2012-03-08

Accuracy of Automated Grammatical Tagging of Narrative Language Samples from Spanish-Speaking Children

Tyson Gordon Harmon
Brigham Young University - Provo

Accuracy of Automated Grammatical Tagging of Narrative Language Samples

from Spanish-Speaking Children

Tyson G. Harmon

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Ron W. Channell, Chair
Christopher Dromey
Shawn L. Nissen

Department of Communication Disorders

Brigham Young University

April 2012

ABSTRACT

Accuracy of Automated Grammatical Tagging of Narrative Language Samples from Spanish-Speaking Children

Tyson G. Harmon
Department of Communication Disorders, BYU
Master of Science

The present study measured the accuracy of automated grammatical tagging software as compared to manual tagging in Spanish-speaking children’s personal and fictional event narrative language samples. Studies have identified articles, clitic (contracted with a verb) pronouns, and verbs as clinical markers for language impairment in Spanish-speaking children. Automated grammatical tagging software may aid in the rapid identification of these grammatical markers. Grammatical morphemes of 30 first and fourth grade children’s personal and fictional event narrative samples were tagged and compared with their respective manually tagged samples. The accuracy of word-level coding averaged 91%, and similar accuracy was found for clinically significant tags. Automated grammatical analysis has the potential to accurately identify clinically relevant grammatical forms in samples from children who speak Spanish.

Keywords: child language, Spanish, narrative, software, grammatical tagging

ACKNOWLEDGEMENTS

I am sincerely grateful to all those who have assisted in any way during this process. I have greatly benefited from the help, advice, encouragement, and patience of faculty, family members, and friends.

I would like to offer a special thanks to Dr. Ron Channell who was constantly available to offer support, direction, and guidance. His availability to assist me in any way made the thesis process not only possible, but enjoyable. Thank you, as well, to committee members Dr. Christopher Dromey for your thorough revisions and attentive suggestions, and Dr. Shawn Nissen for your input and advice.

My wife, Brittney, has been a constant source of counsel and encouragement and I wish to thank her for helping me to keep pushing forward in each stage of the thesis process. There are many others who I will not mention by name but have, nonetheless, helped in some way. Know that your assistance has not been overlooked. Thank you one and all.

Table of Contents

List of Tables

List of Appendixes

Appendix Page

Description of Structure and Content

The body of this thesis is written as a manuscript suitable for submission to a peer-reviewed journal in speech-language pathology. An annotated bibliography is presented in Appendix B.

Introduction

Clinical assessment of the language skills of Spanish-speaking (SpS) children is becoming increasingly important in the United States. In 2000, 16 percent (about one in six) of all children in the United States under the age of 18 were Hispanic, and nearly half (44 percent) of the minorities in kindergarten through 12th grade public schools were Hispanic. Twenty-three percent of 5 to 17 year-old Hispanic children who speak Spanish at home have difficulties speaking English (U.S. Department of Education, 2003).

The language learning environment of SpS children in the United States is highly variable. The U.S Department of Education and the National Center for Education Statistics (2003) reported that "in 1999, over one-half (57 percent) of Hispanic students in kindergarten through 12th grade spoke mostly English at home, one-fourth (25 percent) spoke mostly Spanish, and 17 percent spoke English and Spanish equally" (p. 30).

Perhaps due to the heterogeneity in SpS children’s language environments, samples of a child's spontaneous language (also referred to as spontaneous language samples) are an important clinical tool for language assessment in this population. Many standardized tests, even those which are normed for Latino populations, do not provide sufficient information for this diverse group of SpS children because of how they are normed and the language forms which they test (Gutiérrez-Clellen & Simon-Cereijido, 2009). Guitierrez-Clellen, Restrepo, Bedore, Peña, and Anderson (2000) asserted that the lack of valid standardized language assessment tools for SpS children had underscored the need to use language samples. Restrepo (1998) showed that parental interview and spontaneous language sample analysis accurately discriminated between normal SpS children and those with language impairment (LI). Traditional assessment carries with it many linguistic and cultural biases, whereas the analysis of spontaneous language samples overcomes these biases for multicultural and bilingual (including Spanish-English bilingual) populations due to their naturalistic nature (Heilman et al., 2008). However, in order to be useful in assessment, a representative language sample must be properly analyzed.

Several studies have suggested clinical markers for LI in spontaneous language samples of SpS children. Bedore and Leonard (2001) observed that SpS children with LI incorrectly used plural inflection items (i.e., the children substituted the singular form of the noun for what correctly should have been the plural form), direct object clitics, and articles. The direct object clitics and articles were either omitted or substituted. Clitics are pronouns that are contracted with another grammatical form; thus, these two grammatical forms become part of the same word. For example the verb *mostrar* (‘to show’) may be contracted with the pronoun *lo* to form the word *mostrarlo* (‘to show it’). Gutiérrez-Clellen and Simon-Cereijido (2009) indicated that if clitics, articles, and verbs were less than 90% correct in a narrative sample, language impairment is likely present.

Studies have also examined how to effectively assess grammatical morphemes in language samples of SpS children. Gutiérrez-Clellen et al. (2000) discussed the importance of having a detailed analysis of the number and type of grammatical errors in the language sample of a SpS child. Much of the grammar which can be used to distinguish normal SpS children from SpS children with LI is found in the morphology of the noun phrase (e.g., *un árbol grande* (‘a big tree’) or *los dos hombres* (‘the two men’); Bedore, 2001; Bedore & Leonard, 2001; Gutiérrez-Clellen & Simon-Cereijido, 2009; Gutiérrez-Clellen, Restrepo, & Simon-Cereijido, 2006; Simon-Cereijido & Gutiérrez-Clellen, 2007). Thus, language sampling coupled with knowledge of such clinical markers can provide important information in the language assessment of SpS children.

Language sample analysis, however, requires considerable time on the part of the clinician. Long (2001) found that few grammatical analysis procedures were time-efficient for clinicians. Paul (2007) reasoned that although analyzing speech samples is more time-consuming than scoring standardized tests, the information an analyzed speech sample provides is “richer and more valid” (p. 347). Paul suggested that increasing the efficiency of language sample analysis could be accomplished by either shortcutting some of the steps involved in traditional analysis or by using computer-assisted procedures. Long (2001) concluded that “language analysis software saves time for every clinician who uses it” (p. 414).

