Contrast, Sufficiency, and the Acquisition of Morphological Marking

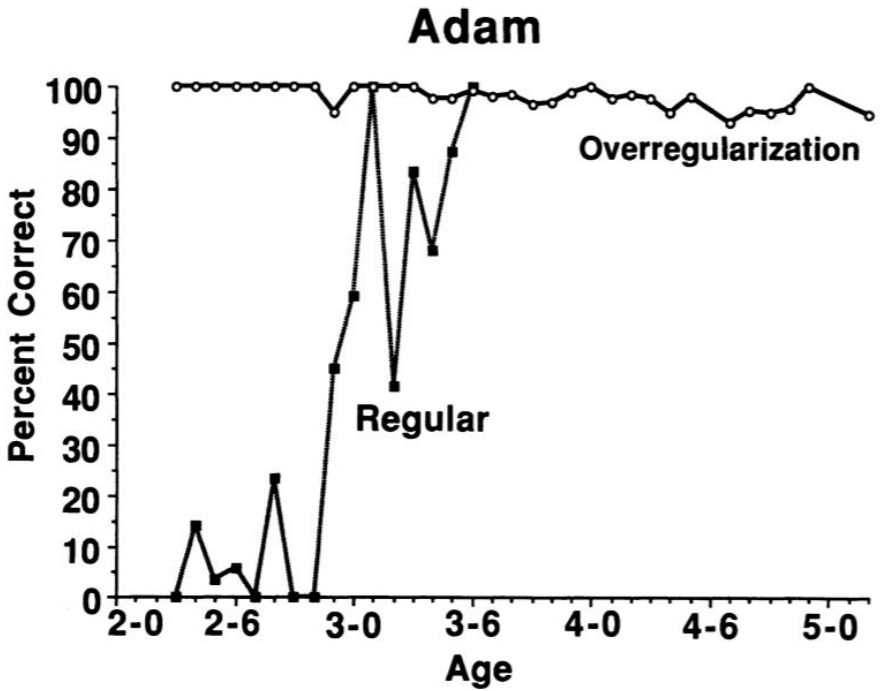
Sarah Brogden Payne
Stony Brook University
sarah.payne@stonybrook.edu

BUCLD 2022

Background

- **Previous work:** how do children learn to map morphosyntactic features to form?
  - English Past Tense
  - German Noun Plurals
  - Spanish Verbal Inflection
  - Hebrew Verbal Inflection
- ***But how do children learn which morphosyntactic features are marked in their language to begin with?***

Background: English Past Tense Acquisition



Order of Acquisition:

- ***-ing*** early
- ***-s*** later
- ***-ed*** usually last, by 3;0

Berko, Jean. 1958. “The child’s learning of English morphology.” Word 14 (2-3): 150–177.

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

Marcus, Gary et al. 1992. “Overregularization in language acquisition.” Monographs of the society for research in child development, i–178.

