in daughter-tree phrases that did not form a clear trend.

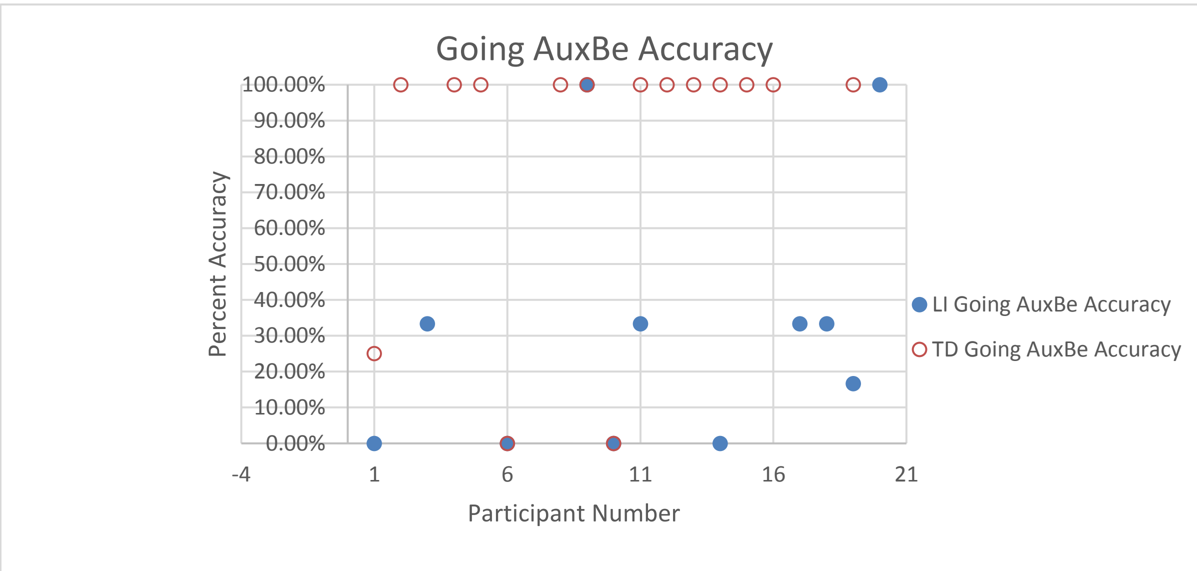

Figure 10: Going AuxBe Accuracy.

MORPHOLOGICAL ACCURACY OF AUXILIARY BE VERB USE

Nested going verb phrases

However, in terms of accurately using the correct tense of an Aux-Be verb before "going", there were marked differences between TD and LI groups. The majority of AuxBe errors in nested going verb phrases were omissions. Additionally, subject omissions were common errors.

AuxBe use in nested going verb phrases

While there were clear general differences between the TD and LI groups, there were outliers in both groups; two TD children had 0% accuracy, and two LI children had over 80% accuracy; one LI child had 100% accuracy. LI children had an average of 32% accuracy in AuxBe use in nested going phrases, while TD children had an average of 82% accuracy.

Potential diagnostic utility of AuxBe use in nested going phrases

Using a cut score of 40% accuracy, this analysis, if used to diagnose language impairment, would have a specificity of 80%, and a sensitivity of 82%. 11 LI children had analyzable going-phrase samples, while 15 TD children had analyzable going-phrase samples; of all of the individual verb-phrase analyses, this was the highest number for both groups.

COMPARATIVE DIAGNOSTIC UTILITY OF DIFFERENT VERB PHRASE ANALYSES

After conducting these initial analyses on nested going phrases, similar analyses were conducted on nested verb phrase structures using looking and getting, in addition to going.

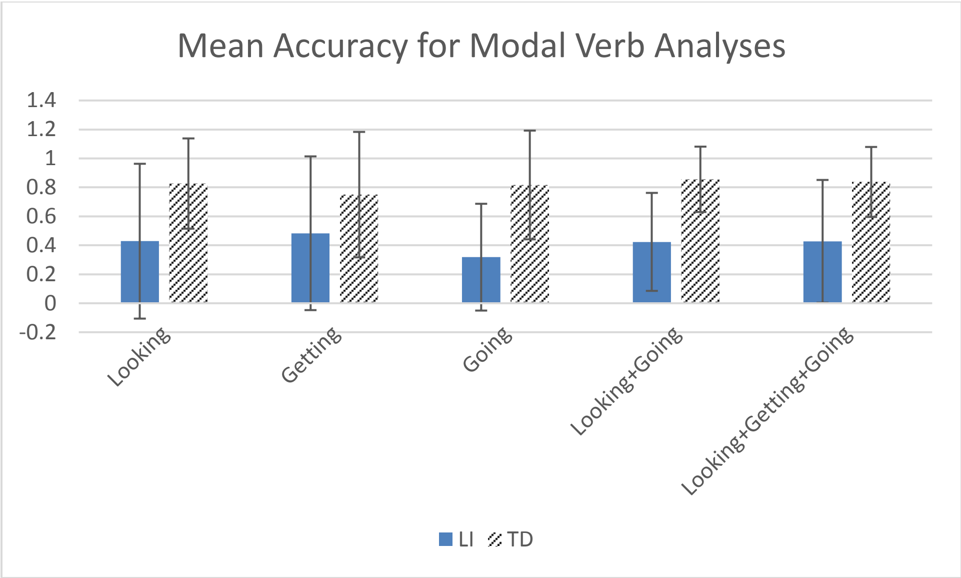

Figure 11: Mean accuracy of AuxBe measures for modal verb analyses.

LOOKING AUXBE ACCURACY

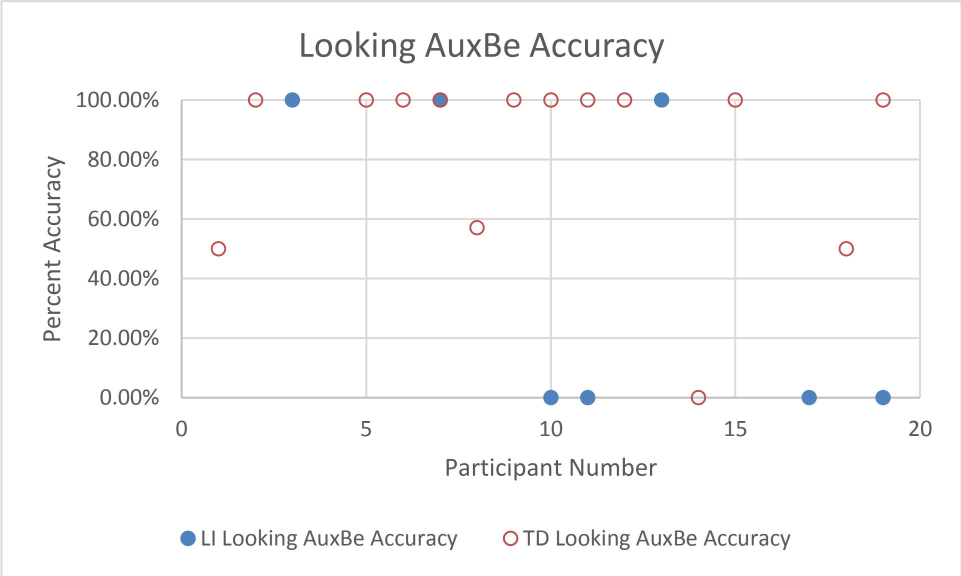

Figure 12: Looking AuxBe accuracy.

