

When Collisions are a Good Thing: the Acquisition of Morphological Marking

Sarah Brogden Payne

Departments of Linguistics and Computer Science
University of Pennsylvania
paynesa@sas.upenn.edu • paynesa.github.io

LING300
December 2, 2021

- Much work on inflectional morphemes \longleftrightarrow semantic features

- Much work on inflectional morphemes \longleftrightarrow semantic features
 - English past tense

- Much work on inflectional morphemes \longleftrightarrow semantic features
 - English past tense
 - German noun plurals

- Much work on inflectional morphemes \longleftrightarrow semantic features
 - English past tense
 - German noun plurals
- How do children learn what semantic features are marked?

- Much work on inflectional morphemes \longleftrightarrow semantic features
 - English past tense
 - German noun plurals
- How do children learn what semantic features are marked?
- **Hypothesis:** learning *what* is marked as a prerequisite to learning *how* it is marked

- Child utilizes **collisions**: single word appearing in multiple different inflected forms

- Child utilizes **collisions**: single word appearing in multiple different inflected forms
 - e.g. the child hearing both *walk* and *walks* may cause them to learn that English marks the third singular

- Child utilizes **collisions**: single word appearing in multiple different inflected forms
 - e.g. the child hearing both *walk* and *walks* may cause them to learn that English marks the third singular
- Combine this with the **Tolerance Principle** (Yang, 2016): children generalize rules when it's more efficient to do so

- Child utilizes **collisions**: single word appearing in multiple different inflected forms
 - e.g. the child hearing both *walk* and *walks* may cause them to learn that English marks the third singular
- Combine this with the **Tolerance Principle** (Yang, 2016): children generalize rules when it's more efficient to do so
 - Single collision insufficient evidence for marking

- Child utilizes **collisions**: single word appearing in multiple different inflected forms
 - e.g. the child hearing both *walk* and *walks* may cause them to learn that English marks the third singular
- Combine this with the **Tolerance Principle** (Yang, 2016): children generalize rules when it's more efficient to do so
 - Single collision insufficient evidence for marking
- Recursively subdivide input to determine which features are marked

- Child utilizes **collisions**: single word appearing in multiple different inflected forms
 - e.g. the child hearing both *walk* and *walks* may cause them to learn that English marks the third singular
- Combine this with the **Tolerance Principle** (Yang, 2016): children generalize rules when it's more efficient to do so
 - Single collision insufficient evidence for marking
- Recursively subdivide input to determine which features are marked
- This approach shows promise on typologically diverse morphologies:
 - **English verbs**
 - **German nouns**
 - **Spanish verbs**
 - **Hebrew verbs**

1 Background

2 Proposed Method

3 Results

- English Verbs
- German Nouns
- Spanish Verbs
- Hebrew Verbs

4 Conclusions & Future Directions

1 Background

2 Proposed Method

3 Results

- English Verbs
- German Nouns
- Spanish Verbs
- Hebrew Verbs

4 Conclusions & Future Directions

Background: Structural Challenges

- Productivity exists despite exceptions to rules
 - e.g. *-ed* is productive despite *went*, *felt*, *was*

Background: Structural Challenges

- Productivity exists despite exceptions to rules
 - e.g. *-ed* is productive despite *went, felt, was*
- Processes may be productive on subsets of the input
 - e.g. German noun plurals: multiple suffixes productive on phonologically + gender-conditioned subsets

Background: Structural Challenges

- Productivity exists despite exceptions to rules
 - e.g. *-ed* is productive despite *went, felt, was*
- Processes may be productive on subsets of the input
 - e.g. German noun plurals: multiple suffixes productive on phonologically + gender-conditioned subsets
- Frequency \neq productivity
 - e.g. German plural default suffix *-s* is the **least** frequent

Background: Structural Challenges

- Productivity exists despite exceptions to rules
 - e.g. *-ed* is productive despite *went, felt, was*
- Processes may be productive on subsets of the input
 - e.g. German noun plurals: multiple suffixes productive on phonologically + gender-conditioned subsets
- Frequency \neq productivity
 - e.g. German plural default suffix *-s* is the **least** frequent
- **Typological diversity**: morphological acquisition across templatic, agglutinative, and fusional languages

Background: Previous Models

- Previous models have tried to account for these patterns

Background: Previous Models

- Previous models have tried to account for these patterns
- **Past Tense Debate** led to a plethora of models of English past tense acquisition

Background: Previous Models

- Previous models have tried to account for these patterns
- **Past Tense Debate** led to a plethora of models of English past tense acquisition
- **Tolerance Principle (TP)** (Yang, 2016) gives an account of English Past Tense acquisition & other morphological paradigms
 - Payne et al. (2021) apply TP to Spanish and English verbal morphology
 - Belth et al. (2021) apply TP to German nouns and English verbs

Background: Previous Models

As of yet, no single model can account for the acquisition of morphology across a typologically diverse set of languages.

