A Type and Token Account of Past-Tense Morphology Acquisition in L1 Spanish Speakers

Mike Olsen
LING 2149
April 28, 2009

Abstract

Researchers attempting to understand the cognitive processes behind language storage and language use have extensively studied the acquisition of past-tense morphology. There are traditionally two main theoretical propositions regarding theses cognitive processes—single-mechanism and dual-mechanism theories. Single-mechanism theories posit an associative network model for dealing with past-tense morphology while dual-mechanism theories posit an associative lexicon that deals with irregular morphology and a rule-forming mechanism that deals with regular morphology.

Studies that have investigated these theories have contributed important findings regarding the acquisition of past-tense morphology. Many studies have made propositions on how type and token frequencies influence past-tense morphology acquisition. For the most part researchers agree that token frequencies in the input influence the acquisition of irregular verbs. Type frequencies in the input are important for children to begin using past-tense morphology productively.

Both theories also make predictions about the acquisition of past-tense morphology. Single-mechanism models maintain that irregular morphology is acquired quickly due to high token frequencies and that regular morphology will be more gradual because this depends on high type frequencies (which usually mean lower token frequencies). Dual-mechanism models assert that regular morphology is acquired quickly and applied to most contexts (evidenced by overregularization errors).

Investigating past-tense morphology in French-English bilinguals, Nicoladis, Palmer, and Marentette (2007) stated that type and token frequencies are both important for past-tense morphology acquisition. However, they interpreted their data based on an unfounded developmental sequence that is sensitive to type and token frequencies, which had not been empirically tested up to this point.

The current study analyzed a set of corpora of monolingual children acquiring Spanish (between the ages of 1;7 and 10;9) with the purpose of empirically investigating the first three stages of the developmental sequence proposed by Nicoladis, Palmer, and Marentette (2007). Although some evidence of expected phenomena in the second and third stages is apparent, results do not support the specific developmental sequence in question. Results do suggest, however, that type frequencies affect error types and token frequencies affect the accuracy rates in verbs used by children acquiring their first language. The results of this study also suggest that neither cognitive processing model is entirely adequate in explaining past-tense morphology acquisition. I briefly suggest an alternative view that combines single-mechanism and dual-mechanisms theories and more adequately explains the observed phenomena surrounding the acquisition of past-tense morphology in children acquiring Spanish.

Introduction

Many researchers investigating the cognitive processes involved in language use and acquisition have studied past-tense morphology (e.g., Bybee, 1995, Marcus et al, 1992; Pinker &Ullman, 2002; Plunkett & Marchman, 1991; Rumelhart & McClelland, 1986). The majority of these studies have surrounded models and theories of linguistic cognitive processes that fall into two different categories—single-mechanism theories and dual-mechanism theories.

Rumelhart & McClelland (1986) proposed that the acquisition of past-tense morphology can be explained by positing a system they call a *pattern associator* consisting of input and output units that are connected to each other and that can be modified depending on the input. As children hear novel words, the *pattern associator* connects them with similar forms. These forms become productive and apply to novel forms when the new words appear to imitate words already stored in the *pattern associator*. In a similar fashion, Bybee (1995) posits an associative network that links phonological input together. A high frequency in the input creates a high lexical strength and makes words resistant to change while a lower input frequency lends itself to regularization.

These single-mechanism theories predict that learning will be gradual and that high token frequency verbs in the input will be acquired quickly while lower token frequency verbs take longer to emerge in speech production. They also predict that type frequency influences the productivity of morphological forms. Forms with higher type frequencies in the input are likely to apply to novel entities. These theories also predict that errors in child speech will consist of OVERRGUARIZATION, such as *sleeped* and *goed* instead of *slept* and *went*, and IRREGULARIZATION, such as *boke* for *baked* (analogizing from *woke/wake*).

In opposition to single-mechanism theories, Marcus et al. (1992) propose that past-tense morphology acquisition can be explained by the application of a symbol-based rule for regular forms and simply memorizing irregular forms as any other lexical item. Roots and affixes are stored separately. When speakers form an utterance, the appropriate root and affix are activated and an inflectional mechanism creates the correct form for regular verbs. When speakers form utterances involving irregular verbs, the appropriate form is accessed directly from the lexicon. Later versions of dual-mechanism theories posit an associative lexicon (Pinker, 1998) much like the associative network of Bybee (1995) which is still separate from the mechanism that deals with regular verbs. This accounts for the graded irregularity evident in many languages (i.e., the fact that many irregular verbs pattern similarly to each other) that earlier versions could not explain.

Dual-mechanism theories predict that children will rapidly acquire and apply regular verbal morphology because rule application is associated with universal grammar principles. These theories also predict that children will produce many overregularization errors and relatively fewer irregularization errors. Token frequency is important for the acquisition of irregular forms and children are able to acquire regular forms because they have a high type frequency.

Although there is an ongoing debate on which set of theories best describes the cognitive processes, important information regarding the acquisition of past-tense morphology has been gained by research that has used past-tense morphology data to strengthen one theoretical position or another.

Clahsen, Aveledo, and Roca (2002) investigated the possibility of applying the dual-mechanism theory of morphology processing to Spanish. They explored corpora of children in narrative and spontaneous situations and claimed that there is a disassociation in the processing of regular and irregular verbs. Based on evidence in the corpora that they used, they suggest that the sudden onset of overregularizations in the children's speech shows that frequency is not what triggers these overregularizations. They posit a syntactic trigger to explain the existence of overregularizations. This however, would ignore any type of phonological patterning evident in many irregular verbs that the single-mechanism theories would account for.

