WHEN COLLISIONS ARE A GOOD THING:

THE ACQUISITION OF MORPHOLOGICAL MARKING

Sarah R. B. Payne

AN HONORS THESIS

in

Linguistics

Presented to the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the Degree of

Bachelor of Arts in Linguistics with Honors

2022

Primary Advisor:

Dr. Charles Yang

University of Pennsylvania Department of Linguistics

Acknowledgments

It truly takes a village, and my village has been exceptional.

Charles Yang taught me everything I know about language acquisition and conducting linguistic research. It is impossible to fully articulate the influence he has had on my development as a scientist, but I am extremely grateful for his insightful feedback and endless patience over the last four years. Charles believed in me even when I (quite often) didn't believe in myself. This thesis wouldn't exist, and I certainly wouldn't be where I am today, without him.

Jordan Kodner has been a valuable resource for all things morphology; the discussions of Spanish and Hebrew morphology in this thesis would be far worse without his input. Julie Anne Legate was a valuable source of insight and critique throughout this project, which originally developed in her course; I am indebted to her for suggesting the method I put forth here. David Embick taught me everything I know about morphological theory, and never failed to patiently answer my questions.

In linguistics, I am also grateful to have been taught by Jianjing Kuang, Meredith Tamminga, Katie Schuler, Rolf Noyer, and Beatrice Santorini. Florian Schwarz and Amy Forsyth took care of an embarrassing number of logistical concerns for me, and Andrea Beltrama kept me sane and caffeinated; neither of these is an easy task. In psychology, Dan Swingley helped start my journey into language acquisition, and John Trueswell and his lab provided helpful feedback. In computer science, Chris Callison-Burch was my first research mentor and spent countless hours helping me develop as a scientist, even when that development ultimately led me elsewhere. Mitch Marcus provided valuable mentorship, wise advice, and a wealth of anecdotes, and Val Tannen motivated me to stick with Penn during a particularly difficult stretch. Andreas Haeberlen, Thomas Farmer, and Boon Thau Loo made me a better teacher and computer scientist.

Outside of Penn, Roger Levy and Yevgeni Berzak taught me how to run experiments, Marjorie Freedman, Ryan Gabbard, and Deniz Beser introduced me to the world of DARPA research, and Naomi Feldman never hesitated to ask me a challenging question. Bob Berwick and his lab helped me think more critically about the relationship between my work, linguistic theory, and the mind and brain. Caleb Belth brought a new perspective to our collaborations and was a valuable resource for all things computational.

Before Penn, Susie Shelton sparked my fascination with language when she introduced me to the world of Latin morphology eight years ago. In 2017, Sandra Kübler introduced me to computational linguistics, and Emad Mohamed taught me how to code. I am extremely grateful to Sandra, Emad, and Indiana University Linguistics for taking a chance on me as a high school student; I hope they feel it paid off.

My friends have supported me every step of the way. Much of this work was done during boba-and-falafel-fueled study sessions with Emily Hong, Yara Salim, and Anjali Gupta. My linguistics friends have been valuable for both research discussions and commiseration: Annika Heuser, Stefan Pophristic, Kevin Liang, Christine Soh, and Karen Li. My Gregory friends have been with me since freshman year: Raymond Yang, Kay Mayle, Alex Yang, Katherine Hann, and Mariah Van Sciver. Here I must also say a word about a canine friend:

Zenith Mayle was the subject of the first independent research I took on, and I am forever indebted to Kay and Zenith for starting me on this journey. Marynancy Mwakalindile, Dana Reckard, and Ashton Courtney supported me all the way from Indiana — thank you!

And finally, my family. My grandfathers, Larry Payne and Geoff Brogden, have been amongst my biggest sources of inspiration and support. My parents, Chris and Nicola Payne, have never failed to encourage my love of learning (or to remind me to relax sometimes). To them, I am most grateful.

Abstract

Much work on morphological acquisition has examined how children learn to map a morpheme's phonological form to its corresponding meaning. Most accounts of this process, however, take morphological features to be prerequisites to the mapping process, and little is known about how children learn which of these features are marked in their language to begin with. In this thesis, I propose that children make use of their early ability to segment and relate inflected forms in their input to detect ***collisions*** – a single root appearing in multiple inflected forms. I suggest that these collisions provide a crucial cue to the acquisition of morphological marking, and present a model that uses collisions, in conjunction with the Principle of Contrast (Clark and MacWhinney 1987), as a cue to learn which features distinguish pairs of inflected forms. This model generalizes from individual collisions to features that are obligatorily marked across the paradigm by using the Tolerance Sufficiency Principle (Yang 2016) and a recursive search procedure. I show that the results of this approach match well with developmental findings regarding vocabulary size and order of acquisition for the well-known problems of the English past tense and German noun plurals, as well as agglutinative Spanish verbs and templatic Hebrew verbs.

Table of Contents

List of Figures

1. Introduction

Children are able to acquire the inflectional morphology of their native language at a young age despite a number of challenges to the acquisition process. A central challenge is the sparsity of the linguistic input. While children may certainly hear a single root in multiple inflected forms, **paradigm saturation**, or the proportion of possible inflected forms that a single root appears in, follows a roughly Zipfian distribution in child-directed speech (Zipf 1936; Lignos and Yang 2016). This means that while a few words will appear in nearly all of their inflected forms, most will appear in only one or two (Lignos and Yang 2016). In a sample of child-directed Spanish, for example, the form with the highest paradigm saturation appears in 83% of its possible inflected forms, while the average verb appears in a mere 7.9%, or approximately 4 out of the 55 possible inflected forms (Lignos and Yang 2016). Yet despite this sparsity, children learn the inflectional morphology of their native language when their total vocabulary is no more than 1000 words cross-linguistically; most morphological paradigms are successfully learnt by the time the vocabulary is a mere 500 words (Fenson et al. 1994; Hart and Risley 1995; Bornstein et al. 2004; Szagun et al. 2006). Further, this number represents the child's entire vocabulary: Bornstein et al. 2004 find that about 56% of the child's early vocabulary is nouns and about 25% is verbs, meaning that children acquire the paradigms that we will consider here on 300 words or less (§3).

Due to the sparsity of the linguistic input, we cannot expect that every form following a certain rule will be attested following that rule in the input. For example, a child may know *paint* but never hear *painted*, and thus have no way of knowing whether this form follows [+PST] → *-ed* or whether it is an exception. Indeed, a closely-related problem is the existence of exceptions to the productive morphological processes the child will acquire; this problem has received particular attention in the modeling literature, primarily sparked by the Past Tense Debate (see Pinker and Ullman 2002; McClelland and Patterson 2002 for review). Yang (2016) explicitly brought together the problems of exceptionality and limited positive evidence under **Tolerance Sufficiency Principle**. The Tolerance Sufficiency Principle provides a computational threshold for the generalization of a rule beyond the input, based on the principle that children will generalize rules when it is more efficient to do so (Yang 2016). It applies the same computational threshold to handle attested exceptions to a productive rule (tolerance) and to measure how much positive evidence is sufficient to generalize a rule when some forms are unattested and thus could be exceptions (sufficiency).

The Tolerance Sufficiency Principle provides a measure of whether or not a hypothesized rule should be generalized beyond the input, but the child must still exploit the necessary information in their linguistic input in order to develop hypotheses about the generalizations that are made in their native language. In this thesis, we demonstrate that the segmentation-based cue of **collisions** – a single root appearing in multiple inflected forms – provide one such piece of information that, due to the early segmentation of inflectional morphology (Santelmann and Jusczyk 1998; Soderstrom, Wexler, and Jusczyk 2002; Mintz 2013; Kim 2015; Kim and Sundara 2021), the child may exploit during acquisition in order to learn which morphological features are marked in their native language. The **Principle of Contrast** (Clark and MacWhinney 1987), or the hypothesis that two phonologically distinct forms will have distinct meanings, plays an important role in this process: when the child hears and relates distinct inflected forms, they must hypothesize that these forms also differ in meaning if they are to use collisions as a cue to learn which features their language marks. Due to the Zipfian distribution of morphological paradigm saturation in the child's input, however, we cannot expect that every word the child knows will yield all of its possible collisions in the input. Thus, we use the Tolerance Sufficiency Principle as a measure of whether ***enough*** collisions of a certain type are attested in the input for the child to generalize. That is, we determine which morphological features correspond to each collision separately and then form generalizations over the feature sets if there is sufficient evidence to do so.

Our model is motivated by the challenges of the sparsity of the input, limited positive evidence, and exceptionality. However, there are several more challenges to morphological acquisition to be discussed. Firstly, the mapping between meaning and form in inflectional morphology is not always one-to-one (see e.g. Embick 2015 for an in-depth discussion of allomorphy and syncretism within the Distributed Morphology framework, or Baayen et al. 2011 for cross-linguistic examples of both). In particular, allomorphy (multiple phonological forms mapping to one morphological feature set) yields the additional task of learning not only which morphological features are marked, but also how these features are phonologically realized and the conditions that determine their realization. For example, in the case of German noun plurals, each suffix is productive on some subset of nouns conditioned by gender and phonology (Wiese 2000; Kauschke, Kurth, and Domahs 2011), but all realize the feature +PL. Several models of the acquisition of German noun plural allomorphy exist in the literature (e.g. Kirov and Cotterell 2018; McCurdy, Lopez, and Goldwater 2020; McCurdy, Goldwater, and Lopez 2020; Belth et al. 2021; Dankers et al. 2021) but for the purposes of learning what morphological features are marked in our model, allomorphy is not a part of the equation. Since our model determines what morphological features are realized over each collision separately and then generalizes across these features, all that matters for us is that the plural form is distinct from the singular form, not ***how*** it is distinct.

In contrast to allomorphy, syncretism (the same phonological form mapping to multiple morphological feature sets) is directly relevant to acquiring which morphological features are marked in a language. Consider, for example, English present tense verbal morphology, which can be represented roughly as follows:

$$[+\text{PRS},\ +3,\ +\text{SG}] \rightarrow \text{-}s$$

$$[+\text{PRS}] \rightarrow \emptyset$$

Although it would technically be equally accurate to represent the same system with rules of the form:

$$[+\text{PRS},\ +3,\ +\text{SG}] \rightarrow \text{-}s$$

$$[+\text{PRS},\ +3,\ \text{-SG}] \rightarrow \emptyset$$

$$[+\text{PRS},\ +1,\ +\text{SG}] \rightarrow \emptyset$$

...

the latter representation would generally be seen as missing a crucial generalization: that the third singular present is explicitly marked differently than other present tense forms, which fall into an Elsewhere Condition (Kiparsky 1973) (equivalently, blocking and underspecification in Distributed Morphology, see Embick 2015). Indeed, we define morphological features as being *marked* if their corresponding phonological realizations differ from the phonological realizations of other forms in the paradigm. The notion of collisions directly allows for this: our model will learn that [+3, +SG] is marked in the English present tense because its phonological realization, *-s*, is *different* than the phonological realization of other person-number combinations, causing a collision. However, it will not learn that first or second person is marked because the phonological realization of these forms will be largely the same: there will not be *sufficient* positive evidence from the collisions. The intuitions that result from this approach align well with theoretical insights on Elsewhere Conditions, syncretism and underspecification (Kiparsky 1973; Embick 2015). Indeed, syncretisms are necessarily represented as an Elsewhere Condition in our model: failing to further subdivide the input at any point allows the model to group several distinct inflections into one, creating an Elsewhere Condition and accounting for syncretisms in the paradigm. Thus, the model proposed here provides both a possible account of the acquisition of syncretisms and a hypothesis regarding how syncretisms are represented by the child.

We have considered a range of general challenges to the child learner, but perhaps the most important challenge is the typological diversity of systems of inflectional morphology. The child must be able to learn the morphological systems of any natural language to which they are exposed, so they cannot make specific assumptions about the type of system they will be encountering. Thus, the learning strategies that the child employs, and any models of such learning strategies, must be able to successfully segment and learn from agglutinative and fusional, templatic and concatenative languages. Since this has not yet been achieved, current models make simplifying assumptions (e.g. the model of Belth et al. 2021 assumes that all morphological processes are suffixing, although this assumption could be easily modified). A unique contribution of the model presented here is that it relies only on inequality between inflected forms, rather than specific transformations between them. That is, we care about the morphological feature(s) associated with any given collision rather than the exact phonological differences between the colliding forms. Because of this, our model is able to generalize to a typologically diverse set of paradigms as the child learner can: we test it on English verbs, German nouns, Spanish verbs, and Hebrew verbs.

This thesis proceeds as follows: §2-4 provide relevant background. In §2, we provide a theoretical overview of the paradigms we select for evaluation of our model and motivate their selection. In §3, we provide an overview of the acquisition patterns attested on each of these paradigms, and in §4, we provide background on previous approaches to modeling morphological acquisition. Our model and the principles upon which it relies are introduced in §5, and results on each of the paradigms are given in §6. Finally, we provide a discussion of our results and directions for future work in §7 before concluding in §8.

2. Background: Paradigms

In this section, we will provide an overview of the theoretical descriptions of each of the four paradigms we consider. This overview is intended both to familiarize the reader with the issues of interest related to each of these paradigms and demonstrate that, by evaluating our model on these paradigms, we will understand how it handles a number of key morphological issues across a set of typologically diverse paradigms.

2.1 English Verbs

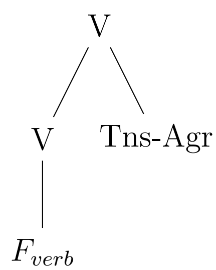

Figure 2.1: The morphosyntactic structure of the English verb, from Halle et al. (1993)

English verbs have typically been described in the literature as consisting of a verbal stem and a single affix that realizes tense and agreement; Figure 2.1 shows such an organization as described by Halle et al. (1993). Since only four productive inflectional affixes are used (*-ed* for +PST, *-s* for [+SG, +3], *-ing* for [+PRS, +PTCP] and *-∅* elsewhere), English provides a relatively simple testing ground for theories of morphosyntactic acquisition. At the same time, however, the ***irregularities*** of English verbs have been a subject of great attention: the Past Tense Debate (§4.1) arose as a result of disagreements regarding how regular and irregular past tense forms are represented in the English lexicon. Indeed, a particularly interesting fact about English verbal morphology is that higher frequency forms are more likely to be irregular (Yang 2016; Lignos and Yang 2016), posing a unique challenge to the child learner. The simplicity of the English verbal paradigm and the massive experimental literature resulting from the Past Tense Debate make English verbs a useful starting place for testing a model of morphological acquisition. At the same time, the distribution of irregulars poses a unique problem to the learner.

2.2 German Nouns

German noun plurals differ from the problem of English verbs in several interesting ways, although both inflectional systems are primarily suffixing. While in the English past tense, the default suffix *-ed* is also the most frequent, for German nouns, the opposite is the case: the least frequent suffix *-s* (occurring with 4% or less of nouns) is often interpreted to be the default plural marker (Mugdan 1977; Marcus et al. 1995; Wiese 2000; Kauschke, Kurth, and Domahs 2011). Further, German noun plurals exhibit allomorphy: each suffix corresponds to the same morphological feature (namely +PL).

Plural nouns in German are marked by either *-n, -en, -e, -s, -er*, or $-\emptyset$. Further, *-e, -er*, and $-\emptyset$ can combine with umlauting, yielding 9 total ways to mark the plural; which one is used depends on the phonology and gender of the stem (Kauschke, Kurth, and Domahs 2011). *-en* and *-n* are typically taken to phonological alternates of one another (Kauschke, Kurth, and Domahs 2011; Laaha et al. 2006), but there is great debate in the literature regarding the exact conditions under which each suffix is productive: see Trommer (2021) for a recent contribution to this debate. Since our model must only discover which features are marked and does not map these features to form, the problem of German noun plurals for our model is simply the problem of learning that plurality is marked in German. By using German noun plurals as a test case, then, we can test the ability of our model to account for an early stage in the acquisition of a hotly debated paradigm.

2.3 Spanish Verbs

Spanish provides an interesting case for the study of acquisition as its verbal morphology is highly regular (Lignos and Yang 2016) and exhibits both agglutinative and fusional behavior. For example, we can contrast the imperfect and perfect past:

(1) *amabas*

ama	ba	s
love	PST, IPFV	2,SG

'You were loving'

(2) *amaste*

ama	ste
love	PST, PFV, 2, SG

'You loved'

Here, the imperfect shows agglutinative behavior, affixing the subject agreement morpheme after the tense-aspect morpheme (ex. 1), while the perfect combines person, number, tense, and aspect into a single fusional morpheme (ex. 2). This contrast makes the learning problem a particularly interesting one, since features that are realized in separate morphemes in some parts of the paradigm will be realized in the same morpheme in other parts. Embick (2015) provides an analysis of Spanish verbs as having linear order:

$$\sqrt{\text{ROOT}} - \text{TH} - \text{T} - \text{AGR}$$

Where $\sqrt{\text{ROOT}}$ corresponds to the verbal root, TH to the theme vowel (either *-i, -e,* or *-a* based on the three verb classes of Spanish, with the *-a* class being productive), T to tense, and AGR to agreement. Embick (2015) notes that in clause structure, $\sqrt{\text{ROOT}}$ undergoes head movement to adjoin to v, which then moves to T. For the purposes of morphological acquisition, the resulting complex head is what is of interest to us; Embick (2015)'s proposal

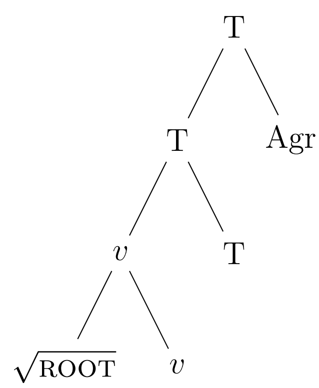

Figure 2.2: The structure of the Spanish verbal complex head proposed by Embick (2015)

for the structure of this complex head is shown in Figure 2.2. The v that is the sister of $\sqrt{\text{ROOT}}$ in Figure 2.2 corresponds to the verbalizing morpheme, which in Spanish is realized as the theme vowel, TH. Spanish verbal morphology is also full of syncretisms: between the second and third plural in the present (except for Peninsular Spanish) and between the first and third singular in many non-present tenses such as the past imperfect and future (Embick 2015). Thus, Spanish provides an interesting test case as a language that exhibits agglutinative and fusional behavior, syncretism, and more complex morphosyntax than either of the other two paradigms discussed thus far. At the same time, its relative regularity (especially compared to English, (Lignos and Yang 2016)) may provide an advantage to learners.