Background: Languages

We choose four representative test cases:

- **English Verbs:** well-studied concatenative, non-agglutinative case

Background: Languages

We choose four representative test cases:

- **English Verbs:** well-studied concatenative, non-agglutinative case
- **German Nouns:** productivity on subsets and most frequent \neq default

We choose four representative test cases:

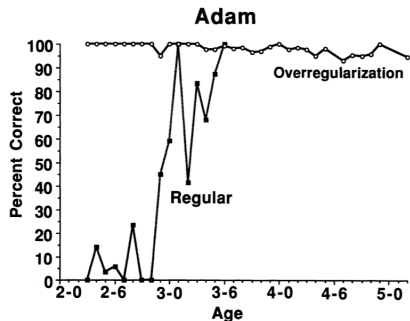
- **English Verbs:** well-studied concatenative, non-agglutinative case
- **German Nouns:** productivity on subsets and most frequent \neq default
- **Spanish Verbs:** exhibits agglutinativity and fusionality

Background: Languages

We choose four representative test cases:

- **English Verbs:** well-studied concatenative, non-agglutinative case
- **German Nouns:** productivity on subsets and most frequent \neq default
- **Spanish Verbs:** exhibits agglutinativity and fusionality
- **Hebrew Verbs:** well-studied templatic morphology

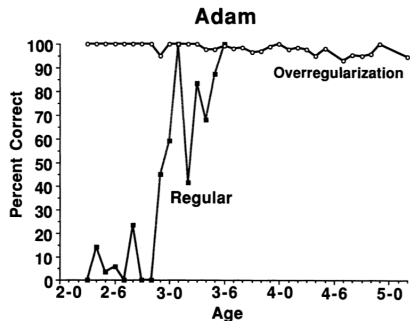
Developmental Regression



Order of acquisition (Brown, 2013; Berko, 1958):

Figure: Adam's Learning of the English Past Tense (Marcus et al., 1992)

Developmental Regression

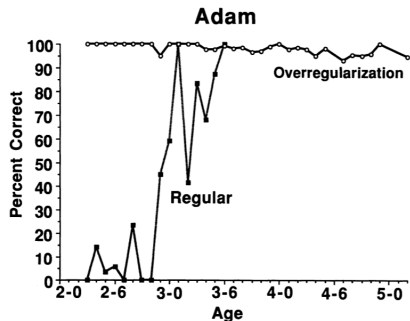


Order of acquisition (Brown, 2013; Berko, 1958):

- 1 Present progressive (*-ing*) early

Figure: Adam's Learning of the English Past Tense (Marcus et al., 1992)

Developmental Regression

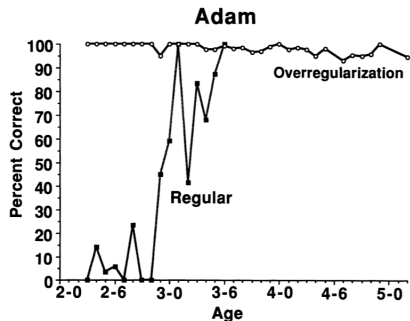


Order of acquisition (Brown, 2013; Berko, 1958):

- 1 Present progressive (*-ing*) early
- 2 third singular later (usually before past but sometimes after)

Figure: Adam's Learning of the English Past Tense (Marcus et al., 1992)

Developmental Regression



Order of acquisition (Brown, 2013; Berko, 1958):

- 1 Present progressive (*-ing*) early
- 2 third singular later (usually before past but sometimes after)
- 3 exception-laden past last

Figure: Adam's Learning of the English Past Tense (Marcus et al., 1992)

- **Different problem:** all suffixes have same semantic value (plural)