The revision of the dual-mechanism model proposed by Pinker (1998) suggests that morphology acquisition is not blind to phonological patterns. An associative lexicon can account for irregular forms that seem to follow the same rule as each other while maintaining lexical strength and not succumbing to the general past-tense formation rule. In fact, Pinker's position relies on the input frequency of suppletive and like-patterned morphological forms. Overregularizations occurs because children have not heard the irregular form enough (i.e., the input frequency is not high enough) to reliably produce it and the rule applies to fill the gap in irregular past-tense morphology. He suggests that as children hear irregular forms enough, they are able to produce these forms reliably and the need for morphology gap filling diminishes.

Other researchers have also discussed the phonological aspects of past-tense morphology acquisition. Marshall and Van der Lely (2006) studied how normally developing children, children with grammatical-specific language impairment, and children and adults with Williams Syndrome use phonotactics to recognize past tense. They discussed the possibility of children recognizing non-permitted consonant clusters in monomorphemic contexts that surface as a result of morphological combinations (such as the [m] [d] sequence in *jammed)*. Because children only see these clusters in past-tense markings, they are able to use this cue to formulate past-tense morphology. Marshall and Van der Lely, following the single-mechanism model, also mention that children are able to distinguish between regular and irregular forms by assessing the phonological overlap manifest in words. There is clearly more overlap in many regular verbs than irregular verbs (e.g., bake-baked [beik-beikt], sleep-slept [slip, slɛpt] where the underlined segments show overlap).

Although phonological information does influence the acquisition and processing of past-tense morphology; Kim, Pinker, Prince, and Prasada (1991) oppose an entirely phonological explanation to past-tense morphology acquisition account. They show that nouns derived from irregular verbs are not stored having regular or irregular past-tense features. When speakers derive these nouns and use them as verbs, they take on regular verbal morphology whereas their irregular verb counterparts maintain irregular verbal morphology. An example that Kim et al. accredit to Paul Kiparsky is *He flied out to center field* (pg. 179). The past tense of the verb *fly* is *flew*, however, in the context mentioned above, *flew* would be ungrammatical. Although phonologically the same, these examples of different syntactic categories behave differently in their past tense formation and show that more than phonological analogizing must be at play in the acquisition of past-tense morphology.

Berent, Pinker, and Shimron (2002) provide another example of how past-tense morphology acquisition cannot rely wholly on phonological patterning. They show that phonological forms of regular and irregular verbs in Hebrew overlap but regular verbs still behave productively as in English. Because there is a distinction between regular and irregular verbs regardless of phonological similarities, something other than phonology must be posited to explain the cognitive processes involving inflections, including past-tense acquisition.

Aside from the influence of phonological patterns, other factors that contribute to the acquisition of past-tense morphology are morpheme recognition and specific developmental patterns that are sensitive to type and token frequency. Aronoff, Giralt, and Mintz (2006) propose a model for morpheme discovery that relies on input frequency. They claim that a morphology recognition system must have four characteristics. They state that such a system must be able to use a plausible learning mechanism, have flexibility in the position in a word where morphemes occur, be able to generate a robust initial list of morphemes on which they can build, and be able to rely heavily on grammatical morphemes because they occur with many other morphemes.

Aronoff, Giralt, and Mintz (2006) tested their model on data from corpora of children acquiring Spanish. They found that when making the model sensitive to the frequency of phonetic strings alone, it was accurate in recognizing morpheme types that paralleled the distribution in the natural language. When they made the model sensitive to syntactic frames as well, it performed equally well with the type distribution, but greatly improved its performance in mimicking the token frequency found in the natural language. This shows that children may use both phonetic and syntactic cues when learning to distinguish phonological strings in their language as morphemes.

Recognizing morphemes is only the first step in past-tense morphology acquisition. After recognizing distinct morphemes, children also have the task of assigning particular meanings to particular forms and learning how these forms that represent meanings distribute in the language (i.e., in what context certain forms are more likely to be found and where there are restrictions on certain forms). Because this is a developmental process, children make mistakes as they are testing phonological generalizations (as a single-mechanism would predict) or filling gaps with default forms (as a dual-mechanism predict).

Bybee (1995) discussed how an associative network would explain developmental patterns in past-tense morphology. This framework predicts that children will learn past-tense verbs that have a high token frequency in the input faster and produce those verbs with a higher accuracy rate than verbs that have a lower token frequency in the input. This is because they are able to form stronger phonological and morphological ties among phonological strings with words that show up more often. This may explain how children gain mastery of supletive forms but not necessarily how they use morphology productively (applying to novel contexts). High type frequency is what determines productivity in child past-tense morphology acquisition. If a large number of different words employ the same past-tense morpheme, children are likely to apply this morpheme to new words that they want to use to refer to the past.

According to this model, overregularization errors in child speech production support the idea that high type frequency is what determines productivity. As children associate the morpheme with the highest type frequency in the past tense as a past-tense marker, they overapply it to novel contexts that require other morphological forms. After gaining more input, children learn to disassociate this morpheme with irregular forms and associate other phonological strings (with lower type frequency) to novel contexts patterning in a similar way. Because children learn phonological patterns and begin associating them with other similar patterns, Bybee's (1995) model would also predict irregularization errors.

