Prior knowledge of distributional information boosts statistical learning in adults and 8-year-old children

Katja Stärk (MPI for Psycholinguistics), Evan Kidd (MPI for Psycholinguistics, Australian National University, & ARC Centre of Excellence for the Dynamics of Language), & Rebecca L. A. Frost (MPI for Psycholinguistics & Edge Hill University)
Katja.Staerk@mpi.nl

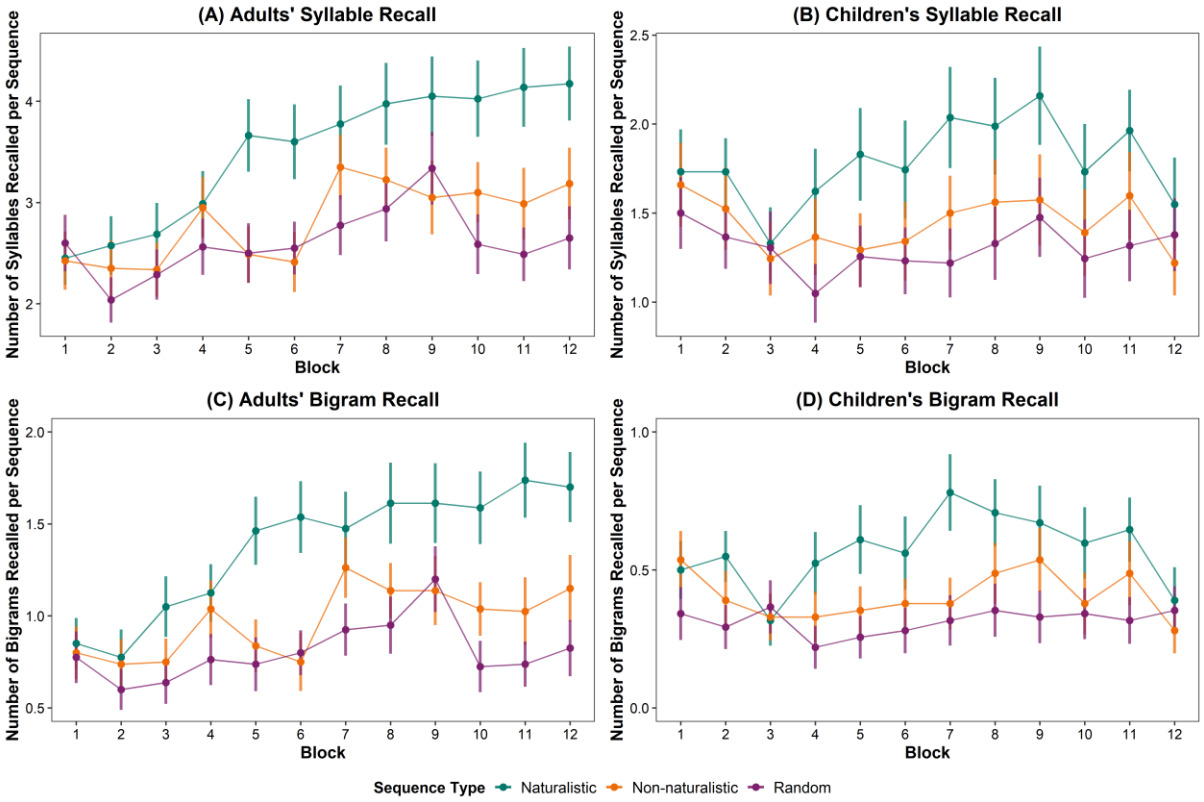
Background. Statistical learning (SL), the ability to implicitly track patterns in the environment, has been identified as a key mechanism for language acquisition (Romberg & Saffran, 2010; Saffran et al., 1996). While there is an underlying assumption that the output of SL is a mental representation of the tracked patterns, studies of auditory SL typically test participants on stimuli without reference to the natural distributions present in the participants' language (see Frost et al., 2015; Siegelman et al., 2018). Here, we investigated how adults' and 8-year-old children's prior experience with sublexical regularities in their native language influences an empirical language learning task when the transitional probabilities (TPs) of the to-be-learned material align with those of the participants' native language (German).