Nested looking verb phrases

Children with LI had an average of .77 nested looking phrases in their narrative samples, while those with TD had an average of 2.14. Some children produced grammatically correct nested verb phrases using looking that used a different auxiliary than AuxBe; these were counted as correct uses of the syntax structure, but were not included in AuxBe accuracy numbers.

AuxBe use in nested looking verb phrases

As with nested going phrases, the most common AuxBe errors in nested looking phrases were omission errors. Two LI children had 100% accuracy in their looking nesting verb phrases, while 1 TD child had 0% accuracy in their looking nested verb phrases. LI children had an average of 43% accuracy for AuxBe use in nested looking verb phrases, while TD children had an average of 83% accuracy.

Potential diagnostic utility of nested looking verb phrase analyses.

Using a cut score of 40% accuracy, this analysis would have a specificity of 93% and a sensitivity of 57%. 7 LI children had analyzable nested looking phrase structures, while 14 TD children had analyzable nested looking phrase structures; some children who did not have analyzable nested going structures were in this group, and vice versa.

GETTING AUXBE ACCURACY

Nested getting verb phrases

Nested getting verb phrase patterns were the least differentiated between LI and TD groups, and, despite the fact that getting was used with high frequency by both TD and LI groups as a whole, of the three verb structures analyzed, getting structures were used by the smallest number of children. This apparent discrepancy is explained by the fact that while the average number of nested getting structures is 1.11 for LI children and .81 for

TD children, the standard deviations of the number of nested getting structures are 2.89 and 1.66. In short, while few children used nested getting structures, those who did use them used them frequently. Another unusual characteristic of the nested getting structures used was that, unlike looking and going structures, where the most common AuxBe error was that of omission, several children made errors in conjugating the correct tense, number, or person while nonetheless filling the AuxBe position.

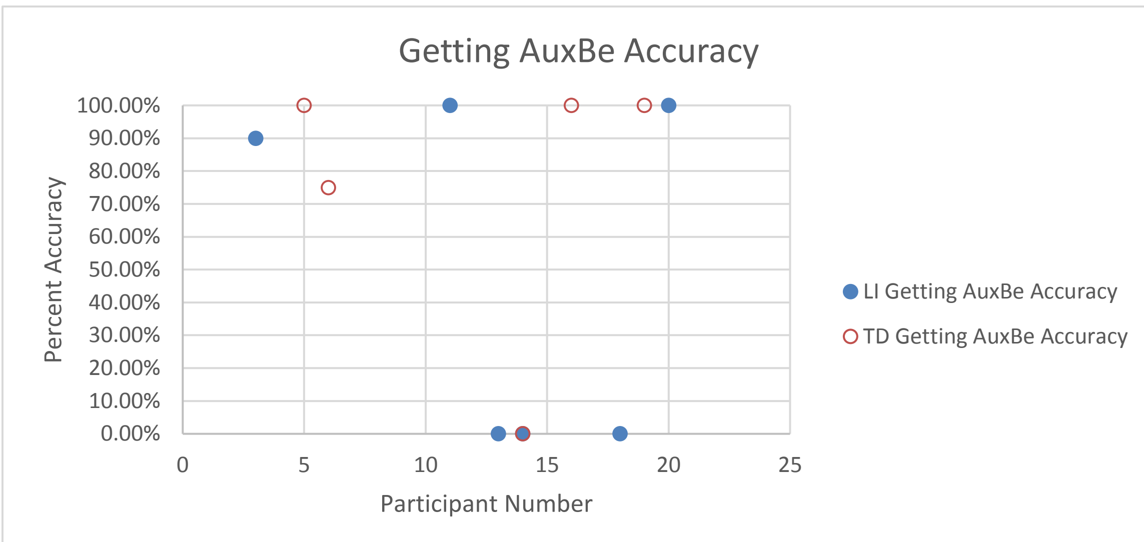

Figure 13: Getting AuxBe Accuracy.

AuxBe use in nested getting verb phrases

LI children had an average AuxBe accuracy of 48.33% in nested getting phrases, while TD children had an average AuxBe accuracy of 75%. One TD child had 0% AuxBe accuracy in nested getting phrases, while two LI children had 100% accuracy and 1 LI child had 90% accuracy. This structure had the lowest TD accuracy rate of the three verb structures analyzed, while also having the highest LI accuracy rate. For the sample analyzed here, there were no significant differences between AuxBe accuracy in LI and TD use of nested getting structures. This unique pattern, of the three verb patterns analyzed, may indicate that these children are using “getting” in a different way than they are using the verbs “going” and “looking”, or analyzing the morphosyntactic features of “getting” differently.

Potential diagnostic utility of analyses of nested getting phrases

Using a cut score of 60% AuxBe accuracy, this measure would have only 80% specificity and 50% sensitivity; flipping a coin would perform as well or better at accurately identifying TD children as being TD as this measure. This structure had the smallest number of individuals with analyzable samples; 6 LI children and 5 TD children had analyzable nested getting structures, and all of the children with analyzable getting structures also had analyzable nested looking or going structures.

MODAL VERB COMPOSITE AUXBE MEASURES

Analyses of language that children choose to use to tell a story are by necessity constrained by both the task itself (using vocabulary particular to the story, using turns of phrase/tenses/person appropriate to the setting and context of the story, etc.) and by the child’s language choices and preferences. A weakness of using an analysis that depends on a child using a specific word and syntactic structure is that not every child will make the same language choices; this was observable in the individual verb-phrase analyses, where different numbers of children used each structure. One way to overcome this weakness is to analyze multiple targets, in the interest of increasing the likelihood that a given child will use at least one of the targets chosen to analyze a specific morphosyntactic feature or structure.

With this in mind, the three verb-specific analyses were used to create composite scores. By combining looking, getting, and going as one composite verb phrase measure, at least one of the three structures could be analyzed for a total of 13 LI children and 17 TD children.

