Forward or Backward? Directional Asymmetries in Statistical Learning for Speech Segmentation: Empirical and Computational Evidence

Corresponding Author (c.bernini@bcbl.eu)[1], Chiara Bernini[1], Marina Kalashnikova[1,2], James S. Magnuson[1,2,3]

[1]Basque Center on Cognition, Brain and Language, [2]Ikerbasque: Basque Foundation for Science, [3]Department of Psychological Sciences, University of Connecticut

Statistical learning (SL) refers to the ability to extract regularities from environmental input and is a fundamental component of human cognition. SL is thought to support infants' segmentation of continuous speech into discrete units, such as words, despite the absence of clear acoustic boundaries. One account posits that learners track transitional probabilities (TPs) between successive elements [1,2]. An alternative account proposes that learners track multi-syllabic chunks based on frequency. Evidence suggests that both forward TPs (fTPs; probability that A is followed by B) and backward TPs (bTPs; probability that B is preceded by A) can support speech segmentation [3,4]. However, studies often confound fTPs and/or bTPs with chunk frequency, making it impossible to disentangle the two primary theoretical accounts.

To address this impasse, we propose a new SL artificial language paradigm consisting of 9 bisyllabic words, with all syllables and multi-syllabic chunks matched in frequency. The grammar makes bTPs more informative (1.0 within words, 0.33 between words) than fTPs (0.33 within, 1.0 between). After exposure, participants chose between words and partwords in a forced-choice task. Crucially, fTP for partwords (final syllable of one word and first of another) was 1.0, placing fTP- and bTP-based learning in direct opposition. Learners attuning to bTPs should prefer words, while those attuning to fTPs should prefer partwords. We tested four theoretically-motivated models of segmentation on this paradigm: a simple recurrent network (SRN [5]), associated with TP-based learning; PARSER [6], a classical chunking model; TRACX2 [7], a connectionist chunking model; and CIPAL [8], a chunking model differing from PARSER in chunk formation and representation. The SRN and CIPAL showed a systematic preference for partwords, indicating fTP sensitivity, whereas PARSER and TRACX2 showed no consistent preference. A behavioral experiment with adult participants (n=55) revealed a mean preference for partwords (36% word choices, p=0.001), aligning with SRN and CIPAL. Interestingly, 49% of subjects showed significant partword preference, while 11% showed significant word preference. Of models with human-like mean preference, only the SRN showed variability, with a small proportion of runs yielding a word preference. Our results suggest that most human learners rely more on fTPs than bTPs when frequency cues are eliminated. The model results suggest this paradigm is outside the scope of earlier chunk-based models, but consistent with CIPAL and a generic, prediction-based mechanism (as in the SRN).

