Set Size Does Not Matter. Entropy Drives Rule Induction in Non-Adjacent Dependency Learning

Ileana GRAMA[a,b], Silvia RADULESCU[b], Frank WIJNEN[b] & Sergey AVRUTIN[b]

[a]University of Amsterdam, Amsterdam Centre for Language & Cognition [b]Utrecht University, Utrecht Institute of Linguistics OTS

From little evidence to abstract rules in language acquisition

(1) statistical learning (Aslin & Newport, 2012)
(2) algebra-like system (Marcus et al, 1999)

Entropy Model

Entropy (input complexity) > Channel capacity (encoding power = entropy/time)

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

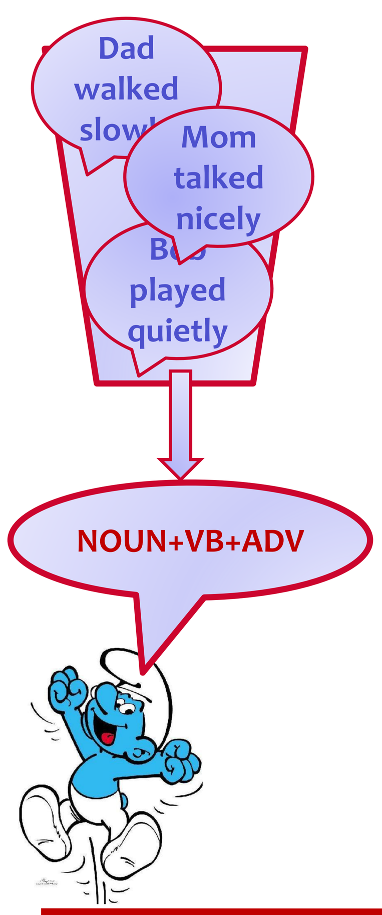
$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$$

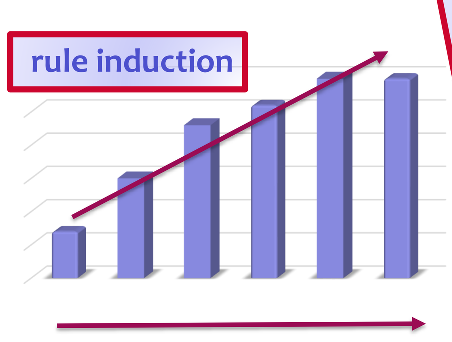
(Shannon, 1948)

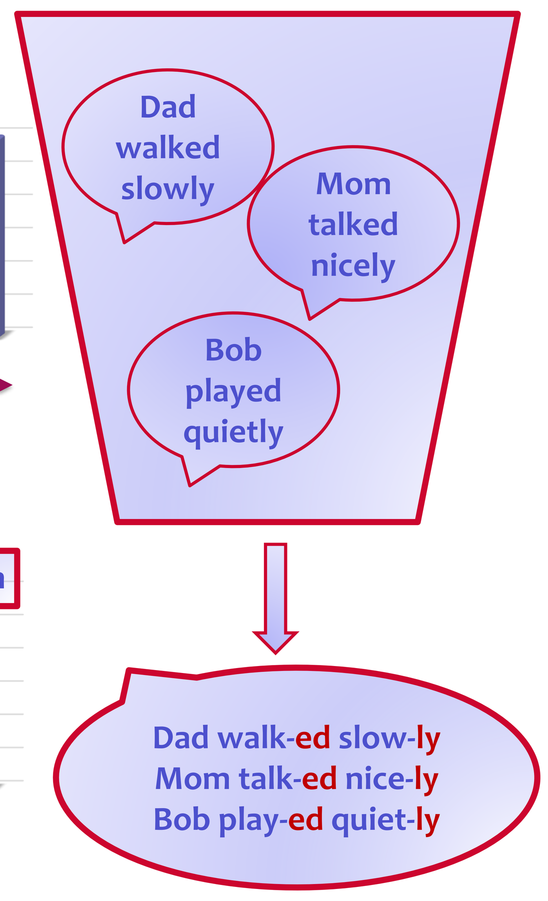
Rule Induction → interaction of *input complexity* (entropy) and *channel capacity*