Background: German Noun Plural Acquisition

Different problem: all suffixes have same feature: +PLURAL

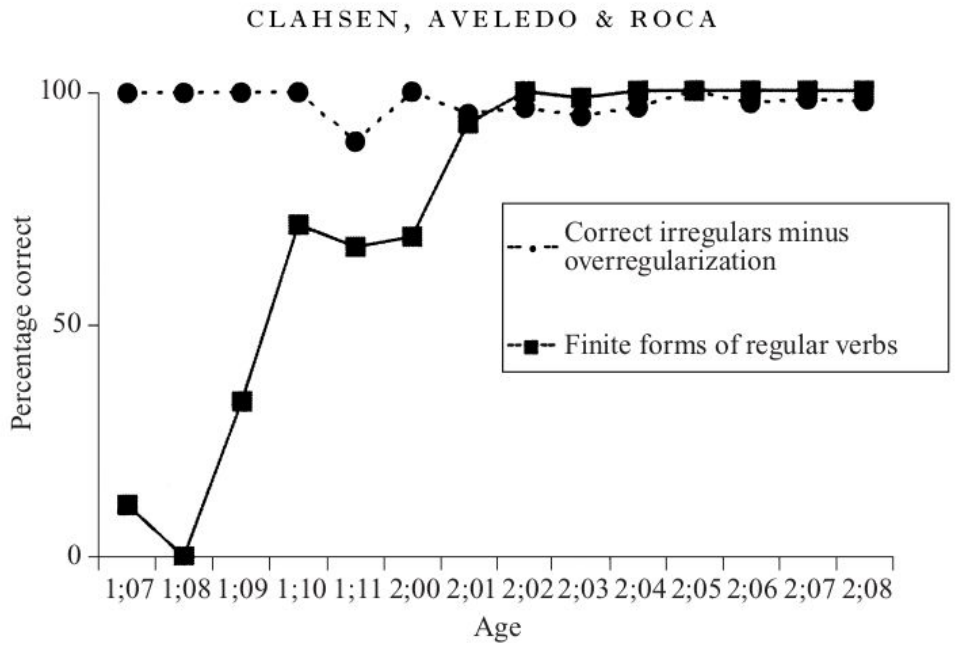
- Productive ***-s*** is **elsewhere condition** but least frequent in input:
  - Interesting implications for modeling work
- No clear order of acquisition

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.

Background: Spanish Verb Acquisition

Order of Acquisition:

- ***Finiteness*** & person marking: 1;7
- ***Number*** marking: 1;7-2;0
- ***Tense:*** 2;0-2;2
- ***Mood*** between 1;7-2;2



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

Background: Hebrew Verb Acquisition

Order of Acquisition:

- ***Person, number & gender*** before tense
- Order of ***person vs. number*** varies
- ***Gender*** appears before or at the same time as number

Lustigman, Lyle. 2013. “Developing structural specification: Productivity in early Hebrew verb usage.” First Language 33 (1): 47–67.
Bat-El, Outi. 2014. The acquisition of hebrew phonology and morphology. Leiden, Netherlands: Brill.

Background

- **Previous work:** how do children learn to map morphosyntactic features to form?
  - English Past Tense
  - German Noun Plurals
  - Spanish Verbal Inflection
  - Hebrew Verbal Inflection
- ***But how do children learn which morphosyntactic features are marked in their language to begin with?***

Contributions

We present a model that learns ***which morphosyntactic features*** are marked across a ***typologically diverse*** set of languages from ***developmentally appropriate*** vocabularies.

We show that this model ***matches well with developmental findings.***

Training Data

- **CHILDES** most frequent forms as ***proxies for early vocabulary***
  - Model takes in forms in order of descending frequency
- **UniMorph** annotated inflectional morphological dataset
- **Input:** (lemma, inflected form, morphosyntactic features)
  - ***Example:*** **English verbs**: (walk, walked, {3, SINGULAR, PAST})
- **Maximum Training Sizes** (number of lemmas):

English	German	Spanish	Hebrew
1280	1444	310	151

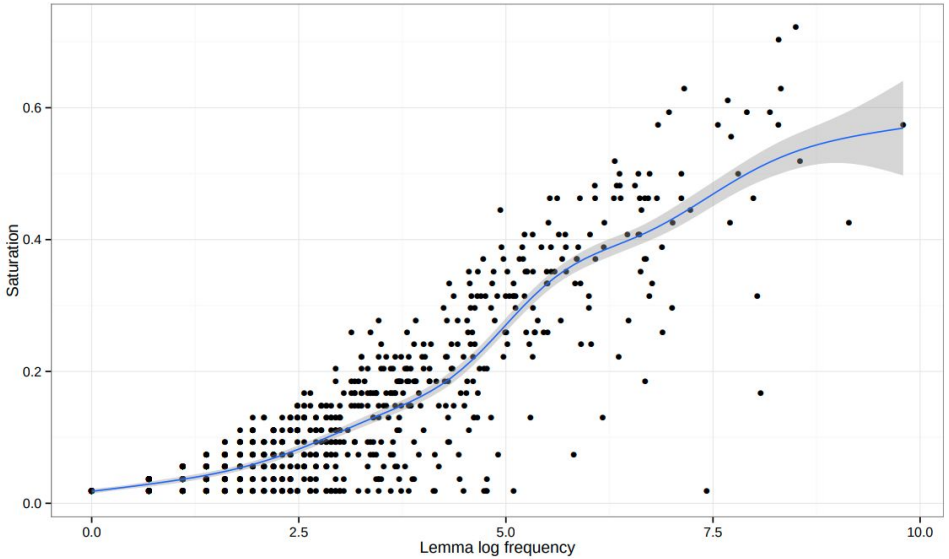
Model: Collisions & The Principle of Contrast

