Tense Marking and Spontaneous Speech Measures in Spanish Specific Language Impairment: A Discriminant Function Analysis

John Grinstead, a Alisa Baron, a Mariana Vega-Mendoza, b Juliana De la Mora, c Myriam Cantú-Sánchez, d and Blanca Flores e

Purpose: To test the proposal that the tense deficit that has been demonstrated for children with specific language impairment (SLI) in other languages is also found in child Spanish and that low performance on tense-related measures can distinguish Spanish-speaking children with SLI from those without.

Method: The authors evaluated evidence from existing spontaneous production, elicited production, and grammaticality judgment studies of finiteness in child Spanish. They measured the relationship of 7 spontaneous speech measures with previous receptive and expressive measures of finiteness and performed a discriminant function analysis, using tense as the target variable, to classify monolingual child Spanish \( n = 55 \) as representing SLI or as typically developing (TD).

Results: Spontaneous speech measures correlated with the results of previous receptive and expressive measures of child Spanish that show a tense deficit. The SLI group was shown to have statistically lower scores than the TD group on 6 of 7 spontaneous speech measures. Multiple discriminant functions, including tense measures by themselves and in combination with spontaneous speech measures, were shown to provide fair to good sensitivity and specificity in the classification of children as having SLI vs. TD.

Conclusion: The findings support the contention that the tense-marking deficit is a plausible clinical marker of SLI for Spanish-speaking children.

Key Words: specific language impairment, Spanish, optional infinitive, interface delay, monolingual

Consistent tense marking develops slowly in a large array of child languages (see Hyams, 2007, and Grinstead, 2010, for reviews). The nature of the phenomenon is that children optionally produce and judge as grammatical finite and nonfinite verbs from the time they begin using verbs with subjects until roughly 4;6 (years;months) in child English (Rice, Wexler, & Hershberger, 1998; Rice, Wexler, & Redmond, 1999). In some cases, the verb lacks bound finiteness morphemes, as in the “a” examples in 1 and 2 from the Brown corpus (Brown, 1973) and the Bloom corpus (Bloom, Lightbown, & Hood, 1975, cited in Harris & Wexler, 1996, p. 11). In other cases, the missing finiteness morphemes would appear to have been auxiliary verbs, as in 3 and 4 from the Suppes (1974) corpus and the Sachs corpus (Sachs, 1983, cited in Vainikka, 1993, pp. 268–272):

Example 1: Eve (2;0—file 14)
   a. It only write on the pad.
   b. My finger hurts.

Example 2: Peter (3;3—file 8)
   a. Patsy need a screw.
   b. This goes in there.

Example 3: Naomi (2;1)
   Me wearing curtain.

Example 4: Nina (2;0—file 5)
   I popping balloons.
Notice that the “a” and “b” examples in 1 and 2 are from the same recording sessions, illustrating the optionality of finiteness marking in what Wexler (1994) dubbed the optional infinitive stage.

Turning to the optional infinitive stage in child Spanish, one finds a common misunderstanding, which is the idea that typically developing (TD) child Spanish speakers are adultlike so early in the development of their tense marking that there is no perceptible optional infinitive stage that they pass through. Early studies, based on spontaneous production data, including studies of child Spanish (Grinstead, 1994), Catalan (Bel, 2001; Torrens, 1995), and Italian (Guasti, 1994), typically have concluded that children’s verbs were adultlike. However, a shift in this view appears to have taken place, in that the authors of more recent work—including Radford and Ploennig-Pacheco (1995); Davidiak and Grinstead (2004); Davidson and Goldrick (2003); Clahsen, Aveledo, and Roca (2002); Liceras, Bel, and Perales (2006); Buesa (2006); and others (although, see Salustri & Hyams, 2003)—have concluded that Southern Romance child languages do pass through an optional infinitive stage.

Because tense in Spanish is expressed in most verb conjugations by a morpheme that simultaneously expresses tense and subject–verb agreement, a nonfinite verb would likely have some non-adultlike form of this portmanteau morpheme, or possibly simply lack it. As a result, it can be difficult to know whether certain verb forms are expressing tense or whether they are untensed, without knowing what the intended subject is. Forms such as habla [speak] could be either third person singular present indicative verbs or nonfinite bare stem verbs. If one assumes that such forms are all present tense, then children look very adultlike. However, there is evidence, even in the spontaneous data, that suggests that this is not so. For example, children occasionally produce them with overt non–third person singular subjects, as in the following, from Grinstead, De la Mora, Pratt, and Flores (2009):

5. Eduardo—3;0.28 (years;months;days)
   Yo quiere hacerlo.
   “I wants to do it.”

6. Carlos—3;3.28
   Yo va a buscar.
   “I goes to look for.”

7. Graciela—2;6.5
   Hace esto yo.
   “I does this.” (p. 242, data are from Grinstead, 1998)

In these examples, we see non–third person singular pronouns being used with apparently third person singular verbs, which suggests that they are nonagreeing forms that are possibly also nontensed, given that the same morpheme expresses both tense and agreement. This problem is exacerbated by the fact that the overwhelming majority of verbal utterances, even in adult Spanish, lack an overt subject (see Otheguy & Zentella, 2012, for a review). The cumulative effect of many ambiguous third person singular present indicative/bare stem forms being used with null subjects, coupled with the fact that researchers often give children the benefit of the doubt, is an overestimation of child Spanish speakers’ competence with tense marking.

