BUCLD 47 Proceedings
To be published in 2023 by Cascadilla Press
Rights forms signed by all authors

Contrast, Sufficiency, and the Acquisition of Morphological Marking

Sarah Payne

A great deal of work on morphological acquisition has examined how children learn to map morphological features to phonological form. Most accounts, however, take morphological features to be prerequisites to the mapping problem. How do children learn which morphological features are marked – and thus must be mapped to form – to begin with? We propose that their early ability to segment and relate inflected forms allows children to track **collisions** – a single stem appearing in multiple inflected forms. We present a computational model that uses collisions as cues to learn inflectional categories, in conjunction with the Tolerance/Sufficiency Principle (Yang, 2016). We apply this model to developmentally-plausible vocabularies of English verbs, German noun plurals, Spanish verbs, and Hebrew verbs, and show that results match well with developmental findings regarding order of acquisition and vocabulary size.

1. Background

Dating back to the Past Tense Debate (Rumelhart and McClelland 1986; Pinker and Prince 1988; Fodor and Pylyshyn 1988; see Pinker and Ullman 2002; McClelland and Patterson 2002 for reviews), an influential line of work has focused on computationally modeling morphological acquisition. This research originally focused on mapping one inflectional category to form (e.g. PAST ⇒*-ed*), but modern Natural Language Processing (NLP) has extended this to arbitrary inflectional

* Sarah Payne, Stony Brook University, sarah.payne@stonybrook.edu. I thank Charles Yang, Jordan Kodner, Julie Anne Legate, Jeff Heinz, Mark Aronoff, Bob Berwick and his lab, members of a seminar course at the University of Pennsylvania, and attendees of the Stony Brook Linguistics Brown Bag for helpful comments and feedback. This material is based upon work supported by the Institute for Advanced Computational Science (IACS) Graduate Research Fellowship and the National Science Foundation (NSF) Graduate Research Fellowship Program under NSF Grant No. 2234683. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the IACS or the NSF.

categories. In the 2022 SIGMORPHON shared task, for example, models are trained on {`stem, inflected form, feature set`} triples drawn from throughout the paradigm and evaluated on a variety of inflections (Kodner et al., 2022). Recent work combines insights from both approaches: Kirov and Cotterell (2018) propose a neural network model of English past tense acquisition with applications to NLP. This model, however, fails to match behavioral results on the English past tense (Corkery et al., 2019) or German noun plurals (McCurdy et al., 2020). At the same time, Payne et al. (2021) and Belth et al. (2021) propose rule-based models employing the Tolerance/Sufficiency Principle (Yang, 2016); the latter matches better with the developmental findings of McCurdy et al. (2020). Despite their differences, these lines of work all assume that inflectional categories are already known and must simply be mapped to form: categories are considered one-at-a-time or encoded explicitly in the input. Yet not all languages mark the same features: while Spanish marks tense on the verb, for instance, Chinese does not. How, then, does the child learn the inflectional categories of her language?

A growing body of experimental evidence shows that infants can relate inflected forms to their stems well before they know the grammatical function of inflection. At 0;6, English-learning infants can relate nonce words suffixed with *-s* to their stems, but not words suffixed with *-ing, -ed*, or the pseudo-morpheme *-sh* (Kim, 2015; Kim and Sundara, 2021). By 0;8, infants are also able to relate nonce words suffixed with *-ing* to their stems (Kim and Sundara, 2021). Similarly in French, infants at 0;11 can relate nonce verbs suffixes with *-e* (infinitive and past participle) to their stems, but not those suffixed with the pseudo-morpheme *-u* (Marquis and Shi, 2012). While these results suggest an early ability to relate inflected forms to their stems, they do not entail that infants know the grammatical functions of inflection: English-learning infants a year older than those studied by Kim and Sundara still do not have a full grasp of the grammatical use of *-s* (Soderstrom et al., 2002). Thus, within the first year of life, English-learning children relate most inflected forms to their stems, but it takes them far longer to learn the grammatical use of these inflected forms. While it has yet to be seen if children learning highly agglutinative or templatic languages show similar behavior, these results strongly support a model of acquisition in which children make use of their early segmentation abilities to learn the inflectional categories of their language; this is the foundation of the model we present in §3.