- **Principle of Contrast:** distinct forms => distinct meanings
  - ***walked*** and ***walk*** can't have the same meaning because they are phonologically distinct
- **Collisions:** instances of lemmas appearing in distinct inflected forms
  - ***walk~walked*** => English marks **±PAST**
- **Infants sensitive to collisions:** can relate nonce inflected words to their stems as early 0;6 (but not for pseudo-suffixes)

Clark, Eve V, and Brian MacWhinney. 1987. "The principle of contrast: A constraint on language acquisition." Mechanisms of language acquisition, 1–33.
Kim, Yun Jung, and Megha Sundara. 2021. "6–month–olds are sensitive to English morphology." Developmental science 24 (4): e13089.

Model: The Tolerance-Sufficiency Principle

- Is a ***single collision*** enough to learn marking?
  - ***I am ~ you are*** => English marks 1 vs. 2 person?
- Do we want ***all items*** to have collisions instead?
  - **Sparsity of the input:** morphological paradigm saturation



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

Model: The Tolerance-Sufficiency Principle (TSP)

- When are there ***enough collisions*** to learn which morphosyntactic features are marked?
- **Intuition:** given a set of items:
  - If we've seen **most** in a certain setting and **they do X** in that setting:
    - **Conclude that all do X**, *even if we haven't seen what the others do in that setting*
  - If we've only **observed a few** in that setting, or observed most but **few do X**:
    - **We can't generalize,** and we lexicalize instead

Model: The Tolerance-Sufficiency Principle (TSP)

- **Threshold** defined by efficiency:
  - Given **N** items, **M** of which we see doing X in a certain setting, all **N** do X in that setting if:

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

  - **For our model:** if enough items appearing in inflection **A** appear in a different form in inflection **B**, then all appear in a different form in inflection **B**