Method. Over two experiments, we tested 40 German-speaking adults and 49 German-speaking 8-year-old children using a speech repetition task modelled on studies of Hebbian learning (see e.g., Page & Norris, 2009), in which participants repeated sequences of eight (adults) or six syllables (children). The study comprised two structured and one unstructured type of sequences. The two structured sequence types contained bisyllabic words (i.e., bigrams), which were either based on frequently occurring German syllable transitions discerned from corpora (*naturalistic sequences*) or were created from unattested syllable transitions (*non-naturalistic sequences*). The unstructured sequence type (*foils*) did not contain any learnable patterns and served as a baseline. Participants received all three sequence types in a pseudorandom order. Importantly, the recall task served as both exposure and test phase, allowing us to examine learning in real time. We predicted that both adults and children would have higher recall accuracy for naturalistic than non-naturalistic sequences, and that recall accuracy would exhibit faster learning rates for naturalistic sequences, with learning facilitated by participants' prior knowledge of German.

Results. Adults' and children's repetition accuracy was higher for naturalistic than non-naturalistic sequences for both syllables and bigrams (see Figure 1; *adults:* syllables: $\beta = -0.34$, $p < .001$; bigrams: $\beta = -0.20$, $p < .001$; *children:* syllables: $\beta = -0.16$, $p = .002$; bigrams: $\beta = -0.08$, $p = .003$), and higher for non-naturalistic than unstructured foil sequences, though this was only significant at the bigram level (*adults:* syllables: $\beta = -0.11$, $p = .07$; bigrams: $\beta = -0.08$, $p = .009$; *children:* syllables: $\beta = -0.07$, $p = .14$; bigrams: $\beta = -0.05$, $p = .02$). Recall accuracy improved throughout the study for adults (*adults:* syllables: $\beta = 0.10$, $p < .001$; bigrams: $\beta = 0.05$, $p < .001$), and this increase was larger for the naturalistic than non-naturalistic sequences (*adults:* syllables: $\beta = -0.04$, $p < .001$; bigrams: $\beta = -0.02$, $p < .001$). Moreover, adults' repetition of naturalistic compared to non-naturalistic sequences improved mostly during the early phases of the experiment (*adults:* syllables: $\beta = -0.18$, $p < .001$; bigrams: $\beta = -0.10$, $p < .001$). Finally, children's recall throughout the study interacted with their performance on a German sentence repetition task, with children demonstrating a higher language proficiency also showing an overall better repetition accuracy (*children:* syllables: $\beta = 0.50$, $p < .001$; bigrams: $\beta = 0.18$, $p < .001$) and a similar pattern of improvement throughout the study as adults (*children:* syllables: $\beta = 0.06$, $p = .05$; bigrams: $\beta = 0.01$, $p = .41$), especially during the early phases of the experiment (*children:* syllables: $\beta = 0.09$, $p = 0.006$; bigrams: $\beta = 0.04$, $p = .02$).

Conclusion. Thus, while adults and children demonstrated learning from both naturalistic and non-naturalistic stimuli, learning was superior for the naturalistic sequences. This indicates that participants drew on their existing distributional knowledge of German to extract naturalistic words faster and more accurately than non-naturalistic words, supporting memory-based theories of SL, where input is chunked and stored in long-term memory (see e.g., Batterink & Paller, 2017; Christiansen, 2019).

Figure 1: Adults' (left) and children's (right) mean recall of syllables (top) and bigrams (bottom) per sequence for the three sequence types, given by experimental block. Error bars indicate the standard error.



References

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468–481. https://doi.org/10.1111/tops.12332

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, *19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, *1*(6), 906–914. https://doi.org/10.1002/wcs.78

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011