The Recursive Search for Morphological Productivity

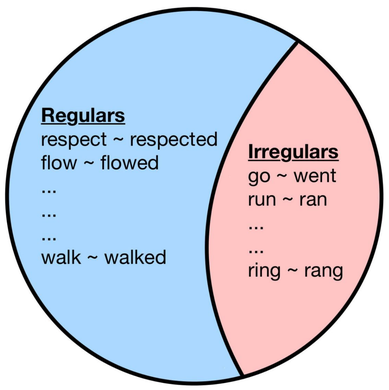
Sarah Payne[1], Caleb Belth[2], Jordan Kodner[3], and Charles Yang[1]

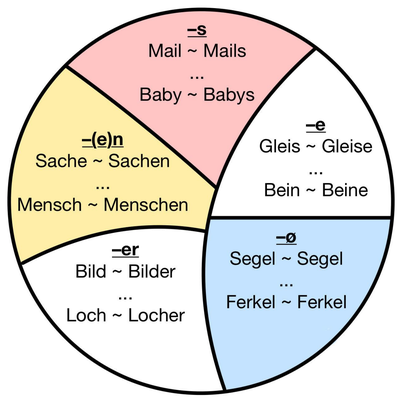
[1]University of Pennsylvania, [2]University of Michigan, [3]Stony Brook University

Background

Morphological productivity

- Arises despite exceptions, as in the **English past tense** (left)
- May be restricted to subgroups, as in **German noun plurals** (right)





Children learn these rules on sparse input

Contribution

We present a **model of morphological learning** capable of extracting ***linguistically interpretable rules*** from ***developmentally plausible vocabularies***.

Selected References

C. Kauschke, A. Kurth, and U. Domahs. Acquisition of German Noun Plurals in Typically Developing Children and Children with Specific Language Impairment. *Child Development Research*, 2011.

S. A. Kuczaj. The acquisition of Regular and Irregular Past Tense Forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600, 1977.

R. Wiese. *The phonology of German*. Clarendon, Oxford, 1996.

C. Yang. *The Price of Linguistic Productivity: How Children Learn to Break the Rules of Language*. MIT press, 2016.

Model

The Tolerance Principle (TP):

- A hypothesized rule ***r*** that may be applied to ***N*** lexical items (types) in the learner's vocabulary is productive iff the number of observed exceptions, ***e***, to ***r*** among those ***N*** items, satisfies $e \leq N/\ln N$

Learning Procedure:

- Hypothesize a rule over ***morphosemantic and phonological features*** and check against TP
- If TP fails, perform ***best-first search*** by subdividing based on the most frequent suffix and recursing

Data

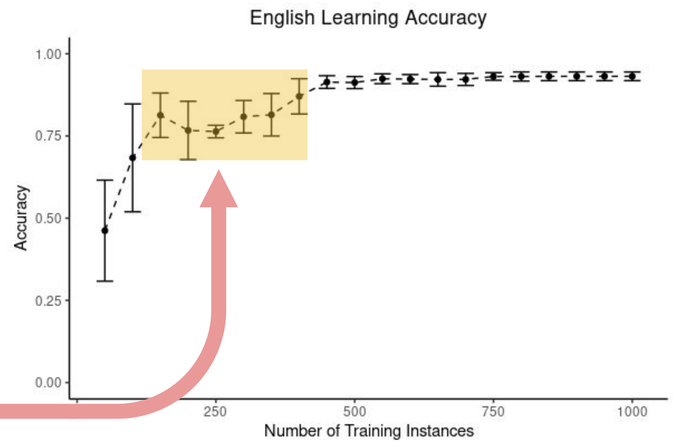
Input: (Lemma, inflected, feature) pairs

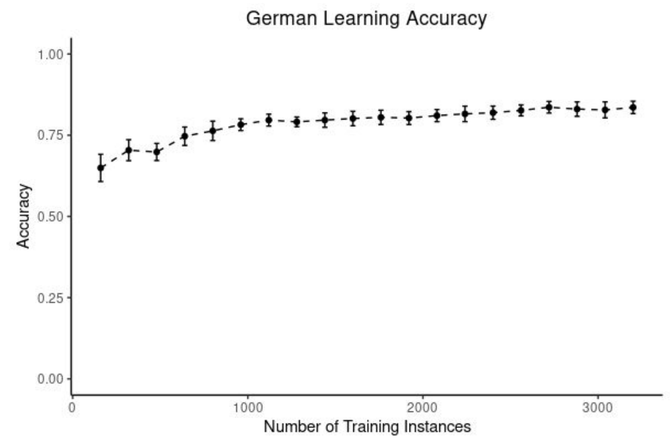
- English: **(walk, {3, SINGULAR, PAST}, walked)**
- German: **(Sache, {FEMININE}, Sachen)**

- **10 splits of stochiastically sampled** nouns from German CELEX and verbs from English CHILDES
- **Morphosemantic features** provided:
  - ***Person, number, and tense*** for English
  - ***Number, case and gender*** for German
- **Phonological features** extracted by learner from ends of lemmas if morphosemantic features insufficient

Results & Discussion

- The model is applied to 200 novel English and 640 novel German lemmas
- ***Our model's learning curves mirror acquisition patterns***, including the **U-shaped English learning curve.**





- The model produces ***linguistically interpretable rules that align with theoretical descriptions:***
  - Rules learned on **300 English words (left)** describe the past tense voicing alternation
  - Rules learned on **800 German words (right)** describe several phonological rules from the theoretical literature

[VERB, PAST] → -d / [n, ɪ, l, m, u, i, z, r, ʒ, v]__
[VERB, PAST] → -t / [k, s, ʃ, p]__
[VERB, PAST] → -ɪd / [t,d]__

[NOM, M] → -e / [d, b, f, g, h, z, s, k, m, tt, il, rt, kt, at, ht, lt, ll, ier]__
[NOM, F] → -en / [r, t, n, g, z, hl]__
[NOM, F] → -n
[NOM] → −∅ / [r, en, um, el]__
[NOM, N] → -e / [r, l, m, t, n, z]__
[NOM, M] → -en / [r, t]__
[NOM, M] → -n / [e]__
[NOM, N] → −∅ / [e]__
[NOM, N] → -se
[NOM, M] → −∅ /[s]__
[NOM, M] → -er

Contact: paynesa@sas.upenn.edu cbelth@umich.edu jordan.kodner@stonybrook.edu charles.yang@ling.upenn.edu