Proceedings of the Society for Computation in Linguistics

Volume 4 Article 18

2021

Learning Morphological Productivity as Meaning-Form Mappings

Sarah R. Payne
University of Pennsylvania, paynesa@sas.upenn.edu

Jordan Kodner
Stony Brook University, jordan.kodner@stonybrook.edu

Charles Yang
University of Pennsylvania, charles.yang@ling.upenn.edu

Follow this and additional works at: https://scholarworks.umass.edu/scil

Part of the Computational Linguistics Commons, First and Second Language Acquisition Commons, and the Morphology Commons

Recommended Citation

Payne, Sarah R.; Kodner, Jordan; and Yang, Charles (2021) "Learning Morphological Productivity as Meaning-Form Mappings," *Proceedings of the Society for Computation in Linguistics*: Vol. 4 , Article 18.
DOI: https://doi.org/10.7275/rbhm-c353
Available at: https://scholarworks.umass.edu/scil/vol4/iss1/18

This Paper is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Proceedings of the Society for Computation in Linguistics by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

Learning Morphological Productivity as Meaning-Form Mappings

Sarah Payne[1] **Jordan Kodner**[2] **Charles Yang**[1]

[1] Departments of Linguistics and Computer and Information Science, University of Pennsylvania
paynesa@sas.upenn.edu, charles.yang@ling.upenn.edu
[2] Department of Linguistics and Institute for Advanced Computational Science, Stony Brook University
jordan.kodner@stonybrook.edu

Abstract

Child language acquisition is famously accurate despite the sparsity of linguistic input. In this paper, we introduce a cognitively motivated method for morphological acquisition with a special focus on verbal inflections. Using UniMorph annotations as an approximation of children's semantic representation of verbal inflection, we use the Tolerance Principle to explicitly identify the formal processes of segmentation and mutation that productively encode the semantic relations (e.g., past tense) between stems and inflected forms. Using a child-directed corpus of verbal inflection forms, our model acquires the verbal inflection morphemes of Spanish and English as a list of explicit and linguistically interpretable rules of suffixation and stem change corresponding to sets of semantic features.

1 Introduction

One of the greatest challenges a child faces as they acquire their native language is to combat the sparsity of the linguistic input. *Paradigm saturation*, the proportion of possible inflectional categories that a lemma realizes in a corpus, is a measure of morphological sparsity. An analysis of child-directed Spanish, for example, shows that each verb lemma only appears in 7.9% of the possible inflectional categories on average; even the most saturated lemma (i.e., *decir* 'to say') only appears in 46 of the 55 inflectional categories that are available in a child-directed corpus (Lignos and Yang, 2016). Similarly, some inflectional categories are used far more frequently than others: the most common, 3rd-person singular present indicative, which is often taken as the default form in linguistic analysis, appears tens of thousands times whereas some (e.g., 2nd-person plural imperfect subjunctive) appear only once, if at all (Lignos and Yang, 2016). We provide an illustration of the sparsity of our own Spanish training data (§ 3.2) in Figures 1 and 2. In order to acquire the verbal morphology of a language, it is necessary for the child to generalize well beyond the input.

The uneveness of morphological distributions entails that children learn morphology in a piecemeal fashion. Brown's (1973) study, for instance, establishes that English-learning children generally acquire the progressive *-ing* before the noun plural *-s*, followed by irregular past tense and then the regular past tense *-ed*. The order of acquisition is related to the input frequency of these morphemes as well as their regularity (Yang 2016). However, children recognize morphological regularity very early and apparently categorically: productive rules are often extended to exceptions whereas analogical errors on the basis of similarity are virtually unattested. For example, as is well known, past tense over-regularization errors (e.g., *go-goed*) are fairly common after *-ed* becomes productive, typically before age 3 (Marcus et al., 1992). But children almost never over-irregularize by analogizing an irregular pattern incorrectly (e.g., **wipe-wope* from *write-wrote*, **beep-bept* from *sleep-slept*; Xu and Pinker 1995). The regular vs. irregular distinction is also clear in Spanish acquisition. Irregularity in Spanish inflection can be seen in both the stem (e.g., diphthongization) and the suffix (e.g., *quis-e* instead of **quis-í*). In a corpus of about 5,700 verb inflections produced by young children (Clahsen et al., 2002), only 3% (168) are errors: of these only 2 are over-irregularization while all others are overregularization. Not a single misuse of diphthongization is found in a corpus of almost 2,000 tokens for which irregularization could have taken place (Mayol, 2007). The asymmetry between the generalization of productive rules and the lexicalization of non-productive forms has been observed in the cross-linguistic study of child morphology (Demuth, 2003; Deen, 2005; Clahsen et al., 1992;

Proceedings of the Society for Computation in Linguistics (SCiL) 2021, pages 177-187.
Held on-line February 14-19, 2021

Caprin and Guasti, 2009), including polysynthetic languages such as Inuktitut (Allen, 1996).

The child acquisition results may seem at odds with gradient claims of morphological productivity (Baayen, 1992; Albright and Hayes, 2003; Seidenberg and Plaut, 2014). But such gradient measures of productivity are definitional, in the form of a ratio between two quantitative measures of the corpus statistics. Furthermore, the psychological evidence adduced for gradience generally involves inherently gradient tasks such as rating which encourage gradient responses even for uncontroverisally discrete categories (e.g., even numbers; Armstrong et al. 1983). Finally, gradient responses to morphological productivity are almost always obtained from adults rather than children. In the original Wug study (Berko, 1958), nonce verbs such as *gling* often lead adults to produce analogical irregularized forms (*glang*, *glung*), which are nearly completely absent in children's responses. Still, adults rate the regularized forms above irregularized forms when both forms are available (Ambridge, 2010). It is worth noting that English has added three novel verbs, *Bing*, *bling* and *gling*, in the past few decades. These verbs belong to the category of irregular verbs most favored for analogy in (adult) experimental studies (Albright and Hayes, 2003; Bybee and Moder, 1983), but both are regular in actual language use.