The overestimation, based on this misunderstanding, persists, in spite of the fact that multiple elicited production studies have shown that child Spanish speakers in Spanish-dominant contexts (Kernan & Blount, 1966; Pérez-Pereira, 1989) are delayed in producing tense and agreement marking on both real and nonce words until roughly 4;0 to 5;0. Although their studies are interesting and informative, neither Pérez-Pereira (1989) nor Kernan and Blount (1966) said much about the types of errors children made, and they tested third person singular past and present only. In contrast, Bedore and Leonard (2001, 2002), as part of a study of specific language impairment (SLI) in Spanish, tested TD U.S. Spanish-speaking children (fifteen 3-year-olds and fifteen 5-year-olds) for their control groups, on first and third person, singular and plural, past and present tense verbs. Their results, consistent with our hypothesis, were that third person singular present indicative verb forms, plausibly bare stems, were the most common error produced by their two control groups of TD Spanish-speaking children, when attempting verbs of other person, number, and tense combinations.

Another way to determine whether Spanish-speaking children view combinations of third person singular present indicative/bare stem verbs with non–third person singular subjects (e.g., “Yo duermo” [I sleeps]) as grammatical is to use the grammaticality choice task (Pratt & Grinstead 2007), which presents them with these potentially nonfinite forms, together with adultlike forms (e.g., “Yo duermo” [I sleep]) and asks them to indicate which they regard as grammatical. If children choose the third person singular/bare stem forms with non–third person singular subjects, then it suggests that their grammars allow nonagreeing forms of the type occasionally found in spontaneous production data (see Examples 1–3 above). If such results are found, then it would seem safe to infer that at least some of the apparently third person singular present indicative forms that occur with null subjects in spontaneous production data are, in fact, nonfinite forms.

Pratt and Grinstead (2007) compared the grammaticality choice task with the more traditional grammaticality
judgment task (e.g., McDaniel & Cairns, 1990) to measure tense marking in Spanish-speaking children and concluded that the grammaticality choice task involves fewer task demands, because children of the same age tested on the same construct scored higher and failed fewer filler items. In more recent work, Vega-Mendoza (2010) and Grinstead, Vega-Mendoza, and Goodall (2010) tested the acceptance, by 44 monolingual 3- to 6-year-old children Spanish speakers in Mexico City ($M_{age} = 4;9$), of two potential nonfinite forms: (a) the bare stem (e.g., canta [sing]) and (b) the morphological infinitive (e.g., cantar [to sing]). The results showed that an array of percentages correct, between 63% and 100%, were correlated with age ($r = .679, p < .001$). This means that when children are asked to choose between finite and nonfinite verbs, occurring with overt subjects, they will choose the nonfinite form up to 37% of the time, as a function of age. Furthermore, they do not begin to make these judgments with 100% accuracy until roughly 5;0, making the optional infinitive phenomenon in Spanish appear to have an endpoint chronologically similar to the one established for child English by Rice and Wexler (1996). In sum, early research into tense marking in child Spanish concluded that children were producing adultlike tensed verb forms because tense and agreement depend on one another for their expression in Spanish and because most verbs in early child Spanish occur without an explicit subject. Also, although earlier work had failed to find nonfinite verbs in child Spanish, we have seen that it is possible to show that child Spanish speakers both produce and judge as acceptable nonagreeing, plausibly nonfinite bare stem as well as morphological infinitive verbs until roughly 4;6 to 5;0, as in other child languages.

**Tense in Spanish SLI**

**Tense in Spanish SLI in language contact contexts.** The two most influential studies of verb inflection in child Spanish speakers with SLI in the U.S. context were conducted by Bedore and Leonard (2001, 2002, 2005). Bedore and Leonard (2001) research is an elicited production study of a wide range of morphosyntactic processes, including verbal inflection. The authors compared 15 children diagnosed with SLI (age range = 3;11–5;6, group mean length of utterance in words [MLUw] = 2.88) with a group of 15 age-matched control participants and 15 control participants matched on MLUw. The children with SLI showed significantly lower scores than age-matched control participants for third person singular and plural, for the present. For the past (preterit), the SLI children had significantly lower scores for third person plural than the age-matched control children. Most interesting for our proposal, according to the error analysis given in the erratum for Bedore and Leonard’s Table 5 (Bedore & Leonard, 2002), the great majority of all verbal inflection errors produced in the test by all children (including the SLI group, which produced the most) consisted of producing a third person singular present indicative verb form when a different person, number, and tense were called for. This is strong confirmation of our bare stem hypothesis because the attempted forms were overwhelmingly replaced by one form, which we believe to be the primary nonfinite form in child Spanish. Bedore and Leonard (2005) analyzed the spontaneous language transcripts of the same group of children and found that the children with SLI were significantly worse than age-matched control participants in their use of third plural present tense. As with other work on spontaneous production data in Spanish, there exists a strong possibility that the data collected in these sessions included a large number of bare stem verbs that were not identified as non-adultlike, though there is no direct evidence in the report to suggest this.

