

How to use Yang's Principles to model acquisition in diachrony The case of psych verbs

Carola Trips¹ & Tom Rainsford²

¹University of Mannheim, ²University of Stuttgart

Any study looking at acquisition in the past must infer aspects of the input to which children are exposed from written texts, yet it is clear that they are not equivalent. As a model of learnability, Yang's (2016) Tolerance and Sufficiency Principles are good candidates for the study of the acquisition of productive rules in historical data, and they have been applied in a number of recent studies (e.g. Kodner 2019, 2020, 2022; Dresher and Lahiri 2022, Ringe and Yang 2022). The model provides a simple but effective algorithm for predicting the point at which language learners will conclude that the number of lexical items belonging to a given class and providing positive evidence for a particular rule is sufficient to conclude that the rule is productive, barring a small number of exceptions which must be memorized. On the basis of child-directed speech data from the CHILDES corpus (MacWhinney 2000), Yang (2016) has shown that the Tolerance Principle is effective in modelling the acquisition of productive rules such as the use of the *-ed* past-tense marker and its corollary, the Sufficiency Principle, is well-suited to the acquisition of argument structure, such as modelling the subset of ditransitive verbs showing double object constructions in modern English (see also Kodner 2019). Not only has the Sufficiency Principle been shown to be effective in correctly predicting the course of acquisition from small amounts of data, similar to those to which a child would be exposed and to the limited data available to historical linguists, the calculation only requires two parameters to be estimated: the total number of lexical items within the class to which the learner is exposed (henceforth N) and the number of these lexical items to which the rule in question can be applied (henceforth M).

However, applying the Sufficiency Principle to historical data brings a number of unique problems not present in the child-directed speech data examined by Yang. In a recent study of the acquisition of psych verbs in Middle English, Trips and Rainsford (2022) identify three central issues: First is the class size problem: how is it possible to estimate the number of lexical items in a particular class (N), in this case psych verbs, from heterogeneous historical corpora? Second is the attestation problem: what is the best way to estimate the positive evidence for a given rule (M), in this case, the use of a subject-EXPERIENCER argument, from historical data? Third is the data compatibility problem: to what extent is data from historical texts comparable to child-directed speech data?

In the present article, we re-examine the validity of the assumptions made by the authors to address these problems. First, contrasting the psych verbs attested in sections M3 (1350-1420) and M4 (1420-1500) of the *Penn-Helsinki Parsed Corpus of Middle English* (PPCME2) and those attested in modern English child-directed speech from the CHILDES corpus, we show that there is broad semantic equivalence between the most frequent verbs in historical texts and those found in child-directed speech, confirming that basing Sufficiency Principle calculations on a "frequency-trimmed" subset of verbs from historical corpora is the best approach to ensure data comparability (see Kodner 2019). Second,

contrary to Trips and Rainsford (2022), we advocate using corpus data in addition to lexicographical resources to address the attestation problem, showing that this prevents the analysis being affected by hapax constructions recorded in historical dictionaries which are very unlikely to have formed part of the learner's input. We conclude by suggesting a new template for researchers working with models of learnability in diachrony, in which a comparison with modern child-directed speech data forms an essential guide to the correct interpretation of the historical data.

References

- Dresher, B. Elan, and Aditi Lahiri. (2022) "The Foot in the History of English: Challenges to Metrical Coherence." In *English Historical Linguistics: Change in Structure and Meaning: Papers from the XXth ICEHL*, edited by Bettelou Los, Claire Cowie, Patrick Honeybone, and Graeme Trousdale, 42–59. Current Issues in Linguistic Theory, volume 358. Amsterdam ; Philadelphia: John Benjamins Publishing Company.
- Kodner, Jordan. (2019). "Estimating Child Linguistic Experience from Historical Corpora." *Glossa. A Journal of General Linguistics* 4, no. 1: 122.
- Kodner, Jordan. (2020). "Language Acquisition in the Past." Doctoral dissertation, University of Pennsylvania.
- Kodner, Jordan. (2022). "What Learning Latin Verbal Morphology Tells Us about Morphological Theory." *Natural Language and Linguistic Theory*. <https://doi.org/10.1007/s11049-022-09553-2>.
- Kroch, Anthony, and Ann Taylor, (eds). 2000. *The Penn-Helsinki Parsed Corpus of Middle English, Second Edition (PPCME2)*. Philadelphia: University of Pennsylvania.
- B. MacWhinney. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates; CHILDES -- NICHD HD082736.
- Ringe, Don, and Charles Yang. (2022). "The Threshold of Productivity and the 'Irregularization' of Verbs in Early Modern English." In *English Historical Linguistics: Change in Structure and Meaning: Papers from the XXth ICEHL*, edited by Bettelou Los, Claire Cowie, Patrick Honeybone, and Graeme Trousdale, 91–112. Current Issues in Linguistic Theory, volume 358. Amsterdam ; Philadelphia: John Benjamins Publishing Company.
- Trips, Carola, and Tom Rainsford. (2022). "Tolerating Subject-Experiencers? Yang's Tolerance Principle Applied to Psych Verbs under Contact in Middle English" Volume 6, Article 12: 1–43.
- Yang, Charles. (2016). *The Price of Linguistic Productivity*. Cambridge, MA: MIT Press.