Taken together, these results suggest that (a) children form morphological generalizations on a relatively small amount of data, (b) children grasp the semantic features of morphological processes as indicated by the rarity of misapplications, and (c) children – and probably adults too, when experimental confounds are eliminated – draw a near-categorical distinction between productive and unproductive processes, with the former occasionally over-used but almost never the latter. The last part, in particular, has been a challenging problem for computational morphology (Gorman et al., 2019). Such developmental findings may serve as design specifications for a psychological theory of morphological acquisition.

1.1 Related Work

Many systems for automatic morpheme segmentation have been proposed in the NLP community: see Tzoukermann and Liberman (1990); Klenk and Langer (1989); Méndez-Cruz et al. (2016) for Spanish, Cotterell et al. (2015); Sirts and Goldwater (2013); Ruokolainen et al. (2014) for English, and Hammarström and Borin (2011) for a review. Of particular relevance here are the Morpho-Challenge tasks (Kurimo et al., 2010) and resulting unsupervised morphological learning models (e.g., Creutz and Lagus 2005, 2007; Monson et al. 2007; Lignos 2010). More recently, Xu et al. (2020) provide a method of unsupervised morphological analysis exploiting a universal framework for morphological typology. The Morpho-Challenge-inspired models rely only on a set of word pairs as their input, making them better-suited for low-resource languages and more cognitively plausible than models that rely on larger or more-saturated data. However, these models focus exclusively on segmentation.

To the best of our knowledge, our model is the first to acquire morphological mappings in a manner consistent with the developmental constraints from child language acquisition reviewed earlier. While there is a large literature on the modeling of child morphological learning stemming from the so-called Past Tense Debate, these models only learn a single category (e.g., English past tense). Our model, by contrast, is designed to handle complex processes that realize multiple semantic features in a single inflected form.

1.2 Outline

In this paper, we present an unsupervised learning model that identifies form-meaning mappings in Spanish and English inflectional morphology and is intended to model the process of child language acquisition. The input to the model is a sample of words from child-directed speech in these languages (MacWhinney, 2000) comparable in size to those acquired during the first years of morphological learning. Children's accurate understanding of the semantic features of morphology is approximated by supplying the words in the input with annotations from the UniMorph project (Sylak-Glassman et al., 2015; Kirov et al., 2018).

The goal of our model is to find the morphological processes – affixation and stem changes – that provide systematic mappings between the semantic features of the stem (e.g., *walk*) and their inflection forms (e.g., *walked*). The most critical aim of our model is to acquire the distinction between productive and unproductive morphological processes, which young children almost never fail to recognize. Similarly, our model is able to learn both the productive processes, which can be applied to novel forms beyond the training data, as well as the unproductive processes (e.g., the stem change in *think-thought*, *sweep-swept*, and *sing-sang*), which will be restricted to the attested items in the training data and memorized accordingly. Productive patterns may be “broad,” widely applicable and possibly default, or “narrow” in that they supersede a broader pattern under specific conditions. In sum, our model acquires broad defaults, narrowly-conditioned productive processes, and unproductive exceptions that must be memorized.

A central component of our learning model is the Tolerance Principle (TP; Yang 2016), a simple mathematical model that specifies the threshold for (morphological) productivity, one which been extensively used in the empirical and experimental studies of child language acquisition (e.g. Schuler et al. 2016). As in our model, the TP makes a categorical distinction between productive and unproductive rules and supports a notion of broadly applicable and conditioned productive processes.

Our model is presented in § 2, with the algorithm laid out in the Appendix. In a nutshell, our model extracts the morphological processes that map the semantic features of the stem to those of the inflected form (e.g., Spanish *ama-rá-n* ‘love-FUT-3.PL’). These processes are then subjected to the TP’s productivity test. If a process is deemed productive, it will be recorded as such while lexically listing its exceptions. If a process fails the TP test, the model subdivides the words into subclasses delimited by finer-grained semantic features and applies the TP test recursively to find narrower productive rules within. The experiments reported in § 3 show that our model successfully acquires almost all of the inflectional rules of English and Spanish available in the training data along with their exceptions, thereby providing a reasonable psychological account of young children’s morphological acquisition. § 4 discusses related work and considers directions for future research.

2 Model Description

2.1 Linguistic background

In the present study, each input item is a tuple that consists of the orthographic form of a verb stem, an inflected form of the stem, and its UniMorph annotation (although other conventions adequately describing the semantic aspects of inflectional morphology could also be used). Table 1 provides some example annotations in both languages. We also assume that the model, like a child language learner, has knowledge of feature *categories*: person (1, 2, 3), number (SG, PL), tense (e.g., present PRS, past PST), aspect (e.g., participle PTCP), mood (e.g., indicative IND, subjunctive SBJV), and so on. These semantic aspects of morphology are well under control for children acquiring English and Spanish before the age of 3 (Brown, 1973; Kvaal et al., 1988). Note that although Spanish and English are typically considered to be fusional languages, Spanish inflection is also agglutinative, realizing person/number after tense/aspect (again, *ama-rá-n* ‘love-FUT-3.PL’).