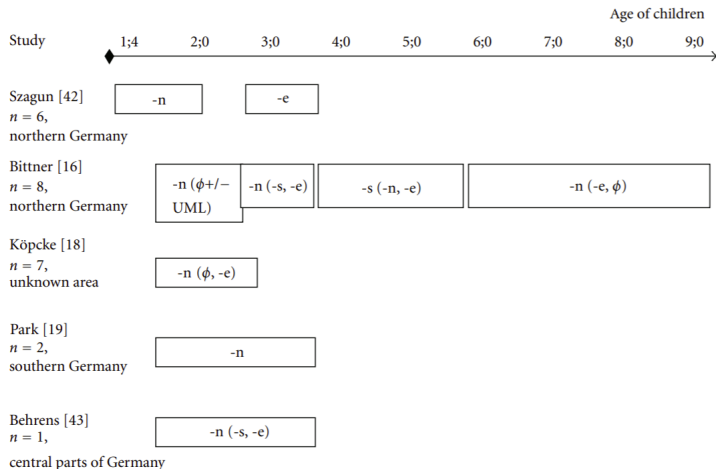
Background: German Nouns

- **Different problem:** all suffixes have same semantic value (plural)
- Default -s least frequent in input

Background: German Nouns

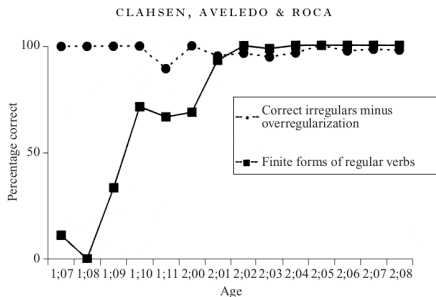
- **Different problem:** all suffixes have same semantic value (plural)
- Default -s least frequent in input
- No clear order of acquisition

Background: German Nouns



Review of overgeneralization by Kauschke et al. (2011)

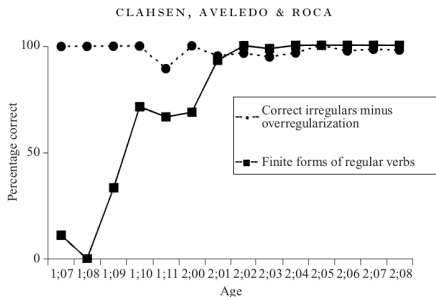
Developmental Regression



Order of acquisition (Clahsen et al., 2002; Montrul, 2004):

Figure: Maria's Finiteness marking in Spanish (Clahsen et al., 2002)

Developmental Regression

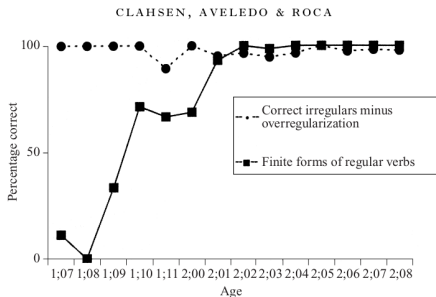


Order of acquisition (Clahsen et al., 2002; Montrul, 2004):

- 1 Finiteness & person marking around 1;7

Figure: Maria's Finiteness marking in Spanish (Clahsen et al., 2002)

Developmental Regression

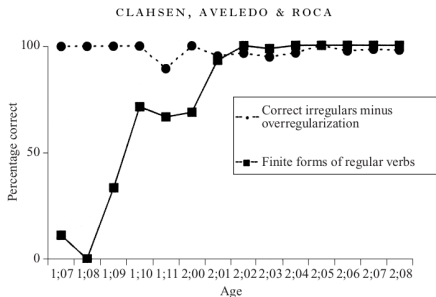


Order of acquisition (Clahsen et al., 2002; Montrul, 2004):

- 1 Finiteness & person marking around 1;7
- 2 Number marking around 1;7-2;0

Figure: Maria's Finiteness marking in Spanish (Clahsen et al., 2002)

Developmental Regression

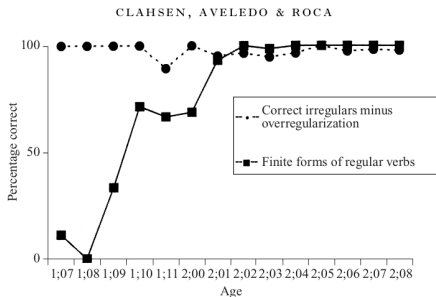


Order of acquisition (Clahsen et al., 2002; Montrul, 2004):

- 1 Finiteness & person marking around 1;7
- 2 Number marking around 1;7-2;0
- 3 Tense by 2;0-2;2