Looking, getting, going composite score AuxBe accuracy

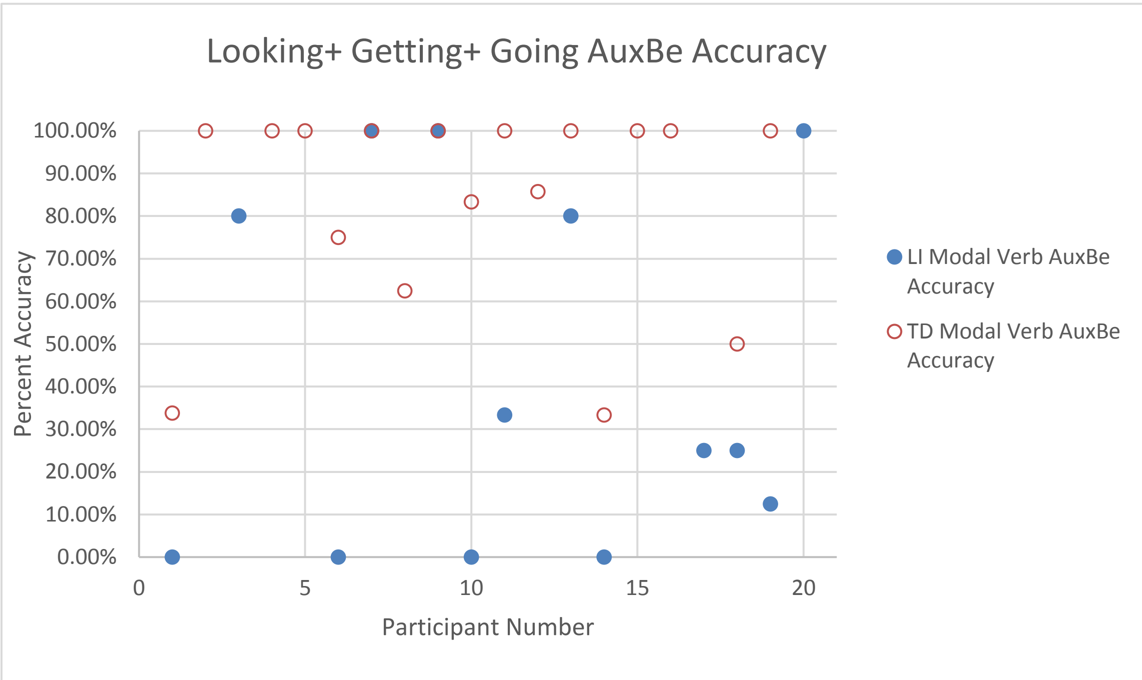

Figure 14: Composite looking, getting, going AuxBe accuracy scores.

The average AuxBe accuracy for the looking, getting, going verb phrase composite was 43% for LI children and 84% for TD children. One LI child had 100% accuracy for the composite score, and no TD child scored below 33%; three LI children scored between 30% and 90% accuracy.

Looking, getting, going composite score diagnostic utility

A composite score brings characteristics of different sub-scores together, which can have benefits, i.e., expanding the number of individuals or utterances that can be analyzed, but using a composite can also mute the strengths of each sub-score. Additionally, when this type of analysis is used to plan treatment, analyzing performance on multiple verb or feature patterns can help the clinician choose appropriate targets for intervention; i.e. if a child is using “going” phrases correctly but not using “getting” at all, “getting” structures could be targeted in treatment, while using “going” structures as a model.

Specificity and sensitivity of individual nested verb phrase analyses

At a cut score of 40% accuracy, AuxBe analyses of nested looking phrases have a specificity of 93% and a sensitivity of 57%, while analyses of nested going phrases have a slightly lower specificity of 80% and a higher sensitivity of 82%. Analyses of nested getting phrases, at a cut score of 60% AuxBe accuracy, have a specificity of 80% but a sensitivity of only 50%.

Table 1: Diagnostic specificity and sensitivity of individual modal verb AuxBe measures.

	Cut Score	**Specificity**	**Sensitivity**	**LR+**	**LR-**
Looking	0.4	0.928571	0.571429	8	0.384615
Getting	0.6	0.8	0.5	2.5	0.375
Going	0.4	0.8	0.818182	4.090909	-0.02273

Specificity and sensitivity of individual nested verb phrase analyses

After these scores are combined into a composite, the specificity and sensitivity average out slightly. If it were especially important to diagnose LI children correctly, sensitivity could be maximized by choosing a cut score of 81% accuracy, for a sensitivity of 77%, but this would bring specificity down to 71%. A cut score of 40% accuracy would increase specificity to 88%, but bring down sensitivity to 62%. Moving the cut score to 33.4% accuracy increases specificity to 94%, making accurate diagnoses of TD more likely, without a corresponding drop in sensitivity; for this composite, a cut score of 33.4% is the most useful score.

Table 2: Diagnostic specificity and sensitivity of looking, getting, going composite score.

	Cut Score	Specificity	Sensitivity	LR+	LR-
Looking+Getting+Going	0.81	0.705882	0.769231	2.615385	-0.08974
Looking+Getting+Going	0.4	0.882353	0.615385	5.230769	0.302564
Looking+Getting+Going	0.334	0.941176	0.615385	10.46154	0.346154

Weaknesses of the looking, getting, going composite.

Interestingly, all of the children with analyzable samples for getting nested verb phrases were also represented in one or both of the looking nested verb phrase and going nested verb phrase sets. Additionally, the specificity of the standalone going analysis was no better than the going analysis, and the sensitivity was no better than random chance. Taking this into account, a new composite was calculated using only looking and going AuxBe accuracy.

Looking, going composite score

The average AuxBe accuracy for the looking, going composite score was 42% for LI children and 86% for TD children; dropping getting values had no impact on the average LI accuracy score and only slightly increased the TD accuracy score. Three LI children had 100% accuracy for this composite score, and all TD children scored above 33%. Only one LI child scored between 33.33% and 100% on this composite.

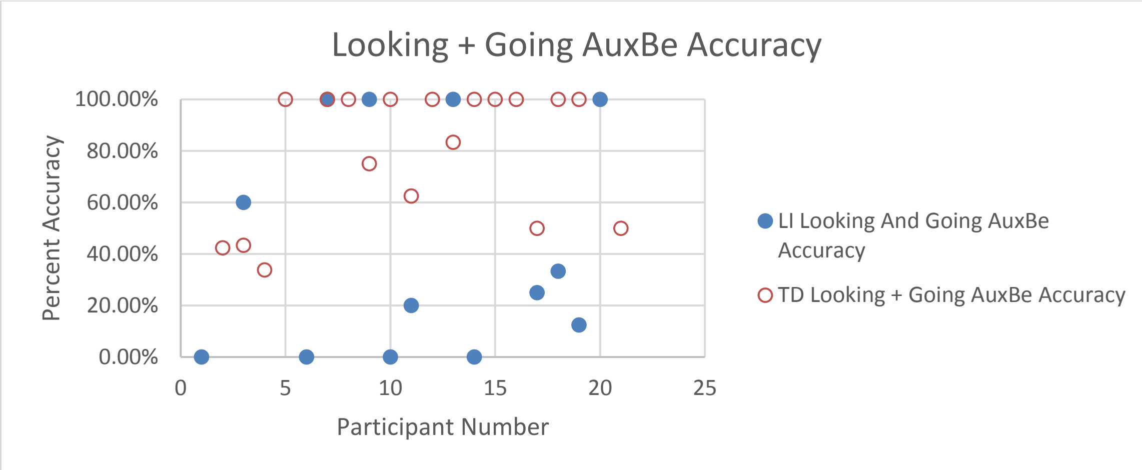

Figure 15: Composite looking, going AuxBe accuracy scores.

Specificity and sensitivity of the looking, going composite. With a cut score of 33.4% accuracy, this composite has a specificity of 100%, meaning that all TD children are accurately identified as such. With this same cut score, this composite has a sensitivity of 62%. Increasing the cut score to 61% accuracy increases the sensitivity to 69%, but brings down the specificity to 82%. Given the small possible gain in sensitivity, 33.4% is probably the most useful cut score for this measure.

Table 3: Diagnostic specificity and sensitivity of looking, going composite score.

	Cut Score	Specificity	Sensitivity	LR+	LR-
Looking+Going	0.334	1	0.615385		0.384615
Looking+Going	0.61	0.823529	0.692308	3.923077	0.159341
Looking+Getting+Going	0.334	0.941176	0.615385	10.46154	0.346154

Comparison of NLTK extension composite scores

Dropping the nested getting verb phrase measure from the composite allowed the looking, going composite to increase specificity to 100% at the same cut score of 33.4%, without dropping sensitivity from 62%.