Lemma	Form	Features
English		
find	find	2 SG PRS
fall	fallen	PTCP PST
call	called	3 PL PST
Spanish		
poder	podría	COND 3 SG
imaginar	imaginar	NFIN
quedar	quedaron	IND PST 3 PL PFV
mirar	mirad	POS IMP 2 PL
caer	caigo	IND PRS 1 SG

Table 1: English and Spanish UniMorph annotations

A morphological process is defined as the mapping that manipulates the stem into the inflected form. For example, an input item may be (*walk*, {*walking*, {PTCP, PRS}}), and the morphological process -ing, which must be learned, is regarded as the realization of the features {PTCP, PRS}. In some cases, the morphological process may be conditioned on the stem. For example, the morphological process defined over the pair (*sing*, {*sang*, {1, SG, PST}}) will be i → a, which realizes the feature {1, SG, PST}. Such stem-conditioned processes are all unproductive in English, as indicated by the virtual absence of over-irregularized forms in child productions (Xu and Pinker, 1995). However, some of the productive morphological processes in Spanish are conditioned on the stem (i.e., the conjugation classes). Our model uses simple string edit methods to extract these morphological processes, which roughly correspond to the familiar suffixation and stem changes in the Indo-European languages. We emphasize that the extraction of the morphological processes and the evaluation of these processes, which is the core component of our learning model, are completely independent. Should we encounter a language with other morphological properties (e.g., prefixation, harmony, reduplication), or indeed from the derivational do-main, the evaluation component of our model can apply without modification.

2.2 The Tolerance Principle

The central component of our learning model is the Tolerance Principle (TP; Yang 2016), which asserts that a rule is productive if and only if:

$$e \leq \theta_N = \frac{N}{\ln N}$$

where N is the number of words a rule can apply to and e is the number of words which do not follow the rule.[1] The value θ_N is thus the threshold for generalization. For example, say we have 100 first person singular present forms, 90 of which are realized in Spanish with the process o. Because $100/\ln 100 \approx 21.7$ which is greater than the 10 exceptions, o is deemed as the productive mapping for $\{1$, SG, PRS$\}$. It is worth stressing that the TP operates on type frequencies. The token frequency of words does not enter into the TP calculation or any other component of our learning model; its only role is that a high frequency word is more likely to be sampled in the training data, thereby contributing as a single type count.

A crucial feature of the TP is its recursive application. In a case like the English past tense, a semantic feature (PST) may be realized by a single productive process (*-ed*), as long as the number of exceptions (i.e., the irregular verbs) does not exceed the threshold. In other cases, however, a semantic feature may not be productively realized by a single process. For example, the semantic feature (PRS) is realized as *-s* for 3rd person singular (3 SG PRS), *-ing* for the participle (PRS PTCP), and null for all other feature sets that contain (PRS). None of the three options can survive as the productive process for (PRS) as the other (two) competitors would collectively exceed the threshold in any reasonable sample. Thus, the learner will subdivide the verbs into more-specific feature sets and search for productive rules within each. In this case, this leads to a set of complementary mappings at varying levels of specification: $\{$PRS$\}$ = $\emptyset$, $\{$3 SG PRS$\}$ = s, and $\{$PTCP PRS$\}$ = ing, where the first mapping may be considered a broad mapping and the second two narrow mappings. Note that, when considered in order of decreasing specificity, these mappings function identically to the disjoint mappings given by: $\{$PRS $\wedge\overline{\text{PTCP}}\wedge\overline{\text{3 SG}}\}$ = $\emptyset$, $\{$3 SG PRS$\}$ = s, and $\{$PTCP PRS$\}$ = ing. The search for productivity over a feature set terminates when the productive mapping is found.

2.3 The Search for Productive Mappings

For each instance of (lemma l, inflected form i, feature-set F) in its input data, our learner applies the TP to the mappings given by:

$$\mathcal{P}(F) \times \sigma(i)$$

where $\mathcal{P}(F)$ is the power-set of the feature-set and $\sigma(i)$ is the set of all substrings at the end of the inflected form. Here we assume $\sigma(i)$ contains non-empty substrings of at most 4 characters unless $i = l$, in which case $\sigma(i) = \{\emptyset\}$. For example, the substrings considered for (*l*=*walk*, $\{$*i*=*walk*, *F*=$\{1$, SG, PRS$\}\}$) are just $\sigma(i)$=$\{\emptyset\}$ since $l = i$, while for (*l*=*walk*, $\{$*i*=*walking*, *F*=$\{$PTCP, PRS$\}\}$), they are $\sigma(i)$=$\{$king, ing, ng, g$\}$. Because we may generate multiple possible suffixes from each inflected form, we define e – the number of exceptions in TP calculation – to be the number of strings that are not equal to, but are also not a sub- or superstring, of the suffix in question.

The model calculates the frequencies of each of the feature categories $\{$person, number, tense, mood, aspect$\}$ on its input, and iterates through them in order of decreasing frequency, following the well-attested frequency effects in language acquisition (Brown, 1973; Yang, 2016). At each pass, it attempts a mapping from the feature space to the suffix space using the Tolerance Principle, constraining the feature space to the features currently under consideration. If this mapping is empty, additional feature-categories are added in order of decreasing frequency until a mapping constrained to these categories is non-empty. That is, we apply the TP at each pass to:

$$\bigcup_{(l,i,F)\in\text{input},\ F\cap C\neq\emptyset} \mathcal{P}(F \cap C) \times \sigma(i),$$

where C is the set of feature categories under consideration at this pass, so $F \cap C$ constrains the pass to only consider the relevant features for any given item. F and $\sigma(i)$ are defined as above.

Because our learner considers all feature combinations in the power-set of the relevant features for a given word, it can yield multiple mappings

[1] θ_N is almost always a small fraction of N unless N is very small: for $N = 5$, one rule covering 2 items and the other rule covering 3 are *both* productive. We are not aware of any empirical studies attesting to this effect but will assume that in such cases, the more dominant rule is the productive one.

corresponding to the same morphological process. For example, in Spanish, {1, PL} maps to mos, but so does {IND, PRS, 1, PL }, {PRS, 1, PL}, {POS, IMP, 1, PL} and others. To remedy this, once our learner has a non-empty constrained mapping at a given pass, it finds the intersection of all feature-sets that have mapped to each morphological process. If the mapping from this intersection to the morphological process in question passes the tolerance threshold, our learner is finished for this mapping. Otherwise, features are added by frequency to increase specification until the mapping either runs out of features or passes the tolerance threshold. This allows our model to acquire the minimal set of features that productively maps to a given morphological process.

