Modeling Substitution Errors in Spanish Morphology Learning

Libby Barak[1] **(libby.berk@gmail.com), Nathalie Fernandez Echeverri**[2] **(nfern@umd.edu),**
Naomi H. Feldman[2,3] **(nhf@umd.edu), Patrick Shafto**[1,4] **(patrick.shafto@gmail.com)**
[1]Department of Mathematics and Computer Science, Rutgers University, Newark, NJ, 07102 USA
[2]Department of Linguistics, University of Maryland, College Park, MD, 20740 USA
[3] Institute for Advanced Computer Studies (UMIACS), University of Maryland, College Park, MD, 20740 USA
[4] School of Mathematics, Institute for Advanced Study, Princeton, NJ, 08540 USA

Abstract

In early stages of language acquisition, children often make inflectional errors on regular verbs. For example, Spanish-learning children produce *–a* (present tense 3rd person singular) when other inflections are expected, while English-learning children often produce the bare form when an inflectional *–ed* or *–s* is expected. Most previous models of inflectional morphology learning have focused on later stages of learning, when children produce regular verbs correctly but have trouble with irregular verbs. In this work, we examine the earlier stage of learning in the acquisition of Spanish. We propose a computational model of Spanish inflection learning and present a novel data set of gold-standard inflectional annotation for Spanish verbs. Our model replicates data from Spanish-learning children, capturing the order in which children acquire different inflections and correctly predicting the substitution errors they make. Analyses of the learned grammar show that the learning trajectory can be explained as a result of the gradual acquisition of inflection-meaning associations. Ours is the first computational model to provide an explanation for this acquisition trajectory in Spanish, and represents a theoretical advance more generally in explaining substitution errors in early morphology learning.

Keywords: Computational modeling; Morphology; Spanish; Language acquisition

Children's use of inflectional morphology changes dramatically over the course of the first few years of language acquisition. At early stages, children may produce a verb form in a semantic context that obligates the use of a different inflection. For example, English-learning children use the bare form when *–ed* for past tense is expected, Spanish-learning children use 3rd person singular form of present tense verbs when inflection for other person/number combinations is expected. Similar substitution errors are observed in highly-inflected languages such as French, Italian, and Hebrew (Wexler, 1994; Grinstead, De la Mora, Vega-Mendoza, Flores, et al., 2009; Aguado-Orea & Pine, 2015).

Previous computational models of morphology learning have mostly focused on English. However, more complex inflectional systems and their corresponding acquisition trajectories enable a direct analysis of agreement errors within a diverse set of inflections and meanings (Legate & Yang, 2007; Freudenthal, Pine, Aguado-Orea, & Gobet, 2007; Engelmann et al., 2019). The rich morphology system of Spanish, for example, includes over 40 possible inflections for each verb (Bosque, 1999). Studies of Spanish and other Romance languages show few if any productions of non-finite verbs similar to the bare form productions in English (Grinstead et al., 2009). Instead, Spanish-learning children primarily use 3rd person singular (3Sg) for present tense in their earliest verb productions. These include contexts where other inflections are expected, i.e., 1st and 2nd person singular and 3rd person plural (Garcia, 2007). Within a few months, children start using the 1st person singular (1Sg) inflection correctly, but continue to use 3Sg where 2nd person singular (2Sg) or 3rd person plural (3Pl) are expected (Fernández Martínez, 1994). Substitution of 3Sg for 3Pl continues for the longest period.[1] This trajectory is consistent across corpus studies and laboratory experiments (Aguado-Orea & Pine, 2015; Bedore & Leonard, 2001; Grinstead et al., 2009; Rujas, Casla, Mariscal, Lázaro López-Villaseñor, & Murillo Sanz, 2019). While children's overall error rate has been estimated at 5%, this low rate mainly reflects the high overall frequency of 3Sg, which is produced correctly; the error rate for 3Pl is estimated at 30-50% of 3Pl productions (Aguado-Orea & Pine, 2015).

In explaining Spanish-learning children's substitution errors, previous literature offers two generalizations. Some studies have referred to the 3Sg form as the Spanish bare form (Grinstead et al., 2009), implying that substitution errors are governed by a form's default status. Other theories refer to 3Sg substitution errors as 'one-off' errors based on the semantic distance of the obligatory inflection from 3Sg (replacing either the number or the person), suggesting that it is the relationship between the two forms that governs substitution errors (Bedore & Leonard, 2001). Computational modeling can help us understand the reason why one inflection is systematically substituted for another, and whether these errors arise from a single cause or multiple different causes.