Bosch and Serra (1997) and Sanz-Torrent, Serrat i Sellabona, Andreu, and Serra-Raventos (2008) also reported on SLI in Spanish, in a language contact situation, in this case with Catalan. It should be noted that the situation of Spanish in Catalunya, where these studies were carried out, is diglossic. This is distinct from the situation of Spanish in most of the United States, as in Bedore and Leonard’s (2001, 2001, 2005) studies, which is a more standard minority language situation. The average language learner in Catalunya is exposed to both languages, though unevenly. In most of the United States, in contrast, the average language learner is not exposed to both, and English is overwhelmingly dominant. Although one might expect language contact phenomena in both situations, language attrition would be much more likely in the U.S. context than in Catalunya. We note that although one might expect language contact phenomena to cause the data to be very different from monolingual Spanish data, we show that, to the contrary, they preponderantly show continuity with the monolingual studies. Bosch and Serra studied 24 children, 12 of whom had language impairment ($M_{age} = 7;6$) and 12 of whom were TD language controls. The children in the study gave language production samples that were then coded for a range of morphological features, including verbal inflection. Bosch and Serra observed that there were significant differences between the two groups in terms of number marking on third person plural forms, which the authors said could potentially be attributed, to phonological processes. We note that a result of removing the plural marker –n from third person plural present indicative forms leaves an apparently third person singular present form (e.g., hablan becomes habla). As in other cases, it is possible that some percentage of these apparently third person singular present indicative forms are, in fact, nonfinite bare stems, consistent with our proposal.
Sanz-Torrent et al. (2008) studied 18 bilingual Catalan–Spanish-speaking children, six with SLI and six in each of two control groups: (a) one age-matched control group and (b) one MLU<sub>w</sub>-matched control group. They examined the children’s verbal utterances collected through spontaneous speech samples at two time points: Time 1 data were collected when the mean age of the SLI and age control children were 3;9, and Time 2 data were collected when the mean age of the SLI and age control children was 4;9. The investigators concluded, on the basis of few detectable tense errors, that the children’s data did not support Rice and Wexler’s (1996) extended optional infinitive analysis. As with much other work on child Spanish and Catalan, however, there is another way to look at their data.

In Sanz-Torrent et al.’s (2008) study, the majority of the children’s utterances at Time 1 and Time 2 were morphological infinitives or third person singular present indicative forms. The SLI group produced statistically more third person singular present indicative tense forms than did the MLU<sub>w</sub> control group at both times, although the authors suggested that this difference may be related to the structure of the spontaneous data collection sessions of the MLU<sub>w</sub> group, which differed from the other two. Relevant to our proposal is the possibility that some percentage of these third person singular present indicative forms in fact consists of nonfinite bare stems, although exactly what that percentage is might be difficult to discern from spontaneous production data.

In sum, the data reported by Sanz-Torrent et al. (2008) are consistent with the observations made thus far, in the sense that large numbers of third person singular present indicative forms, some of which are likely nonfinite bare stems, were used by the children with SLI, distinguishing them from their TD age- and language-matched controls. Again, the spontaneous production methodology stands in the way of determining the degree to which these forms represent optional infinitive verb forms for the children.

In summary, there appears to be a pattern in the Spanish-in-contact studies that shows both TD children and children with SLI overproducing third person singular present indicative tense/bare stem forms. More important, the children with SLI are distinguishable from TD control children, on the basis of their use of bare stems, as the extended optional infinitive hypothesis would predict.

Spanish-predominant contexts. In an attempt to circumvent the limitations of spontaneous production data, Grinstead, De la Mora, Pratt, and Flores (2009) applied the grammaticality choice task for tense to a sample of 27 children from Mexico City, Mexico. Nine children formed the SLI group with an MLU<sub>w</sub> of 3.0 (range = 2.1–3.9, SD = 0.72) and a mean age of 5;6 (range = 4;10–6;7, SD = 7.4 months). Nine TD children formed the MLU<sub>w</sub> control group, and nine formed the age-matched control group. The results showed that children in the SLI group were less proficient than either the MLU control group (p < .001) or the age-matched control group (p < .001). This was true of overall proficiency, and it was also true of each individual nonfinite verb type tested (bare stems and morphological infinitives).

Working with a larger sample, and using an elicited production task, Grinstead, De la Mora, Vega-Mendoza, and Flores (2009) tested 42 monolingual child Spanish speakers in Mexico City, 21 of whom were diagnosed with SLI (age range = 57–78 months, M<sub>age</sub> = 68 months) and 21 of whom were TD age-matched controls. The results revealed that the children with SLI were significantly worse at producing finite verb forms than were the TD age-matched controls. Furthermore, of the 39 errors produced by the TD children, 23% of them (nine) were bare stem forms produced with a non–third person singular subject, confirming our hypothesis that the bare stem is a nonagreeing, plausibly untensed form that Spanish-speaking children produce. Also noteworthy is the fact that a subset of the children (n = 22) in the study who took the elicited production test also completed the grammaticality choice task for tense, described above, and their scores significantly correlated with their elicited production scores (r = .652, p = .001). The fact that the results of these two methodologically distinct tests of tense marking produced results that correlated is consistent with our contention that the grammaticality choice task is indeed measuring finiteness, as we have argued. Thus, the two available studies of tense marking in monolingual Spanish-speaking children with SLI showed that both receptive and expressive measures of tense marking are sufficient to distinguish children with SLI from TD children and support the claim that bare stem forms constitute a plausible nonfinite form in the developing Spanish of children with SLI and of those developing typically.