Feature categories are removed from our learner's consideration once they have been used to constrain a mapping in a pass. After a pass, all endings that were learned in that pass are removed indiscriminately from the inflected forms in the input. For example, most Spanish {3 PL} forms end in n, so once this ending has been acquired, every inflected form in our data ending with n has this ending removed. We do not check the feature-sets corresponding to the inflected form before removal, which is motivated by children's aggressive segmentation "errors" once they identify productive processes (e.g., removing *-s* from the preposition *versus* to create a verb *vers*; Yang 2016).

2.4 Narrow Mappings and Stem Alternations

Narrow mappings refer to productive processes that reside within exceptions to the more-general mappings learned above. For example, the narrow mapping {3, SG, PFV} = o is a productive pattern that supersedes the more general mapping {3, SG} = ∅ in Spanish, and {PTCP, PRS} = ing is a narrow mapping that supersedes {PRS} = ∅ in English.

To identify potential narrow mappings, at each pass, the learner counts cases where the morphological processes proposed so far do not yield the correct inflected form. Then for each feature-set where there was at least one failure, it checks to see if a TP-majority failed. It then checks if it can find some suffix that correctly inflects enough instances that meet the feature-specification to pass the tolerance threshold. If it can, it takes the corresponding morphological process to be a narrow mapping.

In cases where there is no single process that corresponds to the narrow mapping, our learner attempts instead to categorize the suffixes based on endings of the lemma. We subdivide the instances based on properties of the ending of the lemma and search for productivity recursively until we have a set of morphological processes that productively describe our data or we only have one lemma fitting a condition. For example, the Spanish imperfect suffixes are conditioned on the stem vowel, *-aba* for *-ar* verbs and *-ía* for *-er* and *-ir* verbs. A morphological process that treats ia or aba as the default and learns the rest as exceptions will not pass the tolerance threshold because the opposing stem classes are always too large, but if the data is subset according to the stem endings, it will pass:

$$\texttt{a} \rightarrow \texttt{aba}$$
$$\texttt{i}, \texttt{e} \rightarrow \texttt{ia}$$

If our learner can acquire a set of suffixes that are applied productively based on the ending of the lemma, then it takes the most frequent one to be the default. This yields a morphological process like the one for the Spanish imperfect (Table 3).

Features in the current feature category with no mapping in the current pass are treated in much the same way: the learner finds all mappings in an unconstrained pass for which the feature-set includes the feature(s) in question. The mapping with highest count is added to the set of general mappings for the pass, and the others are considered to be possible narrow mappings. For example, our model learns {1, SG, PRS} = o in the first pass in Spanish, since this is the highest-frequency mapping containing {1, SG}. Other mappings from a super-set of {1, SG} are stored as possible narrow mappings.

All narrow mappings are stored separately and verified at the end of learning. Additionally, the suffixes corresponding to these narrow mappings are not removed after each pass. This filters out those that turn out to be outcomes of agglutinative processes our model has yet to learn, and allows the model to acquire these processes.

2.5 Memorizing Irregulars

As well as acquiring productive morphological processes, our learner must also acquire verbs for which there is no productive morphological process determining their inflection. These include suppletive verbs in Spanish and English (e.g., *ir* ∼ *fui* and *go* ∼ *went*), as well as plausibly but not ultimately predictable stem-changing verbs (e.g., *pedir* ∼ *pide* and *sleep* ∼ *slept*). Such patterns are simply memorized lexically and do not generalize, which is consistent with children's behavior concerning these processes (Clahsen et al., 1992; Xu and Pinker, 1995).

To accomplish this, our learner attempts to inflect every lemma in our training data based on the morphological processes it has acquired after the main iteration of learning is complete. It checks among the failing feature-sets for those where a TP-majority are failing, and if it has memorized a possible narrow mapping for such a feature-set, this mapping is added to the set of verified morphological processes.

Finally, any remaining forms not covered under one of the productive rules (including the recently verified narrow mappings) are committed to memory. This completes a representational system that contrasts regular forms, which can be inflected via productive processes, with memorized irregular ones that cannot.

3 Experiments

For both Spanish and English, our data consisted of the most frequently-occurring inflected verb forms in child-directed speech taken from the CHILDES corpora (MacWhinney, 2000). Because semantic context is not available in these wordlists, we annotated the words with lemmatizations and features provided in UniMorph 2.0, a Wiktionary-derived list of feature-sets for every lemma-inflected pairing (Kirov et al., 2018). These features provide a rough approximation of children's accurate morphosemantic knowledge (children rarely use morphological forms in inappropriate semantic contexts; §1). Following UniMorph's annotation scheme, we included English participles but not Spanish participles. For a fixed training data set, which corresponds to a child's internal vocabulary, the learning model will by design deterministically produce a set of output rules.[2] New training data, as long as they are sampled from the high frequency range of child-directed input, which would correspond to the variation across individual children's input data, generally do not lead to significant variation in the output rules. We regard this as an attractive feature as it would account for the general uniformity in the terminal grammars across individual child learners.