In this work, we propose a computational model of Spanish inflection learning. We adopt Fragment Grammars, which are sufficiently interpretable to allow us to examine the model's developing grammar, and have been shown to replicate a wide range of observations on English morphology acquisition including the parallel *bare* form production in English (O'Donnell, 2015; Barak, Harmon, Feldman, Edwards, & Shafto, submitted). We find that our model reproduces the fine-grained trajectory of agreement errors, and moreover, that it provides a novel explanation regarding the source of the children's erroneous productions. Specifically, our model shows

[1]We do not analyze 1st and 2nd plural inflections because their frequency is too low during this period (Aguado-Orea & Pine, 2015).

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

that these errors arise from an interplay between the need to restrict productive inflections, e.g., 3sg, to appropriate contexts, and the need to associate less frequent inflections with their appropriate contexts despite fewer observations. Our analysis additionally shows a different possible cause for each type of substitution error, which can explain why certain inflections take longer to master. Our results have implications in multiple research domains, including psycholinguistics and speech pathology.

In addition to providing new theoretical results, we contribute a novel, publicly available data set with annotations that list the inflection, irregularity, and modifications to the verb for each verb and set of grammatical features (person, number, and tense).

Models of Morphology Learning

Studies of morphology learning diverge into two main explanations of the observed trajectory. Early production of verbs in their inflected form lead to hypotheses that children know the inflectional properties of their native language from a very early point (Wexler, 1994). This explanation marks *bare* form productions in English, i.e., *jump* instead of *jumps/jumped*, as omission errors where the *–s* or *–ed* inflections are omitted. However, this explanation falls short in explaining subject-verb agreement errors in highly inflected languages such as Spanish, which involve substitution. A second approach suggests early productions are limited to a handful of verbs that are learned in their inflected form as the full inflectional properties of language are gradually learned (Tomasello, 2000). The learning trajectory in highly inflected languages provides a rich opportunity to adjudicate between these views and propose others.

Yet, only a handful of models have considered highly-inflectional languages such as Polish, Finish, and Spanish (Legate & Yang, 2007; Freudenthal, Pine, & Gobet, 2010; Freudenthal, Ramscar, Leonard, & Pine, 2021; Engelmann et al., 2019), and a number of gaps remain in our understanding of how patterns of substitution errors relate to knowledge of subject-verb agreement. Several studies have aimed to replicate the distributional pattern of inflection without fully representing the semantic properties of the subjects, or allowing the model to learn subject-verb agreement from the input (Legate & Yang, 2007; Freudenthal et al., 2010). A recent study presents an analysis of morphology acquisition in English and Spanish using an error-driven model (Freudenthal et al., 2021). This study addresses production rate for specific inflections across languages with different distributional properties for typically developing and children with language disorders. While the results capture earlier acquisition of singular vs. plural inflections in Spanish, the simulations produce only a partial replication of the acquisition order of the four inflection types and do not evaluate substitution behavior or analyze the possible causes. Our study similarly extends a model that has been shown to replicate English inflection trajectory with full consideration of form-meaning representation.

Table 1: An example of the CFG rules that correspond to the verb *hablar*

Non-terminal Rules
$Start \rightarrow AgreementF\ Predicate$
$AgreementF \rightarrow Person\ Number$
$Predicate \rightarrow Stem\ Suffix$
$Suffix \rightarrow Tense\ Inflection$
Terminal Rules
$Person \rightarrow$ 3rd
$Number \rightarrow$ singular
$Stem \rightarrow$ *hablar*
$Tense \rightarrow$ present-tense
$Inflection \rightarrow$ *–a*
Memoized fragments
$Start \rightarrow$ 3rd $Number\ Stem$ present-tense *–a*
$Start \rightarrow$ 3rd singular $Stem$ present-tense *–a*

