Limits and Variations of Linguistic Generalizations

An Artificial Grammar Study with Adults

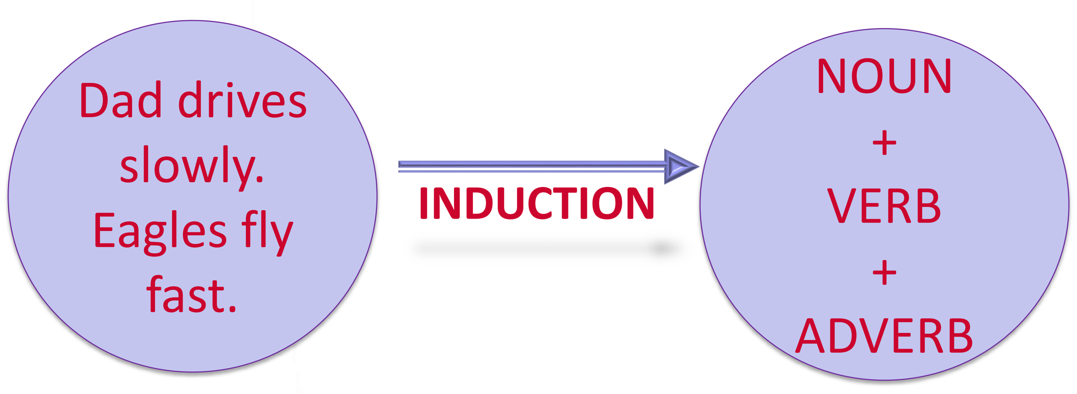
Silvia RĂDULESCU, Frank WIJNEN, Sergey AVRUTIN

Utrecht University, Utrecht Institute of Linguistics (UiL-OTS)

contact: sil.radulescu@gmail.com

From Little Evidence to General Rules

…in child language acquisition



Previous research and accounts

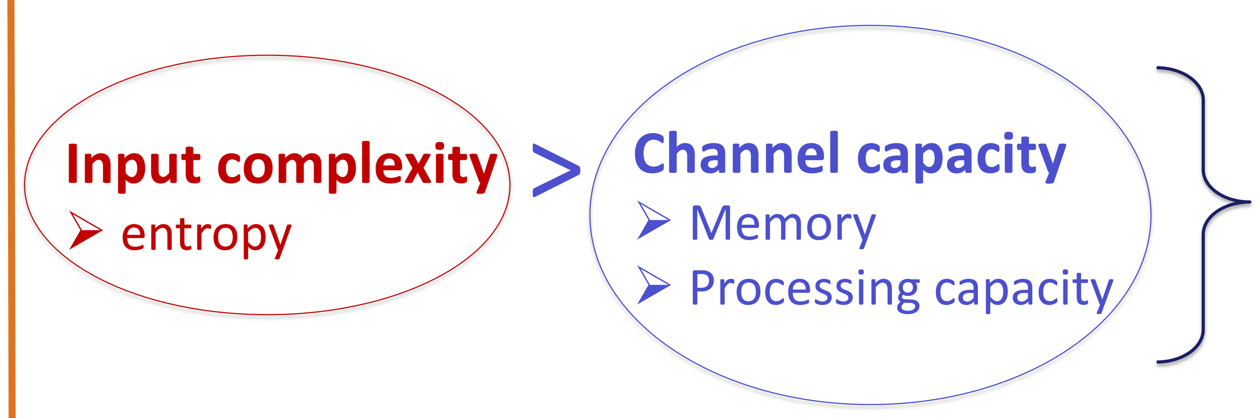
(1) statistical learning → transitional probabilities
e.g. phonotactic regularities (Chambers et al, 2003), word boundaries (Saffran et al, 1996)

- blind to novel items

(2) algebra-like system → algebraic rules that apply to categories (Marcus et al, 1999)
e.g. first item is the same as third item *(li_na_li)*

- but what enables tuning into such rules and what input factors (if any) facilitate or impede this process?