Software for the clinical analysis of syntax has been developed and tested for English (e.g., Channell & Johnson, 1999; Long & Channell, 2001; Sagae, Levie, & MacWhinney, 2005); however, only two software options for the analysis of clinical language samples in Spanish are currently available: the MOR module from Child Language Analysis (CLAN; MacWhinney, 2000) and the Bilingual SE version of the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2010) software. The MOR module of CLAN has vocabulary word sets available in 10 languages other than English, including Spanish, which enables words to be marked for the possible parts of speech (MacWhinney 2008). However, no data have been published as to the accuracy of this automated marking or its clinical use. Recently, Sagae et. al. (2010) reported on current progress on a parser, associated with CLAN that would provide fully automated syntactic annotation for Spanish language samples. However, the accuracy of this parser's analysis of clinical language samples and its use by clinicians has yet to be documented.

In contrast, the SALT software (Miller & Iglesias, 2010) has been used extensively by clinicians; however, SALT tabulates grammatical structures which have been manually coded, rather than automatedly analyzing the grammatical constructions present in the sample. Thus its accuracy is that of the coding clinician. Extensive normative data on SALT-tabulated structures are available, and a comparison of a coded child language sample to these reference data is made as part of SALT's tabulation. However, tasks such as part-of-speech coding and syntactic structural coding must be done manually.

Recently, the software used by Channell and Johnson (1999) to automatedly code grammatical category ("part of speech") information in child language samples in English has been adapted to code the grammatical categories in child language samples in Spanish. This software uses probabilities which were extracted from hand-coded samples and stored in the program's dictionary to grammatically code samples rather than a set of grammatical rules, offering the advantage that the coding 'engine' need not be revised for a different language such as Spanish.

Pilot testing of the program's ability to code language samples in Spanish has yielded promising results. Wilson (2005), who tested the accuracy of the software on samples from native SpS children, found a mean word-level accuracy of 92.4% (*SD* = 2.1). Redd (2006), who tested the accuracy of the software on samples from bilingual children in the United States, found the mean accuracy with which automated tags matched with manual tags in Spanish to be 96.8% (*SD* = 1.8). A summary of the program's accuracy for coding each grammatical category in each of the two pilot studies is presented in Table 1. It can be seen in Table 1 that accuracy levels differed among grammatical categories, sometimes markedly, when few exemplars of a grammatical category were available for coding. This was particularly true in the Redd (2006) study.

Several characteristics of the samples used by Wilson (2005) and Redd (2006) suggest that evaluation of a grammatical coding program needs to be performed using a set of longer,

Table 1

Accuracy of automated tagging in two studies

	Wilson (2005)		Redd (2006)	
	Frequency	Accuracy (%)	Frequency	Accuracy (%)
Adjective (AJ)	916	64	684	64
Plural adjective (AJ.s)	116	52	144	53
Adverb (AV)	3060	89	2517	89
Copula (B)	417	98	857	100
Past copula (B.d)	66	75	506	99
Imperative copula (B.i)	N/A	N/A	1	100
Subjunctive copula (B.s)	170	66	4	75
Conjunction (CC)	3525	96	4233	100
Determiner (D)	5224	95	11227	100
Intensifier/qualifier (IF)	654	56	133	90
Initiator (IN)	83	78	103	78
Common noun (N)	4847	96	10742	98
Plural noun (N.s)	1047	94	1498	95
Negation (NG)	831	93	548	96
Proper noun (NP)	148	100	88	100
Other pronoun (PO)	1399	76	950	95
Personal pronoun (PP)	4782	95	4230	97
Preposition (PR)	3516	96	4682	99
PR+D	320	100	748	100
Question wh-word (Q)	220	91	36	92
Subordinator (SB)	3460	95	2487	100
Infinitive marker (TO)	580	84	438	84
Verb (V)	2470	96	3272	97
Past verb (V.d)	4443	96	4581	97
Conditional verb (V.c)	32	97	1	100
Subjunctive verb (V.s)	310	91	158	94
Imperative verb (V.i)	58	48	18	39
Present participle (V.g)	270	95	1321	98
Past participle (V.n)	265	94	143	77
V+P[a]	156	94	176	90
V+P+P	N/A	N/A	8	75
V.g+P	N/A	N/A	85	32
V.i+P	41	20	13	31
V.n+P	N/A	N/A	1	0
Auxiliary verb (X)	82	89	561	74
Past auxiliary verb (X.d)	345	85	739	93
Imperative auxiliary (X.i)	N/A	N/A	1	100
Subjunctive auxiliary (X.s)	N/A	N/A	5	80

Note. Tag abbreviations are in parentheses.
[a]This abbreviation represents a clitic pronoun, or a pronoun that is contracted with a verb.

more sustained and in-depth samples, which are more representative of those collected by clinicians. Both Wilson and Redd used language samples taken from the child language data exchange system (CHILDES) database. The samples in Wilson's study were narrative in nature; each based on 5 sampling activities or contexts, and averaged 232 utterances per sample. In Redd's study, 80 bilingual children were asked to perform description tasks for their English and Spanish language samples, and 173 children gave a Spanish narration of the wordless picture book *Frog, Where Are You?* (Mayer, 1969); these were short samples, averaging 25 utterances in length. Neither of these two pilot studies used data similar to that which is typically collected by speech-language clinicians.