Nicoladis, Palmer, and Marentette (2007) also discuss the importance of both type and token frequencies in the acquisition of past-tense morphology. They argue that children need to be exposed to a certain level of type and token frequency to be able to acquire past-tense morphology accurately. They compared utterances of English-French simultaneous bilinguals with monolinguals of each language to see how type and token frequency affect accurate past-tense acquisition. Bilinguals in this study were less accurate (comparatively more errors) in producing past-tense morphology in type and token frequencies than their monolingual counterparts. Nicoladis, Plamer, and Marentette attribute lower accuracy rates to lower type and token frequencies in the input compared to the input to which monolingual children are exposed. They also predicted that bilinguals would overregularize in French more than in English because type and token frequencies for regular verbs in the French input were significantly higher than irregular frequencies whereas the English input showed a high type frequency in regular verbs, but a low token frequency. This prediction was not upheld by their study and it led them to conclude that type frequency was not the only factor in regular verb acquisition. They mention that this provides evidence against the dual-mechanism proposal that children overregularize when they lack the irregular form and are filling the gap with the default form.

One weakness in this study that Nicoladis, Palmer, and Marentette (2007) mention is that they interpret their results assuming a developmental sequence that is sensitive to type and token frequencies in the input. This sequence consists of four stages that depend on type and token frequencies in the input. In the first stage, children do not use past-tense morphology to refer to past events; they use the stem of the verb and occasionally an inflected verb that they have learned as part of a chunk ("I ate all of it!"). In the second stage, children start to recognize past-tense morphemes and link them to present verb forms. At this stage, accuracy would be sensitive to token frequency. Children would use and be more accurate with high token frequency verbs more than with low token frequency verbs. In the third stage of past-tense development, children start to become sensitive to phonological patterns that different verb families exhibit. In this stage, overregularization and irregularization errors should be evident as children test different potential past-tense morphemes. In the last developmental stage, children start to generalize across phonological families of verbs. In this stage, children are receptive to high type frequencies and use the regular morphemes in novel contexts.

Although they claim that this developmental sequence emerges from findings of past research, this sequence has not been empirically tested or well defined. The main purpose of this study is to investigate whether the first three stages[1] in the developmental sequence used by Nicoladis, Palmer, and Marentette (2007) can be evidenced by data from children acquiring Spanish as a native language. Other important information that this study provides is the influence that type and token frequencies in the input has on accuracy in production of children acquiring Spanish and some insight regarding the debate over

---

[1] As will be discussed below, the fourth stage in this developmental sequence is not falsifiable using data from pre-existing corpus data

single-mechanism versus dual-mechanism models of cognitive processes. First, an explanation of Spanish past-tense morphology is essential before discussing the details of this study.

Spanish Past-Tense Morphology

All verbs in Spanish fall in one of three conjugation classes. Theme vowels that attach to roots of verbs convey the conjugation class of a particular verb. The theme vowel of the first, and most productive, conjugation class is *a*. The theme vowels of the second and third conjugation class are *e* and *i* respectively. The three stems (combinations of root and theme vowel) in (1) exemplify the three conjugation classes.[2]

(1) *habla*- ‘talk’ *come*- ‘eat’ *asisti*- ‘attend’

Four different manners of conjugation in Spanish, which alternate depending on conjugation class, communicate different aspectual and modal perspectives in the past tense (Alboukrek, 2005)—three simple forms and one compound form. The three simple forms are *imperfecto* (imperfect), *pretérito* (preterit), and *imperfecto de subjuntivo* (imperfect subjunctive). The compound form is *perfecto de indicative* (present perfect), which uses the auxiliary verb *habe*- ‘have’ and a past participle form of the main verb. For

[2] I include hyphens after bare stems herein because they never appear without a suffix or change in stress in the past-tense

the purposes of this study, I did not include any past participle forms because they pattern differently than the other verbs in their formation. I did include all other forms in the study, including auxiliary verbs.

Because Spanish is a morphologically rich language, speakers use a variety of suffixes to communicate the different possible combinations of number, person, tense, aspect, and mode. Although in a morphological analysis, the suffixes that communicate these combinations could be divided into smaller segments to capture generalizations, for the purposes of this study, these divisions are not necessary, and I combine some suffixes for the sake of simplicity. The *imperfecto* is conveyed in regular verbs with the suffixes that attach directly to the root (replacing the theme vowel), in Table 1. The *pretérito* is conveyed in regular verbs with the suffixes that also attach directly to the root in Table 2. Finally, the *imperfecto de subjuntivo* is formed in regular verbs by adding the suffixes in Table 3 to the root.[3]

Table 1: *imperfecto suffixes*

1st conjugation class suffixes	2nd and 3rd conjugation class suffixes	Person and number
ba	*ía*	1st and 3rd person singular
bas	*ías*	2nd person singular
bamos	*íamos*	1st person plural
bais	*íais*	2nd person plural
ban	*ían*	3rd person plural

[3] Alternative suffixes in the *imperfecto de subjuntivo* are used in some dialects. Whether one form or the other is used is irrelevant to the current study.