2.4 Hebrew Verbs

In all paradigms so far, the stem appears somewhere in the input as a result of null suffixing: in English for present tense verbs that are not third singular; in German for singular nouns; and in Spanish verbs, in the third singular present. Yet a unique challenge for children learning templatic inflectional morphologies such as Hebrew and Arabic is that they will ***never*** hear a truly unmarked form in their input (Lustigman 2013).

In modern Israeli Hebrew, verbs are inflected for number and gender in the present tense and for number, gender, and person in the past, future, and imperative (Bolozky 2007). The present tense, also known as ***benoni***, has also been hypothesized to have non-finite properties (see Paltiel-Gedalyovich et al. (2007) for a review), and Shlonsky (1997) suggests that the ***benoni*** is a hybrid which acts both as a finite present tense verb and a non-finite, non-tensed participial.

In the present and past tenses, agreement is marked in the suffix, while tense is marked by the vowels inserted between the consonants of the root (Bolozky 2007). In the future, agreement may be marked in the prefix or suffix (depending on the specifics of the agreement); tense is once more marked in the vowels (Bolozky 2007). Finally, imperative agreement is marked in the suffix, while mood is once more marked with the vowels (Bolozky 2007).

As an example, in the present (or ***benoni***) tense, we have the following four suffixes

(Bolozky 2007):

$$[+\textsc{masc},\ +\textsc{sg}] \rightarrow \emptyset$$

$$[-\textsc{masc},\ +\textsc{sg}] \rightarrow \textit{-et/-a}$$

$$[+\textsc{masc},\ -\textsc{sg}] \rightarrow \textit{-im}$$

$$[-\textsc{masc},\ -\textsc{sg}] \rightarrow \textit{-ot}$$

The present tense is realized as 1*o*2*e*3 in the singular and 1*o*23 in the plural before the application of the agreement suffix, where 1, 2 and 3 correspond to the three consonants in the verbal root (Bolozky 2007). This pattern yields the following inflected forms of the verb with root ***ktv***, meaning 'write':

(3) a. ***kotev***
write$_{[+\textsc{prs},\ +\textsc{masc},\ +\textsc{sg}]}$
'He writes'

b. ***kotevet***
write$_{[+\textsc{prs},\ -\textsc{masc},\ +\textsc{sg}]}$
'She writes'

c. ***kotvim***
write$_{[+\textsc{prs},\ +\textsc{masc},\ -\textsc{sg}]}$
'They (male) write'

d. ***kotvot***
write$_{[+\textsc{prs},\ -\textsc{masc},\ -\textsc{sg}]}$
'They (non-male) write'

One of several templates, or ***binyanim***, may be used for each stem, and there is some semantic regularity in the use of each (Bolozky 2007) – different ***binyanim*** may be used to mark things such as transitivity, reflexitivity, and causativity (Berman 1981).

Hebrew and other non-affixing languages have posed a problem both to linguistic theory and to models of acquisition; how to fit templatic morphology into models of acquisition is an open question. However, since the model we propose here will make use of only string inequality, it does not need to address templatic segmentation. As such, incorporating Hebrew verbs into our test set allows for broader investigation into a paradigm that has often been neglected in the acquisition literature.

2.5 Summary

The paradigms upon which we test our model span a wide range of morphological typologies, including affixing and non-affixing and agglutinative and fusional morphology. They each present unique challenges, including high-frequency exceptions in English, alternations between agglutinativity and fusionality in Spanish, allomorphy in German, and templatic morphology in Hebrew. A plausible model of morphological acquisition must be able to succeed on a typologically diverse set of languages, since the infant has no prior knowledge about the type of paradigm they will acquire. By testing our model on the paradigms discussed here, it is our goal to represent at least some of the typological diversity of the morphological paradigms of the world's languages.

3. Background: Acquisition

In this section, we review the developmental findings for each of the four paradigms we consider. In doing so, we will focus on the order of acquisition and vocabulary sizes, which we will use as a proxy for evaluating the plausibility of our model, and on children's ability to segment and relate inflected forms, a prerequisite for collision-based learning.

3.1 English Verbs

Recent evidence shows that infants are able to segment the inflectional affixes of English verbs very early: Kim (2015) and Kim and Sundara (2021) find that infants can segment -s,*-ing*, *-ed*, and a pseudo-morpheme *-sh* from pseudowords at as early as 6 months. Furthermore, 6-month-olds are able to relate nonce words suffixed with -s to their stems, but not words with any of the other aformentioned suffixes (Kim 2015; Kim and Sundara 2021). By 8 months, children can also relate nonce words suffixed with *-ing* to their stems (Kim and Sundara 2021), and *-ed* follows shortly after.[1] Kim and Sundara (2021) hypothesize that this order of acquisition is related to the overall frequency of the affixes in the input – if we combine the plural noun marker and third singular verb marker, then -s will certainly be the most frequent affix in the child's input (Kim and Sundara (2021), see Marquis and Shi (2012) for similar findings in French). This apparent early blindness to the differences between the two meanings of -s, in addition to the extremely early segmentation and relation between forms, suggests that English-learning children can segment and relate inflected forms well before they understand what meaning is realized by the affixes they segment (Kim and Sundara 2021). Indeed, Soderstrom, Wexler, and Jusczyk (2002) show that English-learning 19-month-olds still do not have a full grasp of the grammatical use of -s. Mintz (2013) shows that 15-month-olds segment *-ing* when processing words, but do not segment endings that are not morphemes. Interestingly, Mintz (2013) does not find the same pattern for 8-month-olds as Kim and Sundara (2021); this is likely due to the increased cognitive load of the within-subjects design that Mintz (2013) utilizes, as well as the shorter familiarization time; see Kim and Sundara (2021) for a full discussion of the differences in approaches.

Santelmann and Jusczyk (1998) show that 18-month-old English learners already have limited knowledge of their language's morphosyntax: when exposing infants to sentences with either well-formed (e.g. "everybody is baking bread") or ill-formed (e.g. "everybody can baking bread") dependencies between the main and auxiliary verbs, they find that 18-month-olds are sensitive to the difference, but 15-month-olds are not (111). They additionally find that 18-month-olds preferred the well-formed dependency only over a limited domain of 1-3 syllables and not over domains of 4-5 syllables, suggesting a limited processing buffer (Santelmann and Jusczyk 1998).

Thus, within the first year of life, children can segment and relate forms using most of the inflectional affixes of English, but it takes them far longer to learn the grammatical use of these affixes. While the distributional information discussed above is certainly a necessary

1. Unpublished findings courtesy of Megha Sundara.

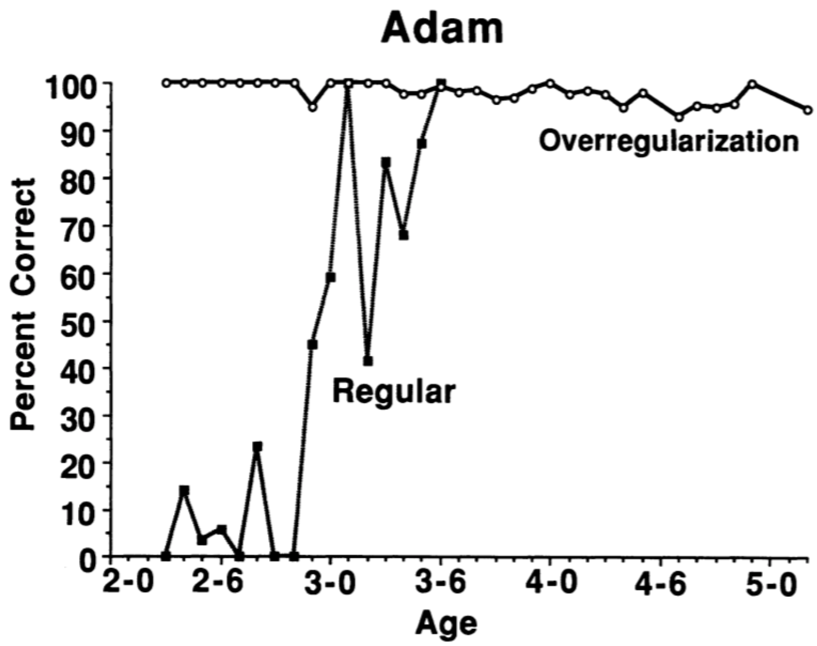

Figure 3.1: Adam's developmental regression in the English past tense, taken from Marcus et al. (1992)

portion of morphological acquisition, it is not by itself sufficient: children understanding that *-ed* is a suffix of English does not entail them knowing which parts of speech it attaches to or its meaning, and this information is acquired later. Learning the mapping between the affix *-ed* and the morphological features that it realizes has been a topic of great discussion, resulting largely from the Past Tense Debate (discussed further in §4.1). One influential finding from the ensuing experimental work is **developmental regression**, in which the child temporarily overgeneralizes *-ed* to irregular verbs when they acquire its productivity as a marker of the past tense, leading to the production of forms such as *goed* and *feeled* (Marcus et al. 1992). Figure 3.1 shows this pattern for the well-studied child "Adam": we see that as Adam's regular production accuracy quickly grows just before 3;0, there is a dip in his irregular production accuracy (Marcus et al. 1992). This dip is caused by his production of *feeled* (Marcus et al. 1992), demonstrating that he has learned the productivity of *-ed* and explaining the subsequent leap in his production accuracy on regular verbs. Clahsen, Aveledo, and Roca (2002) describe the pattern of developmental regression as occurring when markedness of a specific morphological feature becomes sufficiently obligatory for the child, potentially before they have fully acquired all irregulars as exceptions to the resulting rule. Such an interpretation suggests that the child learning that English marks the past tense triggers developmental regression immediately: for Adam, this would suggest that he acquires the obligatory marking of pastness just before 3;0. In the present work, we will take a similar approach to that of Clahsen, Aveledo, and Roca (2002), hypothesizing that children learn that a morphological feature is marked when there is sufficient positive evidence in the input, where we define "sufficient positive evidence" by the Tolerance Sufficiency Principle (Yang 2016). Unlike Clahsen, Aveledo, and Roca (2002), however, we do not necessarily

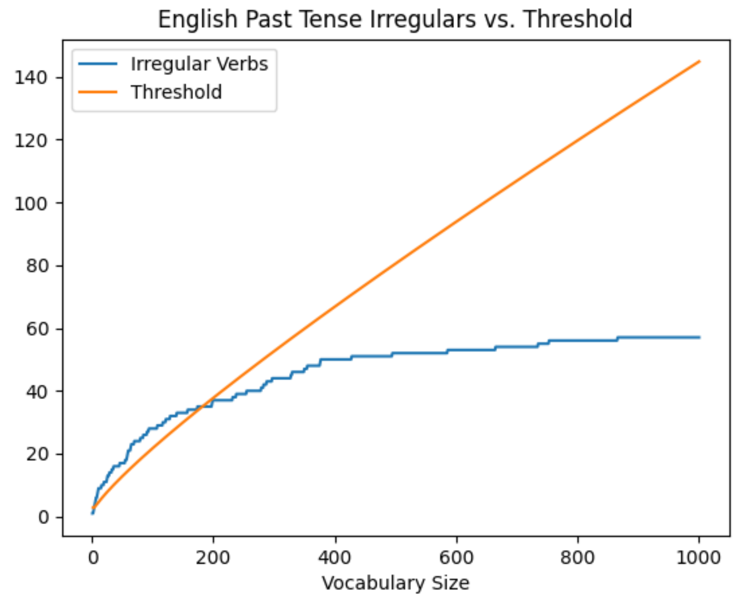

Figure 3.2: The number of irregular verbs vs. the Tolerance Sufficiency Principle threshold on our English training data (§5.4)

interpret this step as a trigger for developmental regression – see §7 for a discussion of how to relate the present models with models that map from morphological features to form.

While much work has focused on the acquisition of the English past tense, other work has looked at the overall acquisition of English verbs, including the present progressive and the third singular present. Brown (1973), for example, establishes that English-learning children generally acquire the present progressive *-ing* before noun plural *-s*, followed by the irregular past tense and then the regular past tense *-ed*. Further evidence from Berko (1958) demonstrates that children typically learn the third singular *-s* between the present progressive *-ing* and past tense, although the third singular emerges later in some children. Yang (2016) demonstrates that the order of acquisition is related to the input frequency of the morphemes (since more frequent words are more likely to be in the child's vocabulary, Goodman, Dale, and Li (2008)) as well as their regularity; the English past tense is acquired so late because of the number of high frequency exceptions to the productive rule. Indeed, if we consider the number of irregular forms in our English training data (see §5.4) compared to the threshold for rule generalization given by the Tolerance Sufficiency Principle (see §5.1), we see that the number of irregular forms dips below the threshold when the child has heard about 200 inflected forms, which corresponds to about 90 unique verbs (Figure 3.2). English-learning children acquire the regular marking of the past tense, the last morphological process to be acquired, between ages 2;0 and 3;0 (Brown 1973; Kuczaj II 1977). At age 2, a child's vocabulary is at most 500 words (Fenson et al. 1994; Hart and Risley 1995). Since about a quarter of the child's vocabulary will be verbs (Bornstein et al. 2004), the 90 unique verbs predicted by Figure 3.2 and the Tolerance Sufficiency Principle match well with developmental findings, corresponding to an overall vocabulary size of under 500.

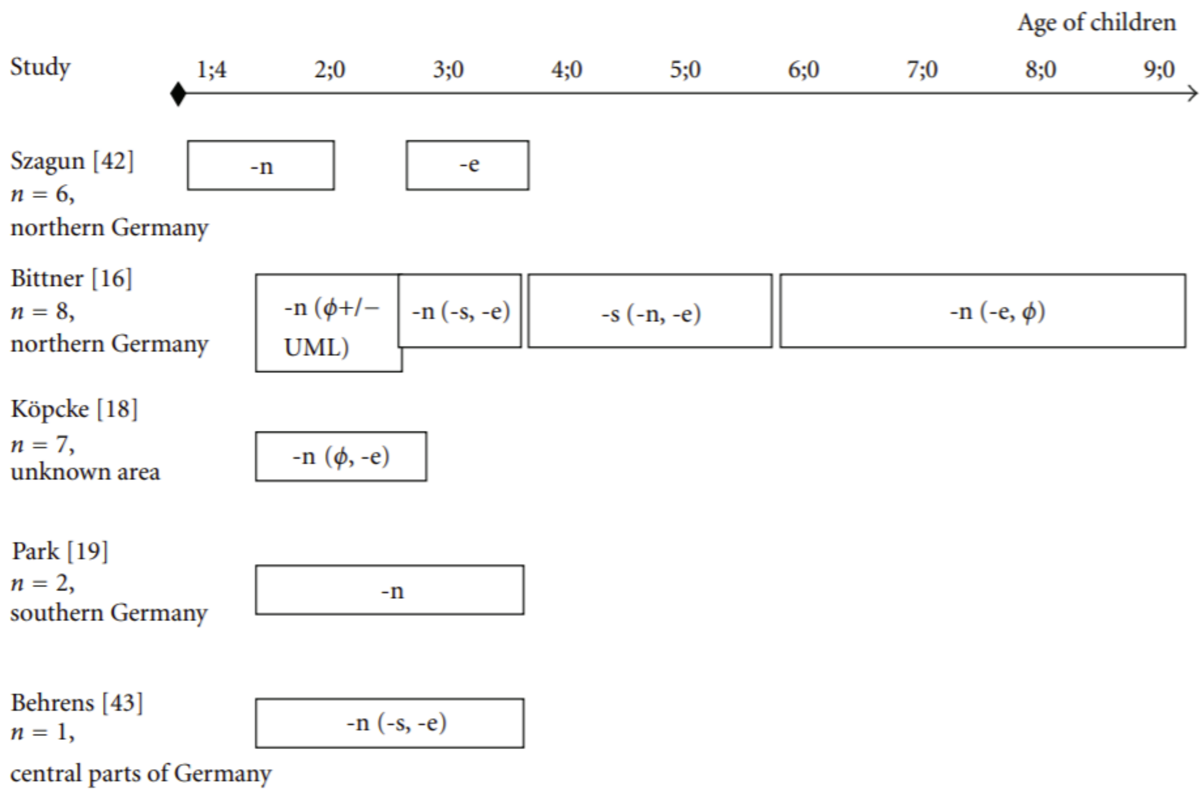

Figure 3.3: A comparison of five developmental studies of overapplication of German noun plural suffixes taken from Kauschke, Kurth, and Domahs (2011)

3.2 German Nouns

While less work has examined the early segmentation of German noun plural morphology, several studies have shown that German-learning infants are sensitive to the morphosyntax of German nouns and verbs at an early age. Similarly to English-learning infants, (c.f. Santelmann and Jusczyk (1998)), German learning 19-month-olds are sensitive to dependencies between main and auxiliary verbs despite the freer word order of German relative to English (Hohle et al. 2006). Further, German learning 14-16-month-olds can use determiners as a distributional cue to categorize nouns as nouns (Höhle et al. 2004), although younger infants cannot. While more work is needed to understand how early German-learning children can relate inflected noun plurals to their corresponding stems, it is clear that morphosyntactic information is available early to the German-learning child.