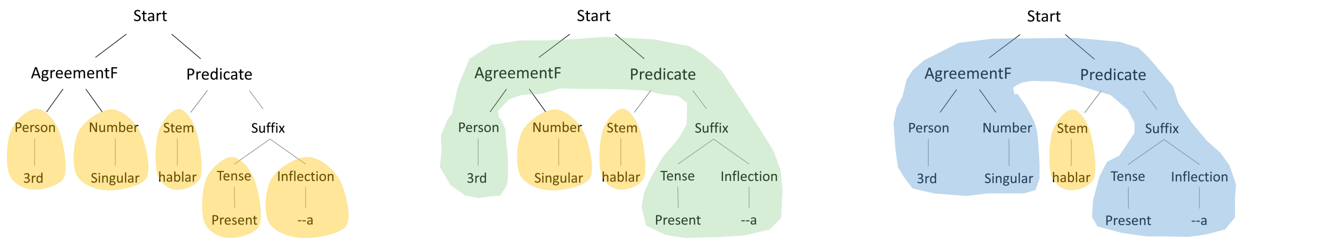
Importantly, our analysis offers an explanation of the source of each substitution mistake for each inflection type that has been missing from previous computational work.

PCFG-based models have been used to study the acquisition of morphology in many languages (Johnson, Griffiths, & Goldwater, 2006; Botha & Blunsom, 2013; Sirts & Goldwater, 2013; O'Donnell, 2015). While the framework has been shown to capture various aspects of the acquisition process including segmentation in multiple languages and overgeneralization in English, to our knowledge it has not been used to analyze subject-verb agreement errors in general. In earlier work, we show that our proposed model produces the shift from *bare* form to *–ed* production in English (Barak et al., submitted). We now extend this framework and adapt its rule composition to capture the key properties for Spanish verb inflections.

Fragment Grammars (FG)

We simulate morphology learning using the Fragment Grammars (FG) model, a Bayesian non-parametric model that learns over a Context-Free Grammar (CFG) (O'Donnell, 2015). The CFG, $\mathcal{G}$, consists of rules of the form $A \rightarrow \beta$ where β is the collection of terminal and non-terminal production rules. The CFG rules represent the knowledge level the learner associates with the input. The top and middle panel of Table 1 provide an example of the rules that generate the verb *hablar* with the 3rd person singular (3Sg) present tense inflection *–a*. The terminal rules imply that the learner can detect the agreement features of the verb (person and number), the tense, and the inflection independently from each other and from the stem.

The model requires the learner to infer the distribution of the rules and their association to each other. The distribution of the CFG rules is learned by updating the counts for terminal and non-terminal rules: (1) for each non-terminal, the model updates the Dirichlet-multinomial pseudocounts $\{\vec{\pi}^A\}_{A \in V_{\mathcal{G}}}$, and (2) for each non-terminal, the model updates the beta-



(a) Terminal-Rule based Fragments (b) Missing-Number Fragment (c) Full-context Fragment

Figure 1: A tree illustration for learning and generating inflected regular verb forms in Spanish, (a) using the CFG rules as listed in the top and middle panel of Table 1, (b) using stored rules and a fragment that associates person, tense and inflection (first fragment in the bottom panel of Table 1), and (c) using rules and a stored fragment that associates number, person, tense, and inflection (second fragment in the bottom panel of Table 1).

binomial distributions $\{\vec{\Psi}_B\}_{B\in\mathrm{rhs}(r\in R_G)}$.

The FG model extends CFG learning by a memoization process that stores fragments of the tree in the memory. This process allows FG to relax the assumption of independence of rules in the CFG. For example, the model learns the association of the rules, e.g., *–a* for present tense 3Sg, by memoizing fragments of the CFG that frequently occur in the input. The bottom panel of Table 1 provides an example of two such fragments. At each iteration, the model updates the parameters for the generating CFG rule and any fragments it uses for the interpretation of the input item. The verb *habla*, for example, can be generated by multiple combinations of the CFG rules and fragments presented in the table as illustrated in Figure 1a.

The model implements the stochastic memoization using the Pitman-Yor process, a generalization of the Dirichlet Process (Pitman & Yor, 1997). Both CFG rules and their combinations are considered fragments. At a first encounter of a fragment, the memoization function will store it in memory. After this, the memoization function either chooses an observed fragment i with the probability $\frac{n_i-a}{N+b}$ or samples a new fragment with the probability $\frac{aK+b}{N+b}$, where n_i is the number of times fragment i has been used, N is the number of input items sampled so far, and K is the number of times a new value has been sampled from the underlying function. The model updates the distribution of fragments using using a vector of hyperparameters for each non-terminal, $\{< a^A, b^A >\}_{A\in V_G}$. For consistency with previous applications of FG to morphology learning, we follow the parameter setting used previously. For full mathematical description of the FG model and parameter setting, please refer to O'Donnell (2015).