2. Developmental Findings

We evaluate our model against two types of developmental findings: order of acquisition and vocabulary size. A well-known pattern in morphological acquisition is that of developmental regression, in which the child temporarily overgeneralizes a regular process (e.g. *-ed*) to irregular words (e.g. *feeled*; Marcus et al. 1992). This pattern has been variously interpreted as occurring when the markedness of a feature becomes sufficiently obligatory (e.g. ±PAST must be marked; Clahsen et al. 1992) or when the child learns that a process realizing an inflectional category is productive (e.g. PAST ⇒ -*ed*; Yang 2016). Under the former interpretation, the order of overregularization should exactly mirror the order in which markedness is learnt. The latter interpretation also predicts a close pairing of these orders: the child can only learn that a process productively realizes an inflectional category when they know that the category is marked. At the same time, it is well known that language acquisition occurs over small, sparse vocabularies (e.g. Lignos and Yang, 2016), and any plausible model should succeed on such input. In the following sections, we thus consider developmental findings regarding order of acquisition and vocabulary size for each of the four paradigms against which we evaluate our model.

2.1. English Verbs

English (Germanic, Indo-European) morphological acquisition has been the subject of a great deal of experimental work. Brown (1973) established that English-learning children generally acquire the present progressive -*ing* before the noun plural -*s*, followed by the irregular past tense and the regular past -*ed*. Berko (1958) demonstrated that children typically acquire the third singular -*s* between -*ing* and -*ed*, although it emerges later in some children. In typically-developing children, the regular past tense emerges between 2;0 and 3;0 (Brown, 1973; Kuczaj, 1977), at which point the child's vocabulary will range from at most 500 (at 2;0) to 1000 (at 3;0) words (Fenson et al., 1994). Since about a quarter of the child's early vocabulary will be verbs (Bornstein et al., 2004), we expect a vocabulary of about 125-250 verbs when the regular past tense emerges.

2.2. German Noun Plurals

German (Germanic, Indo-European) noun plurals have become a topic of particular interest for recent modeling work (e.g. McCurdy et al. 2020; Belth et al. 2021). Five allomorphic suffixes are used to form the nominative plural; umlauting also applies in some cases via a separate process. The least frequent suffix, -*s*, occurs with less than 5% of nouns, but is typically considered the default since it applies to novel nouns (e.g. *die Autos* 'the cars,' Mugdan 1977; Marcus et al. 1995; Wiese 2000). The remaining suffixes apply to subsets of nouns governed by gender and phonology (Wiese, 2000). The acquisition of German noun plurals follows a protracted path with no consistent order of overapplication (Clahsen et al., 1992; Szagun et al., 2006; Kauschke et al., 2011; Mills, 2012). Initial overapplications (typically of the frequent -*e* and -*(e)n*) begin to emerge at 2;0, however, indicating that children at this age are aware that their language marks plurality (Kauschke et al., 2011). At 2;0, the child's vocabulary is typically around 300 words, and no more than 500 (Szagun et al., 2006). Since nouns make up around half of early vocabularies (Bornstein et al., 2004), we expect the that the child's vocabulary will contain around 150 nouns when they learn that ±PLURAL is marked.

2.3. Spanish Verbs

Spanish (Romance, Indo-European) verbs exhibit both fusional and agglutinative morphology. Table 1 shows the paradigm of *amar* 'to love': contrast fusional *-ste* in *amaste* 'you loved,' with concatenative *-ba-s* in *amabas* 'you were loving.'

Table 1: Conjugation of the Spanish verb *amar* 'to love' in Peninsular Spanish

	SINGULAR			PLURAL		
	1	2	3	1	2	3
PRS	am-**o**	ama-**s**	ama	ama-**mos**	amá-**is**	ama-**n**
PFV	am-**é**	ama-**ste**	am-**ó**	ama-**mos**	ama-**ste-is**	ama-**ron**
IPFV	ama-**ba**	ama-**ba-s**	ama-**ba**	amá-**ba-mos**	ama-**ba-is**	ama-**ba-n**
COND	ama-**ría**	ama-**ría-s**	ama-**ría**	ama-**ría-mos**	ama-**ría-is**	ama-**ría-n**
FUT	ama-**ré**	ama-**rá-s**	ama-**rá**	ama-**re-mos**	ama-**ré-is**	ama-**rá-n**
SBJV	am-**e**	am-**e-s**	am-**e**	am-**e-mos**	am-**é-is**	am-**e-n**