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

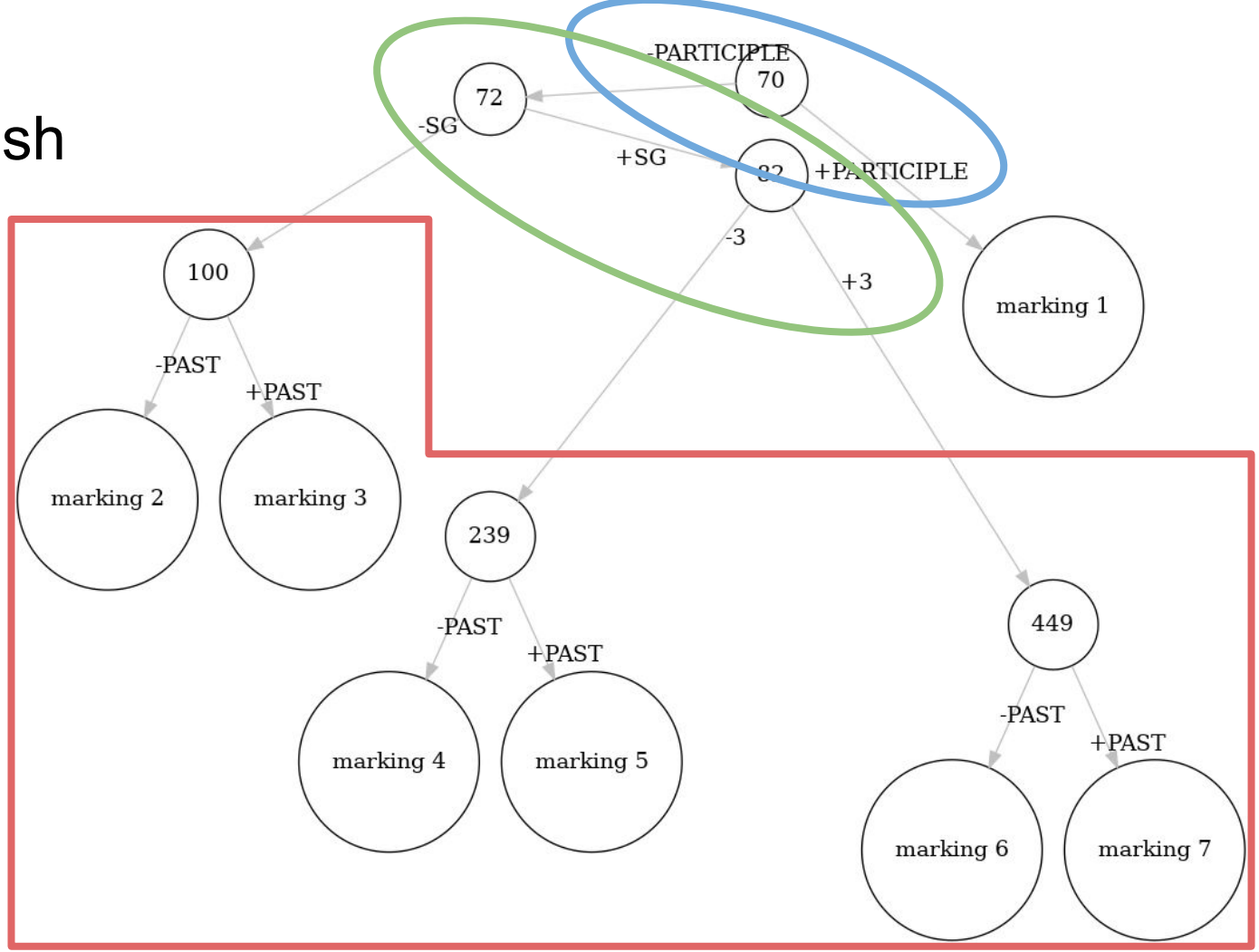
Model: Recursive Subdivision

- Take in input **incrementally** in order of decreasing frequency
- If inflected form ***A*** (less frequent) has a collision with inflected form ***B*** (more frequent)*:*
  - **Do enough verbs which appear in A appear in B *in a different inflected form?***
- Example: collision between ***walked*** and ***walk***
  - **Do enough verbs that appear in +PAST appear in -PAST *in a different inflected form?***

Model: Recursive Subdivision

- If **enough** words have a collision:
  - Subdivide based on the **morphosyntactic difference between *A* and *B***
  - **Recurse on each resulting set**
- Example:
  - Enough collisions between +PARTICIPLE and -PARTICIPLE => **divide lexicon into +PARTICIPLE forms and -PARTICIPLE forms**
  - Recurse on each side; **learn +/-3,SG on -PARTICIPLE branch.**
- Learner produces **binary-branching traces**
  - Each node indicates the vocabulary size when the marking of a given feature set is acquired