Table 2: *pretérito suffixes*

1st conjugation class suffixes	2nd and 3rd conjugation class suffixes	Person and number
é	*í*	1st person singular
aste	*iste*	2nd person singular
ó	*ió*	3rd person singular
amos	*imos*	1st person plural
asteis	*isteis*	2nd person plural
aron	*ieron*	3rd person plural

Table 3: *imperfecto de subjuntivo suffixes*

1st conjugation class suffixes	2nd and 3rd conjugation class suffixes	Person and number
ara/ase	*iera/iese*	1st and 3rd person singular
aras/ases	*ieras/ieses*	2nd person singular
áramos/ásemos	*iéramos/iésemos*	1st person plural
arais/aseis	*ierais/ieseis*	2nd person plural
aran/ase	*ieran/iesen*	3rd person plural

There are only three irregular verbs in the *imperfecto*, they are *se*- 'be', *i*- 'go', and *ve*- 'see'; all other verbs in the *imperfecto* are regular. The *pretérito* and *imperfecto de subjuntivo*, however, have many different irregular verbs, some of which can be categorized into phonological families according to how their stems change and others are supletive. Most of these irregular verbs maintain the suffixes presented in Tables 1-3 (with changes in stress pattern in the 1st and 3rd person singular forms). Prescriptive grammars usually divide these irregular verbs into three different classes, *u-stem*, *i-stem*, and *j-stem* verbs. The last vowel in the stem of *u-stem* verbs changes to *u* in the *pretérito* and *imperfecto de subjuntivo* forms. Consonant changes in these verbs are not predictable. The verb *pone*- 'put' is an example of a *u-stem* verb. Here, the *o* changes to *u* in the *pretérito* and *imperfecto de subjuntivo* forms and results in *puso* 'he/she/it put' or *pusiera* 'I/he/she/it may put'. The *i-stem* verbs are similar to the *u-stem* verbs, but the vowels in the stems change to *i* instead of *u* and the suffixes drop the *i*. *Veni*- 'come' is an example of an *i-stem* verb that is realized as *vino* ‘he/she/it came’ and *viniera* ‘I/he/she/it may have come’ in the *pretérito* and *imperfecto de subjuntivo* forms respectively. The *j-stem* verbs, instead of replacing vowels, replace the last consonant of the stem when it is *c*. The verb *traduci-* ‘translate’ is an example of a *j-stem* verb that is realized as *tradujo* ‘he/she/it translated’ and *tradujera* ‘I/he/she/it may have translated’ in the *pretérito* and *imperfecto de subjuntivo* forms respectively.

Current Study

This study tested the hypothesis that the first three stages of the developmental sequence used by Nicoladis, Palmer, and Marentette (2007) can be evidenced by data from children acquiring Spanish as a native language. Another goal of this study was to advance understand on how past-tense morphological accuracy in monolinguals is sensitive to type and token frequency in the input.

Methods

Participants

The data that I looked at consisted of two different sets of L1 Spanish corpora, all of which can be found on the CHILDES database (MacWhinney, 1995, 2000). One set for the children’s accuracy data and another set for input data. The Diez-Itza (1995), Fernandez-Aguado, Linaza, López Ornat (1994), and Shiro (2000) corpora were used for children’s accuracy data. These corpora were chosen because of the age ranges that they represent. The ages of the children in the first four corpora listed above range from ages 1;7 to 4;11. The Shiro corpus contains data from children range from 6;5 to 10;9 in age. The data include spontaneous child-child and child-adult interactions and oral narratives. The total number of children from whom the data was collected is 185.

The input data was taken from three out of the five corpora listed above; the López Ornat, the Fernandez-Aguado, and the Linaza. These corpora were used because they contain extensive input data from parents, siblings, and other peers. Although these corpora include data from investigators, this data was excluded because of the possible bias toward certain forms that investigators wish to find may show up in their utterances. All other possible input sources were included in the analysis. The Diez-Itza and Shiro corpora were excluded because they only contain data either between target children and peers or between target children and investigators. More detailed information about the elicitation methods used in collecting this data can be obtained from MacWhinney (1995, 2000).

Procedure

All of the past-tense verbs were extracted (excluding past participles) using CLAN software (MacWhinney, 1995, 2000). Of these past-tense verbs, regular verbs and irregular verbs were identified and coded following Clahsen, Aveledo, and Roca (2002). Type and token frequencies of past-tense verbs were calculated for both input and child production groups along with percentages of irregular and regular verbs of all types and tokens. Past-tense verbs coded 'irregular' were verbs that exhibit a change in stem and/or inflectional suffixes displayed in Tables 1-3. Past-tense verbs coded 'regular' were verbs that do not exhibit a change in stem and inflectional suffixes displayed in Tables 1-3.

Error types were also identified and coded following Clahsen, Alveledo, and Roca (2002). Errors were categorized as 'stem overregularization', 'stem irregularization', ‘incorrect conjugation class suffix’, ‘other suffixation error’, and ‘other error’. Table 4 displays an example of each error type.

Table 4: *Error types*

Errors	Example	Correct form
stem overregularization	**ponió*	*Puso*
stem irregularization	**pusía*	*Ponía*
incorrect conjugation class suffix	**saló*	*Salió*
other suffixation error	**vistes*	*Viste*
other error	**tuví*	*Tuve*

Errors were coded ‘stem overregularization’ when verbs that exhibit stem changes in adult speech maintained a regular stem. Errors were coded ‘stem irregularization’ when verbs that maintain regular stems in adult speech change to irregular stems (usually applied from different past-tense forms). Errors were coded ‘incorrect conjugation class suffix’ when the suffix was taken from a different conjugation class than found in adult speech. Errors were coded ‘other suffixation error’ when verbs showed suffixation errors not consistent with other conjugation class. Finally, errors were coded ‘other error’ when a distinction between overregularization and irregularization could not be made.