We focus our acquisition findings on Peninsular Spanish; other dialects exhibit a syncretism between the second and third-person plural. From 1;7 onwards, Spanish-learning children mark person agreement (Aguirre Martínez, 1995; Grinstead, 1998); some children mark all three person contrasts initially, while the second person emerges later for others (Montrul, 2004). Number contrast begins to emerge almost immediately after person, with second plural generally emerging last (Montrul, 2004). Children begin marking tense by 2;0, in the perfect past (PFV, Montrul 2004) and future (FUT, Gaya 2001); the imperfect past (IPFV) emerges up to a year later (Jacobsen and Meisel, 1986). The Spanish subjunctive (SBJV) is syncretic with the negative imperative; both begin to emerge around 2;0 (López Ornat et al., 1994; Aguirre Martínez, 1995; Montrul, 2004), with some variation. The mean vocabulary size for Argentinian-Spanish-learning children at 1;8 is 156 words, with a maximum of just over 400 (Bornstein et al., 2004). Since about a quarter of the vocabulary will be verbs, we expect Spanish-learning children to acquire much of their verbal paradigm on 100 or so verbs.

2.4. Hebrew Verbs

Hebrew (Semitic, Afro-Asiatic) exhibits templatic morphology: verb stems are created by combining roots (typically 3 consonants) with one of seven vowel patterns, which differ in mood, reflexivity, intensification, and causativity. The root כתב (*ktb*), for example, gives rise to the verbs לכתוב (*liχtov*) 'to write,' להיכתב (*lehikatev*) 'to be written,' and להתכתב (*lehitkatev*) 'to correspond,' among others. The full paradigm of לכתוב (*liχtov*) 'to write' is given in Table 2; note that /b/ and /k/ are sometimes realized as [v] and [χ], respectively, due to a separate morphophonological process. Hebrew-learning children begin marking subject agreement consistently between 1;8 and 2;0 (Lustigman, 2013). In a longitudinal study of two children, Bat-El (2014) found that one child produced gender

Table 2: Romanized conjugation of the Hebrew verb לכתוב (*liχtov*) ‘to write’

			PRESENT	PAST	FUTURE
SING	MAS	1	kotev	katav-**ti**	**e**-χtov
		2	kotev	katav-**ta**	**ti**-χtov
		3	kotev	katav	**yi**-χtov
	FEM	1	kotev-**et**	katav-**ti**	**e**-χtov
		2	kotev-**et**	katav-**t**	**ti**-χtev-**i**
		3	kotev-**et**	katv-**a**	**ti**-χtov
PLUR	MAS	1	kotv-**im**	katav-**nu**	**ni**-χtov
		2	kotv-**im**	ktav-**tem**	**ti**-χtev-**u**
		3	kotv-**im**	katv-**u**	**yi**-χtev-**u**
	FEM	1	kotv**ot**	katav-**nu**	**ni**-χtov
		2	kotv**ot**	ktav-**ten**	**ti**-χtov-**na**
		3	kotv**ot**	katv-**u**	**ti**-χtov-**na**

and number agreement before person, and the other produced gender and person agreement before number. While Bat-El argues that this stems from asynchronous development of morphology and phonology, tense marking likely also plays a role, since person is only marked in the non-present. Consistent marking of the three Hebrew tenses, as well as the infinitive and imperative, occurs by 3;0 (Berman, 1981; Bat-El, 2014). At 1;8, the mean vocabulary size of a Hebrew-learning child is 233.8 words, with a maximum of 431 words (Bornstein et al., 2004). Since verbs make up about a quarter of the early vocabulary (Bornstein et al., 2004), we expect that the child’s vocabulary will contain about 60-110 verbs when subject agreement begins to emerge.

3. Model

Our model is motivated by children’s early ability to segment inflected forms and relate them to their stems (§1). We propose that children use this ability to track **collisions** – a single stem appearing in multiple inflected forms – and use these as a cue to learn inflectional categories. If, for example, the child hears *walk* and *walked* and is able to relate them, this provides a cue that their language marks ±PAST. Such a mechanism can be seen as an extension of the well-supported Principle of Contrast (Clark and MacWhinney, 1987), the hypothesis that distinct forms indicate distinct meanings. A single collision, however, does not provide sufficient evidence that the corresponding morphological feature(s) are marked: we would not want the learner to conclude from *I am - you are* that English productively contrasts first and second person, for example. At the same time, requiring *all* verbs in the lexicon to realize a collision would be too severe: the sparsity of the input necessitates that many relevant stems will appear in only a fraction of their possible inflected forms (Lignos and Yang, 2016). We thus need a notion of what constitutes *sufficient* evidence for marking.