Results: English



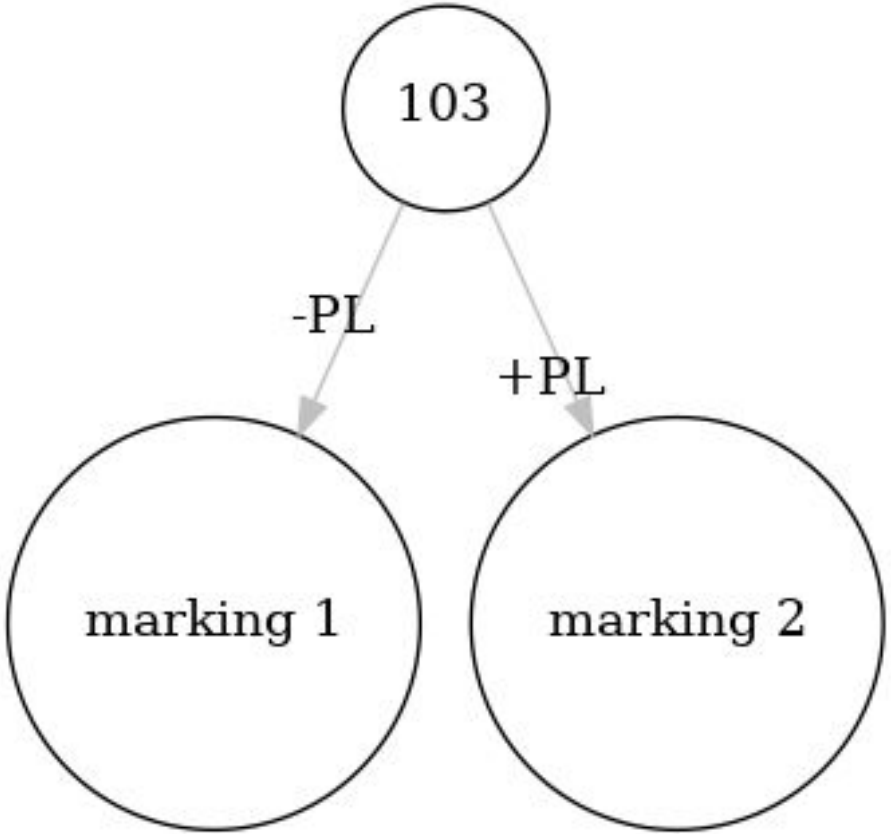
Results: English

- ***-ing*** before ***-s*** before ***-ed***
- Learning done after **188 verb lemmas**
  - Fits with vocab sizes at 3;0
- Past different for each person-number combo?
  - Yang, Ellman, and Legate (2015): past tense acquired later for speakers of AAVE

Bornstein, Marc H et al. 2004. “Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English.” Child development 75 (4): 1115–1139.