Interface Delay and SLI

In the present study, we were concerned with the utility of tense in Spanish as a clinical marker of SLI. A range of explanations for the observation that tense is difficult for children with SLI have been proposed, and although we do not attempt here to empirically distinguish our account from others, we try to state our assumptions clearly. We assume that multiple cognitive domains are involved in using tense (including syntax and discourse pragmatics) and that a principal phenomenon of cognitive development is an initial limitation on the ability of multiple cognitive domains to interface with one another. We term this phenomenon interface delay (Grinstead 1998, 2004, 2010). In earlier work, seeking to explain the early absence of overt subjects in child Catalan and Spanish (Grinstead 2004, Grinstead & Spinner 2009), it was proposed that children fail to use overt subjects, which primarily serve the discourse-pragmatic function
of signaling a change in reference, as a result of assuming that their interlocutors share their perspective as to which referents are salient in what Stalnaker (1978) referred to as the conversational common ground. In the absence of the ability of syntax to use discourse-pragmatic knowledge, children assumed that their interlocutors knew what they were talking about, which would make the attested consistent use of a null subject felicitous. Stalnaker proposed that children had deficiencies neither in syntactic knowledge nor in discourse-pragmatic knowledge but rather that the interface between these domains had not developed to the point at which they could interface with one another in an adult-like fashion.

Other attempts to explain the tense deficit in TD children passing through the optional infinitive stage include Rizzi's (1994) truncation hypothesis, Hyams's (2007) null modal hypothesis, and Waxler's (1998) unique checking constraint (UCC). Attempts to explain the more prolonged and more severe version of this stage, dubbed the extended optional infinitive stage by Rice and Waxler (1996), include Jakubowicz and Nash's (2001) computational complexity hypothesis, which posits that simpler derivations are preserved in the grammars of children with SLI and that present tense forms are simpler than past, by a metric that revives the derivational theory of complexity (e.g., Brown & Hanlon, 1970) in minimalist (Chomsky, 1995) terms. Similarly, Rice and Waxler sought to explain the difficulty that children with SLI have in marking tense by positing that they optionally cannot check more than one syntactic feature. Although both theories claim that children are limited in their capacity to compute syntactic derivations, Jakubowicz and Nash situated the limitation in the quality of the syntactic constituents that children can make use of (features that are always used, e.g., person marking in French, are less costly, whereas those that are less frequently used, e.g., past tense, are more costly), whereas Rice and Waxler situated the limitation in the quantity of features that children can use. The UCC states that Agr and T features are both checked in the adult derivation, but children may check only one of the two, resulting in nonfiniteness, except where a conflicting constraint, minimize violations, overrides the UCC and finite verbs are produced.

The presence of a tense deficit in children with SLI has been corroborated for children learning other, typologically diverse, languages, including French, Dutch, and Hebrew (e.g., Jakubowicz & Roulet, 2004; Leonard, Dromi, Adam, & zadunaisky-ehrich, 2000; Waxler, Schaeffer, & Bol, 2004). Although the representational accounts of SLI mentioned have similarities, they also make distinct predictions, the consequence of which cannot be considered here, for reasons of space. The purpose of the present report is not to attempt to distinguish among these accounts but rather to demonstrate that tense can also be useful in Spanish for the identification of children with SLI.

**Research Questions**

With these considerations in mind, we posed the following research questions:

1. Do spontaneous language measures from the Spanish-speaking children who took the elicited production and grammaticality choice task tests of tense in Grinstead, De la Mora, Vega-Mendoza, and Flores (2009) and Grinstead, De la Mora, Pratt, and Flores (2009) correlate with the results of these studies?

2. Can these same spontaneous language measures distinguish Spanish-speaking children with SLI from an age-matched sample of those without it?

3. Can some combination of these spontaneous language measures and the experimental results alluded to constitute useful discriminant functions capable of identifying Spanish-speaking children with SLI as distinct from children in a TD comparison group with a satisfactory degree of sensitivity and specificity?

**Method**

**Participants**

Fifty-five monolingual Spanish-speaking children participated in this study, each producing a spontaneous speech sample, interacting with adult native speakers of Mexican Spanish, trained as speech-language pathologists, neuropsychologists, or pediatric neuropsychologists. The children in the study came from a day care center/preschool and a speech and hearing clinic that serve a broad socioeconomic spectrum of children in Mexico City. Twenty-six of the children were diagnosed with SLI (age range = 58–82 months, \( M_{\text{age}} = 68 \text{ months}, SD = 6.7 \text{ months} \)), and 29 of them were TD (age range = 37–79 months, \( M_{\text{age}} = 63 \text{ months}, SD = 9.3 \text{ months} \)).

The children with SLI were diagnosed using conventional exclusive and inclusive criteria. Nonverbal IQ was determined using a Spanish translation of the Wechsler Preschool and Primary Scale of Intelligence—Revised (Wechsler, 1989). To be included in the sample, children had to have scores above 85. The Batería de Evaluación de la Lengua Española (BELE; Rangel, Romero, & Gómez 1988) was the standardized language test used. The BELE was locally developed and normed (in Mexico City) and contains seven subtests. We used four of them in our identification process. The criterion used was that children have a score of at most 6 (−1.25 SD below the mean; see Leonard, 1997) on at least one comprehension test (Comprensión Gramatical [Grammatical Comprehension])

\[ 1^\text{a} \text{Agr} \text{ and } T \text{ stand for “element” and “tense,” respectively.} \]
or Adivinanzas [Riddles]) and at least one production test (Producción Dirigida [Elicited Production] or Definiciones [Definitions]) to be included in the SLI sample. Although there are no published external validity studies of the BELE, the four BELE subscores we used for identification correlated with spontaneous speech measures in a sample of 29 children from our study (14 TD, 15 with SLI), including the subordination index (SUB-I; \( r = .570, p = .001 \)), mean length of utterance in morphemes (MLU\(_m\); \( r = .717, p < .001 \)), and number of different words (NDW; \( r = .702, p < .001 \)). See the Measures section for more on MLU\(_m\) and so on.