The order of acquisition of German noun plural suffixes varies greatly among children: Kauschke, Kurth, and Domahs (2011), for example, review patterns of overapplications of each of the plural suffixes across five previous developmental studies and find no consistent order, although the frequent suffixes *-(e)n* and *-e* tend to emerge sooner than others (Figure 3.3). Importantly, more than one suffix is overapplied (Szagun et al. 2006), indicating that multiple plural suffixes are governed by productive processes in German. Mills (2012) and Clahsen et al. (1992) find that overapplication of these suffixes typically occurs at or before

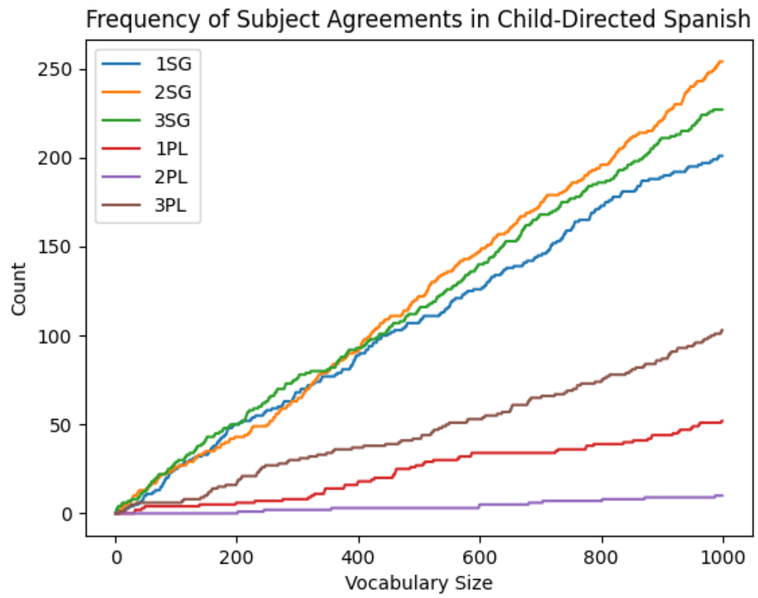

Figure 3.4: Relative frequencies of each subject agreement in our Spanish training data

age 2;0. The child's vocabulary at this age is typically approximately 300 words total, and not more than 500 (Szagun et al. 2006). At this stage, however, children only correctly inflect noun plurals in an average of 3 of the 6 possible inflectional classes (Szagun et al. 2006); accuracy on plural marking continues to increase for several years after this (Kauschke, Kurth, and Domahs 2011).

3.3 Spanish Verbs

Little work has been done on the early segmentation of Spanish verbal morphology, so we will instead focus our experimental review on the subsequent problem of mapping the segmented forms to their corresponding morphological features. We will see that despite the apparent complexity of Spanish morphology reviewed in §2.3, acquisition is quite early, even compared to languages such as English.

3.3.1 Subject Agreement

A number of longitudinal studies (e.g. Aguirre Martınez et al. (1995), Grinstead (1998), and Durán (2000)) have demonstrated that Spanish-learning children begin marking subject agreement quite early. From 1;7 onwards, children mark person contrasts productively across these studies (Montrul 2004). For comparison, the mean vocabulary size for children learning Argentinian Spanish at age 1;8 is 156 words, with the maximum vocabulary size at just over 400 (Bornstein et al. 2004). Interestingly, while some children produce all three person contrasts initially, some only produce a contrast between first and third person, with second person emerging later (Montrul 2004). Number contrast comes soon after person contrast

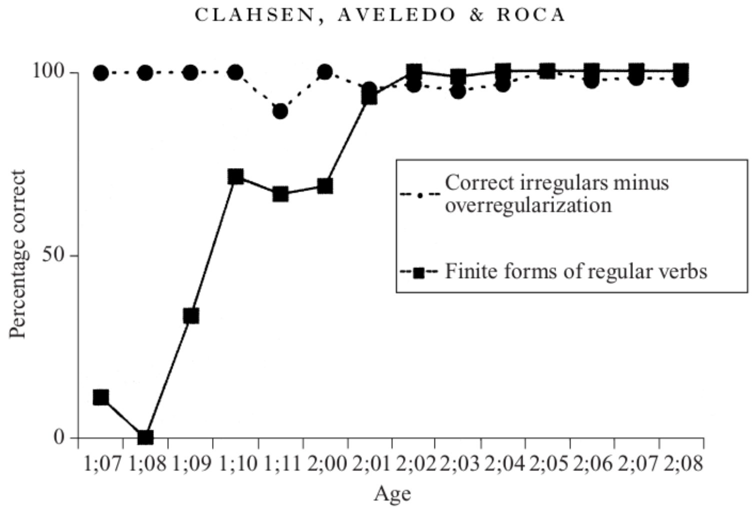

Figure 3.5: Maria's developmental regression in Spanish finiteness marking, taken from Clahsen, Aveledo, and Roca (2002)

is acquired (Montrul 2004), with first and third plural production followed later by second plural, the least produced subject agreement form. Gaya (2001) find that across three children between 1;7 and 2;6, that third singular was by far the most-produced form, making up an average of 60.33% of the children's productions, followed by the first singular at 24.83%, the second singular (6.66%), the third plural (4.93%) and the first plural (3.23%). The children in this study all fail to produce second person plural entirely (Gaya 2001).

Since the third singular in Spanish is marked by the null suffix (henceforth ∅), the high production of apparent third singular forms might lead to the belief that the child is simply failing to add inflectional suffixes rather than productively and correctly inflecting the third singular. However, error rates in subject agreement in child Spanish are extremely low: Gaya (2001) and Durán (2000) both reported error rates in subject marking at or under 5%. These errors are typically either use of the third singular to refer to the first singular or the use of the first singular to refer to the second singular (Montrul 2004). As Montrul (2004) points out, however, the former of these errors is common cross-linguistically (and adults sometimes refer to themselves in the third person) and thus it may not be a ***linguistic*** error in the same sense as the errors we discuss in the present work. Interestingly, an analysis of our Spanish training data (§5.4) demonstrates that the frequencies of each of the three person markings are comparable in the singular (Figure 3.4). In the plural, however, the third plural is most frequent, followed by the first, and finally the second plural. Since this training data consists of Peninsular Spanish, which explicitly marks the second plural, this low rate of second plurals cannot result from annotation errors and instead provides a potential reason for the late acquisition of the second plural.

Interestingly, developmental regression is also present in Spanish finiteness marking: in Spanish, any finite verb must be marked with person and number agreement. Figure 3.5 shows this developmental regression in "Maria" from Clahsen, Aveledo, and Roca (2002): her initially perfect production of subject agreement for irregular verbs regresses as she learns the productive subject agreement processes for the regular verbs. Similarly to Adam (§3.1), this regression on irregulars and leap in performance on regulars indicates the emergence of productive finiteness marking.

3.3.2 Tense and Aspect

By 2;0, Spanish-learning children begin marking verbs in the past tense, exclusively with the preterite (Montrul 2004). The child studied in Jacobsen and Meisel (1986), for example, correctly produces the preterite at the start of the recordings (2;2) but does not consistently produce the imperfect past until nearly 3;0. In the children studied by Gaya (2001), past tense emerged simultaneously with or just after present tense morphology before 2;0, with the future emerging a few months later. Across several studies, children predominately produce present tense forms, followed by past and future forms, but produce all three categories with extremely low error rates (Montrul 2004; Gaya 2001).

3.3.3 Mood

In Spanish, the subjunctive has a syncretism with the imperative, and both emerge early. The subjunctive is produced by age 2;0 (Montrul 2004; Aguirre 2011; López Ornat et al. 1994) and is productive by 2;1 in the child studied by López Ornat et al. The imperative emerges even earlier, with children producing it as early as 1;7 (Montrul 2004). The semantic and syntactic contexts that the subjunctive occurs in are extremely varied in Spanish, however, and can take 6 or more years to fully acquire (Montrul 2004). Since these contexts are largely independent of the morphology of the form, we will only consider the child's productive use of the subjunctive beginning at about 2;0 in the present work.

In summary, we see that in Spanish, subject agreement emerges early, with person marking quickly followed by number marking. Subsequently, tense emerges, and mood emerges sometime between person and tense, with greater variability in its timeline. The overall timeline of the acquisition of Spanish verbal morphology, however, indicates that much of the morphology is learned by around 2;2, on vocabularies of 500 words or less.

3.4 Hebrew Verbs

In a longitudinal study of four Hebrew-learning children, Lustigman (2013) finds that children use one of two alternatives for verbs before learning the productive marking processes in their language: they either produce fully marked forms memorized from their input, or 'bare stem' forms that lack adult-like affixes. In the former case, the fully-marked forms do not obey subject-verb agreement, indicating that the children are not inflecting from known rules but rather repeating forms that they have heard in their input (Lustigman 2013). In the children studied, this period typically occurs between 1;5-1;9. For comparison, at 1;8, the mean vocabulary size for a Hebrew-learning child is 233.8 words, with a maximum of 481 words (Bornstein et al. 2004). Little work has been done on the early segmentation of templatic morphology, so it is unclear how children represent the roots or templates of Hebrew morphology at this stage.

Consistent subject-verb agreement, in which inflectional affixes occur only in obligatory contexts, was first observed at 1;8-2;0 in the children studied by Lustigman (2013). Bare stem forms are still relatively common at this stage, but there is no overlap in the production of memorized forms that violate subject-verb agreement and productive affixation (Lustigman 2013). At this stage, the children produce mainly infinitives and ***benoni*** present forms (Lustigman 2013). Infinitives are, of course, unmarked for number, gender, and person, while the benoni present are marked only for number and gender and not person (Lustigman 2013) and have both a finite present and non-finite participial interpretation (Shlonsky (1997), §2.4). In a further longitudinal study of two Hebrew-learning children between 1;2 and 2;5, Bat-El (2014) finds that one child produces gender and number agreement before person agreement, as expected based on the findings of Lustigman (2013). However, the other child produces gender and person before number, which does not fit with these findings. Bat-El (2014) argues that this difference is due to asynchronous development of morphology and phonology in the two children.

Bat-El (2014) notes early marking of tense in the two children in their study, but these markings are rare; Berman (1981) finds that the emergence of the three Hebrew tenses (past, present, future), as well as the infinitive and imperative, occurs by around 3;0. At this age, children do not yet alternate a given root verb in different ***binyan*** conjugations to mark transitivity, causativity, voice, reflexivity, and reciprocality (Berman 1983). In summary, we expect that gender and number will typically occur before person in Hebrew, followed by tense, the infinitive and imperative, and finally the ***binyan*** conjugations, which emerge at or after 3;0 (Berman 1983).

3.5 Summary

We have reviewed the developmental findings on order of acquisition for each paradigm, which we will use to evaluate the plausibility of the model presented in §5. In addition, we have reviewed infants' early ability to segment and relate inflected forms, a prerequisite for collision-based learning. Much work on English has shown that English-learning children can segment inflected forms and relate them to their stems well before they understand the morphological features corresponding to this relation, suggesting that collisions are both a valuable and available tool during the acquisition process. However, since the literature on early morphological segmentation is focused heavily on English, it is an open question how much these findings extend to other languages (although see Marquis and Shi (2012) for French). For the present work, we assume that collisions are a cue that is available to all learners, but future work should investigate the early segmentation of children learning other paradigms.

4. Background: Modeling

An extremely productive line of work computationally modeling morphological acquisition was sparked by the Past Tense Debate (see McClelland and Patterson (2002) and Pinker and Ullman (2002) for a review). In this section, we will begin by reviewing the main themes and takeaways of this debate as it took place in the 1990s. We will then turn to more modern models of morphological acquisition, both connectionist and rule-based, with the understanding of how the themes of the Past Tense Debate resurface in many of these models.

4.1 The Past Tense Debate

The Past Tense Debate began with Rumelhart and McClelland (1986)'s proposal that a connectionist network was capable of exhibiting rule-like behavior in the English past tense, and Pinker and Prince (1988)'s ensuing critique. The model proposed by Rumelhart and McClelland (1986) is a **single-route** model in that it proposed that regular and irregular English past tense verbs were processed by the same mechanism. The model maps a word's stem to its past tense form (e.g. mapping *walk* to *walked*) through the same parallel distributed processing network irregardless of whether the past tense form is regular (as in *walked*) or irregular (as in *go-went*). The architecture of this network can be seen in Figure 4.1.

An obvious appeal of connectionist models is their closer match to what we know to be true about neuronal computation: activation of each unit in such a model is computed "through a process of weighting each of their input signals by the strength of the connection along which the signal is coming in, summing the weighted input signals, and feeding the result into a non-linear output function" (Pinker and Prince 1988, 75). Many approaches proposed in the neuroscience literature follow similar computational principles (see e.g. Izhikevich (2004) and Long and Fang (2010) for a review). While such approaches match well with single cell recordings (Izhikevich 2003), the algorithms use to train such computational networks have no "known neurophysiologial analogue" (Pinker and Prince 1988, 75). That is, while the method of computation in trained neural networks may mirror that of the human brain, the methods used to train the network do not mirror learning in the human brain (Pinker and Prince 1988). Further, by mapping directly from the stem to the past-tense form, the model of Rumelhart and McClelland (1986) "collapses the major distinctions" of linguistic theory (Pinker and Prince 1988, 88), such as the distinction between phonology and morphology that is standardly accepted (Halle et al. 1993; Embick 2015).

Indeed, Pinker and Prince (1988) show that the model of Rumelhart and McClelland (1986) has several shortcomings. Perhaps most famously, they note that the pattern of developmental regression Rumelhart and McClelland (1986) claim to model is actually a result of the nature of their input data: in the beginning, the model is exposed to a set of 10 verbs, 2 of which are regular, 10 times (Pinker and Prince 1988). Subsequently, it is exposed to a set of 420 verbs, 334 of which are regular, 190 times; the 10 verbs are a subset of the 420 verbs (Pinker and Prince 1988). This matches with an assumption by Rumelhart and McClelland (1986) that "the child learns first about the present and past

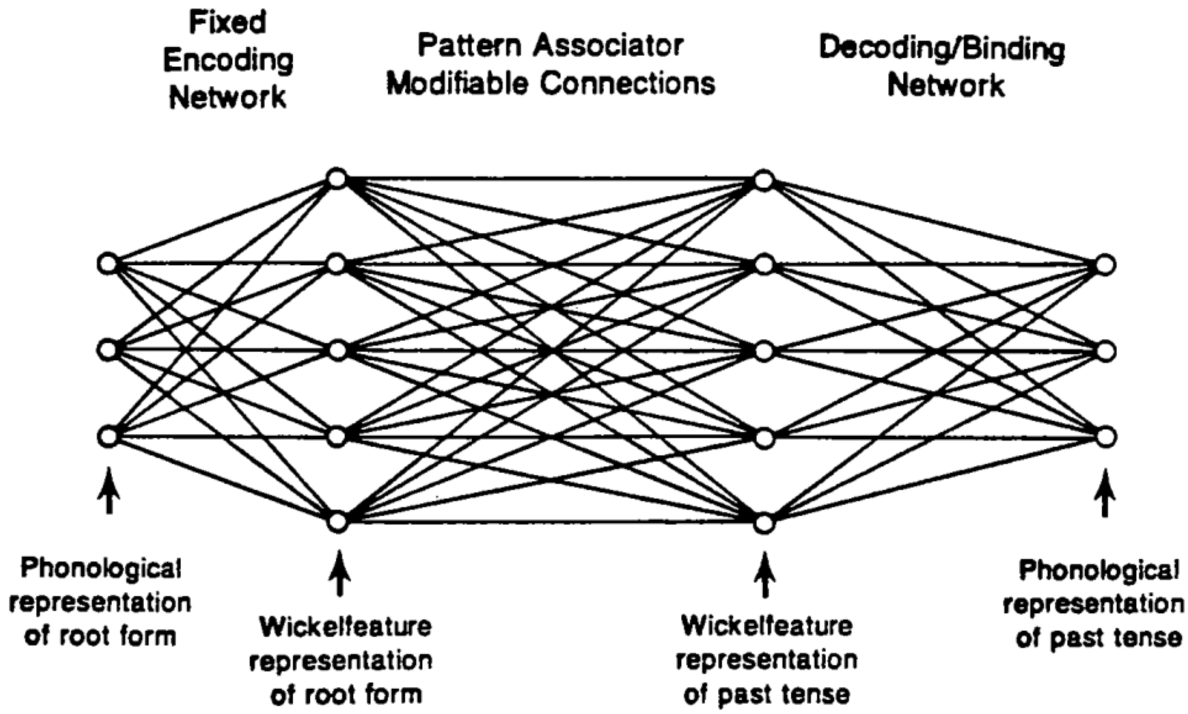

Figure 4.1: The parallel distributed processing model of English past tense acquisition proposed by Rumelhart and McClelland (1986)

tenses of the highest frequency verbs; later on, learning occurs for a much larger ensemble of verbs, including a much larger proportion of regular forms" (Rumelhart and McClelland 1986, 241). However, Pinker and Prince (1988) demonstrate that such a sudden change in the percent of regular forms in the child's vocabulary is simply not realistic: "the percentage of children's verbs that are regular is remarkably stable ... across stages" when considering the children in Brown (1973)'s study (Pinker and Prince 1988, 140). Additionally, the model produces odd and irregular past tense forms for regular stems, e.g. by overapplying vowel changes (*sip-sept*), double-marking regular past tense forms (*type-typeded*), and producing forms with no direct explanation (*mail-membled*) (124–125). Both of these factors, as well as many others discussed by Pinker and Prince (1988), led Rumelhart and McClelland (1986)'s model to generally be rejected as a plausible model of morphological acquisition.

As an alternative to the single-route model of Rumelhart and McClelland (1986), Pinker (1998) proposed a **dual-route** model in which "regular inflection applies freely in any circumstance in which memory fails because regular inflection is computed by ... a symbol-processing rule" (240). Such a pattern, Pinker (1998) argues, would better account for developmental patterns such as U-shaped learning and over-regularization. Under this model, developmental regression occurs as a result of an initial stage in acquisition in which the past tense forms of verbs are simply memorized, followed by the acquisition of a productive rule, and the overapplication of this rule to irregular forms for which the past tense forms have not yet been memorized or for which memory access of the past tense form fails (Pinker

1998). In addition to being able to account for the developmental patterns discussed in §3, Pinker (1998) also argues that his **words-and-rules** model provides a more psychologically plausible account of morphological acquisition. However, this model does not provide an explicit account of how words and rules are learned, or of how the child should discriminate between productive processes and exceptions to these processes.