Errors that could be considered strictly phonological errors, *contré* for *encontré* ‘I found’ for example, and errors involving incorrect inflection for person and tense (using a 1st person singular instead of 3rd person form for a 3rd person referent or using the *pretérito* when the *imperfecto de subjuntivo* was needed) were excluded from the analysis. This is because these types of errors, although interesting in their own regard and perhaps helpful in understanding derivational morphology acquisition, do not necessarily provide insight into the acquisition of past-tense morphology.

To see if the type and token distribution of past-tense verbs in the input was similar to the type and token distribution in target child production, I calculated the percentages of regular and irregular verbs (types and tokens) in each of the groups. I calculated the percentage of regular verbs (types and tokens separately) in the past-tense by dividing the total number of regular verbs by the total number of past-tense types and tokens produced in the input and the child speech data and multiplied the results by 100. I also calculated the percentage of irregular verbs by dividing the total number of irregular verbs by the total number of past-tense types and tokens produced in the input and the child speech data and multiplied the results by 100.

To find overall error rates, I divided the total number of errors found in the child production data by the total number of past-tense verbs produced. I then calculated the error rates for each type of error in the data by dividing the total number of each error by the total number of errors included in the analysis (excluding phonological errors). I then multiplied these results by 100 to convert the rate into percentages.

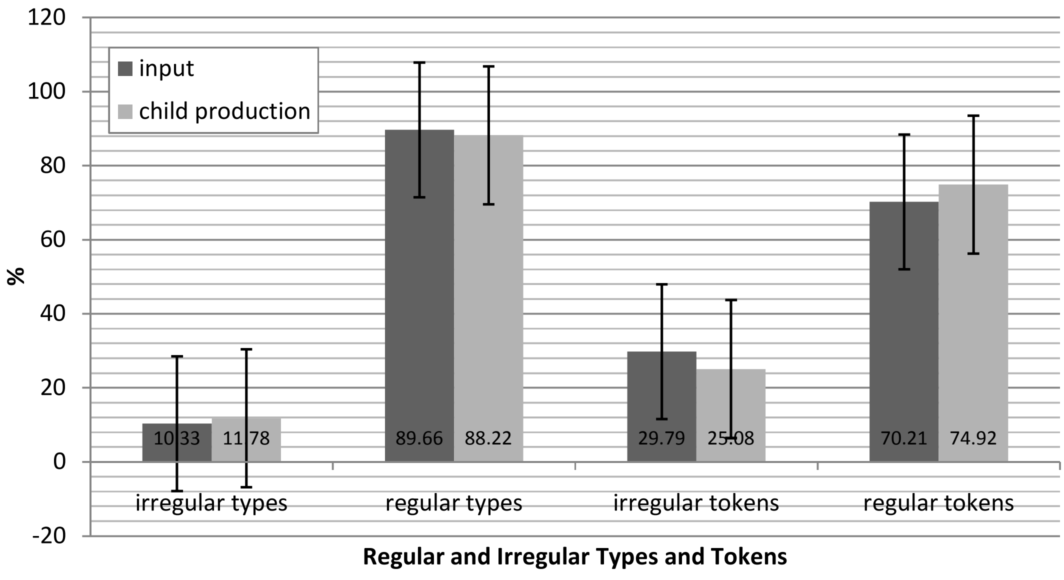
I also compared error types, error accuracy rates, and type and token frequencies to age to test the developmental sequence proposed by Nicoladis, Palmer, and Marentette (2007). I extracted the forty most frequent verbs in the input (20 tokens and above) and compared the verb types and accuracy in production with these input verbs to test the second stage in the developmental sequence. I looked at the types of stem irregularization errors in the child production data to test the third stage in the developmental sequence. Finally, I also calculated the proportion of stem irregularization and stem overregularization errors in the child production data out of the irregular and regular verbs in the input data also to test the third stage. I did this by dividing the number of errors for each error category by the total number of verbs of each category in the input and multiplying by 100 to convert the rate into a percentage of proportionate errors.

Results

Regular and Irregular Type and Token Distribution

There were 503 types and 4022 tokens of past-tense verbs found in the input data. Among the total number of types and tokens, there were 52 (10.33%) irregular types, 451 (89.66%) regular types, 1198 (29.79%) irregular tokens, and 2824 (70.21%) regular tokens. There were 433 types and 3645 tokens of past-tense verbs found in the child production data. Among the total number of types and tokens, there were 51 (11.78%) irregular types, 382 (88.22%) regular types, 914 (25.08%) irregular tokens, and 2731 (74.91%) regular tokens. Figure 1 illustrates the distribution of irregular and regular types and tokens in the input and child production data.