Studies of SpS children suggest that narrative samples are practical to collect and that they offer valid insight into the child's grammatical system. Uccelli (2006) studied two young SpS children and suggested that narrative discourse may lead to expansion of grammatical forms and that grammar and narrative discourse progress should be studied in an integrated manner. Spanish oral narratives highlight some important grammatical features. For example, verb tense markers are used to differentiate between background and foreground events and to mark changes of state and locations (Gutiérrez-Clellen, 2002). Simon-Cereijido and Gutiérrez-Clellen (2009) elicited narrative samples in both English and Spanish to analyze grammatical forms in bilingual children. Fiestas and Peña (2004) used two different tasks to elicit narrative samples in bilingual children. The first method was a wordless picture book and the second was a culturally relevant picture that the child would tell a story about. Neither of these tasks was found to affect the grammaticality of the bilingual children's utterances. Heilman et al. (2008) found that clinicians can produce accurate and reliable transcripts of oral narrative samples in both Spanish and English. McCabe and Bliss (2004) elicited personal event narratives from Spanish-English bilinguals with and without LI. Among other things, these personal event narrative samples were analyzed according to specific syntactic productions involving subject, verb, and object relationships. Personal narratives have been found to be of greater quality than fictional narratives (McCabe et al., 2008) and seem to be culturally appropriate for Spanish speakers whose cultural tradition often involves narratives that focus on experiences (McCabe & Bliss, 2004). Thus “analysis of children’s oral narratives is a viable option for clinicians working with bilingual children” (Heilman et al., 2008, p. 186).

In summary, there is a need for language sample analysis in the clinical assessment of SpS children. Narratives, especially personal narratives, offer excellent data for doing so. Computer software might aid in the analysis of these samples, performing such tasks as automated part-of-speech coding. New software offers such analysis for samples in Spanish, but its accuracy has yet to be measured when coding samples collected using an accepted clinical technique such as high quality event narrative language samples. Acceptably high accuracy in automatically analyzing those grammatical categories pertinent to the diagnosis of SpS children with LI (i.e., articles, clitic pronouns, and verbs) would be useful in clinical practice. The current study measures the accuracy of the automated grammatical category coding software in analyzing personal and fictional event narrative samples of SpS children. Special attention will be given to the results of those grammatical categories which act as clinical markers for LI.

Method

Participants

Shiro (1997) had collected language samples from 113 typically developing children. These children were first or fourth grade students between the ages of 6 to 10 years who attended school in Caracas, Venezuela and were native Spanish speakers. The children had been selected from three public schools and three private schools. Those selected from private schools came from high socioeconomic status (SES) families, and those selected from public schools came from low SES families. Each child produced four event narratives, which were then combined into one sample for each child. The samples had been donated to the online CHILDES database (MacWhinney, 2000). Thirty of these samples were randomly selected for use in the present study. Of the 30 samples randomly selected, 11 were from children in public schools and 19 were from children in private schools.

Procedure

Narrative language sampling. Four narratives had been elicited from each of the subjects. The narratives had been collected for a doctoral dissertation by Martha Shiro of Universidad Central de Venezuela. Before the narrative elicitation tasks, the clinician had a warm-up conversation with the child. During this time the child being interviewed was asked about his or her personal background. Shiro then used four tasks to elicit a variety of personal narratives, including personal-event narratives and fictional narratives.

The first task was a personal event narrative elicited using an open-ended prompt. The elicitor asked the child (in Spanish) to tell a story about a frightening experience. The second task used a structured prompt, similar to that reported by McCabe and Bliss (2004). In Spanish, the elicitor first described a memorable anecdote such as seeing a snake, cutting their finger, or going to the hospital. After a brief description the child was asked, "*¿a ti te pasó algo parecido?*" ('Did something like this ever happen to you?'). Once the desired response had been obtained, the child was asked to describe his/her experience. Three different prompts were used to ensure that each child produced at least one story (Shiro, 1997). Several studies (e.g., McCabe

& Bliss, 2004; McCabe et al., 2008) have used a similar method of eliciting personal event narrative language samples from children with similar cultural and linguistic backgrounds to those in the current study.

The third task used an open ended prompt, wherein the child was asked to tell about a favorite film, video, or TV program. The fourth elicitation task incorporated a structured prompt to help the child produce a narrative retelling: The child was first shown a wordless animated video and then asked to tell the story to the elicitor, who was not present when the video was shown and pretended to not be familiar with the story (Shiro, 1997).

Transcription. The language samples had been previously transcribed into the CLAN format and put onto the CHILDES database. Minor adjustments were required in order to make the samples analyzable by the software program. The marker (“CHI:”) beginning each child utterance was deleted. Mazes (fillers, false starts, repetitions, or interjections) were placed in parentheses to be ignored by the program. Each utterance began on a new line. The transcriptions were then prepared for tagging.

Manual tagging. After transcription had been completed, each word in every child utterance in each sample was manually tagged for grammatical category by a graduate student in speech-language pathology who was fluent in Spanish. The set of grammatical categories used (and listed earlier in Table 1) was based on those categories used in the Language Assessment Remediation Screening Profile (LARSP; Crystal, Garman, & Fletcher, 1989) procedure, adapted to Spanish grammar as described in Butt and Benjamin (2000). A second person, also fluent in Spanish, independently coded 10% of the utterances in the samples. These two codings were compared on a tag-by-tag basis, and interrater reliability was calculated and found to be 93%.

Tagging software. The tagging software uses two types of probability to grammatically code each word in an utterance. The first is the likelihood of the different code options for a single word. For example, in English the word *can* is far more likely to be a modal auxiliary (*she can see you.*) than a noun (*put it in the can.*), and twice more likely to be a noun than a verb (*we're gonna can peaches.*).

The second probability data used is the likelihood that a given grammatical tag will follow the two previous grammatical tags in an utterance. For example, a tag of *noun* is far more likely to follow the tags *determiner* and *adjective* than is the tag *modal auxiliary*. The program thus reconciles these two forms of probability to determine the most likely sequence of tags for a set of words. Technically, this is termed a tri-gram Hidden Markov model (Jurafsky & Martin, 2000; Manning & Schütze, 1999), and the model seems to work well even when faced with the "messy" data implicit in unedited oral child language (Channell & Johnson, 1999).

The program contains a dictionary containing word tag option frequencies and tag-transition frequencies, which were extracted from a manually tagged training corpus. The frequency information was derived from a training corpus which included a manually-tagged Spanish book of children's stories, various speeches in Spanish, and 10 child language samples (Benedet, Cruz, Carrasco, & Snow, 2005) from children growing up in Spain. The dictionary does not contain proper nouns (which are transcribed with an initial capital letter, if the user wants them correctly coded) or interjections (which are placed in parentheses and ignored). Unknown words are guessed as nouns, or as plural nouns if ending in -*s*.