4.2 Connectionism Today

While it is generally accepted today that the model of Rumelhart and McClelland (1986) does not provide a plausible account of morphological acquisition, the ability of connectionist models to account for the patterns of acquisition discussed here remains an open question, perhaps because of the neurological appeal of such models discussed above. Kirov and Cotterell (2018) recently returned to the question of connectionist models as an account of morphological learning by proposing an encoder-decoder (ED) neural network to learn mappings between English verb roots and inflected forms. In their first experiment, Kirov and Cotterell (2018) present the model with only root-past pairs; in the second, they present it with the past tense, gerund, past participle, and third person singular of the verb. They demonstrate that the ED model is able to achieve high accuracy in both experiments, and conclude that "the empirical performance of neural models is no longer an issue" (661). They note, however, that their model is not meant to serve as a direct model of acquisition in children. Here again, the implausibility of connectionist training methods discussed by Pinker and Prince (1988) resurfaces: Kirov and Cotterell (2018) train their model for 100 epochs, meaning that the model makes 100 ***complete*** passes over the training data. This, of course, stands in contrast to the incremental nature of the input to the child. Furthermore, the data used in Kirov and Cotterell (2018)'s first experiment consists of 4,039 verb types, split into 80-10-10 training/development/test sets, meaning that the model sees over 3500 verb types during each epoch of training (658). In Kirov and Cotterell (2018)'s second experiment, the model sees each verb type in its ***complete*** morphological paradigm: root, past tense, gerund, past participle, and third singular (661). The size of both training sets lies in contrast to the child's early acquisition of the past tense on vocabularies of under 1000 words (Fenson et al. 1994; Hart and Risley 1995; Bornstein et al. 2004), while the appearance of each word in its full inflectional paradigm is in contrast with the Zipfian distribution of paradigm saturation found by Lignos and Yang (2016) (see §1 for further discussion). Indeed, of the 3144 most frequent verb stems in English CHILDES, only 656, or less than a quarter, appear in their complete paradigm.[1] By extension, we would expect that for the child to have heard 3500 verb stems in their complete paradigms, they would have to hear over 14,000 verb stems total, which is two orders of magnitude larger than the number of verbs the English-learning 3-year-old knows (§3.1).

In spite of the large, saturated data to which the ED model is exposed, several lines of work have demonstrated that it still struggles to match developmental findings. Kirov and Cotterell (2018) themselves note that the model does not display developmental regression (660). Further, Corkery, Matusevych, and Goldwater (2019) demonstrate that different

1. CHILDES counts courtesy of Charles Yang.

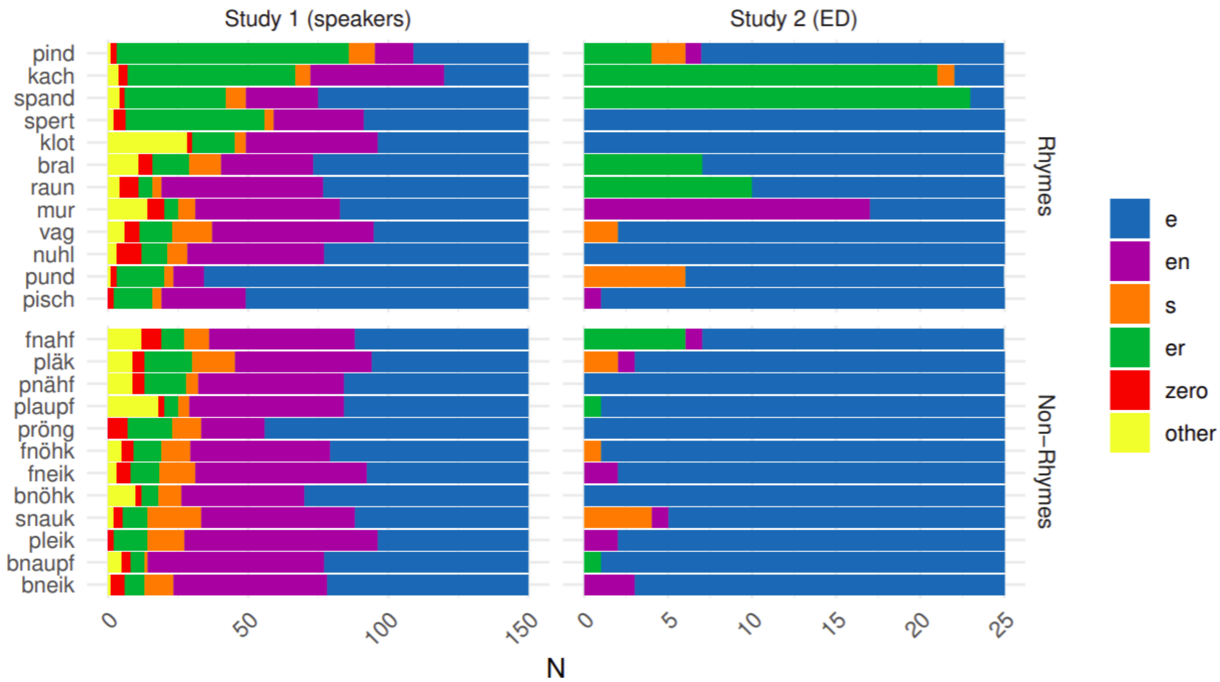

Figure 4.2: Distribution of predicted plural suffixes for human participants vs. the ED model of Kirov and Cotterell (2018), taken from McCurdy, Goldwater, and Lopez (2020)

random initializations of the ED model cause great variation in its correlation with human productions on the English past tense data used by Kirov and Cotterell (2018): from 0.15 to 0.56 for regular verb forms and from 0.23 to 0.41 for irregular verb forms (Corkery, Matusevych, and Goldwater 2019, 3872). Since the model exhibits unstable behavior across different simulations, Corkery, Matusevych, and Goldwater (2019) attempt a second analysis in which each simulation is interpreted as an individual participant and results are averaged across simulations as they are across participants in human data. However, they still find a low correlation between aggregated human and model performance: 0.45 for regulars and 0.19 for irregulars (3873); this correlation is poorer than that of the earlier model of Albright and Hayes (2003) discussed below. They also find that the ED model overproduces irregulars on nonce verbs when compared to humans.

Recent work has also demonstrated that the sensitivity to frequency of the ED model causes it to struggle with morphological paradigms in which the default form is infrequent in the input. As discussed in §2.2, in German noun plurals, the least frequent suffix (namely -s) is usually taken to be the default (Marcus et al. 1995; Wiese 2000; Kauschke, Kurth, and Domahs 2011). This stands in contrast to the English paradigms on which Kirov and Cotterell (2018) test the ED model: as discussed in §2, simply picking the most frequent form as the default in these paradigms will yield the correct results. To better understand the implications of the ED model's frequency sensitivity, McCurdy, Lopez, and Goldwater (2020) and McCurdy, Goldwater, and Lopez (2020) test its ability to match human performance on the case of German noun plurals. McCurdy, Goldwater, and Lopez (2020) test the ability of the ED model to account for German noun plural inflection by comparing human-predicted inflections and ED-predicted inflections on 24 nonce words taken from Marcus et al. (1995). Adult participants were shown the nonce word with a determiner that provided a cue to grammatical gender and asked to produce the plural form; which determiner was used was varied across participants (McCurdy, Goldwater, and Lopez 2020). The authors then presented the same forms to the ED model and compared its predicted plurals with the plurals predicted by human subjects. They find little correlation between speakers' predictions and predictions made by the model: the highest Spearman's correlation is .33 (McCurdy, Goldwater, and Lopez 2020). Figure 4.2 provides a visualization of the distribution of suffixes predicted by the ED model and by humans, taken from McCurdy, Goldwater, and Lopez (2020). They additionally examine the top 5 most likely plural forms predicted by the ED model, and find that the lower-ranked predictions are dominated by forms that "do not cohere to standard plural production," which they take to mean that the "ED model scores may not be cognitively plausible analogues to speaker behavior" (7–8). Building on this, McCurdy, Lopez, and Goldwater (2020) investigate the role of grammatical gender in determining which plural suffix is applied in both humans and the ED model: they again collect human predictions for the 24 nonce words of Marcus et al. (1995), which were designed to lack strong phonological cues to gender. McCurdy, Lopez, and Goldwater (2020) demonstrate that both the speakers and ED model show effects of grammatical gender on the plural form productions, but the effects are significantly greater for the ED model than for the human subjects, suggesting that the mechanism the model uses to hypothesize inflected forms is different from the mechanism used by humans; they suggest that humans pay greater attention to phonological cues.

The most recent proposal for a connectionist model of morphological acquisition is that of Dankers et al. (2021), which proposes a recurrent neural network (RNN) as a model of the acquisition of German noun plurals. Dankers et al. (2021) show that their model correctly learns grammatical gender and end-of-word phonology as predictors of which plural suffix is applied. However, they also demonstrate that the RNN does not handle nouns in low-frequency plural classes well, likely because it relies too heavily on length as a predictor of these suffixes (Dankers et al. 2021). Additionally, the model is biased towards the -e suffix, which does not match with the generalization patterns of humans (Dankers et al. 2021).

4.3 Alternatives to Connectionist Models

Despite their appeal, discussed above, the ability of connectionist models to account for patterns of morphological acquisition is still an open question. In contrast to the connectionist accounts outlined above, however, several other models of morphological acquisition have been proposed in recent years. Albright and Hayes (2003) propose the Minimal Generalization Learner as a non-neural probabilistic account of morphological acquisition. This model uses inductive learning to propose multiple stochastic morphophonological rules that are assigned confidence scores based on their performance in the lexicon (Albright and Hayes 2003). The learning process involves rules being gradually built up through iterative generalization: at the beginning, every present-past pair is memorized as a separate rule (e.g. for *shined*, $\emptyset \rightarrow$d/[ʃaɪn__]$_{[+past]}$). Such a rule could then be combined with other rules (e.g.

the rule corresponding to *robbed*) in which the first portion of the rule is also $\emptyset \rightarrow$d, to form broader generalizations. Albright and Hayes (2003)'s model finds "the tightest rule that will fit both cases" (124) by iteratively generalizing and rewarding rules based on their accuracy. The model is additionally supplied with the phonological features of each segment in the input in order to allow for this tight generalization: the alternations between [t], [d], and **[@d]** in the regular English past tense, for example, are all conditioned on such phonological features. During generation, the model hypothesizes and probabilistically weights multiple potential output forms (Albright and Hayes 2003).

Although originally intended as a model of morphological processing in the visual modality, the Naïve Discriminative Learning (Baayen et al. 2011) and Linear Discriminative Learning (Baayen et al. 2019) models have also been employed in modeling morphological acquisition. Both models are motivated by the assumption that morphology is not its own level of representation separate from phonology and semantics, but rather arises purely as a combination of the two. Both models are trained using the Rescorla–Wagner equation of classical conditioning to relate form to meaning (Rescorla 1972). The input to the Naïve Discriminative Learner is a word form and its corresponding meaning: for example, (*hands*, {HAND, +PL}) and the task of the model is to learn rules of the form $-s$ = +PL. However, as Marantz (2013) points out, providing such features as +PL essentially encodes the morphosyntactic information that Baayen et al. (2011) claim to avoid, thus providing an explicit level of morphological representation in the model. To address this, Baayen et al. (2019) utilize distributional vector-based semantics, with the goal of demonstrating that morphology can be reduced to phonology and semantics. However, a large body of work has rejected such an explanation of morphology based on theoretical (e.g. Embick (2015) and Marantz (2013)) and experimental (e.g. Mintz (2013), Kim (2015), and Kim and Sundara (2021)) grounds. A particularly compelling piece of evidence against morphology as a combination of phonology and semantics is the early segmentation of inflectional morphology reviewed in §3.1: children are sensitive to morphological structure well before they know what the morphemes mean, but this sensitivity is greater than purely phonological controls (Kim and Sundara 2021), implicating structural knowledge beyond just phonology and semantics.

In contrast to the models of Albright and Hayes (2003) and Baayen et al. (2019), which both employ probabilistic learning and generation techniques (albeit with very different assumptions), the Tolerance Sufficiency Principle (Yang 2016) provides a threshold for the generalization of a rule beyond the input, supporting binary claims of productivity such as those put forth by Pinker (1998). Yang (2016) demonstrates that the Tolerance Sufficiency Principle can be applied to a number of morphological paradigms and matches well with developmental findings on each of these. The Tolerance Sufficiency Principle has also found support in a number of artificial language studies; for a further review of this experimental evidence, see §5.1. Payne, Kodner, and Yang (2021) and Belth et al. (2021) utilize the Tolerance Sufficiency Principle and recursive search procedures to develop models of morphological acquisition. Both models take as input a set of (stem, inflected form, feature set) triples (e.g. *walk*, *walked*, [3, SG, PAST]). Payne, Kodner, and Yang (2021) apply the Tolerance Sufficiency Principle and a recursive search procedure to model the acquisition of Spanish and English verbal morphology. Their model is able to extract the productive mappings between meaning and form in both languages on plausible vocabularies, despite the unique challenges of these languages discussed in §2. Belth et al. (2021) also develop a model utilizing the Tolerance Sufficiency Principle and recursive subdivision of the input. This model yields decision-tree-like representations in which each node contains a rule, a set of memorized exceptions, or both. Belth et al. (2021) demonstrate that their model correlates more with the behavioral data of Corkery, Matusevych, and Goldwater (2019) and McCurdy, Goldwater, and Lopez (2020) than the ED model of Kirov and Cotterell (2018).

4.4 Summary

The ability of neural vs. rule-based models to account for morphological acquisition remains an open question. However, note that all of the models presented here define the learning problem as acquiring the phonological transformations that realize a given set of morphological or semantic features. The model of Rumelhart and McClelland (1986) simply assumes the morphological features to be the English past tense, while newer connectionist models such as Kirov and Cotterell (2018) and Dankers et al. (2021) encode these features as a part of the input to the model, allowing the model to learn mappings corresponding to multiple morphological feature sets at once. Similarly, the rules in the model of Albright and Hayes (2003) encode both phonological features and morphological features such as [+PAST], and even the models of Baayen et al. (2011) and Baayen et al. (2019) encode the learning process in this way, despite arguing against a separate morphological level of representation (Marantz 2013). The models of Payne, Kodner, and Yang (2021) and Belth et al. (2021) similarly take morphological features as input and learn the processes that realize these features. Since most of these models are intended to provide an incremental account of acquisition, however, by explicitly encoding these features from the earliest stages, they make the crucial assumption that they are available to the child from the earliest stages of acquisition. Yet which morphological features are marked varies across languages, so surely the child must learn *what* features their language marks before they can learn how these features are realized. By providing the features from the beginning of learning, current accounts of morphological acquisition fail to account for this earlier step in the acquisition process. The present work attempts to fill this gap by proposing that children only learn how morphological features are realized in their language once they have learned which morphological features are realized, and thus that the earlier stages of acquisition look different than those hypothesized by the models discussed in this section.

5. Methods & Model

In this section, we provide an overview of our model and training data. We begin by introducing the Tolerance Sufficiency Principle (Yang 2016), the mechanism of generalization employed by our learner. We then expand on the notion of collisions as a cue to morphological marking, and the two pieces of evidence motivating the use of this cue: early segmentation (§3) and the Principle of Contrast (Clark and MacWhinney 1987). We then provide an overview of how our model recursively subdivides its input and selects features to split the input on. We conclude by providing an overview of our training data and its properties.

5.1 The Tolerance Sufficiency Principle

The **Tolerance Sufficiency Principle** (Yang 2016) is a threshold for the generalization of a rule beyond the input, motivated by the hypothesis that children use a process productively when it is more efficient to do so. It thus provides a cognitively-motivated threshold for the generalization of a rule beyond the input: given a rule r_1 that is eligible to apply to N items with e exceptions, this rules is productive if and only if:

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

The above equation assumes that the e items have been attested as exceptions, but due to the skewed nature of the child's input discussed in §1, it is sometimes the case that some forms are not explicitly attested either following or violating a rule. The Tolerance Sufficiency Principle thus applies the same computational threshold to determine when there is ***enough*** positive evidence to generalize a rule beyond the input. Say we have a rule r_2 that is eligible to apply to N items in the input, but only $M, M \leq N$ of these items are attested following N; the rest are simply not attested either following or violating the rule. By applying the same computational principles as above, the Tolerance Sufficiency Principle states that this rule may generalize beyond the input if and only if:

$$N - M \leq \theta_N = \frac{N}{\ln N}$$

Note that this is essentially equivalent to applying the first formula with the assumption that every one of the $N - M$ unattested forms is an exception: it thus provides a threshold based on a potential worst-case scenario in the unattested forms. To summarize, the Tolerance Sufficiency Principle employs the same computational threshold when some number of items have been attested violating a rule as when they have not been attested either following or violating the rule.

An interesting mathematical consequence of the Tolerance Sufficiency Principle is that θ_N is larger relative to N for smaller values of N. For example, when $N = 10$, $\theta_N = 4.3$, or about 43% of N. By contrast, when $N = 100$, $\theta_N = 21.7$, or about 22% of N. As reviewed in §3, children learn the morphology of their native language on extremely small input, so

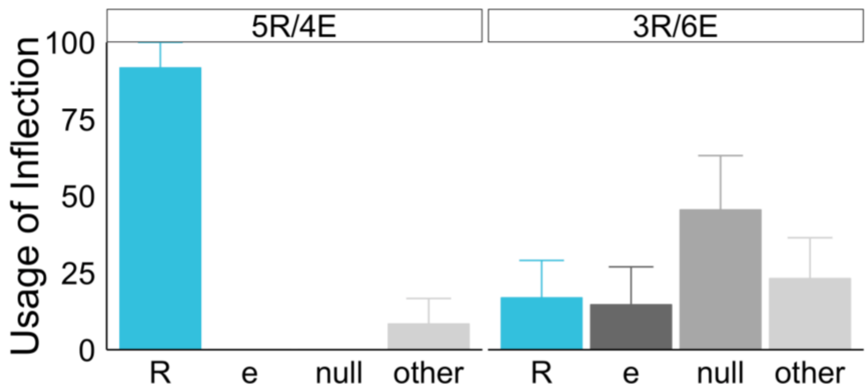

Figure 5.1: The production test results for children in Schuler, Yang, and Newport (2021)'s experiment

this property of the Tolerance Sufficiency Principle is attractive for providing computational accounts of morphological acquisition. Furthermore, the Tolerance Sufficiency Principle can be applied recursively, as exhibited in models such as Belth et al. (2021) and Payne, Kodner, and Yang (2021), and as we will do in the present work.