Figure: Maria's Finiteness marking in Spanish (Clahsen et al., 2002)

Developmental Regression



Order of acquisition (Clahsen et al., 2002; Montrul, 2004):

- 1 Finiteness & person marking around 1;7
- 2 Number marking around 1;7-2;0
- 3 Tense by 2;0-2;2
- 4 Mood between 1;7-2;2

Figure: Maria's Finiteness marking in Spanish (Clahsen et al., 2002)

Background: Hebrew Verbs

Order of acquisition (Bat-El, 2014; Lustigman, 2013)

- Person, number and gender emerge before tense

Background: Hebrew Verbs

Order of acquisition (Bat-El, 2014; Lustigman, 2013)

- Person, number and gender emerge before tense
- The order of person vs. number varies

Order of acquisition (Bat-El, 2014; Lustigman, 2013)

- Person, number and gender emerge before tense
- The order of person vs. number varies
- Gender appears at the same time or before number

1 Background

2 Proposed Method

3 Results

- English Verbs
- German Nouns
- Spanish Verbs
- Hebrew Verbs

4 Conclusions & Future Directions

Proposed Method: The Tolerance Principle

Intuition:

- If many items do something, then all do it

Proposed Method: The Tolerance Principle

Intuition:

- If many items do something, then all do it
- If few items do something, memorize the ones that do

Proposed Method: The Tolerance Principle

Formalize this by computational efficiency: a rule R applicable to N items with e exceptions is productive iff:

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

Proposed Method: Applying the Tolerance Principle

- **Idea:** utilize collisions: a single word appearing in multiple forms.

Proposed Method: Applying the Tolerance Principle

- **Idea:** utilize collisions: a single word appearing in multiple forms.
 - e.g. a child hearing *walk* and *walked* may be a cue that their language marks the past tense

Proposed Method: Applying the Tolerance Principle

- **Idea:** utilize collisions: a single word appearing in multiple forms.
 - e.g. a child hearing *walk* and *walked* may be a cue that their language marks the past tense
 - A single collision is not sufficient evidence

Proposed Method: Applying the Tolerance Principle

- **Idea:** utilize collisions: a single word appearing in multiple forms.
 - e.g. a child hearing *walk* and *walked* may be a cue that their language marks the past tense
 - A single collision is not sufficient evidence
 - Define sufficient with **Tolerance Principle**

Proposed Method: Collisions & Recursion

- Input taken in incrementally

Proposed Method: Collisions & Recursion

- Input taken in incrementally
- When there is a collision between inflected form A and B:

Proposed Method: Collisions & Recursion

- Input taken in incrementally
- When there is a collision between inflected form A and B:
 - n = number of words appearing in the less-frequent inflected form (say A)

Proposed Method: Collisions & Recursion

- Input taken in incrementally
- When there is a collision between inflected form A and B:
 - n = number of words appearing in the less-frequent inflected form (say A)
 - e = number of those words that do not appear in form B with a **different inflected form**

Proposed Method: Collisions & Recursion

- Input taken in incrementally
- When there is a collision between inflected form A and B:
 - n = number of words appearing in the less-frequent inflected form (say A)
 - e = number of those words that do not appear in form B with a **different inflected form**
- Another way to say this: words that follow the rule = ones that appear in both A and B and the A inflected form \neq the B inflected form

Proposed Method: Collisions & Recursion

- If the collision passes the TP, then we divide the input based on the difference in semantic features between the two inflected forms.

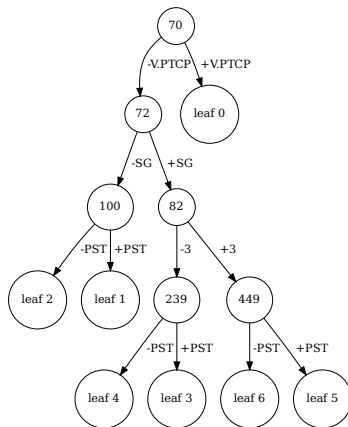


Figure: Example tree from our model

Proposed Method: Collisions & Recursion

- If the collision passes the TP, then we divide the input based on the difference in semantic features between the two inflected forms.
 - e.g. if there is a collision between PST,3,SG and PRS,3,SG, we subdivide based on PST vs. PRS

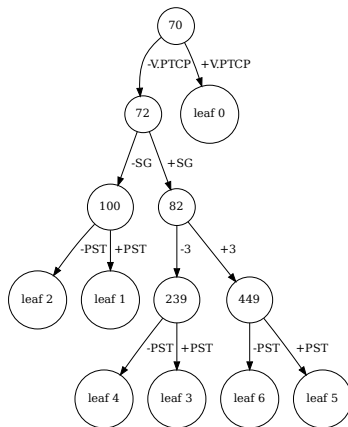


Figure: Example tree from our model

Proposed Method: Collisions & Recursion

- If the collision passes the TP, then we divide the input based on the difference in semantic features between the two inflected forms.
 - e.g. if there is a collision between PST,3,SG and PRS,3,SG, we subdivide based on PST vs. PRS
- We do this recursively and incrementally, creating a tree

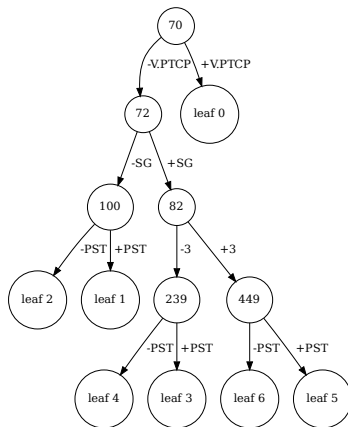


Figure: Example tree from our model

Proposed Method: Data

- Extract most frequent forms from CHILDES
 - Most frequent forms more likely to be added to vocabulary early

- Extract most frequent forms from CHILDES
 - Most frequent forms more likely to be added to vocabulary early
- Intersect these with semantic features from UniMorph (Kirov et al., 2018)

Proposed Method: Data

- Extract most frequent forms from CHILDES
 - Most frequent forms more likely to be added to vocabulary early
- Intersect these with semantic features from UniMorph (Kirov et al., 2018)
- Example input: *go*, *going*, V;V.PTCP;PRS

- Extract most frequent forms from CHILDES
 - Most frequent forms more likely to be added to vocabulary early
- Intersect these with semantic features from UniMorph (Kirov et al., 2018)
- Example input: *go, going*, V;V.PTCP;PRS
- Challenging for Hebrew

1 Background

2 Proposed Method

3 Results

- English Verbs
- German Nouns
- Spanish Verbs
- Hebrew Verbs

4 Conclusions & Future Directions

Results: Interpreting Recursion Trees

- Numbers in nodes = vocabulary size at time of subdivision

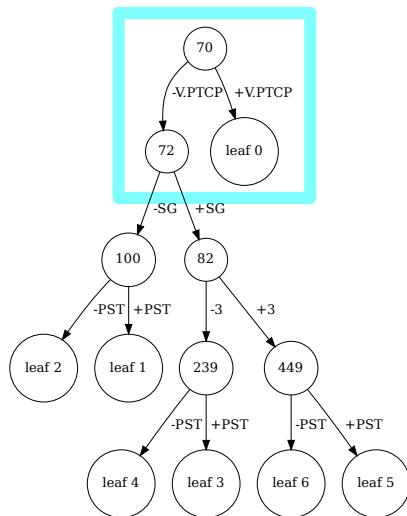
Results: Interpreting Recursion Trees

- Numbers in nodes = vocabulary size at time of subdivision
- +/- features = split features used

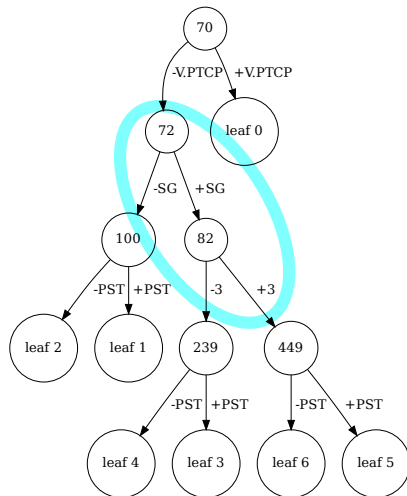
Results: Interpreting Recursion Trees

- Numbers in nodes = vocabulary size at time of subdivision
- +/- features = split features used
- “Leaf” label = no more possible subdivisions found