Automated tagging. Software analysis was completed as the transcribed language samples were run through the grammatical coding software. The accuracy of grammatical coding was defined as the level of agreement of the software with the manual coding; this level was tabulated by a utility program.

Results

Word-level Accuracy

The word-level accuracy with which the software tagged each sample ranged between 86.5% and 95.0% with a mean accuracy level of 91.0% (*SD* = 1.7). This accuracy was obtained by comparing the automatedly tagged samples with the author's manually tagged samples. The word-level accuracy of each individual sample is listed in Table 2 along with the corresponding child's name, age, the number of utterances in the sample, and total number of tags in the sample. It can be seen in Table 2 that samples ranged in length from 113 utterances up to 442 utterances, with an average length of 205 utterances (*SD* = 76). The number of tags used ranged from 579 up to 2614, with an average of 1254 (*SD* = 459) tags produced in these samples.

Correlations were examined among the variables of age, number of utterances, total number of tags, and tagging accuracy. These correlations revealed a statistically significant ($p < .05$) link between age and number of tags ($r = 0.345$) and between age and accuracy ($r = -0.346$). Other correlations were not significant. These correlations indicate that older children used more words and thus grammatical tags in their sample, but that samples from older children were tagged with a slightly lower level of accuracy.

Tag-by-Tag Accuracy

Percentages of tag-by-tag accuracy and percentages of tag confusions based on all 30 samples are shown in Table 3. Manual tags are presented with their frequency of occurrence in addition to the percentages with which these tags were automatedly tagged or confused with

Table 2

Individual sample accuracy of language samples from 30 children

Child	Age	N of Utterances	N of Tags	Accuracy (%)
Ewar	7;2	190	1021	88.1
Gabriela	9;9	181	961	92.1
Katy	9;8	277	1635	92.5
María	10;4	226	1261	90.3
Alberto	7;9	113	610	90.3
Juanfran	10;8	164	1297	90.4
Juan	10;8	148	1005	92.2
María	10;9	201	1701	86.5
Patricia	7;10	178	1235	89.1
Rodrigo	7;6	219	1431	90.9
Sara	10;7	122	878	91.0
Franyu	7;5	275	1557	91.5
Jordan	7;1	199	1236	89.9
Solimar	9;10	114	579	95.0
Wilson	10;9	165	809	90.6
Alfredo	7;0	139	798	93.9
Alicia	10;3	185	1270	88.7
Annette	7;3	256	1442	91.5
César	10;4	442	2614	90.7
Eduardo	10;7	389	1937	89.9
Anger	9;9	161	1410	91.8
Douglas	6;6	138	974	90.1
Nicolás	10;5	303	2132	88.9
Clio	9;9	198	1253	92.0
Cristobal	9;10	246	1395	91.5
Francisco	7;0	180	955	92.3
Iván	6;6	153	836	90.4
Maried	10;6	170	973	91.8
Oriana	9;10	266	1669	92.3
Pablo	7;3	151	745	93.0

Table 3

Frequency, accuracy, and confusions for each automated tag

Tag	Frequency	Accuracy (%)	Confusions (%)
AJ	554	55	AV (4) D (1) N (25) NP (2) PO (1) V (3) V.i (1) V.n (6) V.s (1)
AJ.s	86	41	D (1) IF (1) N (2) N.s (40) NP (3) V (1) V.i+P (9) V.n (1)
AV	2710	89	AJ (2) IF (1) N (1) PO (1) PR (4) SB (1)
B	449	95	X (3)
B.d	470	72	V.d (24) X.d (4)
CC	2785	99	N (1)
D	4742	96	PO (1) PP (1)
IF	629	38	AJ (1) AV (2) CC (1) D (3) N (23) NG (18) PP (14) SB (1)
IN	88	69	CC (1) D (24) IF (2) NG (1) PO (1) V (1)
N	4656	92	AJ (2) N.s (1) NP (1) V (2)
N.s	1062	91	AJ.s (3) N (1) NP (1) V (1) V.n (1) V.s (1)
NG	453	94	D (1) N (4)
NP	755	98	N (1)
PO	1164	70	AJ (1) AJ.s (1) AV (1) D (18) IN (1) N (4) N.s (1) NP (1) V (2)
PP	4127	94	D (5) PO (1)
PR	2971	96	N (1) SB (2) V (1)
PR+D	256	93	N (6)
Q	100	82	D (4) IF (7) N (2) SB (5)
SB	1992	97	PR (2) Q (1)
TO	414	77	PR (23)
V	2519	93	N (3) V.d (2) V.i (1)
V+P	119	96	N (3) N.s (1) V.i+P (1)
V.d	3211	93	B.d (1) N (3) V (1) V.s (1)
V.g	514	96	N (2) V.n (1)
V.n	179	82	AJ (2) N (11) N.s (1) NP (1) V (1) V.d (1) V.i+P (2) X (1) X.d (1)
V.s	84	57	B.s (36) N.s (2) V (1) V.d (1) V.i (1) X.s (1)
X	92	86	B (4) D (1) NG (1) PP (1) V (4) V.s (1) X.d (1)
X.d	355	84	B.d (9) V.d (7) X.s (1)

Note. Tag abbreviations are used in this table; they represent the following grammatical categories: adjective (AJ), plural adjective (AJ.s), adverb (AV), copula (B), past copula (B.d), conjuction (CC), determiner (D), intensifier (IF), initiator (IN), noun (N), plural noun (N.s), negation (NG), proper noun (NP), other pronoun (PO), personal pronoun (PP), preposition (PR), question wh-word (Q), subordinator (SB), infinitive marker (TO), verb (V), past verb (V.d), present participle (V.g), imperative verb (V.i), past participle (V.n), subjunctive verb (V.s), auxiliary verb (X), past auxiliary verb (X.d).

other tags. Only the tags that appeared more than 30 times in the combined 30 samples are listed in Table 3. It can be seen in Table 3 that the accuracy of these grammatical category tags ranged from 38% for intensifier/qualifiers (IF) up to 99% for coordinating conjunctions (CC). In addition, Appendix A presents data on the more clinically salient categories which have been taken from Table 3 and combined with data on those same categories from the studies by Wilson (2005) and Redd (2006), which had been presented in Table 1.