Broad Mappings				
Features	**Defau.**	**Alternations**	**Ct.**	**Ex.**
First Pass				
PRS	∅		2573	*walk*
Second Pass				
3	∅		1717	*walk*
2	∅		571	*walk*
1	∅		554	*walk*
Third Pass				
PL	∅		1454	*walk*
SG	∅		1422	*walk*
Fourth Pass				
NFIN	∅		22	*walk*

Narrow Mappings				
Features	**Defau.**	**Alternations**	**Ct.**	**Ex.**
First Pass				
PTCP, PRS	`ing`	`e → ing`	643	*pleasing*
3 SG PRS	`s`		372	*walks*
Second Pass				
3 PL PST	`ed`	`y→ied,e→ed`	367	*pleased*
3 SG PST	`ed`	`y→ied`	139	*tried*
2 SG PST	`ed`	`y→ied,e→ed`	203	*walked*
1 SG PST	`ed`	`y→ied,d→t`	136	*built*
1 PL PST	`ed`	`y→ied`	67	*cried*

Table 2: Learned mappings for English

3.1 English

For English, we combined CHILDES data from the Manchester, Wells, and Belfast corpora, which together contained 3953 unique inflected forms corresponding to 1285 verb lemmas. We considered the lemma to be the plain infinitive, and removed (cliticized) contractions. We also removed gemination (e.g., converting *put* ∼ *putting* to *put* ∼ *puting*) because this orthographic pattern has no (synchronic) morpho-phonological relevance and is also subject to spelling variation (e.g., *traveling/travelling*). The English dataset is larger than Spanish because irregular verbs in English are notoriously frequent: the *-ed* rule would not emerge as productive if the vocabulary size was too small. The irregulars in Spanish, by contrast, show no obvious frequency effects (Fratini et al., 2014), so a more modest vocabulary is sufficient to identify the major productive processes. This mirrors the considerably earlier acquisition of Spanish inflectional morphology (González, 1978).

Results for English can be seen in Table 2. Each row in the table lists a set of features, the form they are mapped to, any active alternations, the number of times that set is attested in the data, and an example form. The result of each individual pass is indicated and broad and narrow mappings are listed separately. We acquire {PRS} = ∅ in the first pass because, as discussed in § 2.2, we only

[2] The algorithm is actually quite fast. Loading the data for both languages, running the models, and outputting the results in PyCharm takes 6.6 seconds on the first author's consumer-grade laptop.

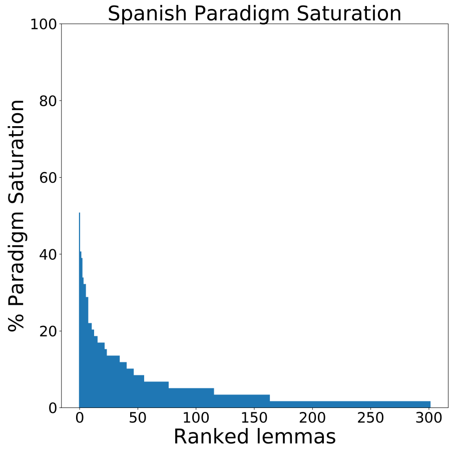

Figure 1: Spanish lemmas by proportion of attested inflectional categories

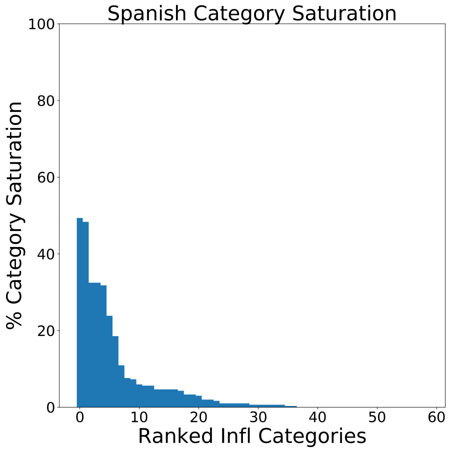

Figure 2: Spanish inflectional categories by proportion of attested lemmas

consider a suffix to count towards e in the TP if it is not a sub- or super- string of another suffix. Since $\emptyset$ is a sub-string of both `s` and `ing`, {PRS} = $\emptyset$ passes the TP. However, when we check for narrow mappings (§ 2.3), we find that both {PTCP, PRS} and {3, SG, PRS} fail to be correctly inflected by $\emptyset$. These feature sets are subsequently learned as productive narrow mappings, shown in Table 2.

For English, the alternations in the narrow mappings are almost entirely orthographic in nature. For example, orthographic `e` must be removed when forming the `ing` participle, and `y` must be converted to `i` before `ed`. Ultimately, there are 765 items that follow spelling alternations which are not sufficiently frequent in the data and thus must be memorized as exceptions (e.g., *flies* and *lying*), or are simply irregular (e.g. *went*). These 765 items correspond to 461 lemmas; identical inflected forms realizing different features are counted as separate exceptions since they are presented separately in the training data (e.g. *went* is counted twice if it appears in both {3, SG, PST} and {2, SG, PST}). Obviously, these orthographic issues do not arise for children acquiring morphology from speech, and we discuss extensions of our model in § 4. These items are stored; the associated patterns are deemed unproductive and do not generalize.

3.2 Spanish

For Spanish, we took the 1,000 most-used inflected forms from the CHILDES FernAguado Corpus and removed cliticized reflexives, resulting in 989 inflected forms corresponding to 302 verb lemmas. We removed orthographic accents from Spanish verbs for ease of processing and considered the Spanish lemma to be the infinitive form with the `r` removed. This yielded an extremely sparse dataset as expected from Lignos and Yang (2016). The 302 lemmas exhibit a highly skewed distribution with most verbs appearing in only one or two inflected categories and none achieving even 60% saturation (Figure 1). Additionally, no inflectional category is attested with more than half of all known verbs, and only 37 are attested at all (Figure 2). With no direct evidence at all in their favor, the remaining endings must be inferred from the rest of the data.