Figure 1 *Percentages of past-tense verbs in the input data*

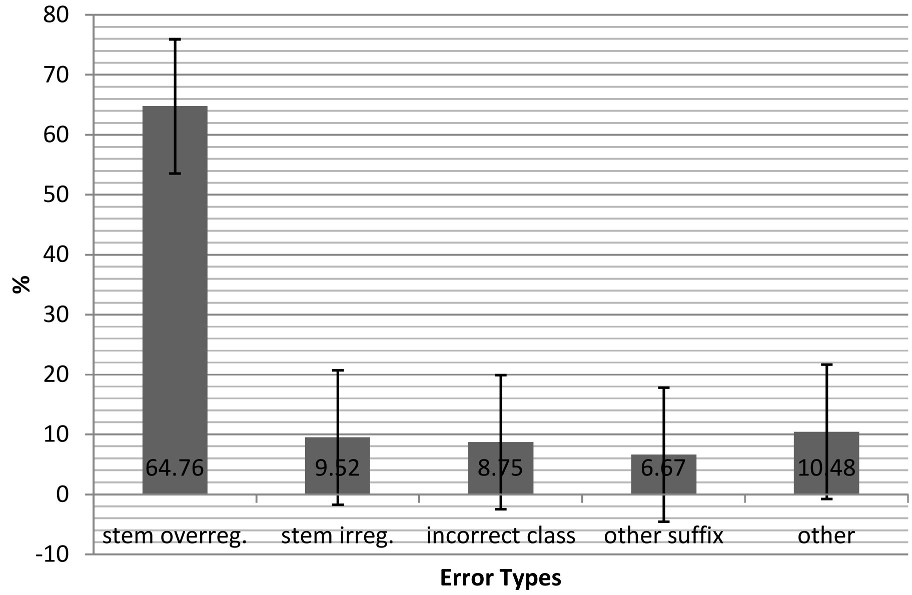


As seen in Figure 1, regular type and token frequencies are much higher than type and token frequencies for irregular verbs in both the input and the child production data. The standard error bars show that the two groups are not significantly different in their type and token productions. In other words, the frequencies in child production data closely resemble frequencies in the input.

Error Rates

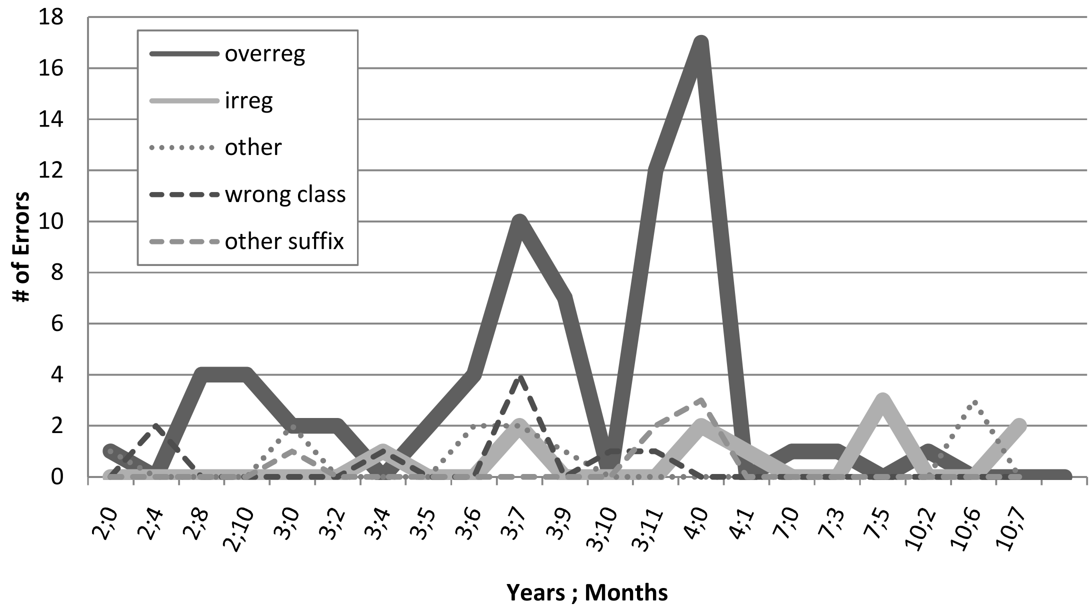
Out of the total number of past-tense verbs in the child production data, there were 165 past-tense errors. Errors that were strictly phonological made up 36.81% (60) of all of the past-tense errors. The total of the remaining errors that were analyzed was 105 verbs. The distribution of the remaining errors consisted of 64.76% (68) stem overregularization errors, 9.52% (10) stem irregularization errors, 8.75% (9) incorrect conjugation class errors, 6.67% (7) other suffixation errors, and 10.48% (11) other errors. Figure 2 shows the distribution of error rates. The standard error bars show that the amount of stem overregularizations is significantly more than the other errors.

Figure 2 *Distribution of error rates*



Type of errors and the amount of times that each error occurred in the child production data were compared to examine possible tendencies among children at certain ages. Figure 3 shows the types of errors that children made over time and by the number of each type of error that was made.

Figure 3 *Error types by number and age*



Starting at age 1;7 there were no utterances that referred to the past in the data that lacked past tense morphology. María (from the López Ornat corpus) used past tense-morphology at the outset of data collection and Juan (from the Linaza corpus) only referred to the past using past-tense morphology starting at age 2;0.

The first recorded morphological errors were both made by María at age 2;0. There was one instance of an overregularization and one instance of an 'other' error. Other overregularization errors began at age 2;8, and the numbers of these errors gradually rose to peak at age 4;0. Before age 2;8 there were 39 types (94 tokens) of verbs produced in the data. Of these verbs, only 46.15% (18 types, 64 tokens) were high token verbs in the input data. Out of all the errors that occurred in the child production data that were errors of high token verbs in the input data, there were only errors concerning four verb types—two 'stem irregularizations', two instances of the same 'other suffixation error', and eleven instances of the same ‘stem overregularization’ error with the verb *hace-* ‘make/do’. There were thirty-two instances of the 3rd person singular form of *hace-* in the input data and it was the twenty-sixth most frequent verb.