Note that the full set of CFG rules also includes non-terminal and terminal rules to generate irregular verbs. These fragments are not relevant to the analysis in this paper, since they cannot be used to generate a verb with a regular inflection. However, the model observes irregular verbs in the input similar to Spanish-learning children and generates corresponding fragments appropriately.

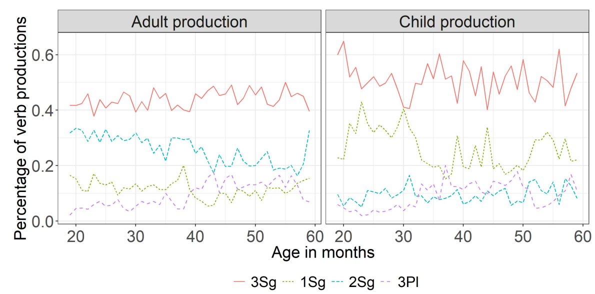


Figure 2: The distribution of verb productions by adults and children in the Spanish portion of CHILDES (529 children over 22 corpora). Percentage of use of 3SG, 1Sg, 2Sg, and 3Pl, over the total number of verb produced in present tense for the age group. Other person, number, and tense combinations are omitted for simplicity to focus on the analyzed cases.

Data Annotation and Methods

Existing data sets for Spanish verb inflections list the grammatical form for each verb in a given context (e.g. Kann & Schütze, 2017). To fully represent the richness of the morphological system, we require annotation of the phonological transformation the verb undergoes and the corresponding grammatical class for every irregular verb. We annotated a new data set of Spanish verbs to identify regular and irregular verbs and their corresponding inflection for each possible context.[2]

The dataset includes 1086 verb types. We include all the verbs produced by any speaker in the Spanish portion of the CHILDES data set (MacWhinney, 2000). Each verb is listed with its stem and all possible tokens, inflected form, English meaning translation, irregularity class for irregular verbs, and inflection for each person, number, and tense combination. The full dataset includes over 40,000 verb tokens that span over 45 property combinations for a full paradigm. This

[2] https://github.com/CoDaS-Lab/SVMorph

dataset allows for future analysis of regular verbs in cases that are not observed in CHILDES and acquisition process of irregular inflection classes. The irregular verb classification corresponds to the notation used in the Spanish conjugation tables provided by Wiktionary, a large multi-lingual crowd-sourced dictionary.[3] A native-speaker of Spanish marked each irregular modification as affecting pronunciation vs. only orthography. The context for each verb is provided along with the morphological notation used in the CHILDES and SIGMORPHON datasets.[4] For completeness, we add frequency data for each verb-context combination from the Spanish portion of CHILDES.

We used these annotations, in conjunction with verb distributions in the Spanish CHILDES corpus (MacWhinney, 2000), to create input data for our models. Following O'Donnell (2015), we extracted all verbs in the child-directed portion of Spanish CHILDES for ages 18-50 months using consecutive months to simulate progressive data. We modeled development using datasets of increasing size, each subset of which is drawn from the respective month of child data and appended to the preceding months' data. Verbs were sampled according to their frequency in the data while preserving overall distribution of semantic and syntactic properties.

The CHILDES Spanish data include $69,514$ verb productions by adult speakers. $23,276$ of these usages are of regular verbs. The present tense amounts to $18,072$ of the regular verb productions (77.6%). The distribution of present-tense regular verb productions over person and number is illustrated in Figure 2. Given that Spanish grammar allows subjects to be omitted, $4,959$ verb productions in CHILDES can be either 1st or 3rd person based on the use of inflection alone. CHILDES marks these cases as 1st or 3rd person, e.g., "13S PAS" for Singular 1st or 3rd person in past-tense. To align the data with the morphology notation used in conjugation tables and other datasets, we adopt the following methodology to disambiguate these tokens. We assign each of these verb uses to either 1st or 3rd person proportionally to how many times each is used in sentences that do have overt subjects, to maintain the overall 1st/3rd person distribution.

We trained the model using 10 independent samples, each representing a hypothetical child using the consecutive sampling data. The samples have the same distribution over the linguistic properties, but differ in the observation of specific stems and their combination with the agreement features. To represent naturalistic complexity, the data include any combination of tenses, person/number, and regular/irregular observed in the child-directed portion of CHILDES. Thus, to memoize a present-tense fragment, the model must observe sufficient number of input items and may store fragments for other forms that are not analyzed in the scope of the current paper. A random seed was used to initialize each simulation. Every simulation was run for a total of 50 sweeps.

