What drives predictive accuracy and response times in incremental sequence learning?

Jessie S. Nixon (jessie.nixon@uni-bielefeld.de)[1], Erdin Mujezinovic[2], Ruben van de Vijver[2]
[1]Dept. Linguistics, University of Bielefeld, [2]Dept. Phonetics, HH University of Düsseldorf

Speech consists of a dynamic signal that unfolds over time, with multiple levels of temporally organised structure. Previous work has shown that humans have a remarkable ability to learn from statistical information in such temporal sequences [1]. Despite its central role in a wide range of linguistic processes and representations, relatively little attention has focused on what mechanisms might underlie sequence learning. An exception is [2], which investigated sequence learning in typing. The present study aims to investigate a) whether we find evidence of *incremental learning during exposure* and b) which *learning mechanisms* best predict accuracy in predicting upcoming syllables (Exp 1) and response time (RT) in syllable identification (Exp 2).

[1] conducted a forced-choice task after training. As we were interested in *incremental* learning, we measured responses throughout the experiment: typing the *upcoming* syllable (Exp 1; measure: accuracy) and typing a key corresponding to the presented syllable (Exp 2; measure: RT). Exp 1 (16 L1 German-speakers) was based on [1] with the same stimuli: 6 trisyllabic words made up of 11 spoken CV syllables, combined into a running speech stream; Exp 2 was based on [2], but with spoken rather than visual cues: 16 different German-speakers heard one of six syllables (la, je, fo, sa, ke, do; e.g. 'sa') and typed the corresponding key (e.g. s) as quickly and accurately as possible.

We expected that learning would occur, such that responses would be more accurate (Exp 1) and faster (Exp 2) with increasing predictability. To measure predictability, we used two learning mechanisms proposed to underlie sequence learning: transitional probability (TP) and error-driven learning (EDL), both based on the preceding syllable for compatibility with [1]. TP was the measure proposed by [1] (TP of Y given X = frequency of bigram XY / frequency of X); EDL used the Rescorla-Wagner equations in the `edl` package[3] with preceding syllables as *cues* and upcoming syllables as the *outcome* to be predicted.

Mean accuracy across all participants and trials in Exp 1 was above chance (22%; one-tailed, $t=16.4$, $p<.001$), demonstrating that incremental learning took place during exposure, even for this difficult task of predicting upcoming syllables. To test the learning mechanism, we generated the two predictability measures (TP and EDL activation) for each individual participant on each trial for each experiment and analysed the results using GAMMs [4]. We included a smooth for the predicability measure (TP or activation) and random effects for participants, with response accuracy (Exp 1) or RT (Exp 2) as dependent variables (Fig 1). Model comparisons using `compareML()` found that activation predicted accuracy better than TP (fREML difference scores 12.4, Exp 1; 3.1, Exp 2). In addition, the fluctuations in accuracy (top right) and the upward deflection in RT after TP=0.2 (bottom right) are *not expected theoretically* for TP; positive and negative correlations are expected here, respectively. Note that with [1]'s design and stimuli, there are TP ranges with sparse data (e.g. between TP 0.8-0.99), likewise for Exp 2 (TP $>$ 0.5), leading to large confidence intervals for these values (values with fewer tick marks, Fig 1 right panels). Nevertheless, a check of the raw data confirms that accuracy was actually higher (38%) with TP=0.5 than with TP=1 (18%) in Exp 1. As these were among the most frequent TP values, sparse data is not an issue here. This result suggests that high TP does not always lead to high accuracy. Although TP may provide a good overall prediction on relatively large amounts of summed TP data, as in [1], the present results suggest it may not be a good predictor of data obtained *incrementally* during learning.

The present results provide insights into the mechanisms underlying sequence learning and extend the findings of [1]. While [1] conducted a forced-choice task after training, the present study found evidence for incremental learning. In addition, we found that TP seems to overestimate predictability in certain cases. This makes sense intuitively, as high TP (e.g. TP of 1) can occur in early trials, when accuracy is intuitively not expected to be high. Participant responses were better predicted here by error-driven learning, which takes into account important factors, such as the history of learning (which here includes effects akin to frequency) and cue competition.

