Rule Learning in Humans and Animals

Raquel G. Alhama, Remko Scha, Willem Zuidema
Institute of Logic, Language and Computation, University of Amsterdam

Introduction: a separate rule learning mechanism?

Peña et al. [2] present human adults with sequences of syllables composed of triplets of the form **AxC**, where A and C consistently co-occur and X varies.

- stream: pulikiberagatafodupurakibefogatalidu...
- partwords: (CxA, xAC): kibera, ragata, ...
- rulewords: (AyC): pumoki, besuga, tanedu

Interesting results:

- Micropause Effect: with only two minutes of exposure, if adding subliminal pauses between items, participants choose rulewords over partwords.
- Time Effect: while showing no preference between rulewords and partwords after 10 minutes of exposure, participants prefer partwords after 30 minutes.

Toro and Trobalón [3] performed similar experiments with rats, and report qualitatively different results. Although the rats learn to discriminate between stimuli on the basis of co-occurrence frequencies, they don't display any form of rule learning.

The time and the micropause effect are often taken as evidence of two learning mechanisms (e.g. see [1]):

- a statistical mechanism that tracks transitional probabilities
- a rule mechanism for structure detection (perhaps human specific)

Do we really need both?

R^2: the Retention-Recognition Model

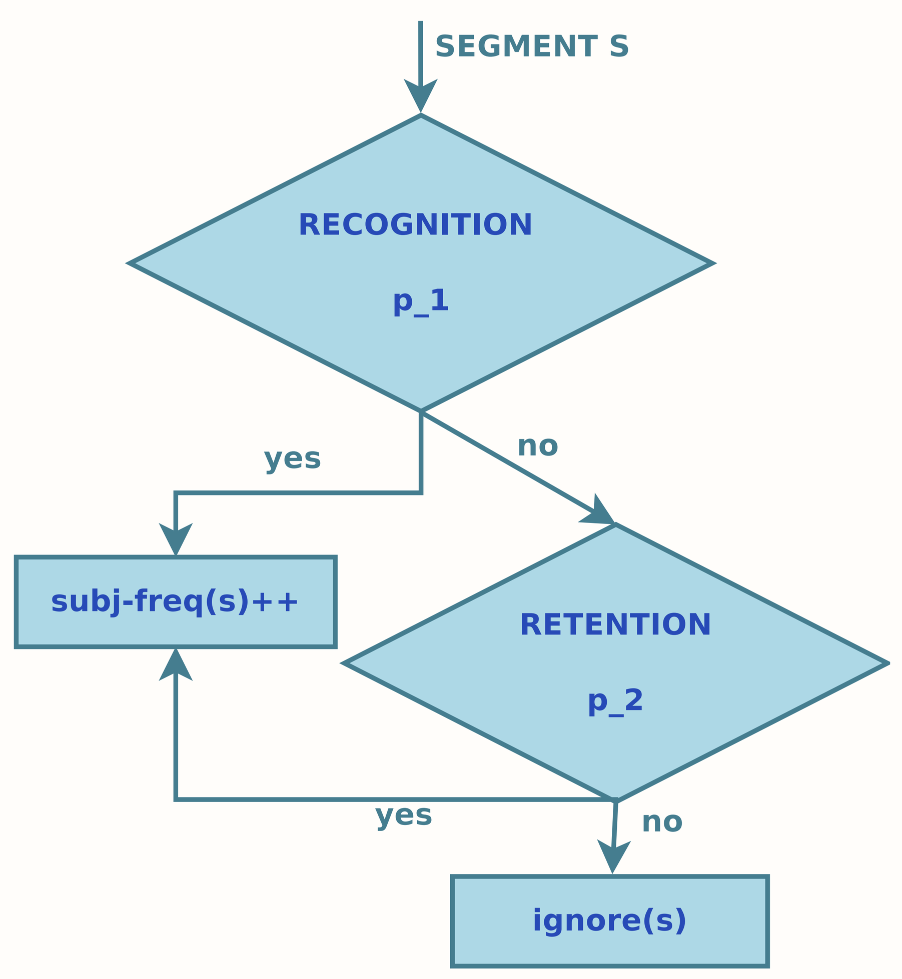

Figure 1: R^2:The Retention-Recognition Model

$$p_1(s) = (1 - B^{\text{activation}(s)}) * D^{\#\text{types}} \quad (1)$$

$$p_2(s) = A^{\text{length}(s)} * C^{\pi} \quad (2)$$

$$0 \leqslant A, B, C, D \leqslant 1$$

$$\pi = \begin{cases} 0 \text{ after pause} \\ 1 \text{ otherwise} \end{cases}$$

We model the probability that a subject memorizes a particular subsequence or increases its subjective count. The model involves free parameters that determine memory constraints which may be fitted to empirical data.