All irregularization errors are errors within the same verb paradigm. The error **acuerdaron* for the correct form *acordaron* ‘they agreed’ is an example of this phenomenon. This verb diphthongizes from *o* to *ue* in other contexts like the simple present tense (*acuerda* ‘he/she/it agrees’). In the *pretérito* however, this verb is regular in that it does not undergo a change in the stem. Irregularization errors do not coincide with phonological forms that span different phonological families. There were no errors like **aquierdaron* which would be patterning like another phonological family of verbs that undergo diphthongization from *e* to *ie*.

The amount of stem irregularization errors (10) was .83% of the total amount of irregular tokens in the input data (1198) while the percentage of stem overregularization errors (68) of the total amount of regular tokens in the input data (2824) was 2.41%. The proportion of stem overrgularization errors to regular verbs in the input was much higher than the proportion of stem irregularization errors to irregular verbs in the input.

Discussion

Distribution

The analysis comparing the type and token frequencies for regular and irregular verbs in the past-tense between the input and child production data shows that the distribution of type and token in the input plays a strong role in and can predict the distribution of the child production. This result is similar to what Nicoladis, Palmer, and Marentette (2007) found in bilinguals' type and token distribution.

The error rates of the child production data show that children overall are committing overregularization errors at a much higher rate than any other error. This is probably because, at least in Spanish, there are significantly higher type and token frequencies of regular verbs in the input that children are receiving. Because both type and token rates are higher in the input, these results do not aid in determining whether type or token frequencies contribute more to the overregularization phenomenon.

That these children produced a considerable amount of phonological errors is noteworthy. Although these errors did not directly relate to past-tense morphology, they are, as this data shows, a significant part of the acquisition of phonology and morphology. More research focusing on these types of errors should be carried out to explore the possible source of these errors and how they affect L1 acquisition.

Developmental Sequence

The current study also shows that data from monolingual children acquiring Spanish does not completely support the developmental sequence posited by Nicoladis, Palmer, and Marentette (2007). There are a few claims the sequence makes that the results of this study do not confirm. The first stage of the developmental sequence in question suggests that children do not use past-tense morphology to refer to the past at first. There is no evidence of such a phenomenon in the child production data. Juan (from Linaza corpus) did not clearly refer to the past until past-tense morphology was present in his speech, at which point he used the morphology to refer to the past. María (from López Ornat corpus) started using past-tense morphology from the outset of the data collection at age 1;7. There is a possibility that children refer to the past without past-tense morphology earlier than 1;7. This, however, would be difficult to empirically study because utterances occurring any earlier than this age are not easy to correctly decipher because phonetic production is not developed enough to be certain of children's true intentions in their speech. Many utterances produced by María even at 1;7 were almost unintelligible.

Although the first stage in the developmental sequence was unfounded in the data, the data did provide some evidence for what Nicoladis, Palmer, and Marentette (2007) propose would be evident in errors in the second stage. Although this may be the case, there is also evidence that production predicted in the second stage is instantiated. This stage predicts that accuracy and production rates are sensitive to input token frequencies. Accuracy rates are sensitive to token frequencies in the input. For the most part, children only made errors with verbs that were not in the forty most frequent verbs in the input. There were, however, a few exceptions; one of which was the verb *hace*- 'make/do'. Two of the other errors were only made once each by different children. The other error was made twice by another child. That *hace*- is found in the most frequent verbs in the input yet represents eleven errors in the child production data is interesting. This particular error was made by five different children with ages ranging from 2;10 to 3;11. This indicates that perhaps there is something inherent about this particular verb that makes it prone to overregularization regardless of the influence of input frequencies.

Although children were more accurate with verbs that are high token verbs in the input they receive, this does not necessarily mean that they produce more verbs that are high token verbs in the input. They actually produced slightly more verbs that were not high token verbs than verbs that were high token verbs in the input. Therefore, high token frequencies do not affect production as is predicted in the second stage of the developmental sequence in question. However, high token frequency does help children recognize and accurately produce past-tense morphology.

The third stage of the developmental sequence predicts that children make overregularization and irregularization errors because they become sensitive to different phonological families among verbs. The data in this study does provide some evidence for this claim. All irregularization errors could be considered to align with other phonological families. This phenomenon however, seems to be restricted to changes within verb paradigms. In other words, a phonological closeness factor prevents analogizing distant phonological families. Overregularization errors are also much higher than any other errors (paralleling regular verb frequencies in the input), which may indicate that children are sensitive to phonological families.

The fourth stage in the developmental sequence predicts that children begin analogizing across phonological families and applying regular morphology in novel contexts. This prediction was not falsifiable using the data in this study because there was no means of deciding whether children were using the past-tense morphology in novel contexts or if they had already acquired all of the verbs produced. A longitudinal study focusing on novel contexts in natural child language would provide evidence for or against stage four.

Another important point to consider regarding the past-tense morphology developmental sequence is that Nicoladis, Palmer, and Marentette (2007) predict that the phenomena that define each stage should be evident before or after the phenomena that define the other stages. In other words, overlap between stages should be minimal. The data in this study does not provide evidence for semi-discrete stages that follow a sequence. As discussed above, stage one was not evident at all in the data. As shown in Figure 3, the different overregularization and irregularization errors do not necessarily pattern in a way predicted by the developmental sequence; only overregularization errors show a strong effect over time. One possible weakness of this study regarding the testing of this sequence that should be mentioned is that it does not include data from the ages 5;0 thru 6;4. It is possible that these ages are key ages in proving evidence for the developmental sequence proposed.