The Tolerance Sufficiency Principle has found support in several experiments with precisely controlled conditions. Schuler, Yang, and Newport (2021) exposed children and adults to an artificial language consisting of 9 nouns, in which the regular plural marking was *-ka*. Since $\theta_9 = 4.1$, they designed one case with 5 regulars and 4 exceptions, for which we would expect generalization under the Tolerance Sufficiency Principle, and another with 3 regulars and 6 exceptions under which we would not predict generalization. Exceptions took one of several plural affixes, and all forms were attested in the singular and plural in the experiment (Schuler, Yang, and Newport 2021). The input was designed to follow a roughly Zipfian frequency distribution, since the Tolerance Sufficiency Principle assumes such a distribution. After exposure, participants were given a *wug*-style production test in which they were exposed to nonce words they had not heard during training and asked to produce the plural form. Schuler, Yang, and Newport (2021) found that children used the productive suffix *-ka* almost 100% of the time in the 5/4 case, while in the 3/6 case, the most common response was an unmarked stem (see Figure 5.1). This suggests that children did not know what to produce in the 3/6 case, since no productive form was learnt, while they had no trouble generalizing in the 5/4 case as predicted by the Tolerance Sufficiency Principle.

Koulaguina and Shi (2019) exposed 14-month old non-Russian-learning infants to a Russian word order rule via passive listening. The input to the children was 16 sentences of the form ABC, of which 8 were attested following a word order change (e.g. ABC–BAC) and 8 were not. During testing, Koulaguina and Shi (2019) then exposed children to sentences that either followed the attested rule (in this case, ABC-BAC), or some other rule (e.g. ABC–ACB). They found no significant difference in looking time between the attested and unattested rule, suggesting that children had not generalized the rule beyond the input (Koulaguina and Shi 2019). This is consistent with the Tolerance Sufficiency Principle: here we have $N = 16, M = 8$, and thus $\theta_{16} = 5.7$. Clearly $16 - 8 \not\leq 5.7$, and so we would not expect the child to generalize a rule applicable to 16 items but only attested with 8 under the Tolerance Sufficiency Principle. Koulaguina and Shi (2019) then tested 14-month olds on 10 sentences of which 8 followed the rule and 2 were not attested following it, and found a significant difference in looking time between the attested and unattested rule: $\theta_{10} = 4.3$ and $8 - 6 \leq 4.3$, so we would expect children to generalize this rule under the Tolerance Sufficiency Principle. Thus, the performance of 14-month-olds on Koulaguina and Shi (2019)'s tests matches with the predictions of the the Tolerance Sufficiency Principle. However, other mechanisms could also explain the behavior attested by both Koulaguina and Shi (2019) and Schuler, Yang, and Newport (2021), such as a simple majority preference. As such, Emond and Shi (2020) utilize the same Russian word-order change data to test the predictions of the Tolerance Sufficiency Principle more exactly: of the 16 sentences to which they exposed the infants, either 5 or 6 were not attested following the rule. Since $\theta_{16} = 5.7$, we would expect infants to generalize in the 5 case and not the 6; this is exactly what Koulaguina and Shi (2019) find. This pattern cannot be explained by simple majority preference and thus provides stronger experimental evidence in favor of the Tolerance Sufficiency Principle.

5.2 Collisions and Recursion

To learn which morphological features are marked in their native language, we propose that the child makes use of the notion of **collisions**, or a single word appearing in multiple inflected forms. If, for example, the child hears *walk* and *walks* and understands that they correspond to different inflected forms of the same stem, they can use this information as a cue to learn the features that distinguish these forms (in this case, 3,SG, ¬3,SG). As reviewed in §3, English-learning children are able to segment inflected forms and relate these forms to their stems very early in development.[1] This suggests that they are able to make use of collisions by relating, for example, *walk* and *walks*: once they know that *-s* is an English affix, the shared stem between the words is transparent. Furthermore, the **Principle of Contrast**, or the hypothesis that any two distinct phonological forms will be distinct in meaning (Clark and MacWhinney 1987), can provide a cue to learn the morphological features distinguishing the two "colliding" forms. Once infants have related these forms, the Principle of Contrast would cause them to hypothesize that the forms have different meanings, and thus provide a cue to learn the morphological features that distinguish the two forms. The Principle of Contrast is well-supported in the developmental literature; see Clark and MacWhinney (1987) for a review.

However, a single collision is likely not sufficient evidence to determine that a set of morphological features is always marked: in English, for example, we do not mark first vs. second person on most verbs, but do on the verb *to be* (compare *I am* vs. *you are* to *I walk* vs. *you walk*). We would not want the child to extrapolate from the collision they see with *to be* and learn that English contrasts the first and second person on regular verbs. Thus, we use the Tolerance Sufficiency Principle as a measure of when the child has enough positive evidence that a feature set is marked in their language. Our model

1. As discussed in §3, more work must be done to determine at what age children learning other languages are able to accomplish this. For the present work, our model assumes that collisions are available to children acquiring any morphological paradigm.

takes in input incrementally, and for each form in the input, it checks to see if that word has appeared in multiple inflected forms. If so, it considers the two inflected forms that the word has appeared in, and determines the morphological features corresponding to the difference between these inflected forms. It then checks across other words in its lexicon to see if a sufficient number also mark these features. If there is sufficient evidence, as defined by the Tolerance Sufficiency Principle, the model subdivides the input based on the collision into forms having the given morphological features and forms not having those features. Subsequent subdivisions will occur within only one of the resulting branches, leading to a binary branching tree representing which features are marked in the language.

We can formalize the conditions for subdivision as follows: if the model encounters a word in inflected form A and inflected form B, where A is less frequent, then we determine the morphological features corresponding to the difference between A and B. We then search the lexicon for other words that have appeared in inflected form A. If a sufficient number of these words also appear in a *different* phonological form, and the difference in morphological features between these forms is the same as for A and B, then we learn that these morphological features are marked in the language. We focus on the less frequent of the two inflected forms because the number of collisions between A and B will be higher relative to the total count of A than the total count of B, giving it the best chance of passing the Tolerance Sufficiency Principle. For example, if we encounter a collision between 3,PL and 3,SG in English, presuming that 3,SG is the more frequent of these two, if a sufficient number of words that appear in the 3,PL also appear in the 3,SG with a different inflected form, then we learn that English marks $\{3,\text{PL}\} \setminus \{3,\text{SG}\} = \pm\text{PL}$. Note that this model will never learn person distinctions such as 1 vs. 2 in English because a sufficient number of words will not have *different* inflected forms across these two feature sets: our model relies only on (phonological) inequality between inflected forms to generalize markedness.

Since this process will be applied recursively, once we have learned that English marks ±PL, we will learn within the -PL branch that it also marks ±3, i.e. that the third singular is marked. This recursive, incremental application of our model yields a binary-branching, tree-like representation of marking (e.g., see Figure 6.1).

5.3 Selecting Split Features

Note that in the case of 3,PL and 3,SG discussed above, we can either consider the split feature to be $\{3,\text{PL}\} \setminus \{3,\text{SG}\} = \pm\text{PL}$ or $\{3,\text{SG}\} \setminus \{3,\text{PL}\} = \pm\text{SG}$. That is, we can either take the marked features to be the difference of the less and more frequent features, or the more and less frequent ones. Since we divide the vocabulary when there is sufficient evidence for marking, the difference between these approaches has a number of interesting implications. Note that in our example, it does not matter which heuristic we use for forms in which number is marked: plural forms will either be sent to the +PL branch or, equivalently, the −SG branch. However, for forms that do not mark number, such as infinitives, the choice of heuristics will cause different behavior: if we choose ±SG, then such forms will be sent to the -SG = + PL branch, but if we choose ±PL, such forms will be sent to the -PL = + SG branch. Such differences are particularly interesting in more complex morphological systems such as Spanish and Hebrew.

Further, note that we are not guaranteed to have a singular feature as our set difference. There may instead be a collision between 3,SG and 2,PL, which would lead to a set difference of either ±3,SG or ±2,PL. In these cases, we may either consider all features in the set difference together when determining if there is sufficient evidence for marking, or we may iterate through each feature and consider it separately. During model development, we tested both approaches with both orders of set difference, and found that the approach in which we consider only a single feature at a time prevented the model from learning as many correct markings. We hypothesize that this is due to the fact that considering a single feature will lead to larger values of N in the Tolerance Sufficiency Principle, which, as reviewed is §5.1, will allow for fewer relative exceptions. As such, we consider the entire set difference in the present work.

When deciding between possible orders of features in the set difference, we found that the difference of the least frequent \ most frequent provided the best results for our model. The opposite approach often yielded extremely long set differences that led to small values of N in one of the resulting nodes after branching. However, these results are purely qualitative, and future work should investigate other potential approaches to selecting split features. Additionally, since our model still relies on explicit specification of morphological features in the input, future work should consider how children go about extracting these features from their environment and input once the phonological collision provides them with a cue to do so; see §7 for further discussion.

5.4 Training Data

For our training data, we intersect the most frequent forms in CHILDES (MacWhinney 2000) for each language with the morphological features provided by UniMorph (Kirov et al. 2018). We choose the most frequent forms in CHILDES to simulate the child's vocabulary since words that are frequent in the input are likely to be learned earlier (Goodman, Dale, and Li 2008). Intersection with UniMorph leads to input that is a triple of the stem, inflected form, and corresponding features, for example: (*walk, walks*, {VERB, PRESENT, 3RD, SINGULAR}).

Our English data is taken from the Manchester, Wells, and Belfast corpora (Theakston et al. 2001; Wells 1981; Henry 1995) and contains 4196 inflected verb forms corresponding to 1280 unique verbs. It is the same data used by Payne, Kodner, and Yang (2021). Our German data is taken from the Leo corpus (Behrens 2006) and contains 1787 inflected noun forms (either singular or plural) corresponding to 1444 unique nouns.[2] Note that for English and German, the number of unique words in the training data exceeds the number that the child will know before acquiring these systems. However, the model does not make use of all of these forms, as we will see in §6.1 and §6.2; learning is complete well before 500 inflected forms in both cases.

Our Spanish data is taken from the Fernández/Aguado corpus and contains 1000 inflected verb forms corresponding to 310 unique verbs. This data is also taken from Payne, Kodner, and Yang (2021). Finally, our Hebrew data comes from the Bat-El, Berman, BSF, Levy, Na'ama, and Ravid corpora (Berman 1990, 2017). Due to the challenges of differing

2. I am grateful to Jordan Kodner for providing this data.

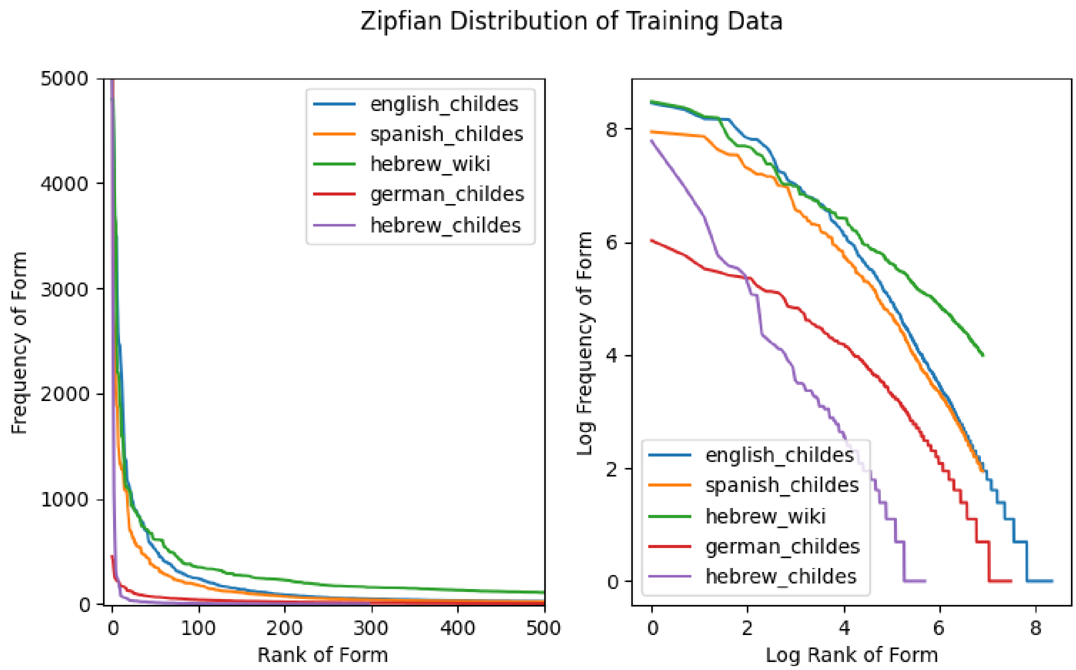

Figure 5.2: A visualization of the Zipfian frequency distribution of the training data in each of the four paradigms considered here. The left plot shows raw rank vs. frequency and the right shows the log-transformed result.

transcription systems across CHILDES and UniMorph, this data contains only 298 inflected verb forms corresponding to 151 unique verbs.

To augment our Hebrew data, we utilize the 100k sentence 2021 Hebrew Wikipedia dump provided by the Leipzig corpus collection (Goldhahn, Eckart, and Quasthoff 2012). Although this corpus does not represent child-directed speech, Kodner (2019) shows that type frequencies of morphophonological and syntactic-semantic patterns are comparable for high-frequency items in child-directed corpora and literary corpora, meaning that the high-frequency vocabulary in the Wikipedia corpus can effectively be used as a proxy for the early Hebrew lexicon. Intersecting the word frequencies of this corpus with UniMorph yields a set of 9,535 inflected forms corresponding to 990 unique verbs. From these, we take the 1000 most-frequent inflected forms to be our Hebrew training set, corresponding to 373 unique verbs. While the UniMorph+CHILDES Hebrew data utilizes ***niqquds*** to explicitly mark vowels, the UniMorph+Wikipedia data does not, and is instead written in ***ktiv hasar niqqud*** format, a standard format for writing Hebrew without the explicit specification of vowels provided by niqquds. The difference in these transcription styles has several important implications discussed in §6.4: since the UniMorph+Wikipedia data does not contain niqquds, there are cases where multiple inflected forms that are phonologically distinct have identical transcriptions. In the future, we hope to develop and use corpora which make use of niqqud or phonological transcription in order to avoid this issue, but due to limited resources and the standardization of ktiv hasar niqqud transcription in Hebrew, this was not possible in the present work.

As discussed in §1, it is well known that the frequencies of words in child-directed-speech

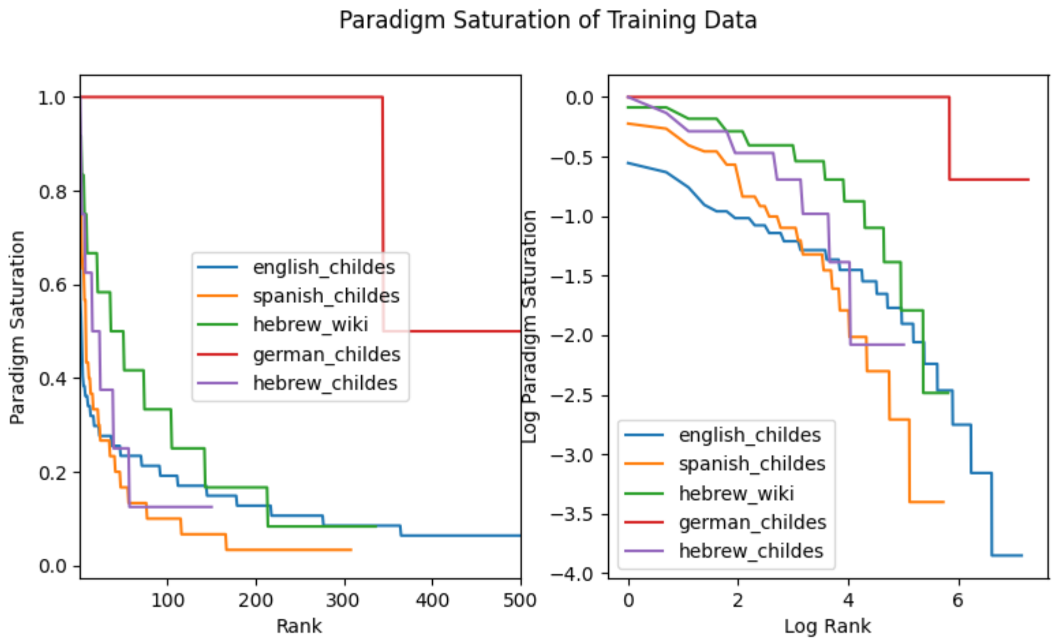

Figure 5.3: A visualization of the Zipfian paradigm saturation distribution of the training data in each of the four paradigms considered here. The left plot shows raw rank vs. frequency and the right shows the log-transformed result.

follow a Zipfian distribution, with the frequency of a word being inversely proportional to its rank (Zipf 1936; Yang 2016; Lignos and Yang 2016). This can be visualized in Figure 5.2: the left subplot shows the raw rank vs. frequency of items in each of our training sets. As we can see, only a few of the inflected forms in the input occur thousands of times; most lie in the long tail of forms of rank 200 and above for which frequency is quite low. In addition to the overall Zipfian frequency distribution of the input, the paradigm saturation, or number of possible inflected forms that a given stem appears in, is also known to follow a Zipfian distribution (see §1, Lignos and Yang (2016)). This means that a few stems will appear in close to all of their inflected forms, but most will only appear in one or two of their possible forms in the child's input. We can see this pattern in our training data in Figure 5.3. Paradigm saturation in Figure 5.3 is normalized to be proportional to the maximum saturation in the training set in order to plot the paradigms together: otherwise, maximum saturation is much lower for morphologically rich languages such as Spanish and Hebrew than languages such as English, in which there are only a few forms in which a word could appear. In the left subplot of Figure 5.3, we can see that only a few stems are highly saturated in the training data, and most have saturations of well under a half, especially for the long tail of our English input. Our German input is an outlier simply because we are considering a paradigm in which there are only two potential forms: singular and plural. The higher saturation observed here is thus unsurprising.