Because it has been shown to have good sensitivity and specificity in identifying Spanish-speaking children with SLI, we also gave children Restrepo’s (1998) family questionnaire. Finally, we gave children a phonological screen in which they were asked to repeat 24 nonce words that included the segments used in Spanish to represent tense in word-final position, with appropriate stress. To be included in the study, children had to produce at least four out of five correctly from each category. Children were given thorough hearing tests and had to pass them at conventional levels. Also, parental report and medical history had to suggest no recent episodes of otitis media with effusion in order for a child to be included. In a similar way, neurological tests determined that the children had no frank neurological damage. With respect to oral structure and oral motor function, an initial examination ruled out structural anomalies and ensured normal function. Parental report and family history interviews ruled out concerns pertaining to social and physical interactions.

To eliminate children with language impairment, we also gave the 29 children who formed the control group the standardized language test, and their results were within 1 SD of the mean for their ages. They were also given the phonological screen and must have passed, to eliminate possible skewing of the results for phonological reasons. Our interactions with them—as well as data we received from their parents’ and teachers’ reports—suggested no abnormalities in their speech or language.

**Measures**

Seven distinct indices of grammatical development (IGDs) were calculated for each transcript. These included MLU\(_w\), MLU\(_m\), and mean length of utterance of the verb phrase (MLU\(_{vp}\)), calculated in words over only the verb phrase. These first three IGDs stem from the work of Brown (1973), with the last measure being our own innovation for studying languages with null subjects, because a count of the VP only should control for the absence of subjects. We also calculated mean length of terminable unit (MLTU), which is commonly used in speech and hearing sciences and differs from MLU in only counting clausal utterances (Hunt, 1965), that is, excluding fragments. Errors per terminable unit (ETU) has shown some promise in earlier work (e.g., Restrepo, 1998) on child Spanish SLI and consists of calculating the mean number of errors in a child’s clausal utterances. NDW is a measure of lexical diversity that represents how often each word was used and how many different words there were in each transcript (Miller, 1991; Watkins, Kelly, Harbers, & Hollis, 1995). The SUB-I is calculated by making a ratio of the total number of clauses to the total number of T units (Loban, 1963).

The calculation of NDW and MLU\(_w\) was done automatically by the Computerized Language Analysis programs from the Child Language Data Exchange System (MacWhinney, 2000). All other IGD calculations were done by two native Spanish speakers of the Spanish of Mexico City. Each coder coded all utterances, and the average of their rating was taken.

The indices we chose can be thought of as constituting a spectrum ranging from the most purely syntactic measure, the SUB-I, which is a measure of clausal embedding, to the most purely lexical measure, NDW, which measures lexical diversity. In between, the MLU measures and MLTU are a kind of hybrid measure of the lexicon and grammar. Although tense is clearly a grammatical phenomenon situated at the syntax–semantics interface, it equally clearly has a lexical component to it, especially with the less compositional, suppletive, and irregular verb forms that are unlikely to be the product of syntactic-free combination. In this way, we could expect the development of the lexical-syntactic spectrum of IGDs used in this study to correlate with the development of different lexical and syntactic aspects of tense, measured by our elicited production and judgment experiments.

**Reliability**

Spontaneous speech samples were transcribed by native speakers of the Spanish of Mexico City, the same dialect spoken by the children in the sample. Transcribers were initially normed on a series of common transcriptions. Then, each recording session was transcribed by a single transcriber and checked by a second transcriber. Finally, half of all transcripts were randomly selected to have 10% of their utterances retranscribed by a second transcriber. Agreement between transcribers as to each word transcribed ranged from 90% to 99%, with a mean agreement percentage of 95.4%. A Krippendorf’s alpha interrater reliability coefficient (Hayes & Krippendorf, 2007) was then calculated for the interval data represented by each transcriber’s number of words per utterance, for each transcript. Alpha values ranged between .904 and .998, with a mean of .974.

Coding for MLU\(_w\), MLU\(_{vp}\), MLTU, ETU, and SUB-I was also carried out by native speakers of Mexican
Spanish. Each transcript was given a numerical value for the IGD in question (e.g., for Transcript A, Coder 1, MLU_m = 2.32; for Coder 2, MLU_m = 2.23). Agreement between coders ranged from 97.18% to 99.22%, with standard deviations between .891 and 4.280. Krippendorf’s alpha values ranged from .893 to .997.

**Procedure**

**Spontaneous production data.** A parent of each child signed U.S. and Mexican Institutional Review Board–approved informed consent documents in order to participate. Children’s spontaneous speech was recorded in sessions of 20 to 35 min, using digital video recorders with built-in microphones. The TD children (n = 29, age range = 37–79 months, M_age = 63 months, SD = 9.3) were recorded in day care centers, and the children with SLI (n = 26, age range = 58–82 months, M_age = 68 months, SD = 6.7) were recorded in a speech and hearing clinic. Children were informally asked about their friends, teachers, family, and favorite movies and television shows. The discourse typically consisted of brief questions by the investigators and narrative answers by the children. These data were then analyzed with respect to the seven IGDs.

**Correlations with IGDs.** Our first research question was whether our seven IGDs from the spontaneous speech samples of the Spanish-speaking children who took the elicited production and grammaticality choice task tests of tense, in Grinstead, De la Mora, Vega-Mendoza, and Flores (2009) and Grinstead, De la Mora, Pratt, and Flores (2009), correlated with the results of these studies. If they did, then the correlations would validate the experimental results of these two tests. There were 18 TD children (age range = 58–82 months, M_age = 69 months, SD = 6.6 months) who participated in this analysis and 18 children diagnosed with SLI (age range = 58–76 months, M_age = 65 months, SD = 5.4 months).