Results: German

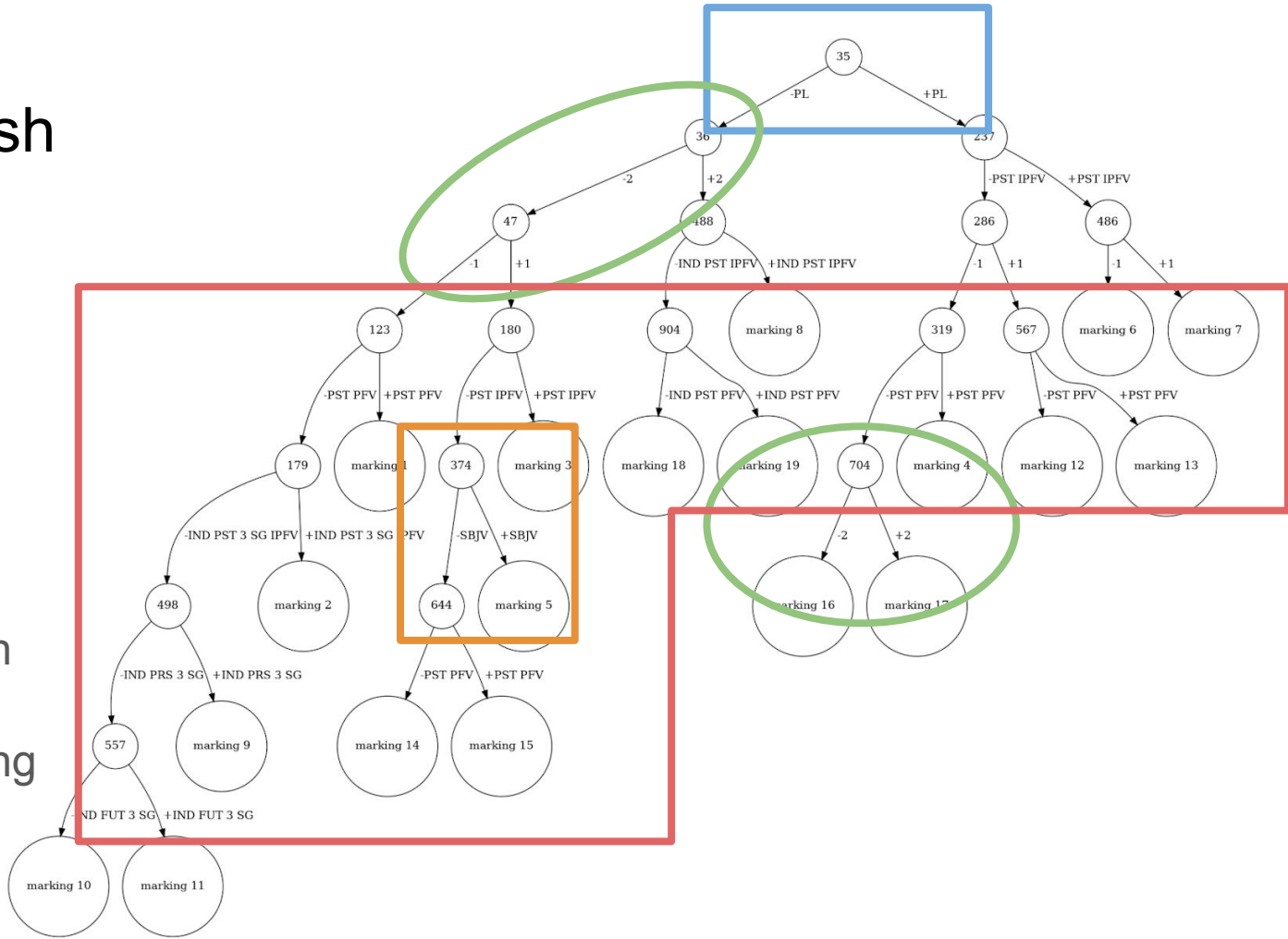
- Learning done after **97 noun lemmas**
- Plural affix overapplication begins when vocab **under 300 words**