New Entropy Model



Less complexity → perceptually-bound learning (*ba* follows *ba*)

Higher complexity → category-based abstractions (Noun-Verb-Noun)

Experiment

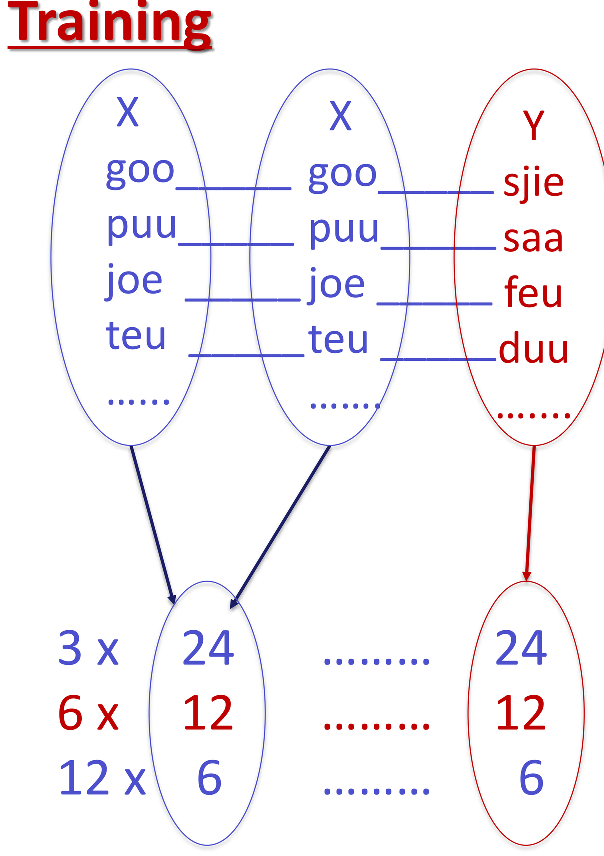
- 35 participants (age 19-26)
- 3-syllable XXY strings (each letter represents a set of syllables)
- three conditions (72 strings each, ~ 4 min)

Entropy → a function of the number of different items in the input and their probability of occurrence (frequency)
→ a measure of input complexity (bits)

$$H(X) = -\sum_{i=1}^{n} p(x_i) log p(x_i)$$

(Shannon, 1948)

Training



High Entropy (H = 4.58 bits)
Medium Entropy (H = 4 bits)
Low Entropy (H = 3.5 bits)

Test

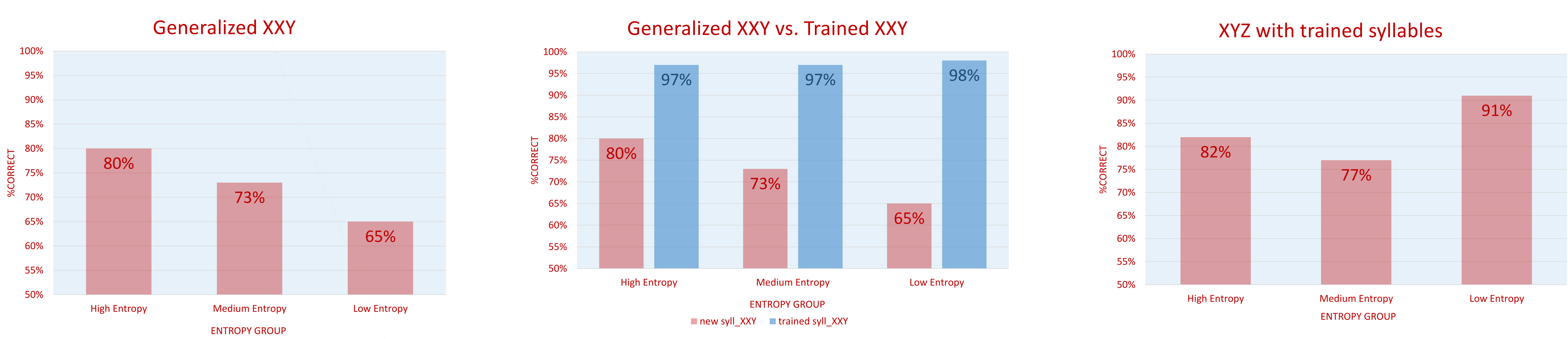
("Could this string be possible in the language that you heard?") – 20 strings