**Contrast between two age-matched groups.** Previous research has shown that the grammaticality choice task and the elicited production tense task can distinguish TD children from children with SLI. In this study, we asked whether our spontaneous speech measures can do the same. To do this, we compared 24 children diagnosed with SLI (age range = 58–76 months, M_age = 66 months, SD = 6.0 months) and 24 age-matched TD children (age range = 55–79 months, M_age = 66 months, SD = 6.0 months). If there are significant differences, they will validate the experimental results of Grinstead, De la Mora, Vega-Mendoza, and Flores (2009) and Grinstead, De la Mora, Pratt, and Flores (2009).  

**Discriminant function analysis.** Although Fisher’s linear discriminant function analysis is commonly used in behavioral science, and in the speech and hearing sciences in particular, it requires multivariate normally distributed data and group variance–covariance matrices to be equal across groups for effective classification. That is not uniformly the case with the data in this study. Consequently, we instead used the nonparametric K nearest neighbor analysis, which requires neither a Gaussian data distribution nor homoscedastic variance between samples.

The nine independent variables, or features, for the K nearest neighbor analysis were (a) NDW, (b) SUB-I, (c) ETU, (d) MLTU, (e) MLU_w, (f) MLU_vp, (g) MLU_m, (h) the elicited production test of tense, and (i) the grammaticality choice task of tense. These features were then used to classify the children in the study into two levels of the dependent variable, or target: (a) SLI or (b) TD. As part of the procedure, the number of optimal neighbors (K) to consider is automatically determined. The children considered for each of the discriminant functions varies according to the tests they took.

**Results**

**Correlations**

In Table 1, we give the r² and p values of the correlations between each IGD and the elicited production scores of the children who took the test reported in in Grinstead, De la Mora, Vega-Mendoza, and Flores (2009), along with the mean IGD value for the sample and its standard deviation. In this table, one can see that IGDs from across the lexical–grammatical spectrum significantly correlated with our elicited production test of tense in our sample, with medium to large effect sizes (Cohen, 1988). Again, the sample included TD children as well as children with SLI. The one measure that did not correlate with our test was ETU. Although this was initially

<table>
<thead>
<tr>
<th>IGD</th>
<th>r²</th>
<th>p</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU_w</td>
<td>.290**</td>
<td>.001</td>
<td>4.2197</td>
<td>1.6188</td>
</tr>
<tr>
<td>MLU_m</td>
<td>2.28**</td>
<td>.001</td>
<td>5.4342</td>
<td>2.3315</td>
</tr>
<tr>
<td>NDW</td>
<td>.298**</td>
<td>.001</td>
<td>163.7778</td>
<td>49.6457</td>
</tr>
<tr>
<td>MLU_vp</td>
<td>.274**</td>
<td>.001</td>
<td>3.9656</td>
<td>1.5709</td>
</tr>
<tr>
<td>MLTU</td>
<td>.253**</td>
<td>.002</td>
<td>4.7489</td>
<td>2.1056</td>
</tr>
<tr>
<td>SUB-I</td>
<td>.152*</td>
<td>.019</td>
<td>1.3186</td>
<td>0.2804</td>
</tr>
<tr>
<td>ETU</td>
<td>.035</td>
<td>.277</td>
<td>0.1386</td>
<td>0.1104</td>
</tr>
</tbody>
</table>

Note. MLU_w = mean length of utterance in words; MLU_m = mean length of utterance in morphemes; NDW = number of different words; MLU_vp = mean length of utterance in words measured on the verb phrase; MLTU = mean length of terminable unit; SUB-I = Subordination Index; ETU = errors per terminable unit.

*p < .05 (two-tailed). **p < .01 (two-tailed).
somewhat mysterious, because it had been shown to be a potentially useful identifier of Spanish-speaking children with SLI in previous research (Restrepo, 1998), it appears that our relatively unstructured spontaneous speech sample was collected differently than the more structured spontaneous production data sample in Restrepo (1998), which may account for the differences.

As with the elicited production task, one can see in Table 2 that, across the lexical–grammatical spectrum, our spontaneous speech measures correlated significantly with grammaticality choice task scores, with large effect sizes, with the exception of ETU, possibly for the reasons alluded to above.

**Comparisons**

Previous research has shown that the grammaticality choice task and our elicited production tense task can distinguish TD children from children with SLI. In this analysis, we asked whether our spontaneous speech measures can do the same. For the five measures listed in Table 3, the distributions of the (continuous) variable values were normal but had heterogeneous variances (with the exception of MLTU, which had homogeneous variances); consequently, their $t$ and $p$ values reflect independent-samples $t$ tests that have been adjusted for unequal variances (the equal and unequal variance versions are identical for MLTU). In all cases, the mean scores of the children with SLI are significantly lower than those of the children in the age-matched control group.

For the two spontaneous speech measures represented in Table 4, the SUB-I and ETU, the distribution of the data was not normal. Consequently, the nonparametric Mann–Whitney $U$ test was calculated. Table 4 shows support for the hypothesis that the SLI children would score lower than the age-matched control children ($z = -4.169, p < .001$) for their performance on the SUB-I. Children with SLI had an average rank of 16.08, whereas children in the age-matched control group had an average rank of 32.92. In contrast, the hypothesis that SLI children would produce more ETUs than the age-matched control children was not supported ($z = -0.072, p = .942$). Children with SLI had an average rank of 24.35, and children in the age-matched control group had an average rank of 24.65.