The relative sparsity of the Zipfian "long tail" in both Figures 5.2 and 5.3 dictates that most inflected forms will appear only a few times in the child's input, and most stems will appear in only a fraction of their possible forms. This is an important problem for the child learner to overcome by generalizing beyond the input (see §3) and is a key motivation of the Tolerance Sufficiency Principle (see §5.1). The necessity to generalize well beyond the input is also a key motivation of the model presented here.

6. Results

We now present the results of our model on each of the four paradigms. These results are in the form of "recursion trees," which allow us to visualize the order and relationship between the features that the model has learned are marked in each language. For example, examining the English recursion tree in Figure 6.1 shows us that the first distinction the model learns is ± V.PTCP. Then, within the verbs that have the feature − V.PTCP, the model recursively applies the same algorithm to learn that ± SG is marked. That is, the model first finds that there are sufficient collisions that realize ± V.PTCP to subdivide its vocabulary with this feature set. Then, within the set of words that are − V.PTCP, there are sufficient collisions to indicate that ± SG is marked (i.e. non-participial verb forms in English mark number), and so on. In these trees, the number on each node corresponds to the total number of inflected forms that the model has taken in when it finds sufficient evidence of marking, and the + and - features on each binary branch correspond to forms that realize these features and forms that do not, respectively. Returning to our English example, we see in Figure 6.1 that the model learns that ±V.PTCP is marked in the language after hearing 70 inflected verb forms, and learns that non-participial verb forms mark number at 72 verb forms, just after learning that ± V.PTCP is marked. Nodes that read "leaf" correspond to nodes at which no more productive collisions were found. For example, in Figure 6.1, we see that leaf 0 corresponds to the participle (i.e. the set of verb forms realizing + V.PTCP), for which no more divisions are possible. The numbering on the leaves is arbitrary and simply provides a label to be used during discussion.

6.1 English verbs

Figure 6.1 gives the results of our model on English verbs. As we can see, despite the large training data, the model does not learn anything further after the first 449 inflected forms, which correspond to 188 unique verbs. This fits well with the developmental findings outlined in §3.1: if the child's vocabulary consists of 500-1000 words at the time they acquire the past tense, we can expect that about a quarter of these will be verbs (Bornstein et al. 2004), and the model learns that English marks the past tense on under 250 unique verbs.

The order of acquisition, which we can visualize by the numbers in the nodes, also matches well with developmental findings: the model first learns that the present participle is marked at just 70 inflected forms, followed by the third singular at 82 inflected forms. This matches well with the findings of Brown (1973) and Berko (1958) that the acquisition of English *-ing* and *-s* precede the acquisition of the past tense. The past tense appears last, with paths to leaves 1-6 marked with ±PST.

Interestingly, our model divides the past between multiple nodes in the tree, despite there being only one productive process for the English past tense. Note that this is necessary under the recursive subdivision of our model: if it first learns that ±3,SG is marked, then it must learn ±PST separately within each of the -3,SG and +3,SG groups based on the recursive nature of the learning procedure. Whether or not this approach can match well with developmental findings is an interesting direction for future work. One fruitful direction

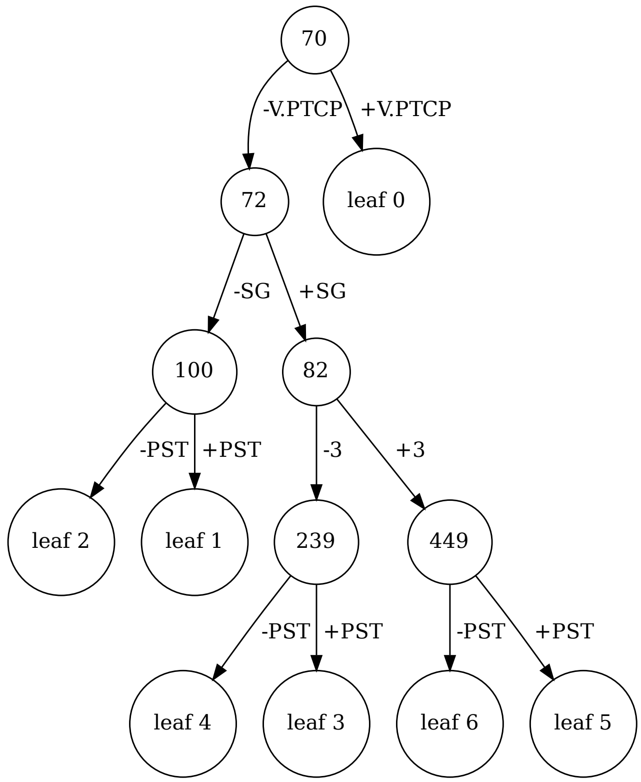

Figure 6.1: The recursion tree produced by our model on English verbs

to pursue would be to consider the effects of these higher morphological features on the lower ones. For example, Yang, Ellman, and Legate (2015) show that the past tense is generally acquired later for learners of African American Vernacular English (AAVE) than those of Standard American English (SAE), despite the fact that past tense marking does not differ across these dialects. The key difference between the inflectional morphology of these dialects is the marking of the third singular (`-s` in SAE and $-\emptyset$ in AAVE), which may shed light on the problem at hand. If the child does not have sufficient evidence to make divisions based on ±SG and ±3, then they will have all non-participial verbs together, leading to a much higher N in a single category, vs. the smaller Ns we see across the three categories in Figure 6.1. As reviewed in §5.1, the Tolerance Sufficiency Principle tolerates fewer relative exceptions for larger values of N. As such, we might expect that learners of AAVE would struggle more to acquire the past tense since they will not subdivide their input based on ±SG and ±3 and thus will be acquiring it over a larger N than our model predicts for learners of SAE, but future work should explore this explanation further.

For the present, we will leave this question open to future research, but note that a large part of answering this question will be incorporating meaning-form mappings into the present work. If, for example, we create mappings between all leaves in the tree corresponding to the form changes between those morphological features, then we can expect that several of

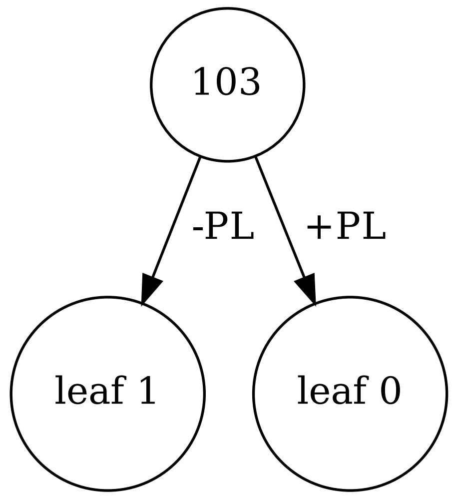

Figure 6.2: The recursion tree produced by our model on German noun plurals

these mappings will simply be redundant in the case of the past tense. Another option is to move away from the hierarchical structure and recursive learning we have posited here, or even incorporate some sort of parallelism or multidimensionality in the learning mechanism. Such endeavors are discussed more in §7 but are beyond the scope of the current work.

6.2 German noun plurals

Figure 6.2 shows the results of our learner on German noun plurals. Since plurality is a single morphological feature, these results are quite simple, with one node corresponding to singular and the other to plural nouns. We see that the learner acquires this morphological distinction when it has been exposed to 103 inflected forms, corresponding to 97 unique nouns. This fits well with the developmental findings reviewed in §3.2. We know that the average vocabulary size when plural overapplication begins is under 300, and we expect nouns to make up around half of this. Our model successfully learns that German marks the plural before it has encountered 150 unique nouns.

Since the tree given in Figure 6.2 gives such a clean categorization of nouns into singular and plural, it would be interesting, and relatively straightforward, to combine this with a model that maps morphological features to form. In particular, the model of Belth et al. (2021) maps German singular nouns to plural nouns, but does not have a way to learn that plurality is marked in German and instead takes this information as a prerequisite. In the future, we plan to combine such approaches with the present model; see §7 for further discussion.

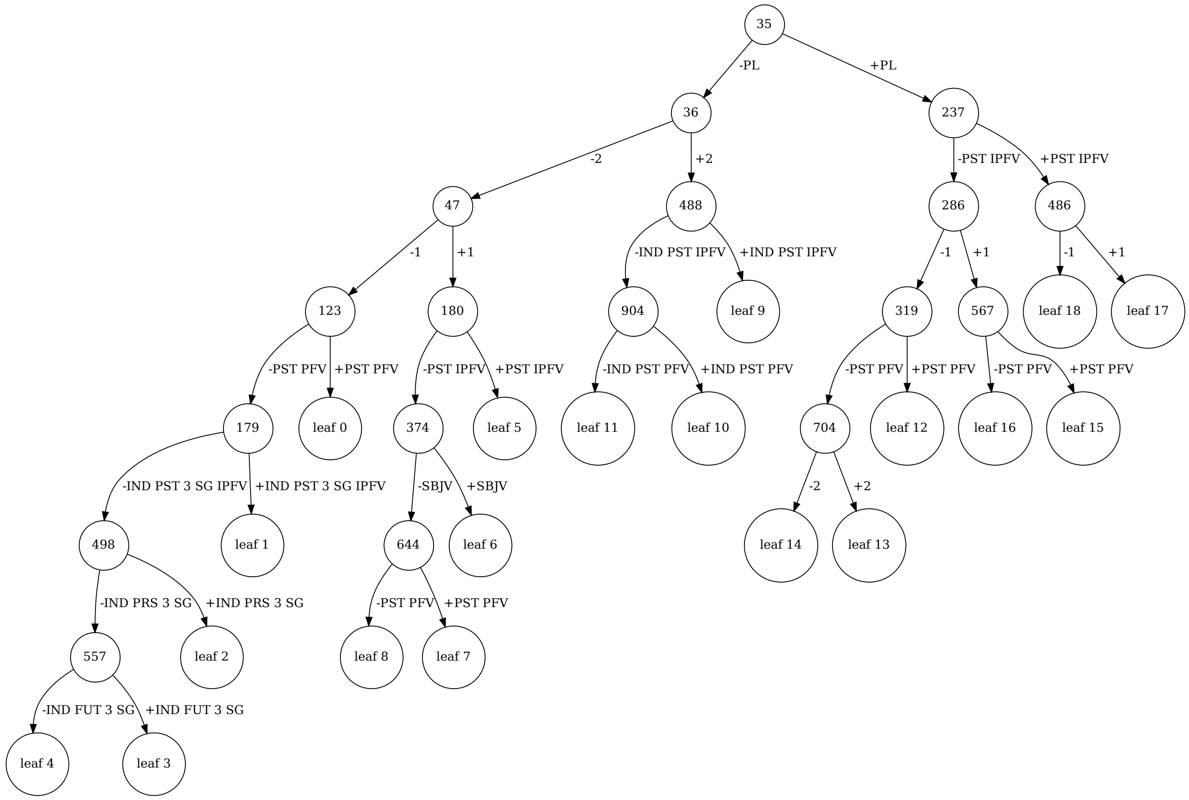

Figure 6.3: The recursion tree produced by our model on Spanish verbs

6.3 Spanish verbs

Figure 6.3 shows the results of our model on Spanish verbs. We see that the model learns that Spanish marks the plural on just 35 inflected forms, corresponding to 17 unique verbs. Since person emerges before number in children as reviewed in §3.3, this might initially seem to indicate that the model is failing to match developmental findings. However, note that person marking emerges in the -PL branch by just 47 inflected forms, at which point the model has learned that the language marks ±2 and ±1. These 47 inflected forms correspond to 19 unique verbs, which matches well with developmental findings. The child's vocabulary will be about 156 words at the time person is acquired and a quarter of this vocabulary will be verbs (Bornstein et al. 2004), and the model successfully learns the person distinctions of Spanish on less than 39 verbs.

In contrast to the singular, person marking in the +PL branch does not emerge until at least 286 inflected forms, with the second plural not emerging until 704 inflected forms (leaf 13), and even then, not emerging consistently across tenses. This relatively late acquisition of the second plural matches well with the findings of Gaya (2001) – that children do not produce any second plural forms before 2;6. Indeed, the second plural begins to emerge when our model has heard 229 unique verb forms, which would approximately correspond to an overall vocabulary size of 687, and this will likely not be achieved until 2;6 or later

(Bornstein et al. 2004). This also provides an experimentally testable hypothesis regarding the acquisition of Spanish subject marking: in contrast with previous work, our model posits that children learn number marking first, but do not learn person marking in the plural until well after they have acquired it in the singular.

With regards to tense, we see that our model begins acquiring tense, marked with ±PST,PFV and ±PST,IPFV, at 179 inflected forms in the third singular (leaf 0), but that this marking comes much later in other person-number combinations (e.g. in the second singular, leaf 10). This fits well with the protracted development of the imperfect noted by Jacobsen and Meisel (1986), but less well with the quick emergence of the preterite at 2;0 noted by Montrul (2004). This deviation from developmental findings may mirror the division of the English past tense discussed in §6.1, and is certainly an interesting direction for future investigation. In fact, this case provides an interesting counterpoint to the English past tense, since the Spanish past is marked across all person-number categories, but marked *di erently* for each, either via agglutination with subject agreement morphemes (§3.3 ex. 1) or fusional suffixes that mark both subject and tense agreement (§3.3 ex. 2). In contrast, the English past is always marked the same regardless of person and number. How this affects the interaction of subject marking and tense in acquisition is an open question and fruitful direction for future research. Despite the piecemeal acquisition of tense our model predicts, however, it successfully acquires tense on less than 300 unique verbs, which matches with the upper end of vocabulary sizes reviewed in §3.3. Further, our model predicts that the future will emerge well after the preterite, which fits well with developmental findings (Gaya 2001).

We see that the subjunctive is only acquired in the first singular at 374 inflected forms (leaf 6). It may be the case that too little data is present for the model to acquire the subjunctive on other subject agreement forms, or that the syncretism that it has with the imperative inhibits our model. A similar issue could also be occurring as with the past tense discussed above. However, note that the subjunctive and imperative follow a much more protracted development (López Ornat et al. 1994; Montrul 2004), so the results here are not unreasonable and reflect that the Spanish-learning child has acquired most of the verbal inflectional morphology of their language by 2;2 (Montrul 2004).

6.4 Hebrew verbs

Figure 6.4 shows the results of our model on the small CHILDES Hebrew training data with *niqqud*. We see that the initial distinction made by the model at 91 inflected forms corresponds to person and gender, which differs from the typical order of gender and number before person but is still attested in the developmental literature (Bat-El 2014). Future tense emerges along with person, number, and gender at 92 inflected forms (61 unique verbs), followed by the infinitive at 126 forms (77 unique verbs, leaf 4). Thus, the tree does show patterns that seem to match with the developmental findings reviewed in §3.4, with subject agreement emerging before tense. We do not see any binyan conjugations emerging, which makes sense given the small nature of our vocabulary and the fact that binyan conjugations do not emerge until later in development (Berman 1983).

Figure 6.5 shows the results of our model on the larger Hebrew Wikipedia training data

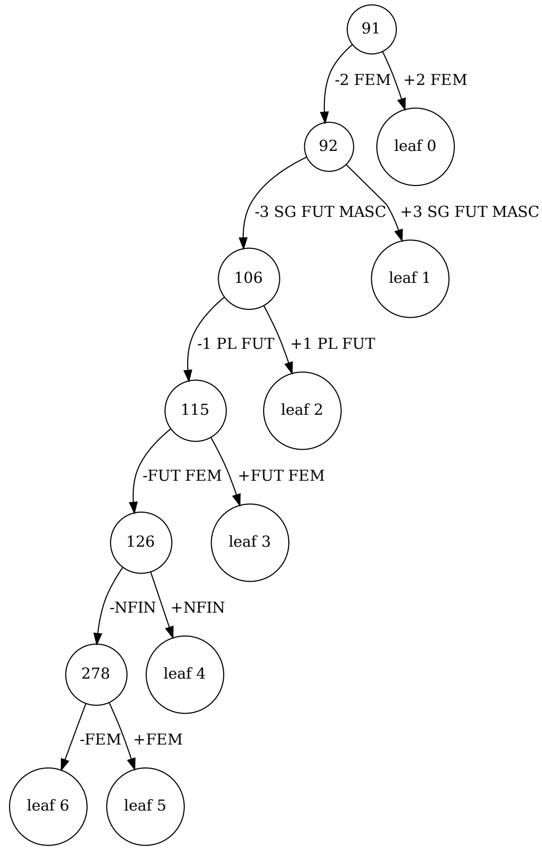

Figure 6.4: The recursion tree produced by our model on Hebrew verbs with the CHILDES training data

without niqqud. As we can see, the first distinctions that emerge are ones of gender (± FEM at 48 inflected forms or 29 unique verbs), followed by number and then person. This matches better with the order of acquisition reviewed in §3.4 than the order on the CHILDES data shown in Figure 6.4. We see that future tense begins to emerge after the model has taken in 485 inflected forms (205 unique verbs), but that the infinitive and imperative do not emerge here; the former is in contrast with Figure 6.4.