[3] https://en.wiktionary.org/wiki/Category:Spanish_verb_forms

[4] See https://talkbank.org/manuals/MOR.pdf and https://github.com/sigmorphon/

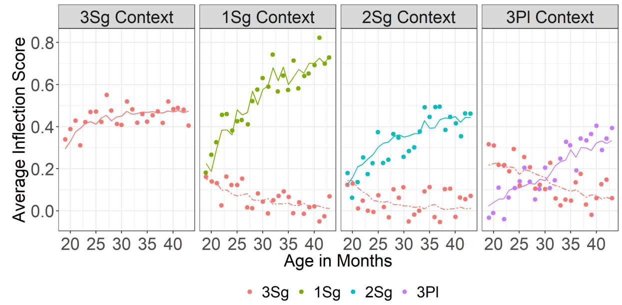


Figure 3: The average score for predicting the grammatical inflections or substituting with 3rd person singular inflections for the present tense.

Simulating Inflection Production

Psycholinguistic studies have observed a consistent substitution error pattern in the production of present-tense inflections by Spanish-learning children. Children use the 3Sg inflection in contexts that obligate the use of other grammatical forms, 1Sg, 2Sg, and 3Pl. Spanish-learning children continue to make errors in the production of 3Pl for the longest period, continuing to use the 3Sg inflection in place of the grammatical form.

To test our model for this substitution pattern, we calculate the probability of producing each regular verb with either the 3Sg inflection or the grammatical inflection for its obligatory context. For example, we calculate the probability of producing *habla* (3Sg) vs. *hablo* (1Sg) when the agreement features are set to be 1st person and singular in present-tense. The probability of each inflected form is measured from the posterior probability of the stored fragments in the grammar. We use the Metropolis–Hastings algorithm to calculate the set of fragments that best describe each predicted item (O'Donnell, 2015). For example, if the grammar contains all the grammars fragments presented in Figure 1, our evaluation process would compare the probability of generating *habla* from each combination of fragments and would record its production probability as the highest probability out of all given options. We present the results for the substitution errors reported in the psycholinguistic literature, i.e., 1Sg and 3Sg for 1Sg context, 2Sg and 3Sg for 2Sg context, and 3Pl and 3Sg for 3Pl context.

Figure 3 presents the results for inflection production for each of the four contexts (3Sg, 1Sg, 2Sg, and 3Pl). Since there are no reports of agreement errors for the 3Sg inflection, we report only the probability of producing 3Sg inflections in a 3Sg context. While not pictured, our results in this context showed that none of the other inflections approaches the probability of 3Sg in production. That is, the model does not produce agreement errors in 3Sg contexts, similar to children.

The two middle panels of Figure 3 present the predictions for 1Sg and 2Sg contexts, respectively. Similar to observations of Spanish-learning children, the model predicts a competing probability of producing the 3Sg inflections given these two context scenarios. The probability of producing the 1Sg inflection in its obligatory context rapidly increases compared

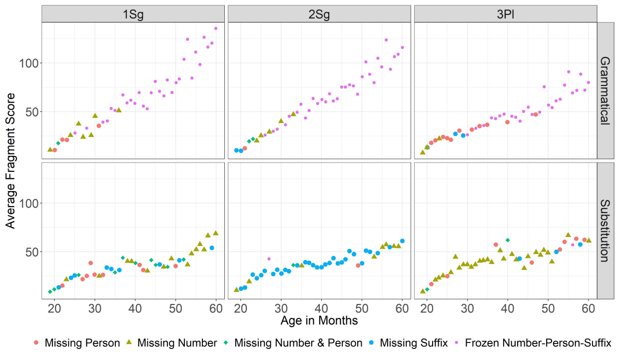


Figure 4: The average scores for stored fragments the model use to predict 3Sg or the grammatical inflections for either 1Sg, 2Sg, or 3Pl. Each fragment includes partial set of features, e.g., 'missing person' fragment associates number and suffix as in Figure 1b.

with that of 2Sg and even 3Sg. This replicates the shortest duration and lowest error rate for 1Sg compared with 2Sg and 3Pl in children (Aguado-Orea & Pine, 2015; Grinstead et al., 2009). The right panel of Figure 3 presents the probability for the 3Pl context. Our results replicate the observation of the longer duration for agreement errors in which 3Sg inflection is produced when 3Pl is expected (Aguado-Orea & Pine, 2015; Bedore & Leonard, 2001; Rujas et al., 2019).