**Discriminant Function Analysis**

What would eventually be most clinically useful, of course, would be if experimental measures, spontaneous measures, or some combination of both could be used to distinguish Spanish-speaking children with SLI from those without. According to Plante and Vance (1994), sensitivity, which in our case consists of the accurate identification of affected children as affected, and sensitivity, the accurate identification of TD children as TD, is considered “poor” between 70% and 79%, “fair” between 80% and 89%, and “good” between 90% and 100%. In this part of the study, we attempted to classify the (already-diagnosed) children in our study as having SLI or TD on the basis of their scores on our two experimental measures (elicited production and grammaticality choice

---

**Table 2.** Mean scores and standard deviations of IGDs and the $r^2$ and $p$ values of their correlations with grammaticality choice task scores.

<table>
<thead>
<tr>
<th>IGD</th>
<th>$r^2$</th>
<th>$p$</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDW</td>
<td>.294**</td>
<td>&lt; .001</td>
<td>160.430</td>
<td>54.0020</td>
</tr>
<tr>
<td>SUB-I</td>
<td>.258**</td>
<td>&lt; .001</td>
<td>1.3050</td>
<td>0.3021</td>
</tr>
<tr>
<td>MLTU</td>
<td>.254**</td>
<td>&lt; .001</td>
<td>5.3978</td>
<td>2.5389</td>
</tr>
<tr>
<td>MLU_m</td>
<td>.249**</td>
<td>&lt; .001</td>
<td>4.1422</td>
<td>1.8387</td>
</tr>
<tr>
<td>MLU_p</td>
<td>.249**</td>
<td>&lt; .001</td>
<td>3.9142</td>
<td>1.7307</td>
</tr>
<tr>
<td>MLTU</td>
<td>.241**</td>
<td>&lt; .001</td>
<td>4.5749</td>
<td>1.2058</td>
</tr>
<tr>
<td>ETU</td>
<td>.026</td>
<td>.280</td>
<td>0.1491</td>
<td>0.1101</td>
</tr>
</tbody>
</table>

**$p < .01$ (two-tailed).**

**Table 3.** The means, standard deviations, $t$ values, $p$ values, and partial $\eta^2$ values for comparisons of the specific language impairment (SLI) and age-matched control groups, using an independent-samples $t$ test with unequal variances assumed, equal variance assumed for MLTU.

<table>
<thead>
<tr>
<th>IGD</th>
<th>Group</th>
<th>$M$</th>
<th>$SD$</th>
<th>$t$</th>
<th>$p$</th>
<th>Partial $\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDW</td>
<td>Age controls</td>
<td>192.1667</td>
<td>55.7110</td>
<td>5.046</td>
<td>&lt; .001</td>
<td>.356**</td>
</tr>
<tr>
<td>MLU_m</td>
<td>Age controls</td>
<td>5.4014</td>
<td>1.7001</td>
<td>6.650</td>
<td>&lt; .001</td>
<td>.490**</td>
</tr>
<tr>
<td>MLU_p</td>
<td>Age controls</td>
<td>4.8329</td>
<td>1.7071</td>
<td>6.456</td>
<td>&lt; .001</td>
<td>.406**</td>
</tr>
<tr>
<td>MLU_m</td>
<td>Age controls</td>
<td>6.9121</td>
<td>2.5042</td>
<td>5.971</td>
<td>&lt; .001</td>
<td>.437**</td>
</tr>
<tr>
<td>MLU_p</td>
<td>Age controls</td>
<td>3.6247</td>
<td>1.002</td>
<td>5.7971</td>
<td>&lt; .001</td>
<td>.427**</td>
</tr>
<tr>
<td>MLTU</td>
<td>Age controls</td>
<td>5.3921</td>
<td>0.9972</td>
<td>5.853</td>
<td>&lt; .001</td>
<td>.427**</td>
</tr>
<tr>
<td>SLI</td>
<td></td>
<td>3.7938</td>
<td>0.8919</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**$p < .01$ (two-tailed).**

**Table 4.** The means, standard deviations, ranks, $z$ values and $p$ values for comparisons of SLI and age-matched control groups, using a nonparametric Mann–Whitney $U$ test.

<table>
<thead>
<tr>
<th>IGD</th>
<th>Group</th>
<th>$M$</th>
<th>$SD$</th>
<th>Rank</th>
<th>$z$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUB-I</td>
<td>Age controls</td>
<td>1.4588</td>
<td>0.3194</td>
<td>32.92</td>
<td>-4.169**</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>SLI</td>
<td></td>
<td>1.1354</td>
<td>0.1439</td>
<td>16.08</td>
<td></td>
<td>.942</td>
</tr>
<tr>
<td>ETU</td>
<td>Age controls</td>
<td>0.1513</td>
<td>0.1152</td>
<td>24.65</td>
<td>-0.072</td>
<td>.942</td>
</tr>
<tr>
<td>SLI</td>
<td></td>
<td>0.1450</td>
<td>0.1080</td>
<td>24.35</td>
<td></td>
<td>.942</td>
</tr>
</tbody>
</table>

**$p < .01$.**
task), our seven spontaneous production measures, or combinations thereof.

In Table 5, in descending order by mean correct classification percentage, we give the number (n) of children classified by each independent variable, or combination of independent variables, and the number of neighbors considered (K), along with the number and percentage of children with SLI correctly identified (sensitivity) and the number and percentage of TD children correctly identified (specificity). Finally, we give the mean correct classification percentage.

