

## Morphological variation in the English past tense: Probabilistic regularities within and across speakers

Péter Rácz<sup>1</sup>, Clay Beckner<sup>1</sup>, Jennifer B. Hay<sup>1</sup>, Janet Pierrehumbert<sup>2</sup>

<sup>1</sup>New Zealand Institute of Language Brain and Behaviour, University of Canterbury

<sup>2</sup>Department of Linguistics, Northwestern University, Northwestern Institute on Complex Systems

We collected data on English past-tense formation from 111 American English speakers on Amazon Mechanical Turk. Participants were presented with a forced-choice task in which they picked either the regular or the irregular past tense form for an English nonce verb, presented in a carrier sentence. The English past tense has been extensively studied with respect to acquisition, cognitive representations, and historical change (Bybee & Slobin 1982, Albright & Hayes 2003), but ours is the first one on a large scale.

The 316 nonce verb stimuli were drawn from 5 morphophonological categories, labeled here with representative examples from real English verbs: (1) SING (e.g. *gink/gank*) (2) BURN (e.g. *gurn/gurnt*) (3) KEEP (e.g. *freep/frept*) (4) DRIVE (e.g. *smide/smode*), and (5) CUT (e.g. *vost/vost*).

The resulting dataset contains representations of many items, distributed across many speakers. This makes it well-suited to ask two innovative questions of considerable theoretical interest, regarding individual variation in past-tense formation preferences, as well as the general relationship between individual-level and group-level variation.

First: To what degree do individual tendencies in one category correlate with tendencies in another? Are participants who are high regularizers for one category, also high regularizers for the other? To answer this question, we investigated pairwise correlations by subjects, for rates of regularization of the different verb categories. All categories were highly correlated (Spearman  $r_s = .5$ ,  $p < .0001$  in all cases), with the exception of the CUT class. Disregarding CUT, the verb classes act together—if a speaker prefers irregular forms in one, they will also prefer irregular forms in the other.

The CUT verbs appear to behave quite differently, and — interestingly — show strong correlation with speaker age. Across the age spectrum (20-65) in our study, younger participants are significantly more likely to choose irregular (no-change) variants for CUT verbs than older participants ( $p < .01$ ). Previous work observes that children are more susceptible to no-change variants in morphology (Berko 1958). However, the broader finding here that age correlates with no-change preferences represents a new discovery, and requires additional investigation into ongoing change.

Second: exactly how are items distributed across speakers? We are concerned with understanding the link between speaker variation and item variation — an issue pertinent to understanding the patterning of all sociolinguistic variables, yet — to our knowledge — never overtly explored.

We ranked items by regularization rate, and then considered the mean participant rank for each item. If items are distributed over people in a structured manner, then there should be a considerable correlation. The most regularized items should have low mean participant ranks (because participants of any rank might use them), and the least regularized items should have high participant ranks (because we only expect them to be regularized by the most prolific regularizers). We find strong correlations, both within verb classes and across them. Statistical comparison with simulated data shows that the observed distributions differ markedly from what would be expected if participants regularized at varying rates, but each over a random subset of items. Items are non-randomly distributed across subjects.

## References

- Adam Albright and Bruce Hayes. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90(2):119-161.
- Berko, Jean. (1958). The child's learning of English morphology. *Word* 14: 150-177.
- Bybee, Joan L. and Dan I. Slobin. (1982). Rules and schemas in the development and use of the English past tense. *Language* 58: 266-289.