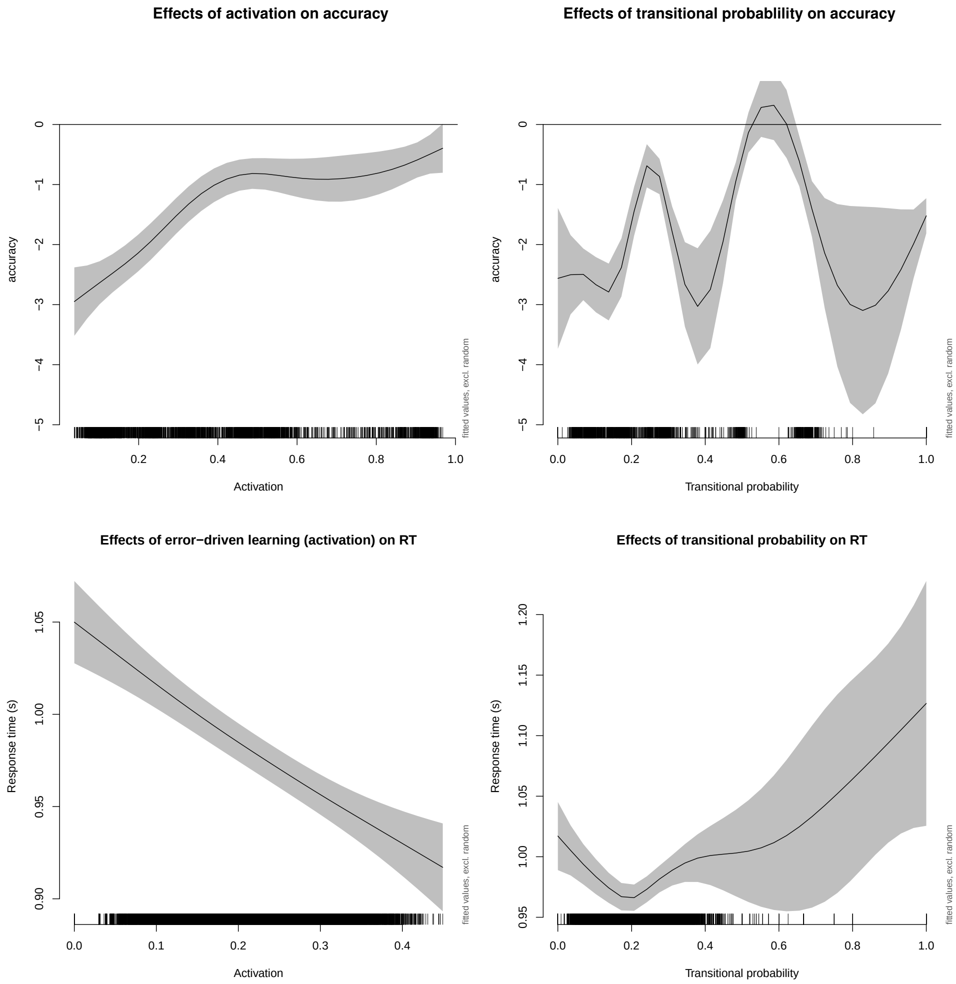

Figure 1: Results of GAMM models of response accuracy (Experiment 1, top) and reaction time (Experiment 2, bottom), with smooths for error-driven (EDL) activation (left) and transitional probability (right). The predictability measure (Activation or Transitional probability) is on the x-axis; the response measure (accuracy or RT) is on the y-axis (accuracy is in log odds scale). Data points are represented by tick marks along the x-axis. Shaded areas represent confidence intervals. Random effects are excluded from the plots.

[1] Jenny R Saffran, Elissa L Newport, and Richard N Aslin. Word segmentation: The role of distributional cues. *Journal of memory and language*, 35(4):606–621, 1996.

[2] Fabian Tomaschek, Michael Ramscar, and Jessie S. Nixon. The keys to the future? An examination of associative versus discriminative accounts of serial pattern learning. *Cognitive Science*, 48(2):e13404, 2024. doi: https://doi.org/10.1111/cogs.13404.

[3] Jacolien van Rij and Dorothee Hoppe. edl: Toolbox for error-driven learning simulations with two-layer networks, 2020. R package version 0.3.

[4] Simon N Wood. *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, 2017.