Also in Table 5, we give the variables and variable combinations that reached at least the 80%, or “fair” level, of either sensitivity or specificity. The most accurate and most balanced of these variables appears to have been the elicited production test, which produced nearly 90% mean correct classification, which would be considered good accuracy for both sensitivity and specificity. Separately, combined functions, such as MLU_m and the grammaticality choice task, achieved 100% correct classification of children with SLI (sensitivity), and the combined function of the elicited production task and MLTU achieved 100% specificity.

### Discussion

We have seen that both our productive and receptive measures of Spanish-speaking children’s tense marking correlate highly with a diverse array of spontaneous speech measures. To the degree that these findings validate the effectiveness of the experimental measures to tell us something about child Spanish speakers’ knowledge of tense, they confirm that the optional infinitive phenomenon exists in child Spanish and can be measured in a variety of ways.

The results of our comparisons of the mean scores of the children in the SLI and TD age-matched control groups show significant differences between the two groups for all measures, except ETU. In the same way that our grammaticality choice task and elicited production studies of tense knowledge in this population have shown that tense can be used to distinguish children with SLI from those without, six of seven of our spontaneous measures (the six that correlated with grammaticality choice task and elicited production scores) are similarly capable of distinguishing children with SLI from those without. These results confirm the distinctions made by the experimental studies in the sense that the distinctions that they made on the basis of tense have been replicated with measures that do not tap the same underlying construct in the same way.

Finally, we have argued that a version of the optional infinitive stage exists in child Spanish, on the basis of the existing literature, and that an extended optional infinitive stage exists for Spanish-speaking children with SLI, on the basis of the literature and our experiments. This argument appears to be confirmed by the fact that our experimental measures of tense, the elicited production and grammaticality choice tasks, can play an important, and sometimes independent, role—in the case of elicited production—in identifying Spanish-speaking children with SLI versus those without.

In this project, we have attempted to demonstrate that both TD monolingual Spanish-speaking children and their counterparts with SLI pass through an optional infinitive stage, with the stage of the children with SLI being more severe and prolonged. We have argued that its existence has been missed because one of the primary nonfinite forms—the bare stem—in child Spanish is difficult to distinguish from an adultlike finite form, third person singular present indicative, especially in spontaneous production data. We have presented evidence that bare stem nonfinite verbs nonetheless occur in spontaneous speech and are sometimes identifiable as such. More convincingly, perhaps, they are produced in elicited production experiments in which non–third person singular subject pronouns are included in the prompt, in our work as well as in the work of others. Furthermore, in a grammaticality choice task, when such pronouns are presented with bare stem verbs, children accept them as grammatical. Finally, Spanish-speaking children with SLI perform significantly worse than do TD children on both receptive and expressive tasks measuring verb finiteness, confirming the existence of an extended optional infinitive stage. This is true in varieties of Spanish that are in contact with either Catalon or English and in varieties that exist in Spanish-predominant, monolingual contexts.

These conclusions have a number of implications for developmental linguistic theories. For Wexler’s (1998)
hypothesis, including the UCC, it should be clear that Spanish, and likely the other Southern Romance languages, should be considered not an exception to the optional infinitive cross-linguistic generalization but rather as a case in which language-particular features require us to measure it differently.

From the perspective of interface delay, it is expected that verbs should be delayed in being expressed as finite, given their sensitivity to discourse pragmatics, because it is also expected that a range of other discourse-sensitive child language phenomena should be delayed in their development, such as the overuse of definite articles (Maratsos, 1974), pronouns (Avrutin, 1999), and null subjects in languages such as Spanish (Grinstead, 2004). Although we argue that the development of the interface among cognitive faculties plays an important role in the development of these constructions, it is certainly not the only relevant factor. Greater regularity in finite verb morphology—for example, the regularity of tense marking in Spanish, as proposed by Legate and Yang (2007)—could also facilitate an earlier end to the optional infinitive stage. Conversely, greater irregularity and increased processing demands from, for example, the discontinuous morpheme in the auxiliary be construction in English, could delay the end of the optional infinitive stage in that language.

In sum, there is a great deal of valuable information that can be gleaned from child Spanish for developmental linguistics, but one must be willing to see the data through more than one methodological lens. There is much more to say about the consequences of slowly developing and impaired tense in the grammatical systems of children speaking the world’s languages, which will wait for further research.

Acknowledgments

This study was supported by an Arts and Sciences Innovation grant from The Ohio State University (OSU) awarded to John Grinstead as well as by an undergraduate research fellowship awarded to Alisa Baron. Preliminary results of this study were presented at the 2011 Symposium on Research in Child Language Disorders, Madison, WI. We are indebted to Steven Naber of the Statistics Consulting Center in the Department of Statistics at OSU; Ann O’Connell of the Quantitative Research, Evaluation and Measurement section of the School of Education at OSU; and Michael Edwards of the Quantitative Program of the Psychology Department at OSU for their help with statistical considerations. We are also grateful for discussion of these issues with the Language Acquisition Discussion Group at OSU. We are especially grateful to Paij Lintz for her generous help with our reliability calculations; to Silvia Romero for her discussion of the Batéria de Evaluación de la Lengua Española; to Yolanda Cabré-Sans for discussion of bilingualism in Catalunya; and to Laida Restrepo for discussions of our errors per T unit results.

References