Results I: model predicts observed skew in response distribution in rats

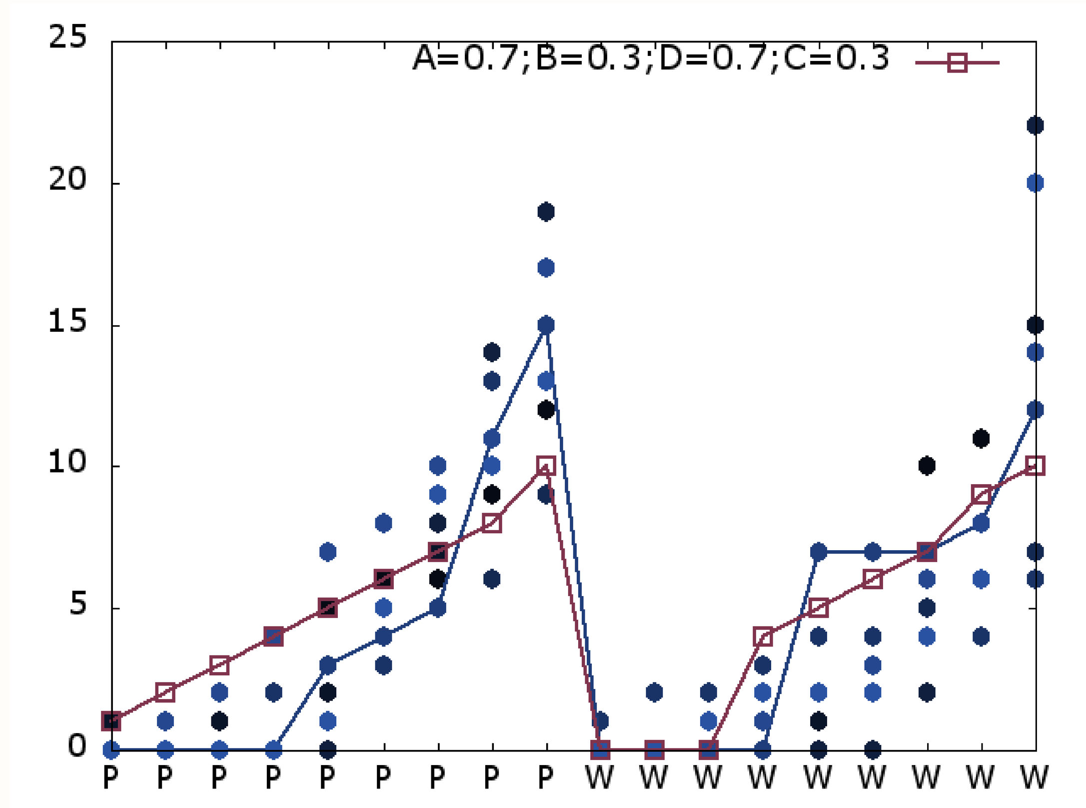

Figure 2: Responses of rats (blue) and subjective frequencies of the model (pink), without pauses. W denotes words; P partwords; both ordered by response frequency.

The model also predicts boost of words over partwords due to micropauses (but incorrectly predicts decreased skew).

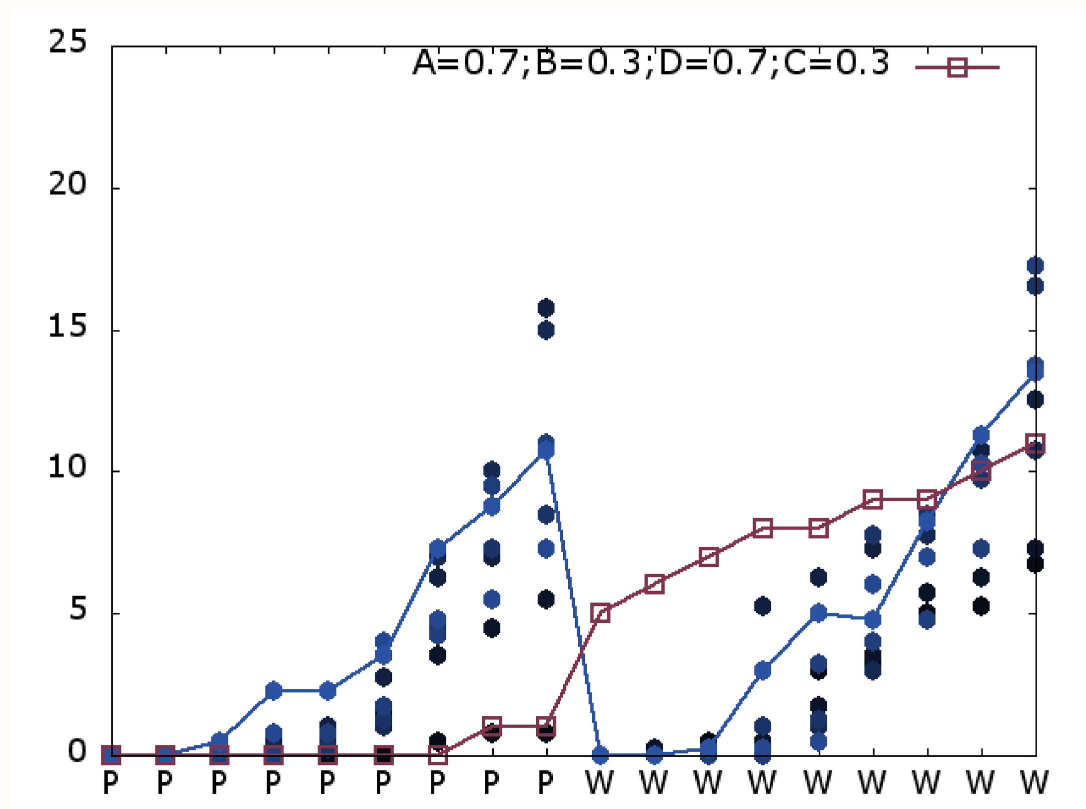

Figure 3: Responses of rats (blue) and subjective frequencies of the model (pink), with pauses. W denotes words; P partwords; both ordered by response frequency.

How do we model the willingness to generalize? Good-Turing

The time effect implies that the subjects' willingness to generalize decreases with the number of exposures. But how do we model generalization? With Good-Turing Smoothing!

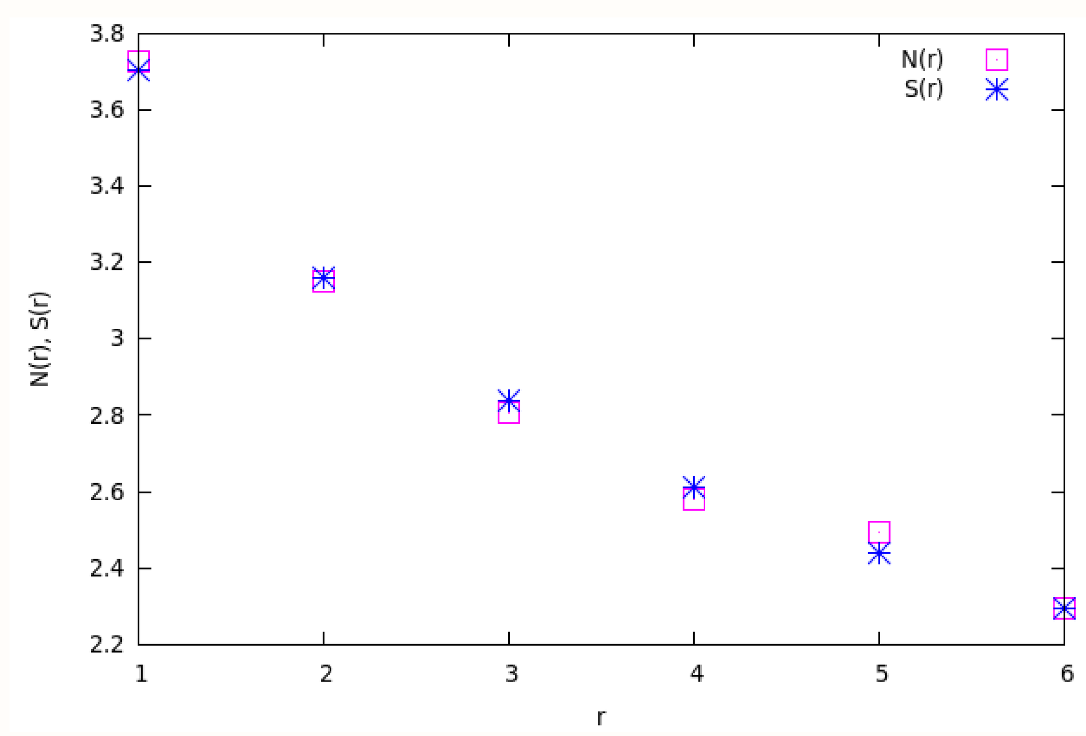

Figure 4: Example of Simple Good Turing.

- Probability of unseen objects: the Missing Mass [4]
- Missing mass: $P_0 = E(N_1)/N$
- Reestimate probabilities:

$$P_r = r'/N$$

- Reestimate frequencies:

$$r' = (r+1)\frac{E(N_{r+1})}{E(N_r)}$$

Results II: model predicts observed decrease in willingness to generalize in humans

Evidence of Time Effect: the missing mass decreases as the number of exposures increases.

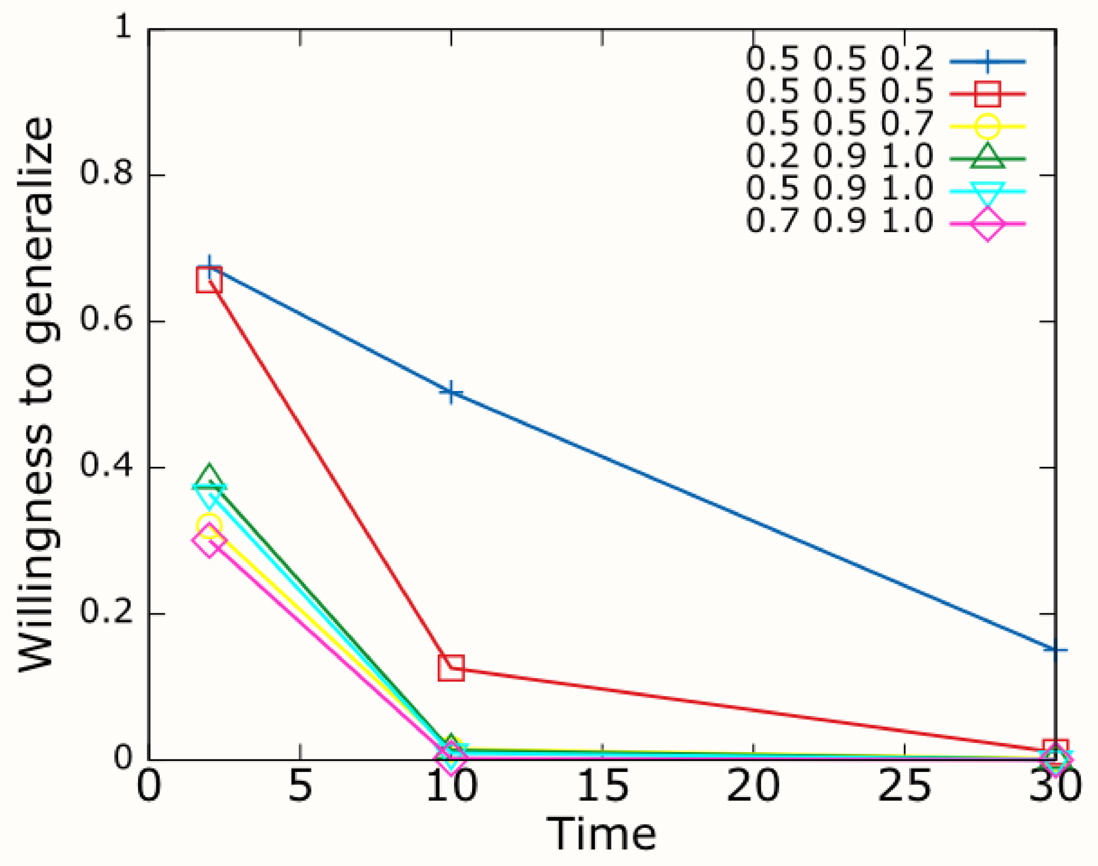

Figure 5: Missing mass for several parameters of the model.

The difference between words and partwords decreases as the number of exposures increases.

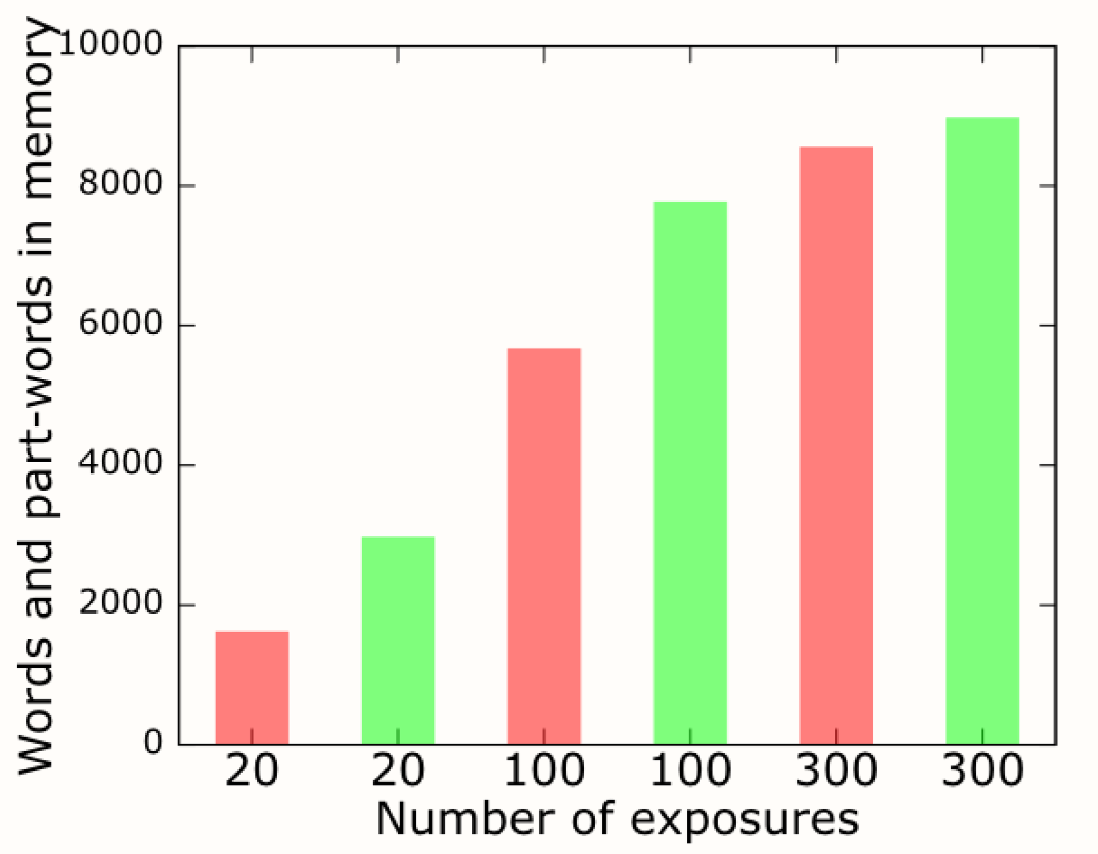

Figure 6: Subjective frequencies of words (green) and partwords (red) in 10K runs of a model with an arbitrary parameter setting (A=B=D=0.5; C=1.0.).

Conclusions

We present a single statistical model that qualitatively accounts for the experimental findings with humans and rats. This suggests that a single statistical mechanism suffices for learning non-adjacent dependencies and hence provides some insight on the mechanisms underlying language acquisition.

References

[1] Endress, A., Bonatti, L. Rapid learning of syllable classes from a perceptually continuous speech stream, *Cognition*, Elsevier, 2006, 105, 247-299

[2] Pena, M.; Bonatti, L.; Nespor, M. Mehler, J. Signal-driven computations in speech processing Science, *American Association for the Advancement of Science*, 2002, 298, 604-607

[3] Toro, J. M., Trobalón, J. B. Statistical computations over a speech stream in a rodent *Perception & psychophysics*, Springer, 2005, 67, 867-875

[4] Gale, W., Sampson, G. Good-turing frequency estimation without tears *Journal of Quantitative Linguistics*, Taylor & Francis, 1995, 2, 217-237