AUTOMATIC IPSYN SCORING USING CLAN

An existing computational method for calculating syntactic complexity uses the CHILDES morphological tagging system and morphological analysis system to calculate IPSYN scores automatically.

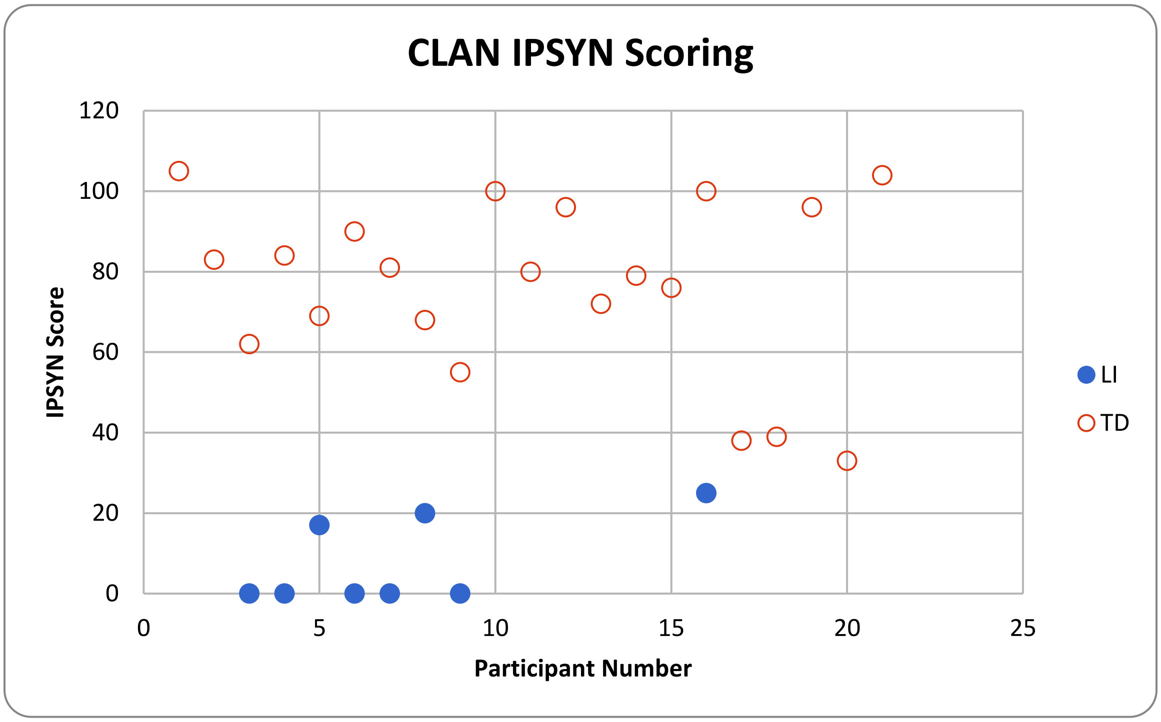

Figure 16: CLAN-native IPSYN scores.

BACKGROUND ON IPSYN

IPSYN normative data is based on 100-utterance samples, but an incomplete score can be calculated on a shorter sample. Additionally, some of the grammatical errors common in children with language impairment impeded the process of calculating this score automatically. Of the LI group, only 8 files were suitable for an abbreviated IPSYN analysis, while 21 TD group narrative samples were suitable for this analysis.

IPSYN COMPOSITE SCORE PATTERNS

The IPSYN scores of each group occur over distinct ranges within this sample, due in part to 5 of the 8 LI children getting an overall score of 0 on the measure. A cut score of 30 using this measure had a sensitivity and specificity of 100% for predicting language impairment.

IPSYN COMPONENT SCORES

IPSYN Q and S scores

Analyzing the four component scores of the IPSYN calculation, the question/negation (IPSYN Q) and sentence (IPSYN S) sub-scores did not significantly differentiate between the two groups. This is likely due to the structure of the elicitation task; since children were retelling a story rather than participating in a conversation, they had no strong reason to pose questions, and the relatively unconstrained task of choosing how to retell the story allowed them to limit the complexity of their sentences if they chose to.

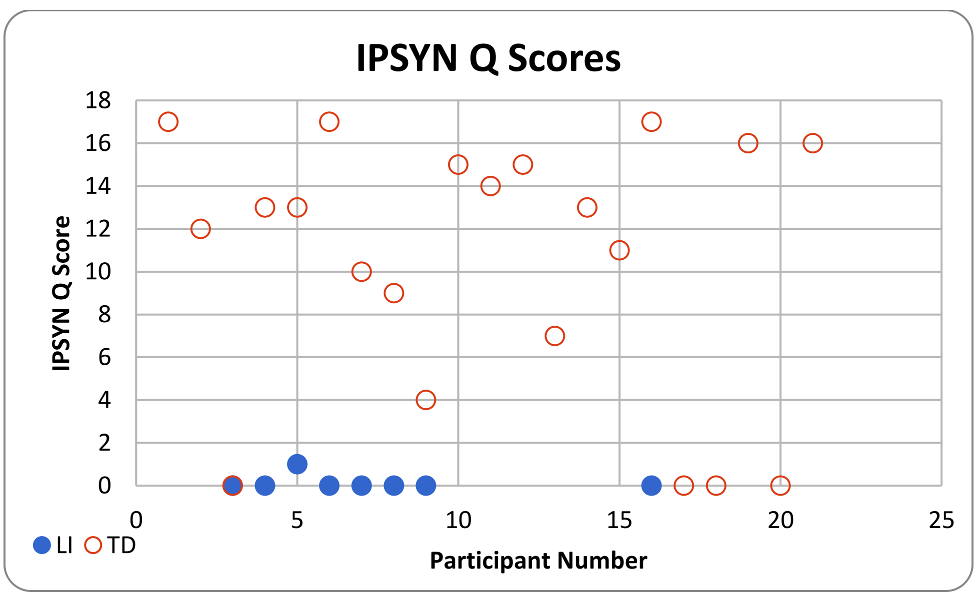

Figure 17: IPSYN Q Scores.