Importantly, the results for each context do not simply reflect the distribution of inflection use in the input. As shown in Figure 2, child-directed speech (CDS) contains predominantly 3Sg inflections for present tense (around 50% for the duration of the recorded data). In addition, CDS contains higher rate of 2Sg and similar rates of 1Sg and 3Pl. The next section reports analyses of the learned grammar to provide an explanation of the learning trajectory in the model, which can shed light on why children might show a similar learning trajectory.

Grammar Analysis

Figure 4 presents the fragments for 1Sg, 2Sg, and 3Pl contexts. We present both the fragments that underlie the production of 3Sg erroneously, and the fragments that enable the model to eventually prefer the expected inflection as a more probable one. As described above, the model can generate inflected verbs using any combination of the fragments memoized in the grammar. These fragments can be of a single CFG rule (yellow shading in Figure 1) or a combination of several rules (green and blue shading in Figure 1). We distinguish 5 types of fragments: (1) Missing person, (2) Missing number, (3) Missing number & person, (4) Missing suffix, and (5) Frozen number, person, and suffix. Figure 1c shows an example of the fifth type of fragment with frozen number, person, and suffix. Figure 1b illustrates one example of a fragment of the third type (missing number). Though the model can generate fragments that include the stem (e.g., "*Start* $\rightarrow$ 3rd singular *hablar* present-tense *–a*"), these fragments can only be used to generate this specific verb. Therefore, these fragments never gain an overall likelihood score as high as more general regular-verb fragments. We average the score for the generated fragments across samples and runs (see Figure 4).

3Sg production. Children produce agreement errors by producing verbs with *–a* or *–e*, the grammatical inflections for 3Sg, when the obligatory context requires the use of a different inflection. The model can produce these errors from the grammar when a fragment that includes *–a* or *–e* inflection is combined with fragments including 1Sg, 2Sg, or 3Pl features. Our analysis shows that the model stores several fragments that associate *–a* and *–e* inflections with all 3Sg features early on. These fragments support the production of 3Sg inflections for the appropriate context, but cannot explain agreement errors, which contain other person and number features.

The model also stores partial fragments, as illustrated in Figure 1b. For example, the missing-number fragment presented in Figure 1b associates *–a* inflection with 3rd person and Present tense. Although such fragment is memoized based on observation of 3Sg verbs, the fragment can be used to produce 3Pl verbs if combined with a fragment of the plural number only and any stem fragment (denoted by yellow shading in the figure). Similar fragments with missing person information can be used to generate an agreement error for 1Sg and 2Sg context. Therefore, agreement errors may arise from fragments that do not restrict the use of 3Sg inflections sufficiently *and* fragments that do not specify the association of 1Sg, 2Sg, and 3Pl to their grammatical inflections.

1Sg context. Our analysis shows that errors of producing 3Sg inflection in 1Sg context mostly result from fragments that recognize *–o* as a frequent inflection without associating it to any number or person features. Indeed, Aguado-Orea and Pine (2015) hypothesize that errors in the 1Sg context may arise either from the distributional properties of 1Sg inflection or from a pragmatic need of children to master the 1Sg context. Our findings support their first hypothesis. Unlike the 2Sg and 3Pl, the *–o* inflection is used for 1Sg context in the same way across verb conjugations. As shown in the top left panel of Figure 4, grammatical prediction of *–o* for 1Sg context is enabled by a mixture of fragments with higher probability that the fragments in the bottom left panel. These fragments denote high probability for *–o* use even if the number feature is missing from the fragment.

Our findings are in line with the more nuanced analysis of the 1Sg errors. Aguado-Orea and Pine find that correct use of 1Sg is mostly with the verb *querer* (*want*), while the use of 3Sg inflections for 1Sg is more frequent with low frequency verbs. Indeed, our analysis focuses on fragments that do not include stems, which will be more influential in the production of low frequency and novel verbs.

2Sg context. The bottom-middle panel of Figure 4 shows that the most frequent type of fragment leading to 3Sg use in 2Sg context is a missing-Suffix fragment. This fragment associates the person and number features (2nd and Singular) without memoizing their association to one of the possible inflections for this context (i.e., *–as*, *–es*, or *-is*). As opposed to the use of a single inflection for 1Sg context, several different inflections can be grammatical for 2Sg context, making it harder for the model to observe sufficient input items for each 2Sg inflection.