Discussion

The automated tagging software program achieved moderate overall accuracy when tagging event narratives of SpS children. The software program tended to tag older children's event narrative language samples less accurately than those of younger children. Furthermore, as would be expected, older children tended to use more tags than their younger counterparts. The present study revealed slightly lower overall word-level accuracy than had been seen in previous similar studies (Redd, 2006; Wilson, 2005). However, the overall level of tagging accuracy is greatly affected by the grammatical composition of the samples. For example, a program that is accurate at tagging nouns will have a higher level of accuracy if the sample contains a greater number of nouns. Thus it is important to examine the accuracy with which each grammatical category was coded, indicated by the tag-by-tag accuracy levels.

The tag-by-tag accuracy of the program, for those tags which occurred 30 times or more, was similar to those of the studies by Redd (2006) and by Wilson (2005), but was most similar to the Wilson study. This was most likely due to the similarities of these studies in terms of sample length and child age. The accuracy with which the program tagged intensifiers and plural adjectives was markedly lower than the already low accuracy reported in the Wilson study.

Examination of the tag confusion data in Table 3 provides one possible reason: each word that the program does not recognize is tagged as a noun, or if it ends in "s," the word is tagged as a plural noun. A total of 23% of intensifier/qualifiers were tagged as nouns by the program, and a total of 40% of plural adjectives were tagged as plural nouns. Apparently, a number of plural adjectives and intensifier/qualifiers are not included in the program's dictionary, leading to a poor level of accuracy for coding these grammatical categories. Another possible explanation for this might be that the Wilson (2005) study and the data upon which the program was originally trained came from children in Mexico, and the children sampled in the present study were in Venezuela. Perhaps training the program on data from a greater variety of countries and thus local dialects might increase the levels of accuracy which may be obtained.

One goal of the present study was to report how accurately the software tagged grammatical forms that help identify LI in SpS children, such as articles, clitic pronouns, and verbs (Bedore, 2001; Bedore & Leonard, 2001; Gutiérrez-Clellen & Simon-Cereijido, 2009; Gutiérrez-Clellenet al., 2006; Simon-Cereijido & Gutiérrez-Clellen, 2007). Although the software cannot currently be used to identify errors in such forms, reports of automated tagging frequency and accuracy of such markers may lead to further discovery of clinical usefulness.

Because clitic pronouns and articles tend to be either omitted or substituted, frequency data of these markers may be beneficial. If a child has omitted a clitic pronoun or article then it obviously would not be tagged and counted; furthermore, if the child substituted one of these with a different grammatical form then it would also not be included in the frequency count. Normative data regarding the frequency of these grammatical forms would allow quick comparisons between the frequency with which a current grammatical form appears in a child language sample and the frequency with which other children the same age use the form. Future studies might collect normative data on these grammatical structures.

The accuracy of the software in tagging those clinically significant markers is comparable to that shown in previous studies. The determiners (which include articles) were tagged with 96% accuracy; the software achieved 95% accuracy or higher at tagging this grammatical category in the Wilson (2005) and Redd (2006) studies. Clitic pronouns appeared in various forms in the present and these previous studies (e.g., V+P, V+P+P, V.n+P, V.i+P); however, the most common and frequently occurring form was V+P. The software tagged V+P with 96% accuracy in the present study, which was slightly more accurate than in the Wilson (2005) and Redd (2006) studies. No other forms of clitic pronouns appeared in the current study more than 30 times, but some were reported in the Wilson (2005) and Redd (2006) studies. The accuracy with which the software tagged these forms in the Wilson (2005) and Redd (2006) studies was low; however, the frequency of occurrence was similarly low. All verbs in the current study were tagged with 82% accuracy or higher except subjunctive verbs and past tense copula verbs. Confusion data showed that the software tagged 36% of subjunctive copulas as subjunctive verbs. The errors in tagging past tense copulas were mainly a result of the software confusing past tense copulas with regular past tense verbs.

In the Wilson (2005) study the software achieved 82% accuracy or higher in all corresponding verb categories except the past tense copula, and in the Redd (2006) study only the past participle and auxiliary verb categories did not achieve this level of accuracy. The accuracy of common and plural nouns was 91% or greater across all studies. The mean accuracy with which the program tagged clinically significant categories (i.e., B, B.d, D, N, N.s, V, V.d, V.g, V.n, V.s, V+P, X, X.d) across all three studies was 91%. Comparison of these three studies reveals high overall accuracy for automated tagging of those grammatical forms which have been shown to be clinically significant in identifying SpS children with LI (Bedore & Leonard, 2001; Gutiérrez-Clellen & Simon-Cereijido, 2009). Thus although some grammatical categories were tagged with low accuracy, tagging accuracy for the more clinically significant categories was above 90%, suggesting an area of clinical application for the tagging program.

The present study and the studies by Wilson (2005) and Redd (2006) represent a preliminary effort to identify clinically useful grammatical categories for SpS children's language samples. The categories used in all three studies were inspired by the English version of the LARSP (Crystal et al, 1989). Recently, developments have been made in adapting the LARSP categories for 13 languages other than English, including an adaptation for Spanish. Further research might profitably use software to automatedly tag language samples using this new and better-documented categorization scheme (Ball, Crystal, & Fletcher, 2012). Also, due to the preliminary nature of this and other similar studies, the manual tagging of grammatical categories was performed by proficient but non-native Spanish speakers. Future research using this and similar software programs might benefit by involvement of native Spanish-speaking researchers.