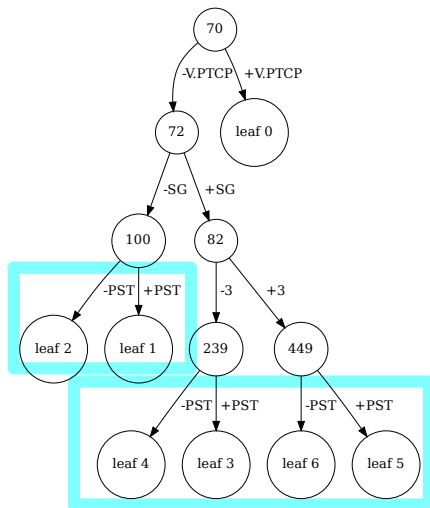
Results: English Verbs



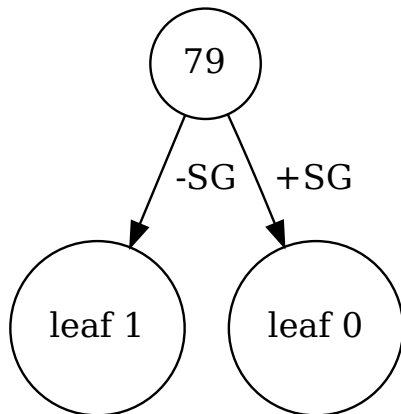
Results: English Verbs



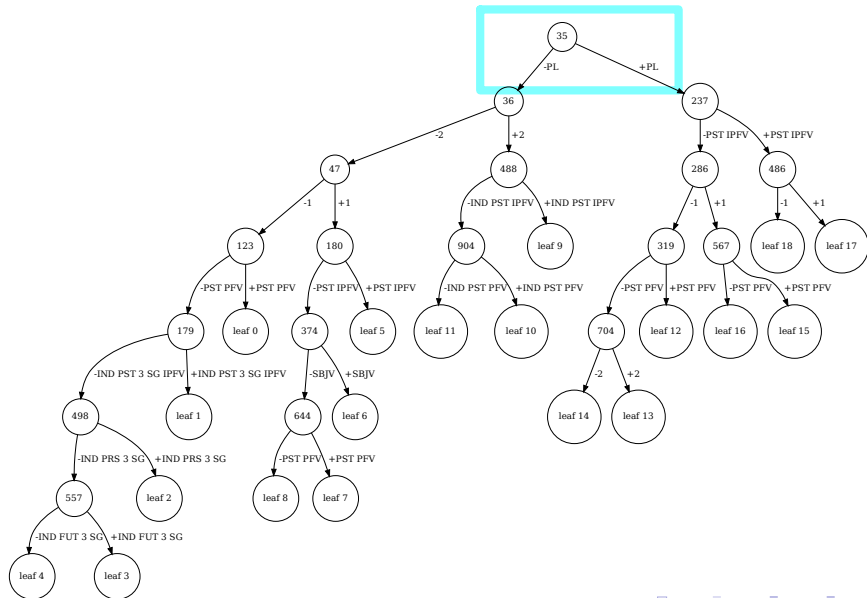
Results: English Verbs



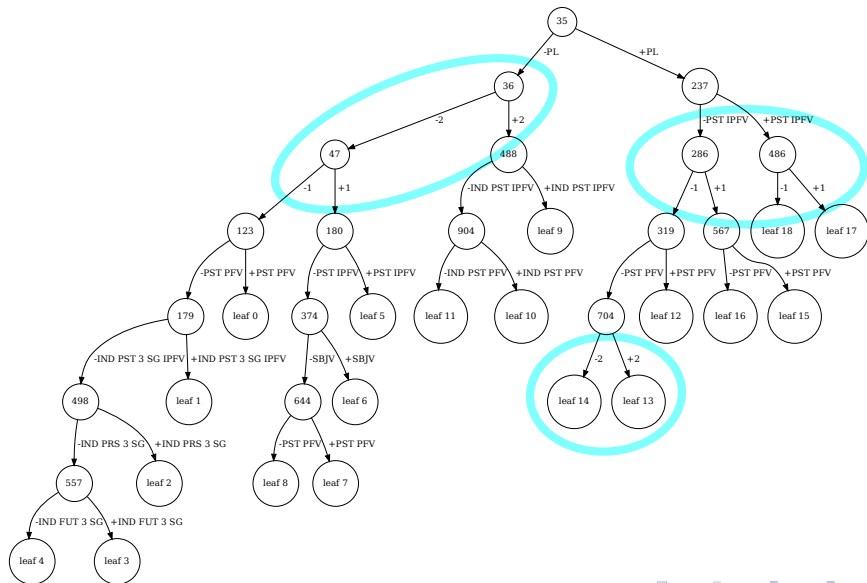
Results: German Nouns



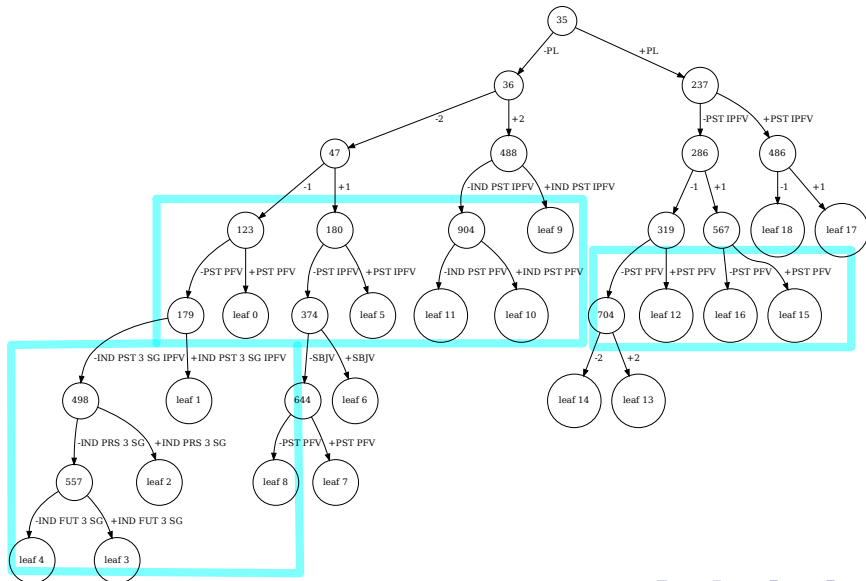
Results: Spanish Verbs



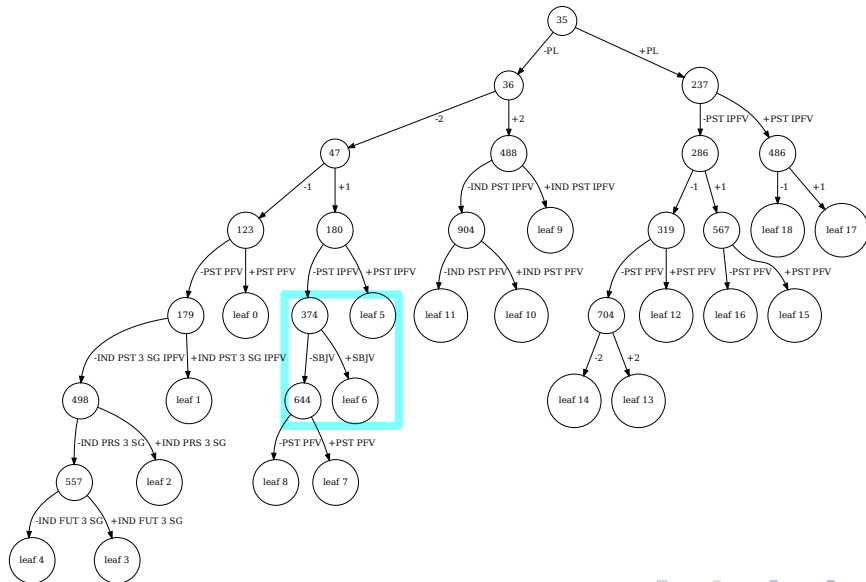
Results: Spanish Verbs



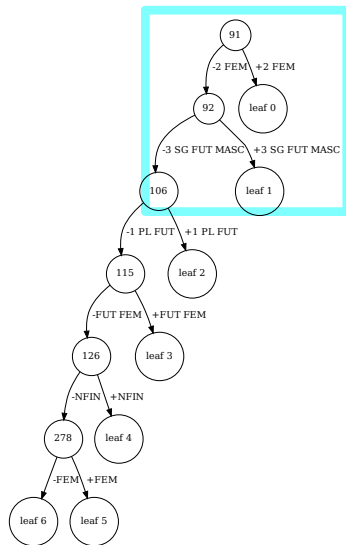
Results: Spanish Verbs



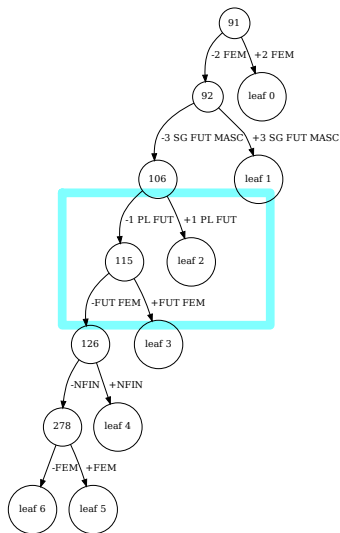
Results: Spanish Verbs



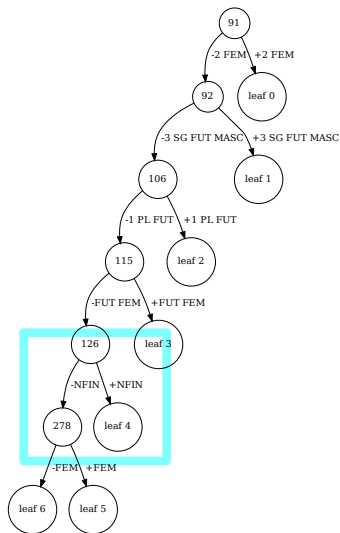
Results: Hebrew Verbs



Results: Hebrew Verbs



Results: Hebrew Verbs



1 Background

2 Proposed Method

3 Results

- English Verbs
- German Nouns
- Spanish Verbs
- Hebrew Verbs

4 Conclusions & Future Directions

- We presented a recursive model able to learn *what* features are marked across a typologically diverse set of languages

- We presented a recursive model able to learn *what* features are marked across a typologically diverse set of languages
 - Works across these languages because only considers inequality between inflected forms.

Conclusions

- We presented a recursive model able to learn *what* features are marked across a typologically diverse set of languages
 - Works across these languages because only considers inequality between inflected forms.