3Pl context. Mistakes for 3Pl resemble what Bedore and Leonard (2001) referred to as a *one-off prediction*. As shown in the right bottom panel of Figure 4, a fragment with missing-number information is used to generate verbs with 3Sg inflections for a 3Pl context by combining these fragments with a number-only fragment (yellow shading in Figure 1).

Correct production. The top three panels of Figure 4 show the probabilities of fragments that enable the prediction of grammatical inflections for 1Sg, 2Sg and 3Pl contexts. By the end of training, the model learns a higher probability of using the fragments in the top three panels compared with the three bottom panels. Moreover, the top panels show that the fragments for all three contexts look the same for correct production. In each case, the model learns a fully memoized a fragment for each context that associates the person, number, and full suffix. The model predicts the production of the grammatical inflection for each context once the score of this fragment is higher that the alternative that offer only partial computation (omitting person, number, or suffix).

Overall, our results suggest that two different accounts in the literature are each at least partially correct. Bedore and Leonard refer to substitution errors in all three contexts as near-misses, hypothesizing that they are caused by replacing a single feature (number or person). This is consistent with our model's behavior in 3Pl contexts, where the model uses a fragment that is missing a context feature. Other studies instead highlight the different rate of productions, error rate, and duration of errors for each of the three substitution contexts and hypothesize either a difference in the cause for the error or differences in the supporting distributional properties (Aguado-Orea & Pine, 2015; Grinstead et al., 2009; Bedore & Leonard, 2001). Consistent with this, our model shows that different causes may arise from a single learning framework as a result of difference in the distributional properties of each inflection pair.

Discussion

In this paper, we present a computational model of Spanish morphology learning using a non-parametric Bayesian framework. We conduct a novel analysis of the substitution errors observed in children and explain them using a developing grammar. Our work provides an important extension of an existing model of past tense acquisition in English to a highly-inflected language. Our results replicate observations from Spanish learning monolinguals while providing a novel explanation of the source for the children's erroneous productions.

Our analysis of the grammar shows that a single type of substitution may be a result of three different learning trajectories. 3Sg substitution arises because of frequency of the correct suffix in the 1Sg case, failure to store the correct suffix in the 2Sg case, and failure to encode number information in the grammar in the 3Pl case. The interplay between distributional properties of use in CDS and diversity across multiple linguistic features results in three pathways to gradual learning.

Our representation of the CFG rules implies the learner is capable of correctly identifying the linguistic properties of the context (person, number, and tense), as well as the stem and inflection from the earliest stage of learning. In that sense, our simulation follows the evidence of early comprehension and production of rich morphology that guided the theories of Wexler (1994). Similar to these theories, our model must learn the correct association of inflections and their meaning over experience. However, our account also differs from these theories in several key points that place it closer to theories of a gradual learning process. First, our data and corresponding simulation starts from 18 months old CDS, which we assume to be an intermediate stage of development; some learning may have occurred to bring children to the knowledge state that we assume. Second, we focus not on the processes by which children learn *–ed* in the first place, but rather, on the processes that allow them to extend this to all of the verbs in their vocabulary. Our model can account for both earlier stages of verb-specific production of *–ed*, through the creation of stem-specific fragments and their use for correct production of specific verbs as a whole, and for later stages through the fragments discussed in this paper. As such, our account is compatible with theories that ascribe early productions as verb-specific instances rather than global understanding of the morphological structure of the language (Tomasello, 2000).

Children with language disorders show lower rates of verbal agreement errors in highly inflected languages, such as Spanish, compared with English-learning children (see Freudenthal et al., 2021 for a review). However, elicitation studies show that the high rate of 3Sg production masks the error rate of agreement error with other lower frequency inflections. Such studies find that verb morphology best predicts group membership of learning trajectory (Castilla-Earls, Auza, Pérez-Leroux, Fulcher-Rood, & Barr, 2020; Bedore & Leonard, 2001). Fragment grammars have been successfully applied to replicate observations from English-learning children with language disorders (Harmon, Barak, Shafto, Edwards, & Feldman, 2022). Our study opens a new venue to understand the source of mistakes and possible interventions for Spanish-learning children (Eaves Jr, Feldman, Griffiths, & Shafto, 2016).