The child language samples for this and the previous studies by Wilson (2005) and Redd (2006) were taken from the CHILDES database (MacWhinney, 2000), which contains a limited set of samples obtained from SpS children. Future studies might well examine samples collected using a variety of elicitation methods from children who represent a greater range of ages, dialects, and ability levels. Samples elicited by speech-language clinicians, for example, might better reflect the samples used to clinically assess SpS children's syntax. Furthermore, samples of younger children would add further insight into the performance and clinical usefulness of the software program. To date, the accuracy of the software program has been measured on Mexican, Puerto Rican, and Venezuelan SpS children. Given the dialectal differences possible in Spanish, testing this grammatical tagging program on a more expansive range of Spanish dialects would be beneficial.

The output of the English version of the grammatical tagging software can be used in a number of ways to derive clinically relevant data (Long & Channell, 2001); for example, the mean length of utterance, the LARSP (Crystal et al, 1989), Developmental Sentence Scoring (Lee, 1974), and the Index of Productive Syntax (IPSyn; Scarborough, 1990). Although these or similar measures are not yet available to be derived from the Spanish version of this study's software, it is anticipated that the data from the present will be more clinically useful as similar software is developed in Spanish. As previously discussed, frequency data of clinically relevant tags may give clinicians rapid access to information regarding a child’s grammatical development. In the present study, the program has shown generally good accuracy in tagging determiners, most verbs, and clitic pronouns contracted with verbs (i.e., V+P). Tagging is, nonetheless, different than identifying errors in the child’s use of grammatical forms. Currently, the program could be used to mark clinically significant forms, after which the clinician would be required to manually identify any errors in the child’s use of these forms. This would save time by helping the clinician skip one step of the process; however, the procedure would remain time consuming and impractical for most clinicians. To allow greater clinical utility, the Spanish grammatical tagging software could identify errors in these clinical markers. The program might be expanded to allow it to identify misconjugated verbs and articles that don’t agree in number or gender with their corresponding noun.

Nevertheless, the results of the present study indicate that Spanish fully automated grammatical tagging software can accurately tag clinically relevant grammatical forms. With further development, the program may become a viable tool for clinicians, helping them in the process of identifying SpS children with LI.

References

Ball, M., Crystal, D., & Fletcher, P. (2012). *Assessing grammar: The languages of LARSP* Bristol, UK: Multilingual Matters.

Bedore, L. M. (2001). Assessing morphosyntax in Spanish-speaking children. *Seminars in Speech and Language, 22*(1), 65.

Bedore, L. M., & Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*(4), 905-924. doi:10.1044/1092-4388(2001/072)

Benedet, M., Cruz, C., Carrasco, M., & Snow, C. (2005). *Becacesno database.* Retrieved May 26, 2005 from http://childes.psy.cmu.edu

Butt, J., & Benjamin, C. (2000). *A new reference grammar of modern Spanish* (3rd ed. ed.). Chicago: McGraw-Hill.

Channell, R. W. (2003). Automated developmental sentence scoring using computerized profiling software. *American Journal of Speech-Language Pathology, 12*(3), 369-375. doi:10.1044/1058-0360(2003/82)

Channell, R. W., & Johnson, B. W. (1999). Automated grammatical tagging of child language samples. *Journal of Speech, Language, and Hearing Research, 42*(3), 727-734.

Crystal, D., Garman, M., & Fletcher, P. (1989). *The grammatical analysis of language disability: A procedure for assessment and remediation* (2nd ed. ed.). London: Cole and Whurr.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*(2), 155-168. doi:10.1044/0161-1461(2004/016)

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13*(2), 175-197. doi:10.1016/S0898-5898(01)00061-4

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L. M., Peña, E. D., & Anderson, R. T. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*(1), 88-98.

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simon-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research, 49*(6), 1209-1223. doi:10.1044/1092-4388(2006/087)

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language,* doi:10.1055/s-0029-1241722

Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders, 28*(2), 178-188.

Lee, L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Long, S. H. (2001). About time: A comparison of computerized and manual procedures for grammatical and phonological analysis. *Clinical Linguistics & Phonetics, 15*(5), 399-426. doi:10.1080/02699200010027778

Long, S. H., & Channell, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology, 10*(2), 180-188. doi:10.1044/1058-0360(2001/017)

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. *In H. Behrens (Ed.), Trends in corpus research: Finding structure in data,*165-198. Amsterdam: Benjamins.

Mayer, M. (1969). *Frog, where are you?*. New York: Dial Press.

McCabe, A., & Bliss, L. S. (2004). Narratives from Spanish-speaking children with impaired and typical language development. *Imagination, Cognition and Personality, 24*(4), 331-346. doi:10.2190/CJQ8-8C9G-05LG-0C2M

McCabe, A., Bliss, L., Barra, G., & Bennett, M. (2008). Comparison of personal versus fictional narratives of children with language impairment. *American Journal of Speech-Language Pathology, 17*(2), 194-206.

Miller, J., & Iglesias, A. (2010). *Systematic analysis of language transcripts (SALT)*. Muscoda, WI: SALT Software LLC.

Paul, R. (2007). *Language disorders from infancy through adolescence assessment & intervention* (3rd ed. ed.). St. Louis, MO: Mosby Elsevier.

Redd, N. (2006). *Automated grammatical tagging of language samples from Spanish-speaking children learning English.* (Unpublished master's thesis). Brigham Young University, Provo, UT.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1398-1411.

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

Sagae, K., Lavie, A., & MacWhinney, B. (2005). Automatic measurement of syntactic development in child language. *Proceedings of the 43rd Meeting of the Association for Computational Linguistics,* Ann Arbor, MI. 197-204.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11,* 1-22.

Shiro, M. (1997). *Getting the story across: A discourse analysis approach to evaluative stance in Venezuelan children's narratives.* (Unpublished doctoral dissertation) Universidad Central de Venezuela, Caracas, Venezuela.

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*(2), 317-339. doi:10.1017/S0142716407070166

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics*, *2009*, *30*(2), 315-337.

U.S. Department of Education, National Center for Education Statistics. (2003). *Status and trends in the education of Hispanics.* Washington, DC: Thomas D. Snyder.

Uccelli, P. (2009). Emerging temporality: Past tense and temporal/aspectual markers in Spanish-speaking children's intra-conversational narratives. *Journal of Child Language*, *36*(5), 929-966. doi:10.1017/S0305000908009288

Wilson, K. M. (2005). *Automated grammatical analysis of children's Spanish language samples.* (Unpublished master’s thesis). Brigham Young University, Provo, UT.