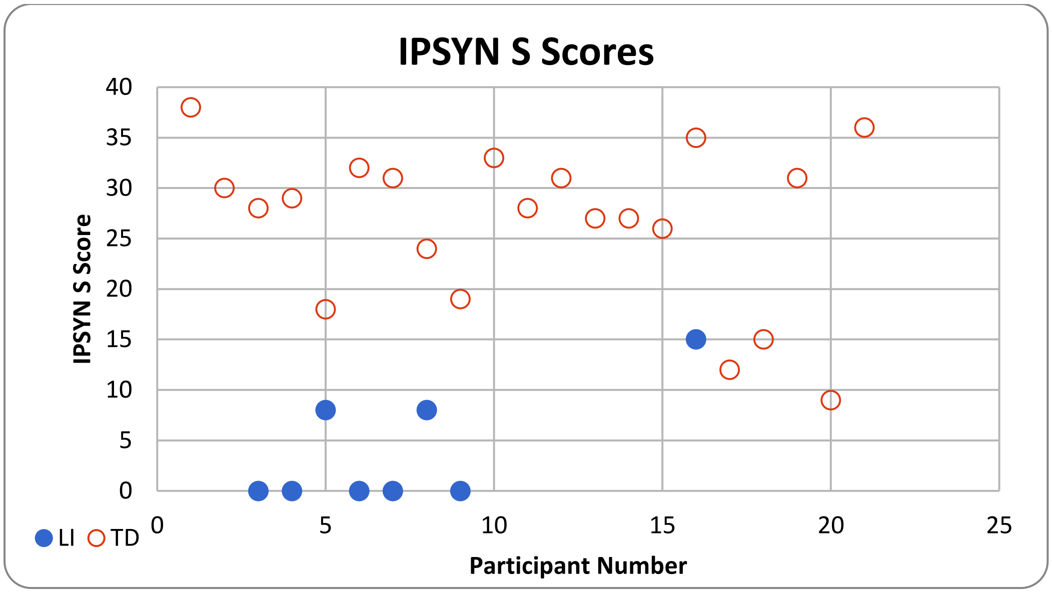

Figure 18: IPSYN S scores.

IPSYN V and N scores

In contrast, the IPSYN noun (IPSYN N) and verb (IPSYN V) sub-scores had completely distinct distributions.

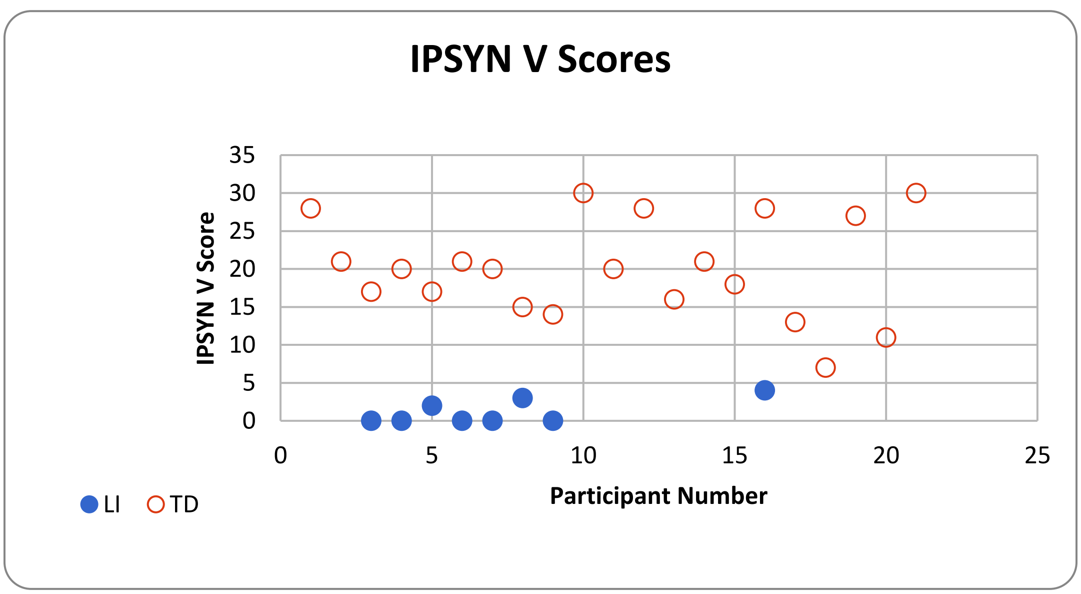

Figure 19: IPSYN V scores.

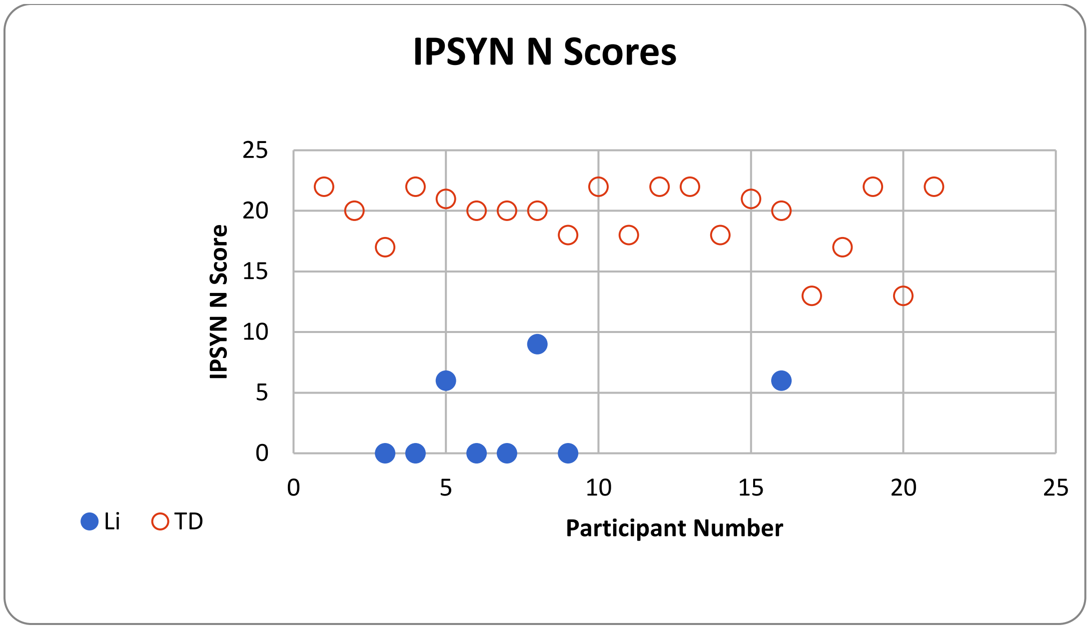

Figure 20: IPSYN N Scores.

Customized IPSYN composite score

A new composite score, the IPSYN N+V score, was calculated; for the samples that could be evaluated using this method, a cut score of 10 using this semi-composite score had a sensitivity of 100%.

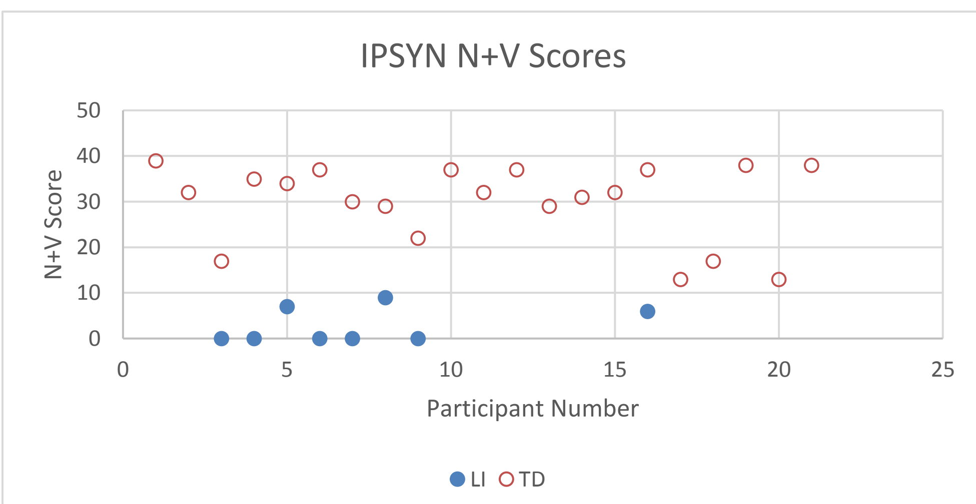

Figure 21: IPSYN N+V Scores.