- XXY_trained_syllables: goo_goo_sjie √
- XYZ_new_syllables: reu_loo_gee *
- XXY_new_syllables: too_too_suu √
- XYZ_trained_syllables: teu_duu_saa *

(XYZ: strings of three different syllables)

Results

– Generalized Linear Mixed Model: Test String Type x Entropy Group interaction - $F(9, 679) = 6.428$, $p = .000$)



Generalization to novel XXY strings

- the tendency to abstract away from the memorized input increases as the *input complexity* increases

Difference: XXY_trained_syll vs. XXY_new_syll

Δ[HiEn] < Δ[MedEn] < Δ[LowEn]

- learners in the HiEn condition had the highest tendency to fully generalize to novel XXY strings

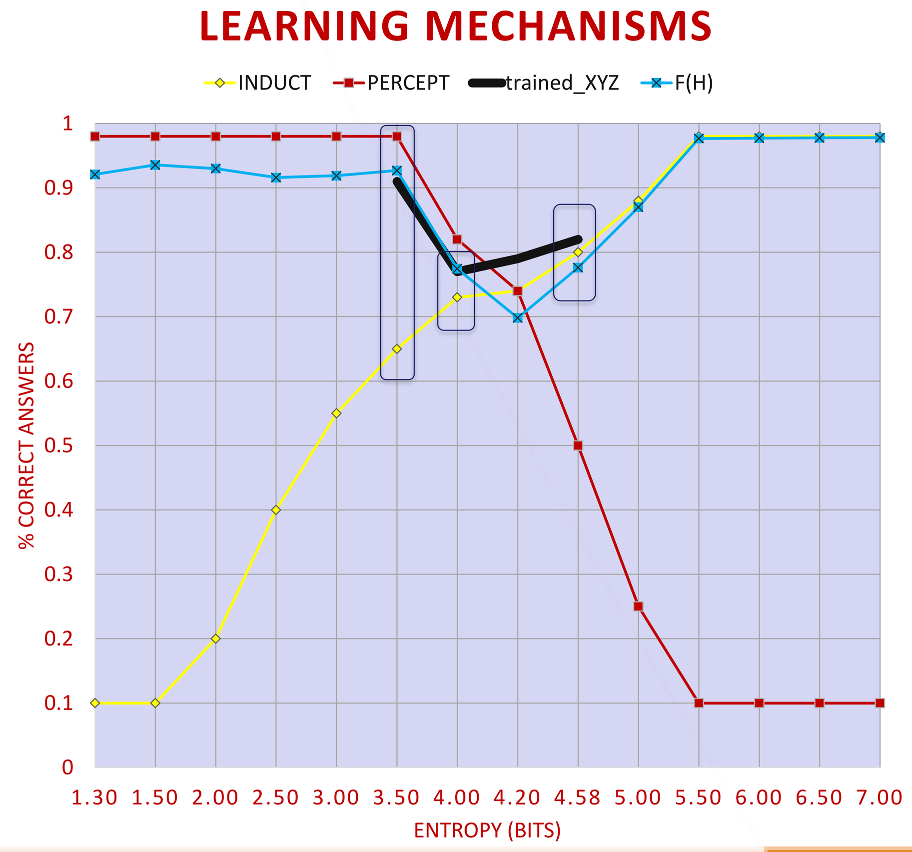
XYZ with trained syllables

- perceptually-& category-based generalization work against each other: memory trace of individual syllables (but not their sequence) prevents rejection

Discussion

A possible model for the interplay between the perceptually-bound and category-based abstractions

Does this model apply to other cognitive processes?



Conclusion

The human brain seems to be sensitive to the amount of information in the environment. A complex linguistic environment triggers the inductive leap from memorizing specific items to extracting generalized rules.

References

Chambers, K., Onishi, K., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*, B69-B77.

Gómez, R.L. & Gerken, L.A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences 4*, 178–186

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*, 77–80.

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

Shannon, C. E. (1948). *A mathematical theory of communication*. Bell System Technical Journal, 27, 379–423.