Appendix A: Frequency and Accuracy of Clinical Markers in Three Studies

	Wilson (2005)		Redd (2006)		Present Study	
	Frequency	Accuracy (%)	Frequency	Accuracy (%)	Frequency	Accuracy (%)
B	417	98	857	100	449	95
B.d	66	75	506	99	470	72
D	5224	95	11227	100	4742	96
N	4847	96	10742	98	4656	92
N.s	1047	94	1498	95	1062	91
V	2470	96	3272	97	2519	93
V.d	4443	96	4581	97	3211	93
V.g	270	95	1321	98	514	96
V.n	265	94	143	77	179	82
V.s	310	91	158	94	84	57
V+P	156	94	176	90	119	96
X	82	89	561	74	92	86
X.d	345	85	739	93	355	84

Appendix B: Annotated Bibliography

Bedore, L. M. (2001). Assessing morphosyntax in Spanish-speaking children. *Seminars in Speech and Language, 22*(1), 65.

This article outlined the nature of the Spanish morphosyntactic system and discussed implications of such for planning clinical assessment tasks. Language sample analysis and structured assessment tasks were discussed. The article suggested that counts of how often children use each article and clitic type be obtained to aid in analysis.

Bedore, L. M., & Leonard, L. B. (2001). Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*(4), 905-924. doi:10.1044/1092-4388(2001/072)

This study observed the use of grammatical morphology in SpS preschoolers with specific language impairment and compared their use of grammatical morphemes to same-age peers or younger children with similar mean lengths of utterance. Language impaired children showed limited use of verb inflections and noun-phrase morphology. They incorrectly used plural inflection items and either omitted or substituted direct object clitics, and articles.

Cereijido, G., & Gutiérrez-Clellen, V. F. (2009). A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistics, 2009, 30*(2), 315-337.

The relationship between lexical and grammatical abilities of bilingual Spanish and English speaking school-aged children was examined. A high correlation within, but not across languages was found. Narrative language samples were elicited using wordless frog stories and were transcribed and analyzed using SALT.

Channell, R. W. (2003). Automated developmental sentence scoring using computerized profiling software. *American Journal of Speech-Language Pathology, 12*(3), 369-375. doi:10.1044/1058-0360(2003/82)

The accuracy of fully automated developmental sentence scoring (DSS) analysis performed by computerized-profiling (CP) software was analyzed. Data was based on language samples from 48 childen, 28 of whom had language impairment. The overall accuracy across all samples was found to be 78.2%.

Channell, R. W., & Johnson, B. W. (1999). Automated grammatical tagging of child language samples. *Journal of Speech, Language, and Hearing Research, 42*(3), 727-734.

This study assessed the accuracy of the English version of the automated grammatical categorization ("tagging") software. Thirty conversational child language samples were tagged using the software program and compared to their manually tagged counterparts. Automated accuracy levels averaged 95.1% on a word-by-word basis.

Fiestas, C. E., & Peña, E. D. (2004). Narrative discourse in bilingual children: Language and task effects. *Language, Speech, and Hearing Services in Schools, 35*(2), 155-168. doi:10.1044/0161-1461(2004/016)

Two different tasks to elicit narrative samples in bilingual children were studied. The first method was a wordless picture book and the second was a culturally relevant picture about which the child told a story. Neither of these tasks was found to affect the grammaticality of the bilingual children's utterances.

Gutiérrez-Clellen, V. F. (2002). Narratives in two languages: Assessing performance of bilingual children. *Linguistics and Education, 13*(2), 175-197. doi:10.1016/S0898-5898(01)00061-4

This study observed the narrative discourse of bilingual children in the process of learning English. It studied the narrative recall and comprehension of these children in both English and Spanish. It showed that there may be differences in what similar narrative tasks demand of bilingual and monolingual speakers.

Gutiérrez-Clellen, V. F., Restrepo, M. A., Bedore, L. M., Peña, E. D., & Anderson, R. T. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*(1), 88-98.

This article discussed how spontaneous language samples can be used in the assessment of children who speak both Spanish and English, and how available procedures can be applied. It suggested that sociolinguistic influences be taken into account and that Spanish grammar and mean length of utterance (MLU) be assessed. It proposed that knowing the number of syntactic or morphological errors within each sentence may be helpful in identifying specific language impairment.

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simon-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research, 49*(6), 1209-1223. doi:10.1044/1092-4388(2006/087)

This study observed how accurately the Spanish Morphosyntax Test (S-MST) could discriminate between SpS children with and without LI in two categories: those who only spoke Spanish, and those who also spoke some English. S-MST was designed to include articles, clitics, and verbs because of their predicted sensitivity to LI in SpS children. Eighty children with LI and 80 without LI were sampled. It was found that bilingual (as opposed to monolingual) children were not more likely to be misclassified and that the measure would, thus, be appropriate in assessing Spanish-dominant or Spanish-only speakers between 4 and 6 years of age.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language,* doi:10.1055/s-0029-1241722

This study suggested that rather than using formal language assessments, informal assessment tools, such as spontaneous language samples, be used to assess Spanish-English-speaking children. Samples obtained in both languages leads to more accurate diagnosis of language impairment in this population. The study observed diagnostic indicators in both English and Spanish language samples. One suggestion given for the assessment of Spanish language samples was analysis of grammatical accuracy on articles, verbs, and clitic pronouns.

Heilmann, J., Miller, J. F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K. D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders, 28*(2), 178-188.

The accuracy and reliability of transcription and analysis of oral narrative language samples taken from bilingual children were measured. There was no significant difference in the accuracy and reliability of transcriptions done by different clinicians. Short oral narratives are reliable and consistent over time, and these samples can and should be used by clinicians to assess bilingual children.