- This model matches well with developmental findings and provides interpretable results regarding order of acquisition.

Now that we know what is marked, learn the processes themselves!

- For languages like German & English, we can apply the model of Belth et al. (2021) to map between e.g. singular and plural

Now that we know what is marked, learn the processes themselves!

- For languages like German & English, we can apply the model of Belth et al. (2021) to map between e.g. singular and plural
- For Spanish and Hebrew, open question

Acknowledgements

I am grateful for the mentorship of Dr. Charles Yang, Dr. Julie Anne Legate, Dr. Jordan Kodner, and Daoxin Li, who provided valuable insight into this work at every stage. I thank the members of Ling-300 for helpful feedback at every stage of this work.

I am also grateful to my friends and family: Anjali Gupta, Emily Hong, and Yara Salim held me accountable during study sessions, and Chris Payne, Nicola Payne, Liz Brogden, Raymond Yang, Ashton Courtney, Marynancy Mwakalindile, Dana Reckard, Alex Yang, and Kay Mayle provided support throughout this project.

References I

- Bat-El, O. (2014). *The Acquisition of Hebrew Phonology and Morphology*. Brill.
- Belth, C., Payne, S., Beser, D., Kodner, J., and Yang, C. (2021). The greedy and recursive search for morphological productivity. *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*.
- Berko, J. (1958). The child's learning of english morphology. *Word*, 14(2-3):150–177.
- Brown, R. (2013). *A first language*. Harvard University Press.
- Clahsen, H., Aveledo, F., and Roca, I. (2002). The development of regular and irregular verb inflection in spanish child language. *Journal of child language*, 29(3):591–622.
- Kauschke, C., Kurth, A., and Domahs, U. (2011). Acquisition of german noun plurals in typically developing children and children with specific language impairment. *Child Development Research*, 2011.

References II

- Kirov, C., Cotterell, R., Sylak-Glassman, J., Walther, G., Vylomova, E., Xia, P., Faruqui, M., Mielke, S. J., McCarthy, A. D., Kübler, S., et al. (2018). Unimorph 2.0: universal morphology. *arXiv preprint arXiv:1810.11101*.
- Lustigman, L. (2013). Developing structural specification: Productivity in early hebrew verb usage. *First Language*, 33(1):47–67.
- Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., and Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, pages i–178.
- Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*, volume 37. John Benjamins Publishing.
- Payne, S., Kodner, J., and Yang, C. (2021). Learning morphological productivity as meaning-form mappings. In *SCIL*.
- Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT press.

Thank you!!