Comparison to CHILDES samples

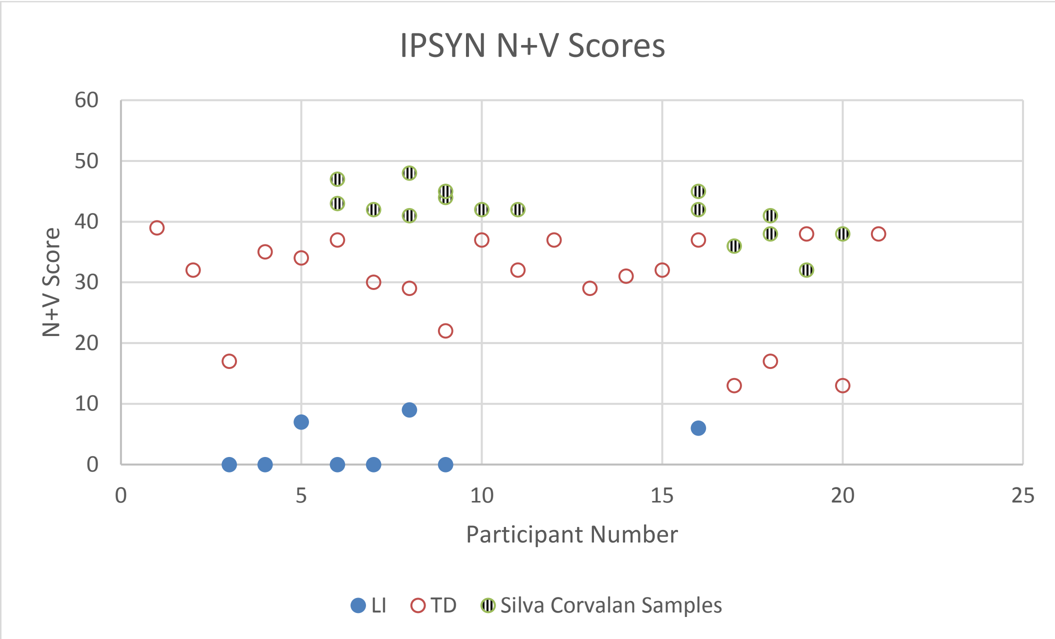

Figure 22: IPSYN N+V Scores with expanded TD bilingual samples.

To further evaluate the use of this tool, the same analysis was run on a set of narratives produced by Spanish-English bilingual children between the ages of 4-7 years old studied by Silva-Corvalan (1989); all of these children scored above the cut score of 10.

Chapter III: Discussion: implications, limitations, and directions for further research

TYPES OF ERRORS

Across all of the analyses performed in this study, the majority of AuxBe errors were errors of omission. Interestingly, children, especially TD children, did make errors in tense and person when conjugating AuxBe verbs in nested getting phrases, and LI children used nested getting phrases more often and with higher accuracy than they used the other analyzed verb phrase structures, while TD children had more errors on nested getting phrases than they did with the other analyzed phrase structures. Additionally, LI children frequently omitted noun phrases that are obligatory in English at the beginning of nested verb phrase structures. While omitting AuxBe in nested –ing phrase structures is a phenomenon that many teachers might attribute to bilingual children's status as second language learners, this analysis was able to show that distributions of AuxBe errors differentiate between Spanish-English bilingual children with and without LI.

POTENTIAL AS FUTURE DIAGNOSTIC MEASURES

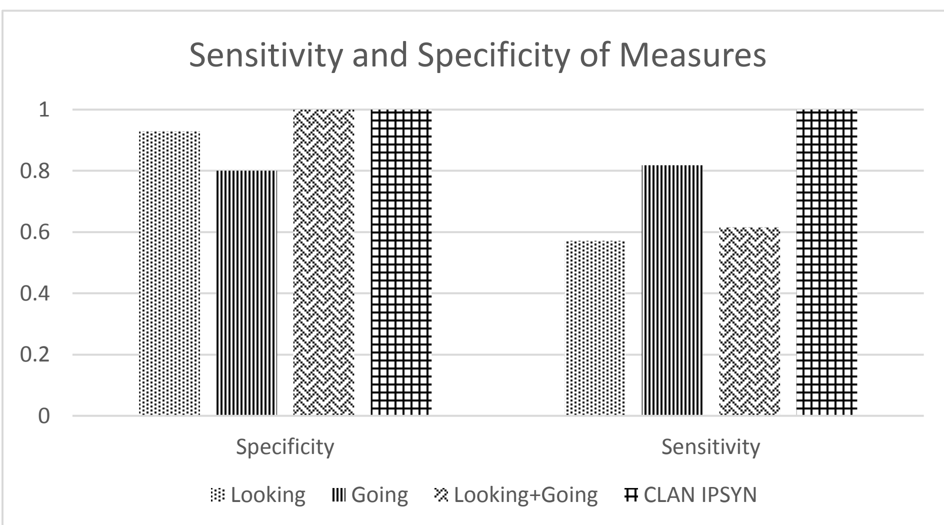

Figure 23: Sensitivity and specificity of NLTK extension and CLAN-native IPSYN.

The CLAN-native IPSYN measures were able to completely differentiate between LI and TD children in this sample. The custom extension of the Python NLTK assessed in this study was able to match IPSYN in specificity when using the looking, going composite AuxBe accuracy score, but lags behind in sensitivity; interestingly, if this were to be used as a diagnostic tool, it might benefit from using the composite to verify TD diagnoses and the nested going measure to verify LI diagnoses, as the going AuxBe measure has higher sensitivity than the composite. On the other hand, the composite score was able to evaluate the largest number of participants; each score system has trade-offs.

LIMITATIONS OF AUXBE ANALYSIS AS A FUTURE DIAGNOSTIC MEASURE

Some of the limitations to these findings are dialectical in nature Using AuxBe deletion as a criterion for diagnosing language impairment also has limitations. While the analysis conducted here suggests that this may be a powerful feature to analyze in order to diagnose language impairment in Spanish-English bilinguals acquiring standard American English, this feature may be of limited use if this application is extended to other linguistic minorities. In particular, many dialects of African American Vernacular English (AAVE) do not require an AuxBe verb in all of the contexts that general American English does (Rickford, Ball, Blake, Jackson, & Martin, 1991). Additionally, many other dialects of English do not require that AuxBe verbs be conjugated in the same ways; i.e. in some dialects it is considered grammatical to say "I be going", or "He been going", "He done gone," etc., without a standard auxiliary. The focused test used in this paper would not accurately evaluate speakers of these dialects, but an expanded system could take these linguistic variations into account.

APPLICATIONS TO THERAPY

While using the analytical methods evaluated here is probably impractical for regular diagnostic purposes, knowing that bilingual LI children frequently omit the auxiliary verb position of the syntax tree, especially when the target is a be verb, may expand possible targets for intervention. After a child has been diagnosed with LI, targeting language complexity tasks that require the child to use nested verb phrases and ensure that they use the appropriate –be verb in that location may be beneficial.