In addition to the experimental results, we present a novel data set that annotates over 1000 Spanish verbs, their English translation, grammatical inflections, and irregular categories. We show the usefulness of our data in creating naturalistic input to our model by combining it with the frequency measures from CHILDES. The data set offers opportunities to study generalization patterns for irregular verbs in a highly-inflected language compared with English. We aim to explore this aspect of the data in future work.

Acknowledgments

We thank Tim O'Donnell for sharing his code and Jan Edwards and Zara Harmon for helpful discussion. This research was supported by NIH R21 DC017217.

References

Aguado-Orea, J., & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PloS one*, *10*(3), e0119613.

Barak, L., Harmon, Z., Feldman, N. H., Edwards, J., & Shafto, P. (submitted). When children's production deviates from observed input: Modeling the variable production of the English past tense. .

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

Bosque, I. (1999). *Gramática descriptiva de la lengua española: Entre la oración y el discurso. morfología* (Vol. 3). Espasa.

Botha, J. A., & Blunsom, P. (2013). Adaptor Grammars for learning non- concatenative morphology..

Castilla-Earls, A., Auza, A., Pérez-Leroux, A. T., Fulcher-Rood, K., & Barr, C. (2020). Morphological errors in monolingual spanish-speaking children with and without developmental language disorders. *Language, speech, and hearing services in schools*, *51*(2), 270–281.

Eaves Jr, B. S., Feldman, N. H., Griffiths, T. L., & Shafto, P. (2016). Infant-directed speech is consistent with teaching. *Psychological review*, *123*(6), 758.

Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., ... Lieven, E. (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology*, *110*, 30–69.

Fernández Martínez, A. (1994). El aprendizaje de los morfemas verbales: Datos de un estudio longitudinal. *La adquisición de la lengua española*, 29–46.

Freudenthal, D., Pine, J., & Gobet, F. (2010). Explaining quantitative variation in the rate of Optional Infinitive errors across languages: a comparison of MOSAIC and the Variational Learning model. *Journal of child language*, *37*(3), 643–669.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science*, *31*(2), 311–341.

Freudenthal, D., Ramscar, M., Leonard, L. B., & Pine, J. M. (2021). Simulating the acquisition of verb inflection in typically developing children and children with developmental language disorder in English and Spanish. *Cognitive Science*, *45*(3), e12945.

Garcia, C. B. (2007). Early root infinitives in a null-subject language: a longitudinal case study of a Spanish child. In *Proceedings of the 2nd conference on generative approaches to language acquisition north america* (pp. 39–50).

Grinstead, J., De la Mora, J., Vega-Mendoza, M., Flores, B., et al. (2009). An elicited production test of the optional infinitive stage in child Spanish. In *Proceedings of the 3rd conference on generative approaches to language acquisition north america (galana 2008)* (pp. 36–45).

Harmon, Z., Barak, L., Shafto, P., Edwards, J., & Feldman, N. H. (2022). The competition–compensation account of developmental language disorder. *Developmental Science*.

Johnson, M., Griffiths, T., & Goldwater, S. (2006). Adaptor Grammars: A framework for specifying compositional nonparametric Bayesian models. *Advances in neural information processing systems*, *19*.

Kann, K., & Schütze, H. (2017). The LMU system for the CoNLL-SIGMORPHON 2017 shared task on universal morphological reinflection. In *Proceedings of the conll sigmorphon 2017 shared task: Universal morphological reinflection* (pp. 40–48).

Legate, J. A., & Yang, C. (2007). Morphosyntactic learning and the development of tense. *Language Acquisition*, *14*(3), 315–344.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. transcription format and programs* (Vol. 1). Psychology Press.

O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. MIT Press.

Pitman, J., & Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability*, 855–900.

Rujas, I., Casla, M., Mariscal, S., Lázaro López-Villaseñor, M., & Murillo Sanz, E. (2019). Effects of grammatical category and morphology on fast mapping in typically developing and late talking toddlers. *First Language*, *39*(3), 249–280.

Sirts, K., & Goldwater, S. (2013). Minimally-supervised morphological segmentation using adaptor grammars. *Transactions of the Association for Computational Linguistics*, *1*, 255–266.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends in cognitive sciences*, *4*(4), 156–163.

Wexler, K. (1994). 14 optional infinitives, head movement and the economy of derivations. *Verb movement*, 305.