Type and Token frequency

The influence of type and token frequency on the acquisition of past-tense morphology in children learning L1 Spanish can be summed up as having two different effects on how acquisition is patterned. The type frequency in the input will regulate the variety of errors evident in child production. In this study, children made significantly more overregularization errors than any other type of errors. This corresponds with the fact that regular verbs have the highest type frequencies in the input data.

The token frequency in the input will influence the overall accuracy rate in child speech production. Children tend to be more accurate with verbs that have high token frequencies in the input than verbs with lower frequencies. Token frequency, however, is not the only factor that contributes to accuracy. Errors in this study with the verb *hace-* 'make/do' show that, although they are rare, errors in high token frequency words in the input do occur in child production. Verb particular peculiarities or syntactic patterning influences may also contribute to past-tense morphology acquisition.

Cognitive processing models

Although it was not the main goal of this study to provide evidence for or against dual-mechanism or single-mechanism models of cognitive processes, the results of this study do provide some insight on the issue. Single-mechanism theories predict that past-tense morphology acquisition will be gradual and that high token frequency verbs will be acquired quickly while lower frequency verbs gradually emerge. Errors rates in high and low token frequency verbs in this study suggest that children do learn high frequency verbs quickly and that low frequency verbs are more gradual. However, single-mechanism models also predict that the overregularization and irregularization errors will be proportionate to the corresponding token frequencies in the input. The results of this study do not support this prediction. Overregularization rates overall represented a higher percentage of the total amount of regular tokens than the percentage of irregular rates represented the total amount of irregular tokens in the input data.

Dual-mechanism theories predict that past-tense morphology acquisition for regular verbs will be relatively quick and that overregularization errors will be dominant in child production. Overregularization rates compared to irregularization rates in this study suggest that dual-mechanism theories may better explain how past-tense morphology is acquired. However, the fact that the amount of overregularization errors gradually grew over time is not predicted by these theories.

Because there are seemingly insuperable problems with each set of theories, perhaps a combination of the theories can provide a better explanation for the cognitive processes involved in past-tense morphology acquisition. It is possible that children are using an associative network to build their lexicons and to associate phonological similarities and form-meaning relationships early on. Over time, they realize that certain forms are dominant in past-tense formation and they begin to form a rule, at which point the rule formation mechanism takes over. In other words, as overall cognitive development advances, so does children's ability to process regular and irregular verbs differently. Research in this area should be carried out to test the possibility of such an explanation.

Conclusion

This study has provided evidence that the past-tense morphology developmental sequence proposed by Nicoladis, Palmer, and Marentette (2007) does not fully describe how monolingual Spanish speaking children acquire past-tense morphology. Although there was some evidence that supported stages two and three, no effect for age was found that suggests a *sequence* that patterns as described by these authors. This study has also provided insight on how type and token frequencies influence morphological error patterns in child speech production and the debate on how to model the cognitive processes behind regular and irregular verb acquisition and formation.

References

Alboukrek, A. et al. (Eds.). (2005). *El Pequeño Larousse Ilustrado: Edición Centenario* (11th ed.). Mexico City: Larousse.

Aronoff, J.M., Giralt, N., & Mintz, T.H. (2006). Stochastic approaches to morphology acquisition. In C. A. Klee & T. L. Face (Eds.), *7th Conference of the Acquisition of Spanish and Portuguese as First and Second Langauges*. (pp. 110-121). Somerville, MA: Cascadilla Proceedings Project.

Berent, I., Pinker, S., & Shimron, J. (2002). The nature of regularity and irregularity: Evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research, 31*(5), 459-502.

Bybee, J. (1995). Regular morphology and the lexicon. *Languages and Cognitive Processes,10*(5), 425-455.

Clahsen, H., Aveledo, F., & Roca, I. (2002). The development of regular and irregular verb inflections in Spanish child language. *Journal of Child Language, 29*(3), 591-622.

Diez-Itza, E. (1995). Procesos fonológicos en la adquisición del español como lengua materna. In J.M. Ruiz, P. Sheerin, & E. Conzález-Cascos (Eds.), *Actas del XI Congreso Nacional de Lingüística Aplicada*. Valladolid: Universidad de Valladolid.

Kim, J.J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive Science, 15*, 173-218.

López Ornat, S. (1994) *La adquisición de la lengua Española*. Madrid: Siglo XXL.

MacWhinney, B. (1995, 2000). *The CHILDES project*. Hillsdale, NJ: Erlbaum.

Marcus, G.F., Pinker, S., Ullman, M.T., Hollander, M., Rose, T.J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57* (4, Serial No. 228).

Marshall, C.R., & van der Lely, H.K.J. (2006). A challenge to current models of past tense inflection: The impact of phonotactics. *Cognition, 100*, 302-320.

Nicoladis, E., Plamer, A., & Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science, 10*(2), 237-254.

Pinker, S. (1998). Words and rules. *Lingua, 106*, 219-242.

Pinker, S., & Ullman, M.T. (2002). The past and futre of the past tense. *Trends in Cognitive Sciences, 6*(11), 456-463.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38*, 43-102.

Rumellhart, D., & McClelland, J. (1986). On learning the past tenses of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 216-271). Cambridge, MA: Bradford Books.

Shiro, M. (2000). Diferencias sociales en la construcción del yo y del otro: Expresiones evaluativas en las narraciones de niños caraqueños en edad escolar. In T. de Bustos et al. (Eds.), *Lengua, discurso, texto*. (Vol. 1, pp. 1303-1318). Madrid: Universidad Complutense & Visor Libros.