It is also worthwhile to note the differing vocabulary sizes at which these patterns are learned: while agreement emerges by around 100 inflected forms in the CHILDES-trained model, it takes almost three times as much data for person agreement to emerge in the Wikipedia-trained model. Tense emerges even later in the Wikipedia-trained model, at close to 500 inflected forms, and is incomplete even here: we only see partial evidence for the future tense and no indication of past tense. While only future tense emerges in the CHILDES-trained model as well, this model is exposed to significantly less data: only 151 unique verbs. While the vocabulary size of the Hebrew-learning child at 3;0, approximately when tense is acquired, is not documented, Bornstein et al. (2004) find that the vocabulary of a Hebrew-learning child at 1;8 is slightly larger than that of the English-learning child,

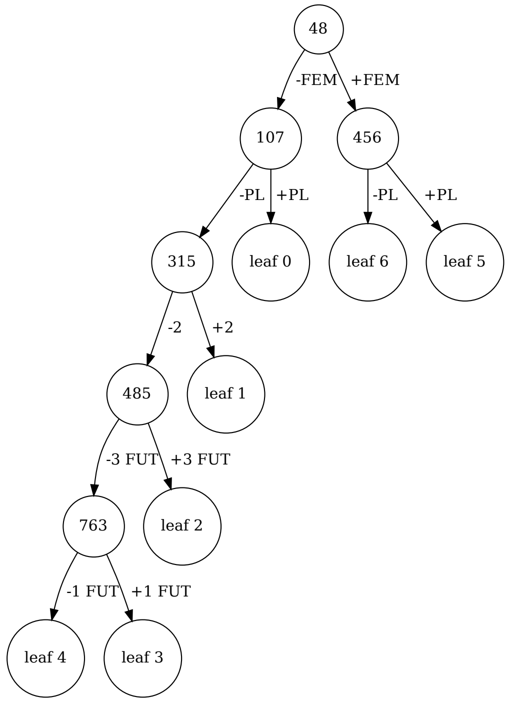

Figure 6.5: The recursion tree produced by our model on Hebrew verbs with the Wikipedia training data

and Brown (1973) shows that English-learning children at about 3;0 have vocabularies of at most 1000 words. Even at the upper limit, we can expect about a quarter of the vocabulary to be verbs, meaning that if we analogize from English to Hebrew, we would expect tense to be acquired in Hebrew well before 300 unique verbs. Yet the Wikipedia-trained model only shows knowledge of the future tense after being trained on 373 unique verbs.

A likely reason for the late acquisition by the Wikipedia-trained model and the difference between the two models is that the CHILDES data marks niqqud and the Wikipedia data does not. Since Hebrew verbs are templatic (§2.4), the vowels carry morphological value, and omission of them may lead to syncretisms in orthography that are not reflected in phonology. For example, the ktiv hasar niqqud transcription כתבת could be read as either *katavta* "you (masculine singular) wrote" or *katavt* "you (feminine singular) wrote." Not only are these forms morphologically distinct, but they are also phonologcally distinct, a distinction that is not reflected in the orthography. Since the present model explicitly relies on orthographic distinctions as cues to marking, it makes sense that an underspecified orthographic system such as ktiv hassar niqqud would cause the model to need more data, since this underspecification would prevent the model from finding sufficient evidence for marking. Furthermore, since the Tolerance Sufficiency Principle tolerates proportionally more exceptions for smaller values of N, this fact, compounded with the orthographic underspecification creating a need for more data, will make it difficult to learn anything on small vocabularies. This matches well with what we observe in Figure 6.5.

Despite the shortcomings of the Wikipedia orthography, however, the model is still able to learn in approximately the correct order in both Figures 6.4 and 6.5, although it certainly performs better with regards to ordering in Figure 6.5. In the future, we hope to obtain phonologically transcribed Hebrew data for which there is not orthographic underspecification. Since our model shows promise on both data sets here, despite issues with each, we are hopeful as to its performance on cleaner data that more accurately represents the collisions in the language.

7. Discussion

The model presented in this thesis provides a possible account of how children learn what morphological features are marked in their language. The model relies on children's early ability to segment and relate inflected forms in their input (§3) in order to detect collisions – a single word appearing in multiple inflected forms. It also relies on the Principle of Contrast (Clark and MacWhinney 1987) as a cue to learn the morphological features distinguishing the colliding forms, and the Tolerance Sufficiency Principle (Yang 2016) to generalize from individual collisions to obligatory marking across the language. The model closely follows developmental patterns regarding order of acquisition and vocabulary size on a typologically diverse set of morphological paradigms (§6).

One immediate future goal for this work is to acquire and run the model on Hebrew data that is more representative of the child's input. Such data could be phonologically transcribed, romanized, or transcribed with ***niqqud***; the important requirement for our model is that orthographic syncretisms reflect phonological syncretisms in the data. Besides this immediate goal, this thesis also points to several broader directions for future work. These include combining our model with models that distributionally learn morphosemantic information from the input and models that map morphological features to phonological form. It also involves experimentally investigating predictions put forth by the present model and revising it based on the experimental evidence.

7.1 Distributionally Learning Features

Currently, our model provides an account of detecting when feature(s) are marked that relies on inequality between inflected forms and children's early ability to relate these forms (Kim and Sundara 2021). We hypothesize that once the child has detected a collision between two inflected forms in the input, that they are then able to use this information, in conjunction with the Principle of Contrast, as a cue to learn the features that distinguish these forms. Yet our model still explicitly relies on UniMorph (Kirov et al. 2018) to provide the necessary features. In contrast to the explicitly specified features our model receives, the child must extract these features from their input and environment. As such, another interesting direction to explore is combining a learning model such as ours with grounded learning models.

The form-based heuristic we use to determine whether there is sufficient evidence for marking could be used in conjunction with models of grounded learning to provide a more plausible account of the acquisition of morphological marking. In particular, grounded word learning models such as Pursuit (Stevens et al. 2017) and the more recent Memory-Bounded implementation of Pursuit (Soh and Yang 2021), provide a promising avenue to explore the extraction of morphosemantic information from the input. Gabbard et al. (2021)'s implementation of Pursuit shows success extracting the necessary semantic features to acquire Mandarin Chinese classifiers, and such approaches may be able to generalize to morphosemantic information such as person and number.

Grounded word learning models map between words in the input and features in the environment that infants may attend to – these include objects, actions, and gaze in the case of Gabbard et al. (2021)'s implementation. A growing body of work shows that infants can also attend to cues such as person and number in their input: conceptual knowledge of the singular-plural distinction, for example, has been shown to emerge at roughly the same age cross-linguistically, regardless of when it emerges in the child's morphosyntactic development (Li et al. 2009). Barner et al. (2007) show that 22-month-old English-learning infants can represent sets of 4 objects as "plural" or "more than one," and Li et al. (2009) extends this finding to Japanese- and Chinese-learning children. Since plural marking of English nouns is acquired around 22-24 months (Brown 1973) but plurality is not obligatorily marked in the morphosyntax of Japanese or Chinese, Li et al. (2009)'s findings suggest that the conceptual knowledge of number emerges independently of language. Before this knowledge emerges, infants are in a stage of "parallel individuation" in which they can represent up to three individual objects at a time in working memory as separate symbols but don't yet have general symbols for "one" or "more than one" (Feigenson, Carey, and Hauser 2002; Feigenson and Carey 2005; Barner et al. 2007; Li et al. 2009). Most studies (see Li et al. (2009) for a review) have found this stage to last until about 22 months, but Barner et al. (2009) suggest that it may only last until 15 months if probed under the right experimental conditions. This is particularly interesting in light of the early emergence of number during the acquisition of Spanish verbs at around 20 months (§3.3); children must have some conceptual knowledge of number to understand that it is obligatorily marked in the grammar.

It is slightly more challenging to understand the conceptual development of person than number, but it has been shown that the child develops an understanding of the self as different from others between 12-24 months. Dixon (1957) tracks infants' developing self-image by observing their interactions with their mirror reflections between 4 and 12 months. He finds that by 12 months, infants show a clear knowledge of self vs. other as demonstrated by their preference to interact with others instead of their reflection (Dixon 1957). However, as Butterworth (1992) points out, self-perception at these stages doesn't necessarily imply self-conception. Amsterdam (1972) famously found that infants between 20-24 months would pay attention to a dab of rouge on their nose when placed in front of a mirror, but younger infants would not, suggesting that awareness of the self as having stable facial features emerges around this time. Lewis and Brooks-Gunn (1979) suggest that this awareness may actually occur as early as 15 months if probed under the right conditions. They suggest that the infant begins understanding the self as an enduring entity with unique qualities by the end of the first year. More recently, Kampis et al. (2022) demonstrate that children between 16-26 months can map from others to themselves by placing a sticker on their own face in the same position that a sticker was placed on a parent's face. Whether or not this understanding of self vs. other is directly related to the acquisition of person is an open question, but it is evident that children have some of the conceptual knowledge in place to extract person information from their environment by the end of their second year, if not earlier.

The evidence presented thus far suggests that person and number are environmental cues to which the child can attend and map phonological form. However, the case of tense and aspect provide a more complex picture, since the cues to tense are typically less direct in the environment. It is clear, however, that infants have an early understanding of temporal ordering: they are able to understand the ordering of actions in service of a goal as early as 8 months (Piaget and Cook 1952; Piaget 2013b). They are also able to partially reconstruct past events without perceptual support by the age of 18 months, including by remembering where a toy was placed or where a caregiver went (Piaget 2013a). Despite this conceptual knowledge, it may be the case that the most salient cues to tense and aspect are distributional and part of the child's linguistic input. For example, if the child has acquired the meanings of *yesterday*, *today*, *tomorrow*, *soon*, and other adverbs that indicate tense, we could expect them to use bootstrapping techniques to understand the tense or aspect that is marked on the verb in question. Extensive evidence has demonstrated that children are capable or these sorts of distributional learning and bootstrapping tasks (e.g. Gerken, Wilson, and Lewis (2005), Stevens et al. (2017), Koulaguina and Shi (2019), and Schuler, Yang, and Newport (2021) for distributional learning, and Landau, Gleitman, and Landau (1985), Naigles and Hoff-Ginsberg (1995), Naigles (1996), and Fisher et al. (2010) for bootstrapping). An exciting future direction for our model, then, is to combine it with grounded word learning – and potentially distributional learning – models in order to extract morphosemantic information from the input and environment rather than explicitly encoding features in the input to the model.

7.2 Mapping Features to Form

As discussed in §4, most previous work on modeling morphological acquisition has focused on mapping from morphological features to their phonological realizations. An interesting direction for future work is thus to use such models to map the features our model acquires to their corresponding phonological realizations. There are a number of possible approaches to integrating our model with models that map features to form, but a particularly straightforward one would be to use the model of Belth et al. (2021) to map between the leaves of the recursion trees produced by our model, as discussed for German in §6.2.

While Clahsen, Aveledo, and Roca (2002) suggest that acquiring markedness immediately triggers generalized mappings between features and phonological realizations (by proxy of developmental regression, §3.1), other evidence suggests that children know features are marked in their language before they know the productive processes that realize them phonologically. For example, Wagner (2001) shows that children as early as 2;9 have an adult-like understanding of tense in comprehension, although most children do not mark tense in an adult-like way for another several months (Brown 1973). Zapf and Smith (2009) demonstrate that English-learning children have knowledge that the plural is marked in English before they productively produce *-s*. When presented with both familiar and nonce items and asked to label them, children produced the singular form significantly less when they were presented with a single object followed by two objects of the same type ($A \rightarrow AA$) than when they were presented with a single object followed by an object pair of different types ($A \rightarrow AB$), but rarely used the plural in both cases (Zapf and Smith 2009). Indeed the most common production by the children in the first case was nothing, which Zapf and Smith (2009) take as evidence that "they know in some way that the two things of the same kind get a different label than does one thing" (1150). That is, English-learning children have some knowledge that the plural is marked *before* they know how to mark it. This difference between conditions for English-learning children is in contrast to Japanese-learning children of the same age: Japanese does not obligatorily mark number on nouns, and the Japanese children in Zapf and Smith (2009)'s study offered the singular form significantly more than English-learning children in the $A \rightarrow AA$ case. This suggests that the difference across the two cases for English-learning children is caused by their linguistic knowledge of marking rather than any other sources.

It is still possible that Clahsen, Aveledo, and Roca (2002)'s hypothesis does hold for some paradigms or some children, and our model could account for this. If, for example, we were to combine our model with the model of Belth et al. (2021), whether or not there would be a productive mapping from features to phonological realization at a given node in our recursion tree would be entirely governed by the nature of the forms sent to that node during subdivision. It is entirely plausible that in some cases, the model would subdivide, for example, based on ±PAST, and that there would be an immediate generalization in one of the resulting branches based on the forms that are sent to that node. For example, +PAST→ *-ed* might immediately be productive at the +PAST node, yielding the behavior Clahsen, Aveledo, and Roca (2002) describe. It is also possible, however, that no generalization is possible at that node until more forms are learned and sent to the node, which would mirror the patterns put forth by Wagner (2001) and Zapf and Smith (2009).

Integrating our models with ones that map features to form may also yield an explanatory description of the widely-attested Root Infinitive (RI) stage in morphological development (see e.g. Legate and Yang (2007) and Kupisch and Rinke (2007)). At this stage in development, children use non-finite verbs in root clauses, despite this pattern being ungrammatical in the adult grammar. Some examples of this phenomena include (taken from Legate and Yang (2007)):

(4) a. Papa have it. (English)

b. Lashevel al ha-shulxan. (Hebrew)
sit$_{[+\text{INF}]}$ on the-table.

c. Dormir petit bébé. (French)
sleep$_{[+\text{INF}]}$ little baby.

d. mein Kakao hinstelln. (German)
my cocoa put$_{[+\text{INF}]}$

This pattern may be explainable if we combine our understanding of the acquisition of marking and the mapping from features to form: at earlier stages in the acquisition of marking, groupings of forms will be far too broad – consider the first few nodes on the English tree in Figure 6.1, which would only separate participial and non-participial forms. It may be the case that at some point during acquisition, an underspecified node in the recursion tree that will later be subdivided contains enough infinitives that their phonological realization generalizes under models such as Belth et al. (2021)'s; this would correspond to a case of weak Blind Alley Development described by Dressler et al. (2019). More likely, the early underspecification during the acquisition of morphological marking causes enough phonologically heterogeneous forms to be grouped together that no phonological realization generalizes and children fall back on the RI. Later cues from collisions and subdivision at the node in question will allow the model to recover from this by grouping more closely related forms as the vocabulary size increases, and forming generalized mappings from features to form over these groups. Thus, understanding the interaction between the acquisition of marking and the mapping from features to phonological form will be an important step in understanding the RI phenomenon.

7.3 Exploring Model Predictions

Our model makes several interesting predictions regarding order of acquisition that can be empirically tested. Firstly, as discussed in §6.1 and §6.3, our model predicts that in cases where subject agreement is acquired before tense agreement, the marking of tense will appear at different points for each subject agreement. It is interesting to contrast the cases of English and Spanish with regards to this attribute of our model, since in the former, the past does not differ based on subject agreement, but in the latter it does. In §6.1, we discussed one possible reason why such a treatment could be appealing with regards to larger N tolerating fewer exceptions, but future work should further explore this question. One way to do this is to further investigate cross-dialectal differences to better understand how overt subject marking in the third singular interacts with the acquisition of tense in English. Another direction for future work is to explore corpora of children's early speech and determine if tense marking emerges at different points for different agreement markers robustly across children. Our model makes the prediction, for example, that the English past tense will emerge in the plural well before it emerges in the singular (Figure 6.1), and this prediction is easily testable with corpora of child speech. Another prediction made by our model is that number marking emerges earlier than person marking in Spanish, but that person marking emerges earlier for -PL than +PL forms. This is again testable: elicitation or comprehension studies could look at when children are sensitive to number and person in Spanish, for example by using a similar paradigm as Zapf and Smith (2009).

If the predictions our model makes regarding more piecemeal acquisition are incorrect, there are several potential revisions that could be made to the model. As discussed in §6.1, this piecemeal prediction derives directly from the recursive nature of our learning algorithm, so this would suggest that something other than recursive subdivision is used by children. A number of other options exist: one could simply not divide the input when there is evidence that a feature is marked so that all features are learned at the same "level" in the recursion tree. This implementation would be challenging under the Tolerance Sufficiency Principle, however, since a lower relative number of unattested forms are tolerated for larger N, meaning that it may be harder to acquire marked features that are learned later. Another option is to propagate up the recursion tree when a split feature is found to see if it applies more broadly. For example, in English, when we learn that ±PAST is marked in the −SG branch at 100 inflected forms (Figure 6.1), we could recurse up the tree and see if it is marked in all -V.PTCP forms, which would allow us to form a broader generalization regarding tense marking. This would prevent our tree from being binary-branching but would likely eliminate the larger N issue discussed above, and is thus a promising direction for future modeling work.

Broader experimental verification of the learning strategies hypothesized here is also an important direction for future work. One potential way to approach this would be through artificial language studies with carefully controlled conditions. Such studies would allow us to verify that children's markedness generalizations are governed by the Tolerance Sufficiency Principle. Li and Schuler (2021) hypothesize a similar distributional learning algorithm for the acquisition of recursive syntactic structures and verify it through such an artificial language study; similar techniques could be used to validate the approach presented here. In addition, future experimental work should investigate infants' sensitivity to collisions, which is a crucial component of our model. Investigating infants' sensitivity to these cues will be especially important for paradigms such as Spanish verbs and Hebrew verbs for which there is little current evidence surrounding early morphological segmentation by infants.

8. Conclusion

In this thesis, I have presented a computational account of how children may learn what morphological features are marked in their language. I proposed that children make use of their early ability to segment and relate inflected forms in their input in order to find collisions – a single stem appearing in multiple inflected forms. I suggest that these collisions, in conjunction with the Principle of Contrast (Clark and MacWhinney 1987), provide a cue for the child to learn which features distinguish individual pairs of inflected forms, and that the Tolerance Sufficiency Principle (Yang 2016), in conjunction with a recursive search procedure, provides a generalization mechanism to learn which features are obligatorily marked across the paradigm. I have demonstrated that this model matches well with developmental findings regarding vocabulary size and order of acquisition on a typologically diverse set of morphological paradigms, including English verbs, German noun plurals, Spanish verbs, and Hebrew verbs. Indeed, because our model relies only on the existence of collisions rather than on any specific phonological transformations between the inflected forms, it provides a promising way to generalize to a diverse set of languages. Since most previous work on morphological acquisition focuses on the mapping from meaning to form, this model also sheds light on a lesser-studied portion of the acquisition process.