Results for Spanish, sorted by pass and mapping type, can be seen in Table 3. Notable here is that the alternations capture differences between conjugations, for example the imperfect for both *-er/-ir* and *-ar* verbs in the fourth pass, and several subjunctives for *-ar* and *-ir* verbs in the narrow mappings. Additionally, the learner successfully acquired mappings that are quite rare, for example the 2nd person plural imperative `-d` which only appears twice in the data in *sed* and *mirad*. As well as these productive results, the learner also acquired 266 memorized exceptions corresponding to 78 lemmas in Spanish, including *ser* and *haber*. This means that just over a quarter of the 302 verb lemmas observed in Spanish were inflected irregularly at least once, contrasting with over a third of the 1285 English lemmas and matching well with the findings of Fratini et al. (2014) discussed above.

In the first pass, we see in Table 3 that we learn the common person+number endings in Spanish.

Broad Mappings				
Features	**Default**	**Alterns.**	**Ct.**	**Ex.**
First Pass				
3 SG	∅		227	*ama*
3 PL	n		103	*aman*
1 PL	mos		51	*amamos*
2 PL	is		10	*amais*
PRS 1 SG	o		163	*amo*
PRS 2 SG	s		129	*amas*
Second Pass				
IND	∅		651	*ama*
IMP	∅		127	*ama*
NFIN	r		146	*amar*
COND	ria		16	*amaria*
Third Pass				
PRS	∅		492	*ama*
FUT	ra		20	*amara*
Fourth Pass				
IPFV	ia	a→aba	65	*amaba*

Narrow Mappings				
Features	**Default**	**Alterns.**	**Ct.**	**Ex.**
First Pass				
SBJV PRS 3 SG	e	i → a	13	*ame*
POS IMP 3 SG	e	i → a	14	*ame*
IND PST 3 SG PFV	o		72	*amo*
SBJV PRS 3 PL	an		2	*coman*
POS IMP 3 PL	an		2	*coman*
IND PST 3 PL PFV	ron		23	*amaron*
POS IMP 1 PL	emos		3	*amemos*
SBJV PRS 1 PL	emos		3	*amemos*
POS IMP 2 PL	d		2	*amad*
SBJV PRS 1 SG	e	i → a	14	*ame*
IND PST 1 SG PFV	e	i → i	18	*ame*
COND 2 SG	rias		2	*amarias*
SBJV PRS 2 SG	es	i → as	33	*ames*
IND FUT 2 SG	ras		3	*amaras*
IND PST 2 SG IPFV	ias		9	*comias*
IND PST 2 SG PFV	ste		10	*amaste*
Second Pass				
IND FUT 1 PL	re		2	*amaremos*

Table 3: Learned mappings for Spanish

For {1 SG} and {2 SG}, we have too many exceptions (namely the 1st person in every non-present tense and the second singular imperative) to acquire generalizable rules based on only person and number, so we follow § 2.3 and take the more-specified rule with highest count. After acquiring these endings, we check for possible sub-regularities. Those that were successfully verified at the end of learning are listed in Table 3, but several others are hypothesized, since the person+number endings learned in the first pass do not fully inflect Spanish verbs in non-present tenses, and the agglutinative endings indicating these tenses are not learned until at least the second pass. Such hypothesized and later-rejected narrow mappings include, for example, {3, PL, FUT} = ran.

4 Discussion

Our model can be seen as an operationalization of the Tolerance Principle which has been used extensively, by the means of manual calculation using corpus statistics, to model morphological acquisition in many languages. It provides a mechanistic account of the developmental findings (§1) that children make a near-categorical distinction between unproductive and productive processes (Tables 2 and 3). Further, since our learner explicitly memorizes the irregular forms to which unproductive processes apply (§ 2.4), it, like children, will not generalize these processes beyond its input (Clahsen et al., 2002; Mayol, 2007).

The iterative processing of feature-categories in a descending order of frequency enables the model to acquire the productive morphological processes of both Spanish and English in approximately the same order as a child. For example, Aguirre and Marrero (2009) show that Spanish-learning children acquire person+number endings early in the acquisition process, and our learner similarly acquires these in the first pass of its learning (Table 3). Likewise, Brown (1973) shows that English-learning children acquire the present participle before the past tense, which is less regular; our model learns the narrow mapping for {PTCP, PRS} in the first pass of learning and the narrow mappings for PST in the second.

4.1 Limitations

Though the learner successfully acquires conjugation-dependent endings such as the Spanish imperfect and most of the present subjunctives, there are a few generalizations that it misses. For example, the imperfect subjunctives (in *-ra-* or *-se-*) are missing simply because they are not attested in the data; these are among the least common forms in Spanish. Indeed, we are not aware of any study of child Spanish that documents such processes. More subtly, there are some instances where *-er* forms are not accounted for, such as the {SBJV PRS 3 SG}, where e→a is not acquired. This likely results from the training data, as many of the *-er* verbs in our data are irregular (e.g. *ser, haber, tener*), which prevents our learner from forming generalizations for such morphological processes. However, this is in line with the relatively late acquisition of the subjunctive by Spanish-learning children (González, 1978), implying a larger dataset may be necessary to properly acquire the subjunctive. A similar case happens for the {FUT 1 SG}, which does not appear in our data and thus is not attested as a narrow mapping. However, we do learn the attested mapping for {FUT 1 PL}; for this case, note that since re was learned as a second pass narrow mapping, we can agglutinate it with first pass mappings to create, in this case, remos.

The limitations of the Spanish data can also be seen in the distribution of the mappings for the 2nd person singular. We only learn s = {2 SG PRS} rather than {2 SG}. There are just 9 instances of {2 SG IPFV}, 3 {2 SG FUT}, and many imperatives (which do not take -s): 101 of our 254 {2 SG} forms are {POS IMP 2 SG}. This means that the s ending cannot generalize to {2 SG} and we must learn separate mappings, in this case, for the present, future, imperfect, and conditional {2 SG}.