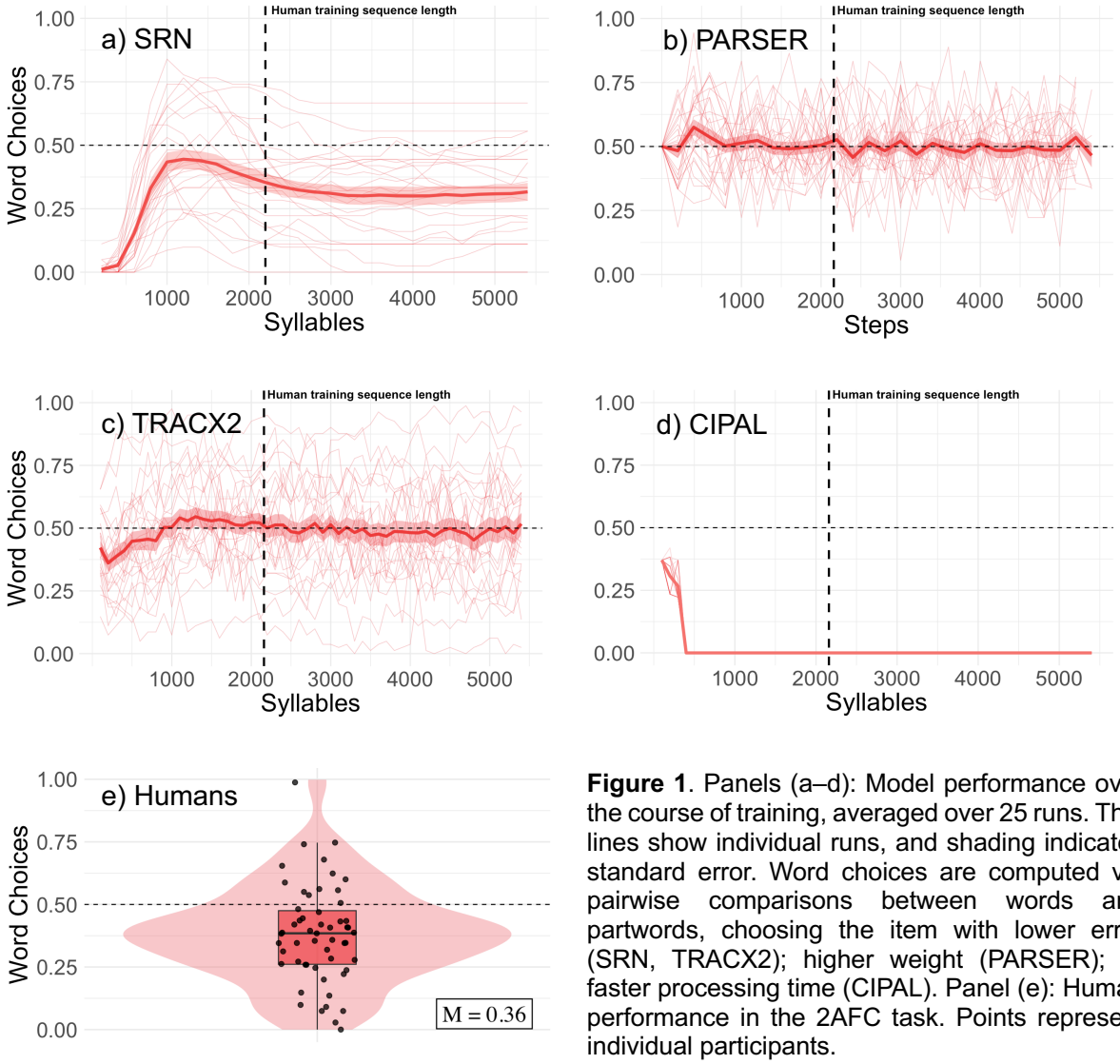

Figure 1. Panels (a–d): Model performance over the course of training, averaged over 25 runs. Thin lines show individual runs, and shading indicates standard error. Word choices are computed via pairwise comparisons between words and partwords, choosing the item with lower error (SRN, TRACX2); higher weight (PARSER); or faster processing time (CIPAL). Panel (e): Human performance in the 2AFC task. Points represent individual participants.

References

[1] Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, *274*(5294), 1926–1928.

[2] Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psy. Sci.*, *9*(4), 321–324. Scopus.

[3] Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Mem. Cognit.*, *36*(7), 1299–1305.

[4] Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.

[5] Elman, J. L. (1990). Finding structure in time. *Cog. Sci.*, *14*, 179–211.

[6] Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *J. Mem. Lang.*, *39*(2), 246–263.

[7] French, R. M., & Cottrell, G. (2014). TRACX 2.0: A memory-based, biologically-plausible model of sequence segmentation and chunk extraction. *Cog. Sci. Proceedings*, *36*(36).

[8] Jessop, A., Pine, J., & Gobet, F. (2025). Chunk-based incremental processing and learning: An integrated theory of word discovery, implicit statistical learning, and speed of lexical processing. *Psy. Rev.*, *132*(6), 1340–1374.