This study also highlighted that children with LI are not only constrained to a relatively small set of vocabulary, they are more likely to depend heavily on a few highly productive words or phrases, like going or getting, to convey as much meaning as possible with only a few words. Words like going and getting, which can be used to perform the same tense-assigning tasks as more difficult morphological features, like past tense, do, were overused by LI children compared to their TD peers.

Potentially, these trends could be used to design intervention strategies or dynamic assessment tasks; a target construction using a word of this type that child uses infrequently could be used in a construction using this nested verb phrase structure, or a word the child uses frequently could be used to model the correct AuxBe use and structure, in order to train the child to acquire this skill, or to expand their vocabulary in a way that capitalizes on the skill of using nested verb phrases to overcome weaknesses in irregular morphosyntax.

STRENGTHS OF THE COMPUTATIONAL METHODS ANALYZED

Existing computational methods, like those available using transcribed and coded files and SALT, or transcribed files and CLAN, can be highly accurate in distinguishing between impaired and typically developing language use. Existing natural language processing techniques, like the base code of the NLTK in Python, can analyze grammatical speech with a high level of accuracy. Expanding natural language processing methods to assess ungrammatical and nonstandard speech may have applications not only in the child language assessment arena, but also in the creation of speech and language recognition tools for non-standard dialects, or for the analysis of speech data gathered in a noisy environment where enough of the signal is corrupted.

STRENGTHS OF NLTK EXTENSIONS

Using an approach that focuses not on word order but on higher-levels of complex syntax trees allows a single analysis to touch on multiple linguistic skills in a familiar way. The high correlation of AuxBe deletion in a relatively commonly-used verb construction- i.e., "going" phrases- to language impairment suggests that asking children to produce sentences that target complex nested verb phrases may be a useful diagnostic task, especially for children who come from a complex linguistic background and for whom existing language assessments are inadequate.

STRENGTHS OF CLAN IPSYN

The CLAN-native IPSYN function was highly accurate at distinguishing between TD and LI narratives, and is executable with minimal formatting; all that is needed to run this analysis on a file is basic transcription in a CLAN-compatible format.

While the N and V sub-scores were informative in this study, the S and Q sub-scores were not; this is likely due in part to the fact that all of the samples in this study were story re-tells. Based on both the content and the social context within which children were re-telling the story, it's unlikely that most people would narrate or answer direct questions in this context.

Limitations to the computational methods analyzed

While the novel analytical method discussed in this paper shows promise as a diagnostic screening tool in the future, existing computational methods, including the CLAN-native automatic IPSYN scoring function, currently have greater accuracy. Additionally, while SALT requires the researcher to hand-code language samples, both CLAN-native analyses and this new method only require transcription; CLAN requires some basic preliminary formatting for accurate identification of speaker and file type. The NLTK-based method described here, however, can analyze the productions of a single speaker from an unformatted plaintext transcript, or from CLAN or SALT-formatted transcripts.

The CLAN-native IPSYN calculating function requires 100 utterances and error-free transcription in order to return useful results; while errors affect the accuracy of results returned by this NLTK extension, transcription errors do not interfere with the general running of the analysis, and since the entire file can still be processed, the user is informed by error messages when the analysis is likely incomplete. While this analytical tool currently only works in English, CLAN can additionally analyze language samples in a variety of languages, including Spanish. This limitation can be addressed in the future; parsers for additional languages can be created or adopted from existing natural language processing tools as they are developed.

This extension of the NLTK is currently only capable of doing rigorous analysis of specific phrase structures, but by performing this analysis on the tree level instead of the word or N-Gram level, even an analysis limited to a small set of verb forms can focus on specific morphological or syntactic processes rather than specific sequences of words.

As such, a tree-level analytical tool, however limited, can provide a more nuanced analysis of language use than a flat N-Gram analysis.

One of the strengths of this NLTK extension lies in the ability to analyze ungrammatical and disfluent speech. Conventional natural language parsers are dependent on words order alone; while this parser uses the native NLTK parsing framework as a stepping stone, when NLTK-native parser creates errors the method jumps to a secondary method that searches for morphemes that indicate specific positions in a syntax tree, and builds the tree around that data. While this parser is currently limited to a few morphemes (i.e., -ing endings) and common morphological data of the words that normally build up the tree levels connected to those forms, with the time the majority of English morphemes could be included. As Spanish has more regular forms than English, and a smaller set of irregular forms, this morpheme-based parser should be easily extendable to Spanish.

Another concern strikes more deeply at the concept of using computational methods for this type of analysis. Is it more useful to choose features to analyze after a careful reading of past research, or to do a computational analysis of a corpus of data to find trends, and then attempt to explain them by going through past research? Beginning from the literature on child language production can make outlier behavior more obvious, and help to constrain later analysis to likely targets; on the other hand, if the targets suggested by this review of the literature are not present in the available sample set, it is possible to miss key characteristics that are present. Beginning from the analysis, on the other hand, can identify trends that are less well known, or trends that are particularly evident in the available sample set; however, this can also put the researcher in the position of forming explanations rather than hypotheses, and weakens the strength of the conclusions that can be drawn from subsequent analyses. This study used a mix of both approaches, performing exploratory analyses and using the literature to guide the choice of targets for subsequent analyses, to highlight significant differences between LI and TD groups in this sample while focusing on phenomena known to be associated with LI.

This attempt to develop a new computational method of analyzing ungrammatical speech has revealed several limitations of using computational methods to diagnose language impairment. Significantly, all of the computational methods discussed here require transcribed language samples; most children with suspected language impairment will have been diagnosed using a standardized measure well before a researcher has the opportunity to transcribe their language sample. However, these computational techniques can identify features that may be especially useful for distinguishing between impairment and normal language acquisition, especially for children who come from a linguistic minority, i.e. being bilingual or speaking a minority dialect.

LIMITATIONS TO NLTK EXTENSION METHODS

Chief limitations to the NLTK extension used here are due to these software tools still being in development; for instance, while some levels of CLAN IPSYN could be conducted with Spanish, since it comes with a basic Spanish parser and dictionary, currently the NLTK extension only has a functioning English parser. While Spanish parsers exist for grammatical content, ungrammatical passages would be read incorrectly, and null elements of syntax trees would not be accurately labeled.

Another weakness is that analyses can currently only be done on one specific verb or structure at a time. With time, this library of verbs that are ready to be analyzed will grow, but at the moment each new verb structure requires a separate analysis.

Finally, while the syntax-tree based search and indexing program is able to address some grammar errors, any automatic analysis will occasionally misinterpret some ungrammatical utterances that a human coder might be able to tag correctly.

Limitations to CLAN IPSYN

Like the NLTK extension, CLAN IPSYN does require that you type up a transcript, which can be time consuming and labor intensive. Additionally, the sub-scores of the IPSYN calculation are of limited use in a constrained activity like a narrative re-tell, since the speaker has limited opportunities to ask questions, and low motivation to use complex sentence structures. Furthermore, IPSYN calculation requires at least 100 utterances, and can only read ungrammatical utterances with limited accuracy. While the CLAN IPSYN scoring method is currently more accurate than the extension of the NLTK used here, these disparities could be addressed in time.

Future directions for research

Next steps for continuing to develop the exploratory analytical tool described here include comparing analyses of highly productive words like “going” and “getting” to other verb structures using modal verbs across TD and LI populations, expanding the parser to analyze verb and noun phrases in Spanish, and documenting and formatting this software tool for use by others to conduct further research. Additionally, while the morpheme-based tree building function can currently work for distinctive verb forms, further work is needed to expand this analysis of non-standard production to a wider set of morphemes, and the method for alternating between the parsers for grammatical and ungrammatical utterances needs to be improved.

References

[Ahadi, H., Nilipour, R., Rovshan, B., Ashayeri, H., & Jalaie, S. (2014). The perception and expression of verb morphology in bilinguals with specific language impairment. Audiology, 23(1), 62-69.

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. Journal of Speech, Language, and Hearing Research, 41(5), 1185-1192. doi:10.1044/jslhr.4105.1185

Bedore, L. M., Peña, E. D., Gillam, R. B., & Ho, T. (2010). Language sample measures and language ability in spanish-english bilingual kindergarteners. Journal of Communication Disorders, 43(6), 498-510. doi:10.1016/j.jcomdis.2010.05.002

Bedore, L., & Leonard, L. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. Applied Psycholinguistics, 26, pp 195-225. doi:10.1017/S0142716405050149.

Beverly, B. L., & Williams, C. C. (2004). Present tense be use in young children with specific language impairment: Less is more. Journal of Speech, Language, and Hearing Research, 47(4), 944-956. doi:10.1044/1092-4388(2004/070)

Bishop, D. (1994). Grammatical errors in specific language impairment: Competence or performance limitations?. Applied Psycholinguistics, 15, pp 507-550. doi:10.1017/S0142716400006895.

Bonet, E. (1995). Feature structure of romance clitics. Natural Language & Linguistic Theory, 13(4), 607-647. doi:10.1007/BF00992853

Brown, R. (1973). A first language: The early stages. London: George Allen & Unwin.

Conti-Ramsden, G. & Jones, M. (1997). Verb Use in Specific Language Impairment. J Speech Lang Hear Res, 40(6), 1298-1313. doi: 10.1044/jslhr.4006.1298.

Crago M, Paradis J. (2003) Two of a kind? Commonalities and variation in Languages and language learners. In: Levy Y, Schaeffer J, editors. Language competence across populations: Towards a definition of specific language impairment. Lawrence Erlbaum Associates; Mahwah, NJ: 2003. pp. 97–110.

Gladfelter, A., & Leonard, L. B. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. Journal of Speech, Language, and Hearing Research : JSLHR, 56(2), 542-552. doi:10.1044/1092-4388(2012/12-0100)

Goldstein, B. (2012). Bilingual language development & disorders in spanish-english speakers (2nd ed.). Baltimore, Md: Paul H. Brookes Pub. Co.

Hadley, P. A., & Holt, J. K. (2006;2016;). Individual differences in the onset of tense marking: A growth-curve analysis. Journal of Speech, Language, and Hearing Research, 49(5), 984-1000. doi:10.1044/1092-4388(2006/071)

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., . . . Samhällsvetenskapliga fakulteten. (2011). Grammar predicts procedural learning and consolidation deficits in children with specific language impairment. Research in Developmental Disabilities, 32(6), 2362-2375. doi:10.1016/j.ridd.2011.07.026

Jordaan, H. (2008). Clinical intervention for bilingual children: An international survey. Folia Phoniatrica Et Logopaedica, 60(2), 97-105. doi:10.1159/000114652

Kohnert, K., (2010) Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions, Journal of Communication Disorders, Volume 43, Issue 6, November–December 2010, Pages 456-473, ISSN 0021-9924, http://dx.doi.org/10.1016/j.jcomdis.2010.02.002. (http://www.sciencedirect.com/science/article/pii/S002199241000016X)

Leonard, L. B., Miller, C. A., Owen, A. J.. (2000). The comprehension of verb agreement morphology by english-speaking children with specific language impairment. Clinical Linguistics & Phonetics, 14(6), 465-481. doi:10.1080/026992000415886

Leonard, L. B., Deevy, P., Kurtz, R., Krantz Chorev, L., Owen, A., Polite, E., ... Finneran, D. (2007). Lexical Aspect and the Use of Verb Morphology by Children With Specific Language Impairment. J Speech Lang Hear Res, 50(3), 759-777. doi: 10.1044/1092-4388(2007/053).

Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. International Journal of Corpus Linguistics, 14(1), 3-28. doi:10.1075/ijcl.14.1.02lu

MacWhinney, B. (2000). The CHILDES Project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, G.,Restrepo, A, & Auza, A. (2013). Comparison of Spanish morphology in monolingual and Spanish–English bilingual children with and without language impairment. Bilingualism: Language and Cognition, 16, pp 578-596 doi:10.1017/S1366728912000697

Paradis, J., Rice, M. L., Crago, M., & Marquis, J. (2008). The Acquisition of Tense in English: Distinguishing child second language from first language and specific language impairment. Applied Psycholinguistics, 29(4), 689–722. http://doi.org/10.1017/S0142716408080296

Peña, E. D., Gillam, R. B., & Bedore, L. M. (2014). Dynamic Assessment of Narrative Ability in English Accurately Identifies Language Impairment in English Language Learners. J Speech Lang Hear Res, 57(6), 2208-2220. doi: 10.1044/2014_JSLHR-L-13-0151.

Perkins, J. (2010). Python text processing with NLTK 2.0 cookbook: Over 80 practical recipes for using python's NLTK suite of libraries to maximize your natural language processing capabilities. Mumbai;Birmingham;: PACKT Publishing.

Pescarini, D. (2010). Elsewhere in romance: Evidence from clitic clusters. Linguistic Inquiry, 41(3), 427-444. doi:10.1162/LING_a_00003

Rice, M. L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. Journal of Speech and Hearing Research, 38(4), 850-863.

Rice, M. L., Wexler, K., & Redmond, S. M. (1999). Grammaticality judgments of an extended optional infinitive grammar: Evidence from english-speaking children with specific language impairment. Journal of Speech, Language, and Hearing Research, 42(4), 943. doi:10.1044/jslhr.4204.943

Rickford, J. R., Ball, A., Blake, R., Jackson, R., & Martin, N. (1991). Rappin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English. Language Variation and Change, 3(1), 103-132. doi:10.1017/S0954394500000466

Ryan, C. (2011) Language use in the United States: 2011. US Census Bureau. Retrieved from: https://www.census.gov/prod/2013pubs/acs-22.pdf

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. Journal of Child Language, 37(3), 705-729. doi:10.1017/S0305000909990407

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. Child Development, 61(6), 1728-1743. doi:10.1111/j.1467-8624.1990.tb03562.x

Silva-Corvalan, C. (1989) “Algunos aspectos de la gramática de los niños bilingües de Los Angeles (E.E.U.U.)", at the III Congreso Internacional del Español de América, Universidad de Valladolid.

Solorio, T, Sherman, M., Liu, Y, Bedore, L., Peña, E., & Iglesias, A. (2011). Analyzing language samples of Spanish–English bilingual children for the automated prediction of language dominance. Natural Language Engineering, 17, pp 367-395. doi:10.1017/S1351324910000252.]

Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. Journal of Communication Disorders, 30(4), 325-344. doi:10.1016/S0021-9924(97)00015-4