Long, S. H. (2001). About time: A comparison of computerized and manual procedures for grammatical and phonological analysis. *Clinical Linguistics & Phonetics, 15*(5), 399-426. doi:10.1080/02699200010027778

This article compared the time it takes to analyze a language sample manually to the time it takes to run a computerized analysis of the sample. Grammatical transcription and analysis of language samples are time consuming. Three child language samples were used. Several phonological and grammatical analyses were performed including LARSP, DSS, and IPSyn. Long found "language analysis software saves time for every clinician who uses it". Exactly how much time analysis took, and how much time was saved, was affected by the type of sample, the type of analysis, and the efficiency of the individual analyzing the sample.

Long, S. H., & Channell, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech-Language Pathology, 10*(2), 180-188. doi:10.1044/1058-0360(2001/017)

Four language analysis procedures were performed using software. MLU, LARSP, IPSyn, and DSS were analyzed using computerized profiling software. It was found that the accuracy of these four automated analyses were comparable to the manual interrater reliability on the same analyses.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

This manual described the CLAN program, which is used to analyze data transcribed from the CHILDES. It explained the features of CLAN and how to use the program.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. *In H. Behrens (Ed.), Trends in corpus research: Finding structure in data*, pp. 165-198. Amsterdam: Benjamins.

An overview of the CHILDES, including a discussion on how it influenced research and its anticipated future contributions. In the system morphosyntactic analysis can be performed through MOR, POST, and GRASP programs. Morphological tagging was performed and was shown to assign tags with 94% accuracy. MOR grammars are available in Spanish as well as 9 other languages. A detailed explanation of the English MOR and POST programs for automated tagging was given.

McCabe, A., & Bliss, L. S. (2004). Narratives from Spanish-speaking children with impaired and typical language development. *Imagination, Cognition and Personality, 24*(4), 331-346. doi:10.2190/CJQ8-8C9G-05LG-0C2M

Personal narratives were elicited in order to analyze the difference in personal narrative production of Spanish-English bilingual children with and without LI. The study found that those with LI had difficulty elaborating narratives in both languages. There were significant correlations between the English and Spanish narratives.

McCabe, A., Bliss, L., Barra, G., & Bennett, M. (2008). Comparison of personal versus fictional narratives of children with language impairment. *American Journal of Speech-Language Pathology*, *17*(2), 194-206.

This study compared production of personal narratives and fictional narratives in children with language impairment. It was found that the quality of personal narratives in these

children exceeded the quality of fictional narratives and suggested that personal narratives be considered in assessment, intervention, and research.

Paul, R. (2007). *Language disorders from infancy through adolescence assessment & intervention* (3rd ed. ed.). St. Louis MO: Mosby Elsevier.

This textbook is a comprehensive guide to assessment and intervention of language disorders from infancy through adolescence. Paul discussed and explained in detail the theoretical and practical implications of working with these populations. Detailed descriptions of assessment and intervention techniques for the prelinguistic, emerging language, developing language, and advanced language periods were given.

Redd, N. (2006). *Automated grammatical tagging of language samples from Spanish-speaking children learning English.* (Unpublished master's thesis). Brigham Young University, Provo, UT.

This study analyzed the accuracy of automated grammatical tagging in Spanish and English language samples of bilingual children. Samples were taken from 254 Puerto Rican children living in the United States. Two different groups of conversational samples taken from the CHILDES database were used. One group, which included 80 bilingual children, performed a descriptive task. The other group, which included 173 bilingual children, gave a narration of a wordless picture book. The mean accuracy with which automated tags matched with manual tags in English was 96.4%. The mean accuracy in Spanish was 96.8%.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1398-1411.

This study identified a set of measures that would discriminate SpS children with normal language from those with language impairment. The four measures that most accurately discriminated between these groups were parent report, mean length of T-unit, number of errors per T-unit, and history of speech and language problems.

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

This article explained and described a project which annotated grammatical relations in the form of labeled dependency structures and used these manually tagged grammatical relation annotations to develop a data-driven parser. The article focused on the English parser, but also explained progress and future plans with the parser in Spanish and Hebrew.

Shiro, M. (1997). *Getting the story across: A discourse analysis approach to evaluative stance in Venezuelan children's narratives.* (Unpublished doctoral dissertation) Universidad Central de Venezuela, Caracas, Venezuela.

This dissertation included the collection and use of 444 personal event narrative language samples from 113 Venezuelan children. The children were given open-ended and structured prompts to elicit production of personal narratives and fictional narratives.

Simon-Cereijido, G., & Gutiérrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28*(2), 317-339. doi:10.1017/S0142716407070166

This study analyzed various ways that spontaneous language samples of Spanish-speaking preschoolers with language impairment may be analyzed to identify the affected children.

Several main analyses were deemed fair discriminators, including determining the correct use of articles, verbs, and clitics. However, this method was found to perhaps miss children whose language is less syntactically complex.

U.S. Department of Education, National Center for Education Statistics. (2003). *Status and trends in the education of Hispanics.* Washington, DC: Thomas D. Snyder.

This report gave information on the demographics of the Hispanic population in the United States. It detailed many characteristics and achievements of this population at different educational levels. Included is a report on family structure and the language spoken at home.

Uccelli, P. (2009). Emerging temporality: Past tense and temporal/aspectual markers in Spanish-speaking children's intra-conversational narratives. *Journal of Child Language, 36*(5), 929-966. doi:10.1017/S0305000908009288

This study observed the verb usage of Spanish-speaking children in intra-conversational narrative language samples. Two young, Spanish-speaking children were followed over a year and their speech samples were analyzed according to verb tense, temporal/aspectual markers and narrative components. The children were found to be able to construct narratives with specific temporal relations by age 3. The relationship between discourse level and grammatical forms was discussed. The results suggested an integrated approach in the study of grammatical and discourse progress.

Wilson, K. M. (2005). *Automated grammatical analysis of children's Spanish language samples.* (Unpublished master’s thesis). Brigham Young University, Provo, UT.

Twenty-four child language samples were tagged using the Spanish version of the

automated tagging software. Automatedly tagged results were compared to manually tagged results of the same sample. The 24 samples were conversational samples taken from the CHILDES database. All children sampled were native Spanish speakers and ranged from six to twelve years old. The mean word-level accuracy for the 24 language samples was 92.4%.