The relatively small size of our Spanish data may be part of the reason that we fail to learn all cases for some of our narrow mappings. Ultimately, a frequency-based sample from child-directed corpora is only a crude proxy: accurate modeling of morphological acquisition must approximate the child's learners of (psychological) vocabulary as closely as possible.

4.2 Future directions

The linguistically interpretable rules which our model learns can be used for morphological segmentation, analysis, and generation. It can be extended to languages with other types of morphological processes, leaving the productivity evaluation component unchanged. Representing word forms as sequences of phonological feature bundles can remove the artificiality of orthography, potentially leading to the discovery of phonologically conditioned morphological rules.

The development of grammar, including morphology, often follows a a U-shaped trajectory marked by near-perfect production, a period of overgeneralization, and a return to near-perfect production (see e.g. Marcus et al. 1992). This has been interpreted as the calibration between rules and exceptions: children gradually accumulate vocabulary items that follow productive rules, which eventually overwhelm the exceptions that do not follow the rules. Our model can straightforwardly account for such developmental process by providing the learner with an incrementally larger amount of data (i.e., more lemma/inflection pairs). The tipping point at which rules become productive would correspond to the dip in the U-shaped learning curve.

Derivational morphology poses interesting challenges for our model. It remains to be seen if simple semantic features such as category transformation are sufficient to enable successful acquisition (e.g., the German DErivBase; Zeller et al. 2013) or finer-grained semantic distinctions need to be made: the English deverbal suffix *-er* can realize the agent (e.g., *kicker*) or the instrument (e.g., *cutter*).

Our model requires explicit representation of semantic features in order to identify morphological processes. Some of these features are likely universal while others may need to be constructed on a language-specific basis. An intriguing direction is to explore the extent to which distributional methods can induce semantic relations that hold for morphologically related words (e.g., Luong et al. 2013), as opposed to relying on annotated data such as UniMorph. The induction of semantic relations needn't be comprehensive as long as it is highly accurate: as our model demonstrates, a very modest amount of high quality data may be sufficient for linguistically accurate morphological learning.

Acknowledgements

We are grateful to Bob Berwick and his lab, Spencer Caplan, Kyle Gorman, Mitch Marcus, Hongzhi Xu, and the anonymous SCiL reviewers for their feedback; this work benefitted greatly from their input.

References

Carmen Aguirre and Victoria Marrero. 2009. Number morphology in spanish first language acquisition. *Development of nominal inflection in first language acquisition: A cross-linguistic perspective*, pages 341–370.

A. Albright and B. Hayes. 2003. Rules vs. analogy in english past tenses: a computational/experimental study. *Cognition*, 90:119–161.

Shanley Allen. 1996. *Aspects of argument structure acquisition in Inuktitut*. John Benjamins Publishing, Amsterdam.

Ben Ambridge. 2010. Children's judgements of regular and irregular novel past-tense forms: New data on the English past-tense debate. *Developmental Psychology*, 46(6):1497–1504.

Sharon L. Armstrong, Lila R. Gleitman, and Henry Gleitman. 1983. What some concepts might not be. *Cognition*, 13(3):263–308.

Harald Baayen. 1992. Quantitative aspects of morphological productivity. In *Yearbook of morphology 1991*, pages 109–149. Springer.

Jean Berko. 1958. The child's learning of English morphology. *Word*, 14(2–3):150–177.

Roger Brown. 1973. *A first language: The early stages.* Harvard U. Press.

Joan Bybee and Carol L. Moder. 1983. Morphological classes as natural categories. *Language*, 59(2):251–270.

Claudia Caprin and Maria Teresa Guasti. 2009. The acquisition of morphosyntax in Italian: A cross-sectional study. *Applied Psycholinguistics*, 30(1):23–52.

Harald Clahsen, Fraibet Aveledo, and Iggy Roac. 2002. The development of regular and irregular verb inflection in spanish child language. *Journal of Child Language*, 29(3):591–622.

Harald Clahsen, Monika Rothweiler, Andreas Woest, and Gary Marcus. 1992. Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, 45:225–255.

Ryan Cotterell, Thomas Müller, Alexander Fraser, and Hinrich Schütze. 2015. Labeled morphological segmentation with semi-markov models. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 164–174.

Mathias Creutz and Krista Lagus. 2005. *Unsupervised morpheme segmentation and morphology induction from text corpora using Morfessor 1.0*. Helsinki University of Technology Helsinki.

Mathias Creutz and Krista Lagus. 2007. Unsupervised models for morpheme segmentation and morphology learning. *ACM Transactions on Speech and Language Processing (TSLP)*, 4(1):1–34.

Kamil Ud Deen. 2005. *The acquisition of Swahili*. John Benjamins Publishing, Amsterdam.

Katherine Demuth. 2003. The acquisition of Bantu languages. In *The Bantu languages*, pages 209–222. Curzon Press Surrey,, United Kingdom.

Viviana Fratini, Joana Acha, and Itziar Laka. 2014. Frequency and morphological irregularity are independent variables. evidence from a corpus study of spanish verbs. *Corpus Linguistics and Linguistic Theory*, 10(2):289 – 314.

Gustavo González. 1978. *The acquisition of Spanish grammar by native Spanish speaking children*. National Clearinghouse for Bilingual Education.

Kyle Gorman, Arya D McCarthy, Ryan Cotterell, Ekaterina Vylomova, Miikka Silfverberg, and Magdalena Markowska. 2019. Weird inflects but ok: Making sense of morphological generation errors. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 140–151.

Harald Hammarström and Lars Borin. 2011. Unsupervised learning of morphology. *Computational Linguistics*, 37(2):309–350.