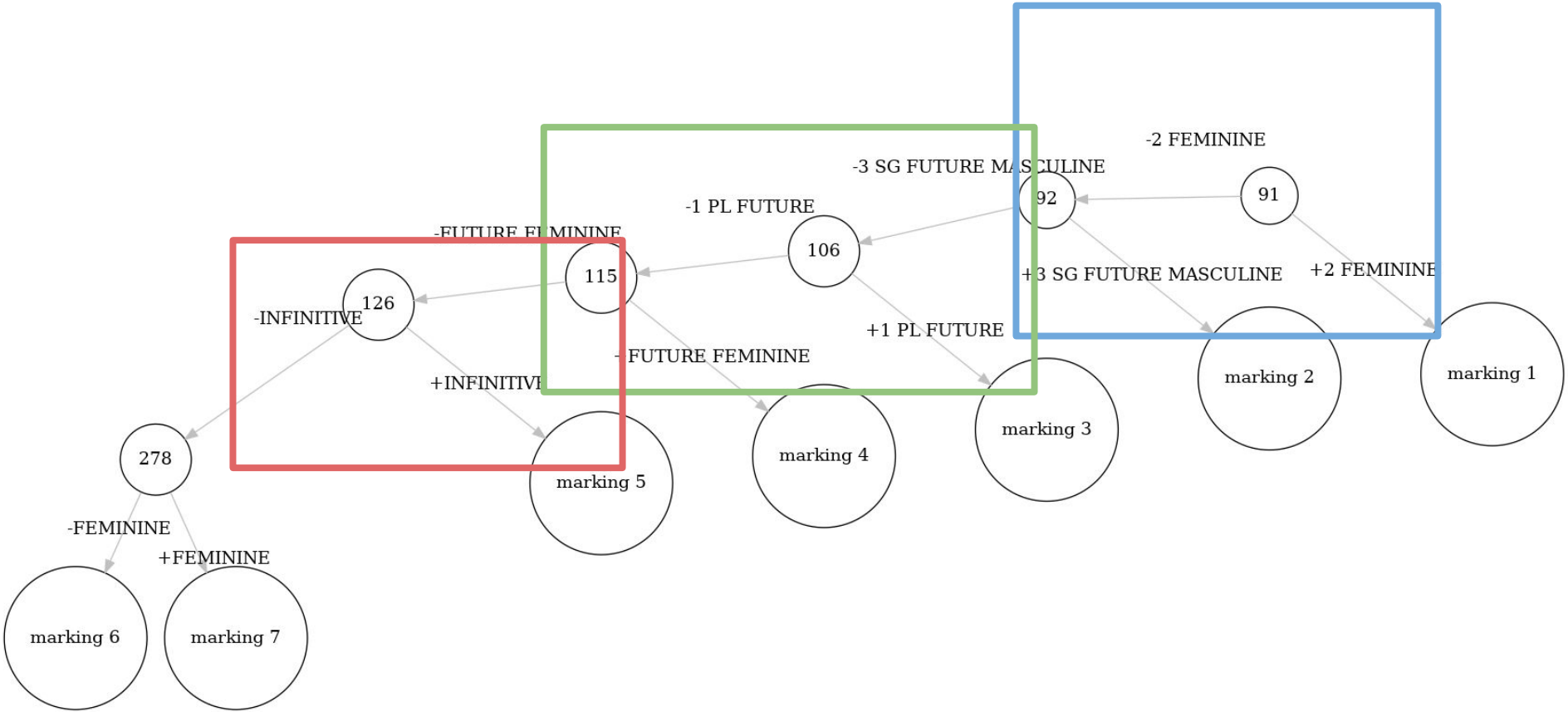
Kauschke, Christina, Anna Kurth, and Ulrike Domahs. 2011. “Acquisition of German noun plurals in typically developing children and children with specific language impairment.” Child Development Research.

Results: Spanish

- Traditional view: ***person before number***
- ***Singular*** person before ***plural*** person instead?
- Learning done on **~300 verb lemmas**, matching developmental findings



Results: Hebrew



Results: Hebrew

- Subject agreement before tense
- Learned from only **151 verbs**
  - Not enough data for complete learning trace

Discussion

- Recursive search based on **Principle of Contrast** and **Tolerance-Sufficiency Principle** learns ***which morphosyntactic features are marked*** in a developmentally-plausible way
- Reliance only on inequality => generalization to morphological marking in **typologically diverse languages**

Future Work: Model Implications

- Investigation of **Spanish person-number ordering**
  - Does person really emerge before number?
  - Or do children know their language marks number but haven't learned how to mark the plural?
- Investigation of **English tense "splitting"**
  - Does past tense emerge at different times for different person agreements?
- Investigation of **Root Infinitives**
  - May emerge before tense marking is acquired

Future Work: Further Modeling

- **Combine with grounded/distributional models** to learn morphosyntactic features rather than providing them explicitly
- Combine with models that **map morphosyntactic features to phonological form** (e.g. Belth et al. 2021)
- Apply model to **case morphology** (e.g. German nouns)

Thank you!!

I am grateful to Charles Yang, Julie Anne Legate, Jordan Kodner, Jeff Heinz, Caleb Belth, and members of the SBU Linguistics Brown Bag and Bob Berwick’s Lab Meeting for their feedback and support.

This work was supported by the Paula Menyuk Travel Award, Institute for Advanced Computational Science Graduate Fellowship, and NSF Graduate Research Fellowship.

This work is taken from my bachelor’s thesis: