Stability in the Production of Syntactic Attachment Across the Lifespan

Ellis Cain, Alton Chao, Rachel Ryskin (University of California, Merced)
ecain@ucmerced.edu

Verbs can appear in multiple syntactic structures, sometimes creating ambiguity. In "pet the frog [with the feather]," the prepositional phrase (PP) could attach to the verb and signify the instrument used for the action, or to the noun, to modify the object of the verb. Individual verbs are biased to appear with one of the two structures, and listeners rely on co-occurrence statistics to disambiguate meanings (Ryskin et al., 2017; Snedeker & Trueswell, 2004). However, corpus analyses indicate that verb biases may shift over time (Cain & Ryskin, 2024) or differ across dialects (Bresnan & Ford, 2010). In a recent study, older listeners did not appear to have the same verb biases as young adults in a comprehension task, and these differences were predicted by changes in usage patterns from diachronic corpora (Cain & Ryskin, 2024). Whether the usage patterns of older adults, in the present, also differ from those of young adults is an open question. In the current work, we collected sentence completion data for *with*-PPs from different age cohorts to see whether verb-specific syntactic attachment biases in language production differ across age groups. We also test the utility of using large language models (LLMs) for annotation.

Methods: Participants (*n=823* Native English speakers recruited on Prolific) were presented with a stem and asked to provide their completion. The stems consisted of 37 unique verbs that were used in the structure "[*verb*] the [*animal*] with the __." 18 filler sentences lacking the *with*-PP were randomly mixed with the critical stimuli. Throughout the experiment, participants were also asked to provide synonyms or antonyms for a given word as an attention check (*n = 15*). After selecting participants who passed all attention checks, the final dataset consisted of 79,791 completions from 537 participants (age range: 25-76 y.o., mean = 49.3 y.o.).

Due to the presence of global attachment ambiguity, the completions cannot be readily parsed (e.g., using a dependency parser) and require other annotation. However, due to the large amount of data, human hand-annotation is impractical. We therefore tested two automated annotation methods: spaCy dependency parsing (as a baseline) and an LLM (GPT 4o-mini, see prompt below) (Bavaresco et al., 2024; Goel et al., 2023). We used data from a previous set of annotated completions (Set 1: *n = 1,600*; 800 per construction), and hand-annotated a subset of this new dataset (Set 2: *n = 2,360*; 1,180 per construction; IAA *r* = 0.68) to evaluate the accuracy of the annotation methods. Between the two annotation methods, the LLM provided the most accurate annotations (Set 1: 92.3%; Set 2: 75.2%) compared to spacy (51.6%; 52.2%). The LLM prompt that was used can be seen below.

Results: Overall, participants were more likely to use instrument constructions than modifier constructions (59,199 instrument, 18,984 modifier). Verb-specific usage proportions were highly correlated across age groups (Fig. 1; *r* > 0.99) and were also highly correlated with previous norms (0.77 < *r* < 0.80).

Conclusion: Contrasting Cain & Ryskin (2024), which found differences in the processing of verb-specific *with*-PP attachment across the lifespan, we found relative stability in the production of verb-specific syntactic attachment for this alternation. The exact reasons underlying this difference are still unclear. Moreover, we developed a new annotation method for productions with globally ambiguous attachment using GPT 4o-mini, which appears more accurate than existing automated annotation methods. The high correlation of the LLM-annotated verb biases with previous hand-annotated norms from Ryskin et al. (2017) indicates that the method holds promise, though future work is needed to further assess its validity.

References:

Bavaresco, A., Bernardi, R., Bertolazzi, L., Elliott, D., Fernández, R., Gatt, A., ... & Testoni, A. (2024). LLMs instead of Human Judges? A Large Scale Empirical Study across 20 NLP Evaluation Tasks. arXiv preprint arXiv:2406.18403.

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. Language 86(1), 168-213.

Cain, E., & Ryskin, R. (2024). Diachronic change in verb usage statistics predicts differences in sentence processing across the lifespan. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (pp. 221–230). Association for Computational Linguistics.

Goel, A., Gueta, A., Gilon, O., Liu, C., Erell, S., Nguyen, L. H., ... & Feder, A. (2023, December). Llms accelerate annotation for medical information extraction. In Machine Learning for Health (ML4H) (pp. 82-100). PMLR.

Ryskin, R. A., Qi, Z., Duff, M. C., & Brown-Schmidt, S. (2017). Verb biases are shaped through lifelong learning. *Journal of experimental psychology. Learning, memory, and cognition*, *43*(5), 781–794.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: the role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology*, *49*(3), 238–299.

GPT Prompt:

Classify the phrase starting with the word "with" in sentences into three categories: Modifier, Instrument, or Neither.

Modifier: The phrase starting with the word "with" attaches to a noun, describing a characteristic or attribute.

- *{Given two examples here}*

Instrument: The phrase starting with the word "with" attaches to a verb performing an action.

- *{Given two examples here}*

Neither: The phrase starting with the word "with" does not fit into the above categories or does not make sense.

- *{Given two examples here}*

Classify the following sentences accordingly. Do not add any additional messages besides the classifications. Print the sentence, verb, noun, phrase starting with the word "with", and the attachment type ("Instrument", "Modifier", or "Neither") for each sentence in the below format.

Look at the man with the hat. - Look - man - with the hat - Modifier
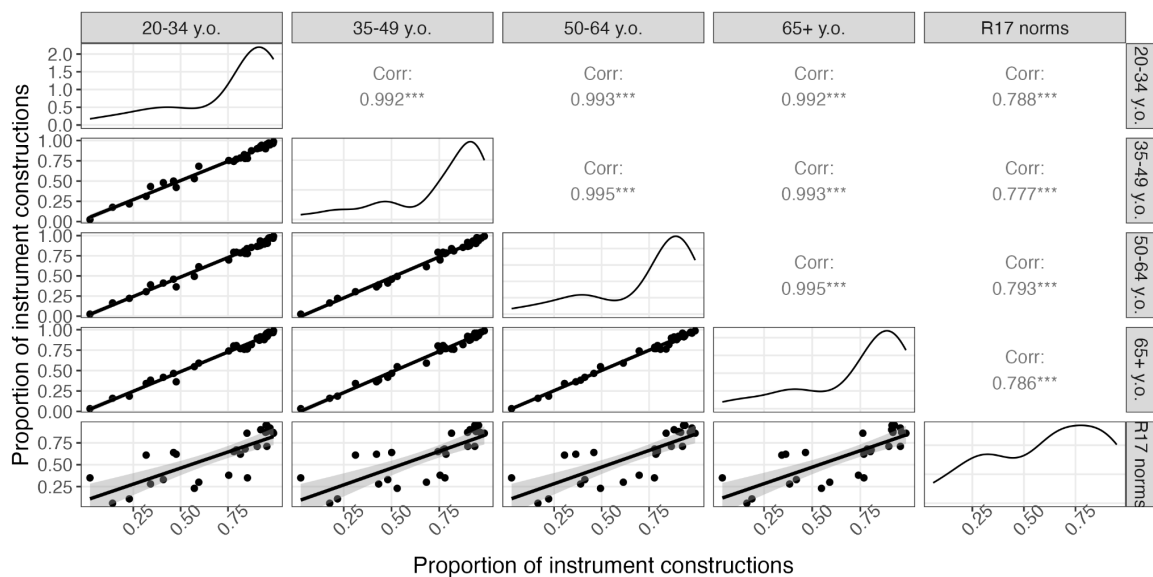
Figures:

Figure 1: Scatterplots and correlation between the proportion of instrument constructions produced by age groups from the newly collected data and production norms from previous research [2]. In the scatterplots, each point represents an individual verb. Diagonal shows the density plots of the distribution of proportions from the group. R17 norms are from Ryskin et al., 2017.