Christo Kirov, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sebastian J Mielke, Arya D McCarthy, Sandra Kübler, et al. 2018. Unimorph 2.0: Universal morphology. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Ursula Klenk and Hagen Langer. 1989. Morphological Segmentation Without a Lexicon. *Literary and Linguistic Computing*, 4(4):247–253.

Mikko Kurimo, Sami Virpioja, Ville Turunen, and Krista Lagus. 2010. Morpho challenge 2005-2010: Evaluations and results. In *Proceedings of the 11th Meeting of the ACL Special Interest Group on Computational Morphology and Phonology*, pages 87–95.

Joy T Kvaal, Nancy Shipstead-Cox, Susan G Nevitt, Barbara W Hodson, and Patricia B Launer. 1988. The acquisition of 10 spanish morphemes by spanish speaking children. *Language, Speech, and Hearing Services in Schools*, 19(4):384–394.

Constantine Lignos. 2010. Learning from unseen data. In *Proceedings of the Morpho Challenge 2010 Workshop*, pages 35–38.

Constantine Lignos and Charles Yang. 2016. Morphology and language acquisition. *Hippisley, Andrew R. abd Stump, G., editor, The Cambridge handbook of Morphology*, pages 765–791.

Minh-Thang Luong, Richard Socher, and Christopher D Manning. 2013. Better word representations with recursive neural networks for morphology. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 104–113.

B. MacWhinney. 2000. The childes project: Tools for analyzing talk. *MIT Press.*

Gary Marcus, Steven Pinker, Michael T. Ullman, Michelle Hollander, John Rosen, and Fei Xu. 1992. *Overregularization in language acquisition*. Monographs of the Society for Research in Child Development. University of Chicago Press, Chicago.

Laia Mayol. 2007. Acquisition of irregular patterns in Spanish verbal morphology. In *Proceedings of the twelfth ESSLLI Student Session*, pages 1–11, Dublin.

Carlos-Francisco Méndez-Cruz, Alfonso Medina-Urrea, and Gerardo Sierra. 2016. Unsupervised morphological segmentation based on affixality measurements. *Pattern Recognition Letters*, 84:127–133.

Christian Monson, Jaime Carbonell, Alon Lavie, and Lori Levin. 2007. Paramor: Finding paradigms across morphology. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 900–907. Springer.

Teemu Ruokolainen, Oskar Kohonen, Sami Virpioja, and Mikko Kurimo. 2014. Painless semi-supervised morphological segmentation using conditional random fields. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 84–89.

Kathryn Schuler, Charles Yang, and Elissa Newport. 2016. Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient to do so. In *The 38th Cognitive Society Annual Meeting*, Philadelphia, PA.

Mark S. Seidenberg and D. Plaut. 2014. Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive science*, 38 6:1190–228.

Kairit Sirts and Sharon Goldwater. 2013. Minimally-supervised morphological segmentation using adaptor grammars. *Transactions of the Association for Computational Linguistics*, 1:255–266.

John Sylak-Glassman, Christo Kirov, David Yarowsky, and Roger Que. 2015. A language-independent feature schema for inflectional morphology. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 674–680, Beijing, China. Association for Computational Linguistics.

Evelyne Tzoukermann and Mark Liberman. 1990. A finite-state morphological processor for Spanish. In *COLING 1990 Volume 3: Papers presented to the 13th International Conference on Computational Linguistics*.

Fei Xu and Steven Pinker. 1995. Weird past tense forms. *Journal of Child Language*, 22(3):531–556.

Hongzhi Xu, Jordan Kodner, Mitch Marcus, and Charles Yang. 2020. Modeling morphological typology for unsupervised learning of language morphology. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6672–6681.

Charles Yang. 2016. *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

Britta Zeller, Jan Šnajder, and Sebastian Padó. 2013. Derivbase: Inducing and evaluating a derivational morphology resource for german. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1201–1211.

A Appendix

Algorithm 1 Overview of our Learning Algorithm

```
$input = [(l_1, i_1, F_1), ..., (l_n, i_n, F_n)]$
for feature category $C$ in order of decreasing frequency do
    $S = \bigcup_{(l,i,F)\in input,\ F\cap C\neq\emptyset} \mathcal{P}(F \cap C) \times \sigma(i)$
    $mappings = TP(S)$
    if $mappings = \emptyset$ then
        for feature category $C_2$ in order of decreasing frequency do
            $mappings =$ TP constraining to $C \cup C_2$
            if $mappings \neq \emptyset$ then
                break
            end if
        end for
    end if
    for any feature-set $c \in C$ without a mapping do
        Do an unconstrained TP pass
        Add superset of $c$ with highest count to $mappings$
        Add all other supersets to possible narrow mappings
    end for
    for $m \in mappings$ do
        for $(l, i, F) \in input$, $m \in F$ do
            if $N/\ln N$ or more instances of $F$ fail to be inflected by $m$ then
                $n_m =$ new TP mapping constraining to $F$
                if $n_m$ correctly inflects $F$ then
                    Add $n_m$ to $mappings$
                else
                    $sub = \{(l_j, i_j, F_j) \in input, F_j = F\}$
                    Divide $sub$ based on endings of each $l_j$
                    while $n_m$ does not correctly inflect $F$ and $\forall s \in sub, |s| > 1$ do
                        Increase specificity of endings in $sub$
                        $n_m = \cup_{s\in sub} TP(s)$
                    end while
                    if $n_m$ correctly inflects $F$ then
                        Add $n_m$ to $mappings$
                    end if
                end if
            end if
        end for
    end for
end for
for $(l, i, F)$ that can't be inflected with learned mappings from passes do
    if $\exists$ narrow mapping $n$ that inflects $N/\ln N$ instances of $F$ then
        Verify $n$ and memorize any exceptions to it
    else
        Memorize$(l, i, F)$ as an exception
    end if
end for
```