The **Tolerance/Sufficiency Principle** (TSP, Yang 2016) provides a threshold for the amount of positive evidence necessary for generalization, and has found support in several experiments with precisely controlled conditions (e.g. Schuler 2017; Koulaguina and Shi 2019; Emond and Shi 2020; Schuler et al. 2021). The TSP is based on the hypothesis that generalization occurs when it is more efficient to do so, and is formalized as follows: given a rule R eligible to apply to N items and attested applying to M of these N items, the rule generalizes if and only if:

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

We apply the TSP to collisions as follows:[1] if the model encounters a collision between inflected forms A and B (e.g. *walk-walked*), it uses this as a cue to learn which features distinguish the two forms (in this case, $\pm$PAST). It then searches its lexicon for other words appearing in inflection B (where B occurs with fewer lemmas than A). If a sufficient (as defined by the TSP) number of words appearing in inflection B also appear *in a different form* in inflection A, then the model learns that the corresponding features are marked. In terms of the equation above, we let N be the number of words appearing in inflection B, and M be the number of words appearing in different forms in B and A. Returning to the example of *I am-you are*, a sufficient number of verbs appearing in the second person in English will not appear *in a different form* in the first person (c.f. *I walk-you walk*, etc.), so first-second contrast will never pass the TSP. By hypothesis, however, a sufficient number of verbs appearing in +PAST will also appear in -PAST in a different form, and the model will learn that $\pm$PAST is marked in English.

Our model takes in input incrementally in the form of {`stem, inflected form, feature`} triples; the stem is encoded in the input to model children's early ability to segment and relate inflected forms (§1). We make the simplifying assumption that features are determinable by the child; the task is only to learn which features – and thus which inflectional categories – are productively marked (but see §6 for future directions). When a new item is encountered, the model searches its lexicon for other inflected forms with the same stem to determine if there is a collision. If so, it considers all pairs of inflected forms in which it has seen the stem in order of decreasing type frequency. For example, if it has seen *walk, walks*, and *walked*, the model considers *walk-walks*, *walk-walked*, and *walks-walked*; the type frequency of a collision is defined as the sum of the number of verbs occurring in inflectional category A and the number occurring in B. To determine the features realized by a collision, the model uses set difference, choosing from $A\backslash B$ and $B\backslash A$ the set with the smallest cardinality. If, for example, the model encounters a collision between {3, SG, PRESENT} and {PAST}, the features it learns will be $\pm$ PAST, since $|\{\text{PAST}\} - \{3, \text{SG}, \text{PRESENT}\}| = 1$ is less than $|\{3, \text{SG}, \text{PRESENT}\} - \{\text{PAST}\}| = 3$. Once the relevant features are

[1] For extremely small values of N, θ_N is not a strict minority of N: θ_3, for example, is 2.7. To handle such cases, we follow Payne et al. (2021) and make the additional requirement that $M > N/2$.

determined, they are tested against the TSP as defined above. When a productive contrast is found, learning stops for the current item and the model learns that the relevant features are marked. It then subdivides its lexicon with these features (e.g. into +PAST and -PAST forms). As further input is taken in, the same method is applied recursively within each resulting set. For example, the model may learn that ±PARTICIPLE is marked in English, and then that ±{3,SG} is marked for -PARTICIPLE forms, and so on.

4. Training Data

To mimic the child's early vocabulary, we extract the most frequent inflected forms from the CHILDES database (MacWhinney, 2000). These forms are annotated with their stems and features by combining them with UniMorph 3.0 (McCarthy et al., 2020), a morphological database designed for NLP tasks such as SIGMORPHON (§1, Kodner et al. 2022). Our English data consists of 4196 inflected verbs (1280 stems) taken from the Manchester, Wells, and Belfast corpora (Theakston et al., 2001; Wells, 1981; Henry, 1995), and our German data consists of 1787 inflected nouns (1444 unique stems) taken from the Leo corpus (Behrens, 2006). Our Spanish data consists of 999 inflected verbs (310 stems) of Peninsular Spanish taken from the Fernández/Aguado corpus , and our Hebrew data contains 3440 inflected forms (524 stems) extracted from the Berman and Ravid corpora (Berman, 1990). Our model takes in input in order of decreasing frequency, since frequency is well-correlated with order of acquisition (Goodman et al., 2008). While this order is meant to model one child's vocabulary acquisition trajectory, future work could model variation in acquisition trajectories by adding random jitters to frequencies as in Belth et al. (2021); see §6 for further discussion.