The results presented in this thesis also raise several interesting questions that should motivate future work on the acquisition of morphological marking. One such question is how children extract morphosemantic information from their input and environment, which may be explored both experimentally and by combining our model with grounded learning models. Another question is how the acquisition of morphological marking interacts with the portion of acquisition that maps morphological features to phonological form, and what potential implications this may have both broadly for theories of acquisition and for theories of the Root Infinitive phenomenon. A final, and particularly interesting, question raised by our model regards the interaction between subject agreement and tense in language acquisition and whether one can influence the other. This question will warrant future experimental investigation and further modeling.

References

Aguirre, Carmen. 2011. "Early verb development in one Spanish-speaking child." In *Development of verb in ection in rst language acquisition,* 1–26. De Gruyter Mouton.

Aguirre Martınez, Marıa del Carmen, et al. 1995. "La adquisición de las categorıas gramaticales, en español."

Albright, Adam, and Bruce Hayes. 2003. "Rules vs. analogy in English past tenses: A computational/experimental study." *Cognition* 90 (2): 119–161.

Amsterdam, Beulah. 1972. "Mirror self-image reactions before age two." *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology* 5 (4): 297–305.

Baayen, R Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan, and James P Blevins. 2019. "The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning." *Complexity* 2019.

Baayen, R Harald, Petar Milin, Dusica Filipović urević, Peter Hendrix, and Marco Marelli. 2011. "An amorphous model for morphological processing in visual comprehension based on naive discriminative learning." *Psychological review* 118 (3): 438.

Barner, D, M Kibbe, J Wood, and S Carey. 2009. "Evidence for a non-linguistic singular–plural distinction in 15-month-old children." *Manuscript submitted for publication.*

Barner, D, D Thalwitz, Justin Wood, Shu-Ju Yang, and Susan Carey. 2007. "On the relation between the acquisition of singular–plural morpho-syntax and the conceptual distinction between one and more than one." *Developmental science* 10 (3): 365–373.

Bat-El, Outi. 2014. *The acquisition of hebrew phonology and morphology.* Leiden, Netherlands: Brill.

Behrens, Heike. 2006. "The input–output relationship in first language acquisition." *Language and cognitive processes* 21 (1-3): 2–24.

Belth, Caleb, Sarah Payne, Deniz Beser, Jordan Kodner, and Charles Yang. 2021. "The Greedy and Recursive Search for Morphological Productivity." *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society.*

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

Berman, Ruth. 1981. "Regularity vs anomaly: the acquisition of Hebrew inflectional morphology." *Journal of Child Language* 8 (2): 265–282.

Berman, Ruth. 1983. "Establishing a schema: Children's construals of verb-tense marking." *Language Sciences* 5 (1): 61–78.

Berman, Ruth. 1990. "On acquiring an (S) VO language: Subjectless sentences in children's Hebrew."

Berman, Ruth. 2017. *The acquisition of Hebrew.* Psychology Press.

Bolozky, Shmuel. 2007. "Israeli Hebrew morphology." *Morphologies of Asia and Africa (including the Caucasus),* 283–308.

Bornstein, Marc H, Linda R Cote, Sharone Maital, Kathleen Painter, Sung-Yun Park, Liliana Pascual, Marie-Germaine Pêcheux, Josette Ruel, Paola Venuti, and Andre Vyt. 2004. "Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English." *Child development* 75 (4): 1115–1139.

Brown, Roger. 1973. *A rst language.* Harvard University Press.

Butterworth, George. 1992. "Origins of self-perception in infancy." *Psychological Inquiry* 3 (2): 103–111.

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

Clahsen, Harald, Monika Rothweiler, Andreas Woest, and Gary F Marcus. 1992. "Regular and irregular inflection in the acquisition of German noun plurals." *Cognition* 45 (3): 225–255.

Clark, Eve V, and Brian MacWhinney. 1987. "The principle of contrast: A constraint on language acquisition." *Mechanisms of language acquisition,* 1–33.

Corkery, Maria, Yevgen Matusevych, and Sharon Goldwater. 2019. "Are we there yet? Encoder-decoder neural networks as cognitive models of English past tense inflection." *Association for Computational Linguistics.*

Dankers, Verna, Anna Langedijk, Kate McCurdy, Adina Williams, and Dieuwke Hupkes. 2021. "Generalising to German Plural Noun Classes, from the Perspective of a Recurrent Neural Network." In *Proceedings of the 25th Conference on Computational Natural Language Learning,* 94–108.

Dixon, James C. 1957. "Development of self recognition." *The Journal of Genetic Psychology* 91 (2): 251–256.

Dressler, Wolfgang U, Anastasia Christofidou, Natalia Gagarina, Katharina Korecky-Kröll, and Marianne Kilani-Schochd. 2019. "Morphological blind-alley developments as a theoretical challenge to both usage-based and nativist acquisition models." *Italian journal of linguistics* 31 (2): 107–140.

Durán, Pilar. 2000. "Verbal morphology in early Spanish." *Spanish applied linguistics at the turn of the millennium,* 36–49.

Embick, David. 2015. *The morpheme.* De Gruyter Mouton.

Emond, Emeryse, and Rushen Shi. 2020. "Infants' rule generalization is governed by the Tolerance Principle." In *45th annual Boston University Conference on Language Development.*

Feigenson, Lisa, and Susan Carey. 2005. "On the limits of infants' quantification of small object arrays." *Cognition* 97 (3): 295–313.

Feigenson, Lisa, Susan Carey, and Marc Hauser. 2002. "The representations underlying infants' choice of more: Object files versus analog magnitudes." *Psychological science* 13 (2): 150–156.

Fenson, Larry, Philip S Dale, J Steven Reznick, Elizabeth Bates, Donna J Thal, Stephen J Pethick, Michael Tomasello, Carolyn B Mervis, and Joan Stiles. 1994. "Variability in early communicative development." *Monographs of the society for research in child development,* i–185.

Fisher, Cynthia, Yael Gertner, Rose M Scott, and Sylvia Yuan. 2010. "Syntactic bootstrapping." *Wiley Interdisciplinary Reviews: Cognitive Science* 1 (2): 143–149.

Gabbard, Ryan, Deniz Beser, Jacob Lichtefeld, Joe Cecil, Mitch Marcus, Sarah Payne, Charles Yang, and Marjorie Freedman. 2021. "ADAM: A Sandbox for Implementing Language Learning." *arXiv preprint arXiv:2105.02263.*

Gaya, Aurora Bel. 2001. *Teoria lingu stica i adquisicio del llenguatge: analisi comparada dels trets morfologics en catala i en castella.* Vol. 44. Institut d'estudis catalans.

Gerken, Louann, Rachel Wilson, and William Lewis. 2005. "Infants can use distributional cues to form syntactic categories." *Journal of child language* 32 (2): 249–268.

Goldhahn, Dirk, Thomas Eckart, and Uwe Quasthoff. 2012. "Building large monolingual dictionaries at the leipzig corpora collection: From 100 to 200 languages." In *Proceedings of* *the Eighth International Conference on Language Resources and Evaluation (LREC'12),* 759–765.

Goodman, Judith C, Philip S Dale, and Ping Li. 2008. "Does frequency count? Parental input and the acquisition of vocabulary." *Journal of child language* 35 (3): 515–531.

Grinstead, John Allen Ray. 1998. *Subjects, sentential negation and imperatives in child Spanish and Catalan.* University of California, Los Angeles Press.

Halle, Morris, Alec Marantz, Kenneth Hale, and Samuel Jay Keyser. 1993. "Distributed morphology and the pieces of inflection." *1993,* 111–176.

Hart, Betty, and Todd R Risley. 1995. *Meaningful di erences in the everyday experience of young American children.* Paul H Brookes Publishing.

Henry, Alison. 1995. *Belfast English and Standard English: Dialect variation and parameter setting.* Oxford University Press on Demand.

Hohle, Barbara, Michaela Schmitz, Lynn M Santelmann, and Jurgen Weissenborn. 2006. "The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities." *Language Learning and Development* 2 (4): 277–300.

Höhle, Barbara, Jürgen Weissenborn, Dorothea Kiefer, Antje Schulz, and Michaela Schmitz. 2004. "Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements." *Infancy* 5 (3): 341–353.

Izhikevich, Eugene M. 2003. "Simple model of spiking neurons." *IEEE transactions on neural networks* 14 6:1569–72.

Izhikevich, Eugene M. 2004. "Which model to use for cortical spiking neurons?" *IEEE Transactions on Neural Networks* 15:1063–1070.

Jacobsen, Teresa, and Jürgen M Meisel. 1986. "¿ Aspecto antes que tiempo? Una mirada a la adquisición temprana del español." In *Adquisicion de lenguaje/Adquisicao da linguagem.* 97–114.

Kampis, Dora, Charlotte Grosse Wiesmann, Sarah Koop, and Victoria Southgate. 2022. "Understanding the self in relation to others: Infants spontaneously map another's face to their own at 16–26 months." *Developmental science* 25 (3): e13197.

Kauschke, Christina, Anna Kurth, and Ulrike Domahs. 2011. "Acquisition of German noun plurals in typically developing children and children with specific language impairment." *Child Development Research.*

Kim, Yun Jung. 2015. *6-month-olds' segmentation and representation of morphologically complex words.* University of California, Los Angeles.

Kim, Yun Jung, and Megha Sundara. 2021. "6–month–olds are sensitive to English morphology." *Developmental science* 24 (4): e13089.

Kiparsky, Paul. 1973. "'Elsewhere' in Phonology." In *A Festschrift for Morris Halle,* 96–103.

Kirov, Christo, and Ryan Cotterell. 2018. "Recurrent neural networks in linguistic theory: Revisiting Pinker and Prince (1988) and the past tense debate." *Transactions of the Association for Computational Linguistics* 6:651–665.

Kirov, Christo, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, et al. 2018. "UniMorph 2.0: Universal Morphology." In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).*

Kodner, Jordan. 2019. "Estimating child linguistic experience from historical corpora." *Glossa: a journal of general linguistics* 4 (1).

Koulaguina, Elena, and Rushen Shi. 2019. "Rule generalization from inconsistent input in early infancy." *Language Acquisition* 26 (4): 416–435.

Kuczaj II, Stan A. 1977. "The acquisition of regular and irregular past tense forms." *Journal of Verbal Learning and Verbal Behavior* 16 (5): 589–600.

Kupisch, Tanja, and Esther Rinke. 2007. "A fresh look at root infinitives from a cross-linguistic perspective." *Nordlyd* 34 (3).

Laaha, Sabine, Dorit Ravid, Katharina Korecky-Kröll, Gregor Laaha, and Wolfgang U. Dressler. 2006. "Early noun plurals in German: regularity, productivity or default?" *Journal of Child Language* 33:271–302.

Landau, Barbara, Lila R Gleitman, and Barbara Landau. 1985. *Language and experience: Evidence from the blind child.* Vol. 8. Harvard University Press.

Legate, Julie Anne, and Charles Yang. 2007. "Morphosyntactic learning and the development of tense." *Language Acquisition* 14 (3): 315–344.

Lewis, Michael, and Jeanne Brooks-Gunn. 1979. "Toward a theory of social cognition: The development of self." *New directions for child and adolescent development* 1979 (4): 1–20.

Li, Daoxin, and Kathryn Schuler. 2021. "Distributional learning of recursive structures." In *Proceedings of the Annual Meeting of the Cognitive Science Society,* vol. 43.

Li, Peggy, Tamiko Ogura, David Barner, Shu-Ju Yang, and Susan Carey. 2009. "Does the conceptual distinction between singular and plural sets depend on language?" *Developmental psychology* 45 (6): 1644.

Lignos, Constantine, and Charles Yang. 2016. "Morphology and language acquisition." *The Cambridge handbook of morphology,* 743764.

Long, Lyle N., and Guoliang Fang. 2010. "A Review of Biologically Plausible Neuron Models for Spiking Neural Networks."

López Ornat, Susana, Almudena Fernández, Pilar Gallo, and Sonia Mariscal. 1994. *La adquisicion de la lengua espanola.*

Lustigman, Lyle. 2013. "Developing structural specification: Productivity in early Hebrew verb usage." *First Language* 33 (1): 47–67.

MacWhinney, Brian. 2000. *The CHILDES project: The database.* Vol. 2. Psychology Press.

Marantz, Alec. 2013. "No escape from morphemes in morphological processing." *Language and cognitive processes* 28 (7): 905–916.

Marcus, Gary F, Ursula Brinkmann, Harald Clahsen, Richard Wiese, and Steven Pinker. 1995. "German inflection: The exception that proves the rule." *Cognitive psychology* 29 (3): 189–256.

Marcus, Gary F, Steven Pinker, Michael Ullman, Michelle Hollander, T John Rosen, Fei Xu, and Harald Clahsen. 1992. "Overregularization in language acquisition." *Monographs of the society for research in child development,* i–178.

Marquis, Alexandra, and Rushen Shi. 2012. "Initial morphological learning in preverbal infants." *Cognition* 122 (1): 61–66.

McClelland, James L, and Karalyn Patterson. 2002. "Rules or connections in past-tense inflections: What does the evidence rule out?" *Trends in cognitive sciences* 6 (11): 465–472.

McCurdy, Kate, Sharon Goldwater, and Adam Lopez. 2020. "Inflecting when there's no majority: Limitations of encoder-decoder neural networks as cognitive models for German plurals." *Association for Computational Linguistics.*

McCurdy, Kate, Adam Lopez, and Sharon Goldwater. 2020. "Conditioning, but on which distribution? Grammatical gender in German plural inflection." In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics,* 59–65.

Mills, Anne E. 2012. *The acquisition of gender: A study of English and German.* Vol. 20. Springer Science & Business Media.

Mintz, Toben H. 2013. "The segmentation of sub-lexical morphemes in English-learning 15-month-olds." *Frontiers in psychology* 4:24.

Montrul, Silvina. 2004. *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition.* Vol. 37. John Benjamins Publishing.

Mugdan, Joachim. 1977. "Flexionsmorphologie und Psycholinguistik (Tübinger Beiträge zur Linguistik 82)." *Tübingen: Narr.*

Naigles, Letitia R. 1996. "The use of multiple frames in verb learning via syntactic bootstrapping." *Cognition* 58 (2): 221–251.

Naigles, Letitia R, and Erika Hoff-Ginsberg. 1995. "Input to verb learning: Evidence for the plausibility of syntactic bootstrapping." *Developmental Psychology* 31 (5): 827.

Paltiel-Gedalyovich, Leah R, Aviya Hacohen, Rachel Eitan, and Jeannete Schaeffer. 2007. "The acquisition of Hebrew tense."

Payne, Sarah, Jordan Kodner, and Charles Yang. 2021. "Learning Morphological Productivity as Meaning-Form Mappings." In *SCIL.*

Piaget, Jean. 2013a. *Play, dreams and imitation in childhood.* Routledge.

Piaget, Jean. 2013b. *The construction of reality in the child.* Routledge.

Piaget, Jean, and Margaret Trans Cook. 1952. "The origins of intelligence in children."

Pinker, Steven. 1998. "Words and rules." *Lingua* 106 (1-4): 219–242.

Pinker, Steven, and Alan Prince. 1988. "On language and connectionism: Analysis of a parallel distributed processing model of language acquisition." *Cognition* 28 (1-2): 73–193.

Pinker, Steven, and Michael T Ullman. 2002. "The past and future of the past tense." *Trends in cognitive sciences* 6 (11): 456–463.

Rescorla, Robert A. 1972. "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement." *Current research and theory,* 64–99.

Rumelhart, David E, and James L McClelland. 1986. "On learning the past tenses of English verbs."

Santelmann, Lynn M, and Peter W Jusczyk. 1998. "Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space." *Cognition* 69 (2): 105–134.

Schuler, Kathryn, Charles Yang, and Elissa Newport. 2021. "Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient."

Shlonsky, Ur. 1997. *Clause structure and word order in Hebrew and Arabic: An essay in comparative Semitic syntax.* Oxford University Press.

Soderstrom, Melanie, Kenneth Wexler, and Peter W Jusczyk. 2002. "English-learning toddlers' sensitivity to agreement morphology in receptive grammar." In *Proceedings of the 26th Annual Boston University conference on language development,* 2:643–652. Cascadilla Somerville, MA.

Soh, Christine, and Charles Yang. 2021. "Memory Constraints on Cross Situational Word Learning." In *Proceedings of the Annual Meeting of the Cognitive Science Society,* vol. 43.

Stevens, Jon Scott, Lila R Gleitman, John C Trueswell, and Charles Yang. 2017. "The pursuit of word meanings." *Cognitive science* 41:638–676.

Szagun, Gisela, Claudia Steinbrink, Melanie Franik, and Barbara Stumper. 2006. "Development of vocabulary and grammar in young German-speaking children assessed with a German language development inventory." *First language* 26 (3): 259–280.

Theakston, Anna L, Elena VM Lieven, Julian M Pine, and Caroline F Rowland. 2001. "The role of performance limitations in the acquisition of verb-argument structure: An alternative account." *Journal of child language* 28 (1): 127–152.

Trommer, Jochen. 2021. "The subsegmental structure of German plural allomorphy." *Natural Language & Linguistic Theory* 39 (2): 601–656.

Wagner, Laura. 2001. "Aspectual influences on early tense comprehension." *Journal of Child Language* 28 (3): 661–681.

Wells, Gordon. 1981. *Learning through Interaction: The Study of Language Development.*

Wiese, Richard. 2000. *The phonology of German.* Oxford University Press on Demand.

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

Yang, Charles, Allison Ellman, and Julie Anne Legate. 2015. "Input and its structural description." *50 Years Later,* 285.

Zapf, Jennifer A., and Linda B. Smith. 2009. "Knowing more than one can say: The early regular plural." *Journal of Child Language* 36:1145–1155.

Zipf, George Kingsley. 1936. *The psycho-biology of language: An introduction to dynamic philology.* Routledge.