5. Results

The recursive subdivision of our model allows us to visualize the learning trajectory as a tree. For example, the learning trajectory in Fig. 1 shows that the model first learns that participles (±V.PTCP) are marked, then recursively learns that ±PRS is marked for +V.PTCP forms, and so on. Each node is annotated with the number of inflected forms in the model's vocabulary when it learns that the feature(s) are marked: in Fig. 1, the model learns that ±V.PTCP is marked on 11 inflected forms. Nodes reading "marking" indicate the point at which no more collisions were found, and the path from the root to the marking gives the features it realizes (e.g. *marking 2.1* realizes {-V.PTCP, -SG, -PST} in Fig. 1). The numbering on the *marking* nodes reflects the order in which the inflectional categories are learnt; we use e.g. *marking 1.1* and *marking 1.2* to indicate cases where two inflectional categories emerge as the result of a single subdivision. For example, the present and past participles both emerge from subdivision on ±PRS in Fig. 1.

We now compare each learning trace to the developmental findings reviewed in §2. When trained on English, our model learns that the present and past par-

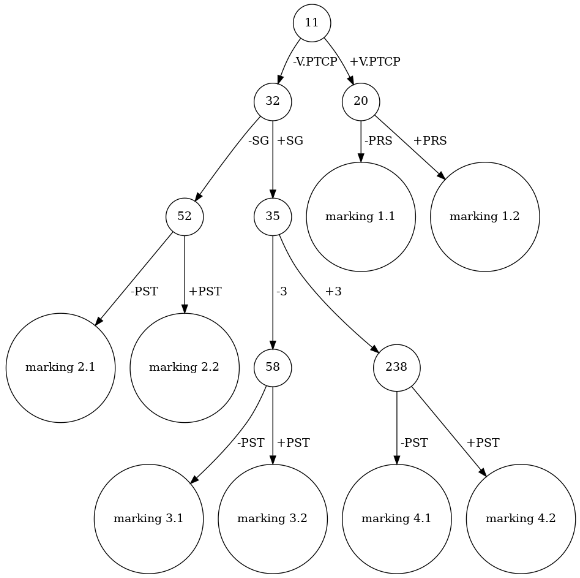

Figure 1: The results of our model on English

ticiples ({+V.PTCP, ±PRS}) are marked on just 20 inflected forms (11 stems, Fig 1). This is followed by the third singular (±SG almost immediately followed by ±3, at 32 and 35 inflected forms respectively, or 17-18 stems). Past tense (±PST) is the last marking to be learnt: it does not emerge across the paradigm until 238 inflected forms (112 stems). This order of acquisition matches well with the developmental findings reviewed in §2.1. Further, learning is complete by 112 verb stems, matching well with vocabulary sizes at 2;0 (§2.1). Due to its hierarchical subdivision and acquisition of ±{3,SG} before ±{PST}, our model predicts that tense marking may emerge *at different times* for each subject agreement: ±{PST} emerges at 52 inflected forms for {-SG} forms but 238 for {+3, +SG} forms in Fig. 1. This prediction may shed light on dialectal variation in acquisition: tense marking generally emerges later in children learning African American English (AAE) than those learning Mainstream American English (MAE), despite no significant differences in tense marking in the input (Yang et al., 2015). However, the third singular is generally unmarked in AAE, meaning that AAE-learning children would not subdivide on ±{3, SG}. Since the TSP allows relatively more exceptions for smaller values of N (Yang, 2016), our model predicts that relatively more evidence will be needed to learn tense if we have not already subdivided on ±{3,SG} as N will be larger. Our model thus provides a possible mechanistic explanation of variation that may be tested against empirical findings.

Because our task for German is simply to learn that ±{PL} is marked, we do not include a learning trace. Our model learns that ±{PL} is marked on just 70 inflected nouns (66 roots), well under the vocabulary size at which overap-

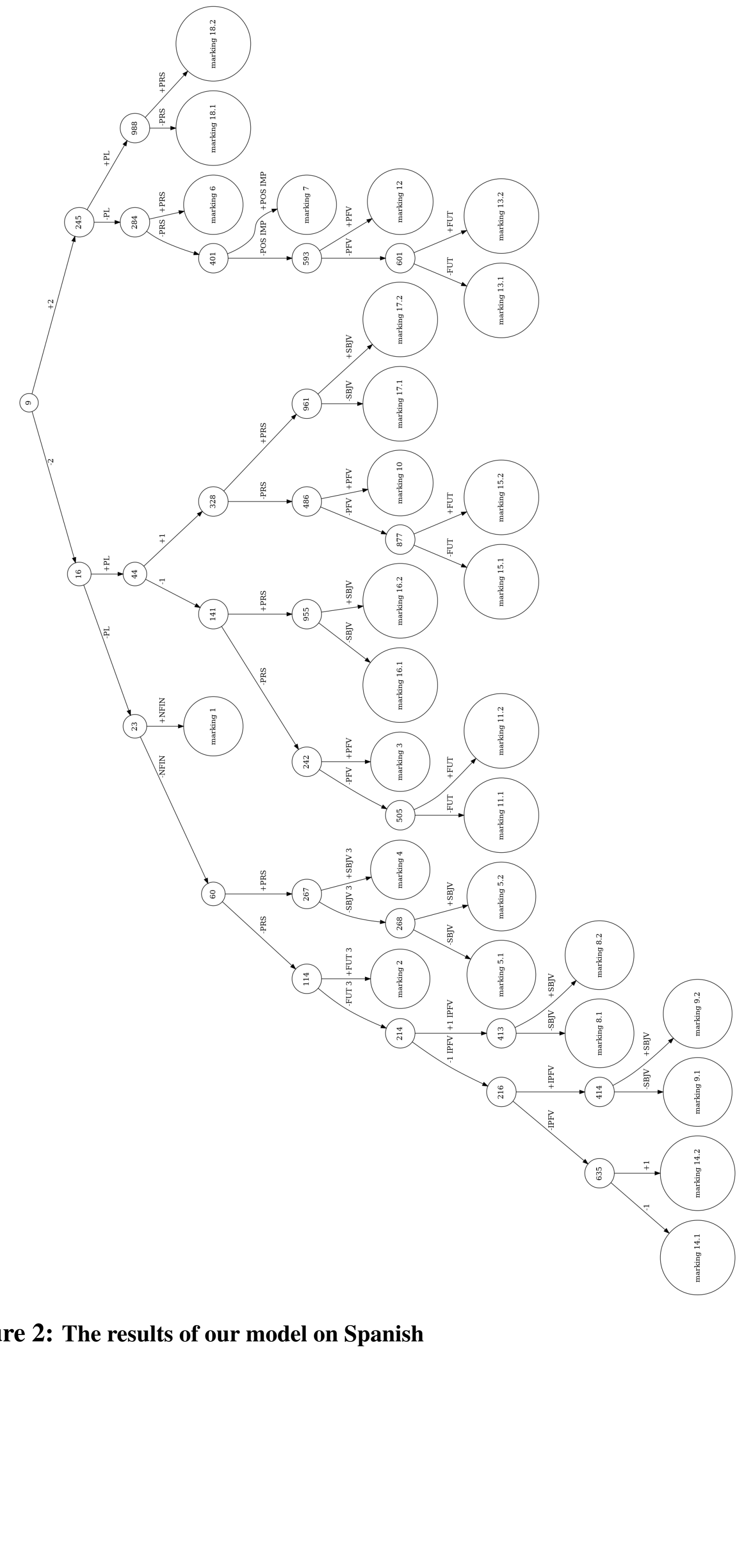

Figure 2: **The results of our model on Spanish**

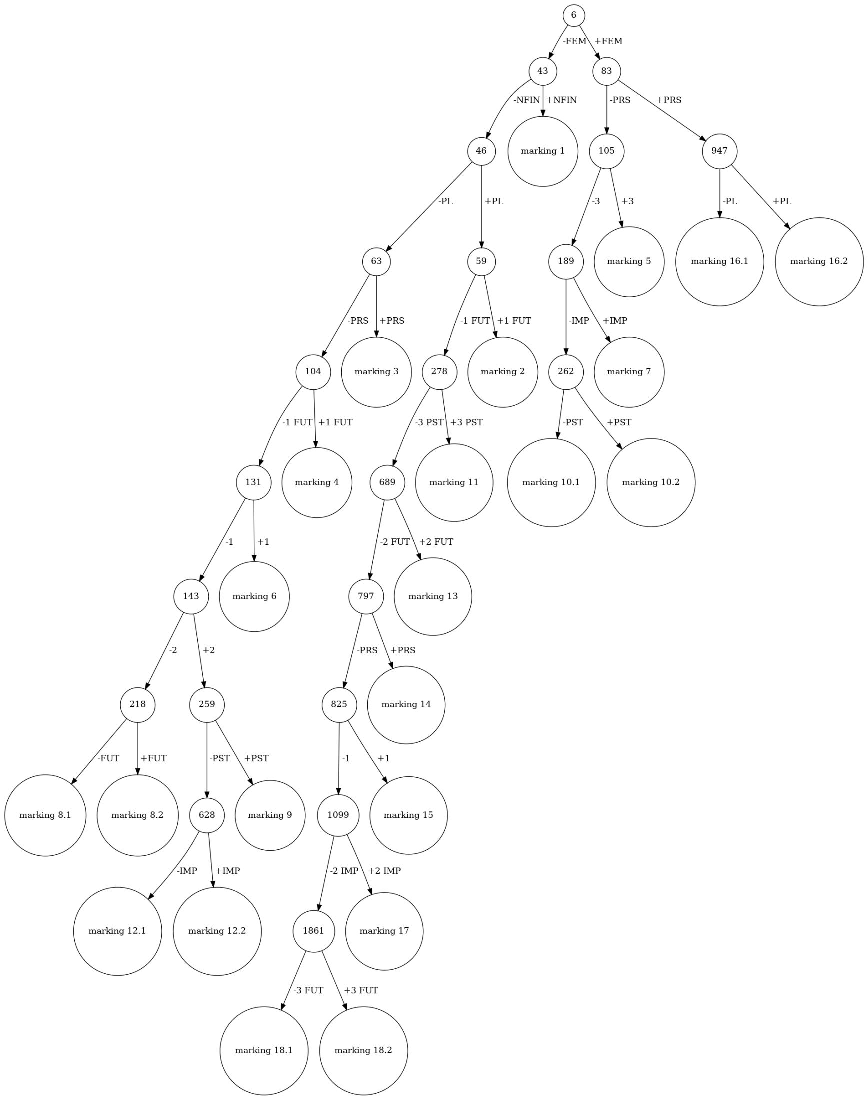

Figure 3: **The results of our model on Hebrew**

plication begins (§2.2). As discussed in §2, overapplication may begin well after markedness is learnt; our model predicts that German-learning children know their language marks $\pm\{\textsc{pl}\}$ well before they overapply any suffix. This makes sense given the protracted development of German noun plurals (§2.2).

The learning trace of our model on Spanish is given in Fig. 2. Person marking emerges as early as 9 inflected forms (7 stems), when the model learns that $\pm\{2\}$ is marked. While this may seem at odds with the later emergence of the second person discussed in §2.3, note that the subsequent marking of $\pm\{\text{PL}\}$ does not emerge until 245 inflected forms (108 stems) in the second person, vs. 16 inflected forms (9 stems) for the non-second. Before the child learns that $\pm\{\text{PL}\}$ is marked in the second person, there is likely no phonological process that is sufficiently dominant to mark the second person. Alternatively, the singular may dominate, causing a lack of second plural productions. Both an overall delay in second person marking and a delay in second plural marking are attested (§2.3). The next contrast to emerge is the present tense (±PRS) at 44-60 inflected forms (18-23 stems) in the non-second, but far later in the second (e.g. marking 18.2). The future tense (±FUT) emerges between 114-877 inflected forms (49-282 stems) across the paradigm. Aspect marking (±IPFV, ±PFV) emerges for past tense forms at 214-593 inflected forms (93-202 stems). If the perfect past dominates over the imperfect, this may explain the late emergence of the imperfect: before subdivision, children would mark all past forms as perfect. Finally, subjunctive (±SBJV) emerges between 267 and 961 inflected forms (117-299 stems), mirroring the longer acquisition trajectory of the Spanish subjunctive reviewed in §2.3.

The learning trace of our model on Hebrew is given in Fig. 3. Gender (± FEM) is the first marking to emerge, at only 6 inflected forms (3 stems). The model then learns that $\pm\{\text{PRS}\}$ is marked for +{FEM} verbs at 83 inflected forms (38 stems). Number contrast is then learnt for +{PRS} forms, and person for -{PRS}. For -{FEM} forms, the infinitive emerges early, at 43 inflected forms (23 stems), matching well with Hebrew-learning children's early production of the infinitive (Lustigman, 2013). Number marking emerges for -{FEM} at 46 inflected forms (25 stems). Tense and agreement emerge in a more piecemeal manner after this, although the model correctly represents the person syncretism in the present tense. The model sometimes acquires a contrast that conflates agreement and tense: for example, marking 4 and marking 13 both subdivide on $\pm\{\text{FUT}\}$ *and* person. A potential remedy and interesting direction for future work will be exploring methods other than set difference for determining relevant features. Nevertheless, our model is done learning by 1861 inflected forms (323 lemmas), at which point it has learnt that all three tenses, as well as the imperative, are marked. While we do not have vocabulary measures at 3;0 for Hebrew-learning children, we know that English learning children's vocabulary at this age contains about 250 verbs, so this result is not unreasonable; much of the paradigm is learnt on far less data.

6. Discussion & Conclusion

In this paper, we proposed that children use their early ability segment and relate inflected forms to track collisions and learn the inflectional categories of their language. We demonstrated that a computational model making use of collisions matches well with developmental findings regarding order of acquisition and vocabulary size, and discussed implications of this model and directions for future work. One area warranting further investigation is variation: in §5, we provided a possible account of the differences in past tense acquisition between learners of AAE and MAE that can be explored empirically. Further, the data used here was meant to model only a single child's vocabulary (§4); future work could investigate the effects of vocabulary variation on our model by introducing random frequency jitters and sampling subsets of the data, as in Belth et al. (2021). The effects of vocabulary variation on the acquisition of morphological marking should also be explored empirically. Finally, the model presented here used set difference to extract the features a collision realizes, but future work could consider how such features may be learnt distributionally from the input instead.

References

Aguirre Martínez, María del Carmen. 1995. La adquisición de las categorías gramaticales, en Español .

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* 21:2–24.

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* 14:150–177.

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

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

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* 75:1115–1139.

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

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

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* .

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

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.

Fodor, Jerry A, and Zenon W Pylyshyn. 1988. Connectionism and cognitive architecture: A critical analysis. *Cognition* 28:3–71.

Gaya, Aurora Bel. 2001. *Teoria lingüística i adquisició del llenguatge: anàlisi comparada dels trets morfològics en català i en castellà*, volume 44. Institut d'estudis catalans.

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

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

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

Jacobsen, Teresa, and Jürgen M Meisel. 1986. ¿Aspecto antes que tiempo? una mirada a la adquisición temprana del Español. In *Adquisición de lenguaje/adquisição da linguagem.*, 97–114.

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:e13089.

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.

Kodner, Jordan, Salam Khalifa ..., and Ekaterina Vylomova. 2022. SIGMORPHON–UniMorph 2022 shared task 0: Generalization and typologically diverse morphological inflection. In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, 176–203. Seattle, Washington: Association for Computational Linguistics.

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

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

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

López Ornat, Susana, Almudena Fernández, Pilar Gallo, and Sonia Mariscal. 1994. La adquisición de la lengua Española.

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

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

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

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* 122:61–66.
McCarthy, Arya D., Christo Kirov ..., and David Yarowsky. 2020. UniMorph 3.0: Universal Morphology. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association.
McClelland, James L, and Karalyn Patterson. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in cognitive sciences* 6:465–472.
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* .
Mills, Anne E. 2012. *The acquisition of gender: A study of English and German*, volume 20. Springer Science & Business Media.
Montrul, Silvina. 2004. *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*, volume 37. John Benjamins Publishing.
Mugdan, Joachim. 1977. Flexionsmorphologie und psycholinguistik (tübinger beiträge zur linguistik 82). *Tübingen: Narr* .
Payne, Sarah, Jordan Kodner, and Charles Yang. 2021. Learning morphological productivity as meaning-form mappings. In *SCIL*.
Pinker, Steven, and Alan Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28:73–193.
Pinker, Steven, and Michael T Ullman. 2002. The past and future of the past tense. *Trends in cognitive sciences* 6:456–463.
Rumelhart, David E, and James L McClelland. 1986. On learning the past tenses of English verbs .
Schuler, Kathryn, Charles Yang, and Elissa Newport. 2021. Testing the Tolerance Principle: Children form productive rules when it is more computationally efficient .
Schuler, Kathryn Dolores. 2017. *The acquisition of productive rules in child and adult language learners*. Georgetown University.
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*, volume 2, 643–652. Cascadilla Somerville, MA.
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* 26:259–280.
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* 28:127–152.
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.