Entropy > channel capacity → category-based generalizations

Entropy < channel capacity → item-bound generalizations







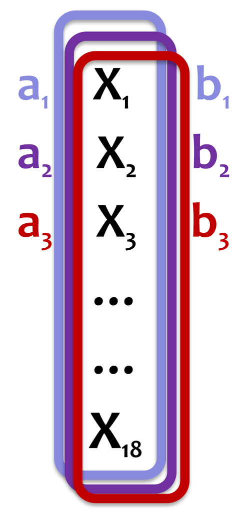
Predictions of Entropy Model for Non-Adjacent Dependency Learning

- Non-Adjacent Dependency Learning (*a X b*): captures the dynamics between item-bound generalizations (a_i predicts b_i) and category-based generalizations ($a_i\_b_i$ generalized over *X*)
- Low entropy drives item-bound generalizations (learning of $a_i\_b_i$ frames and their specific distributional patterns – by memorizing specific items and their combinations)
- High entropy drives category-based generalizations (generalizing over the intervening X category)

Experiment - Effect of Entropy on Rule Induction In Non-Adjacent Dependency Learning

High Entropy

$H_{total} = (H_{[bigram]} + H_{[trigram]})/2 = 4.7$ bits

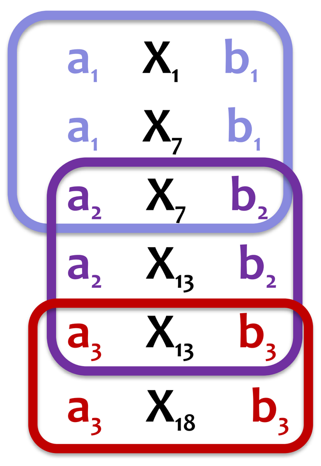


3 a_b / 18 Xs / 3a_b *18Xs - 6 reps

20 participants

Medium Entropy

$H_{total} = 4.27$ bits

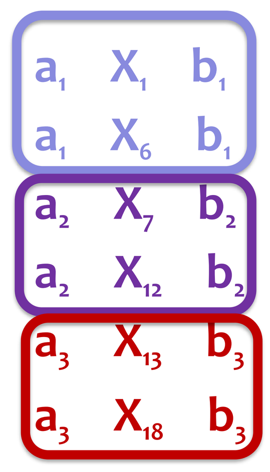


3 a_b / 18 Xs / 3a_b*12Xs - 9 reps

27 participants

Low Entropy

$H_{total} = 3.52$ bits

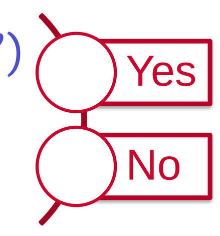


3 a_b / 18 Xs / 3a_b*6Xs - 18 reps

29 participants

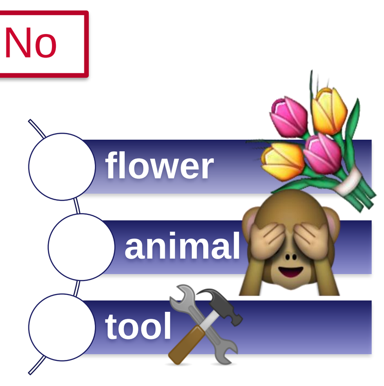
Test 1: NAD-Learning ("Is this string possible in the language that you heard?")

- Consistent: $a_1\ Y_{1,2}\ b_1$, $a_2\ Y_{1,2}\ b_2$, $a_3\ Y_{1,2}\ b_3$
- Inconsistent: $a_1\ Y_{1,2}\ b_2$, $a_2\ Y_{1,2}\ b_3$, $a_3\ Y_{1,2}\ b_1$



Test 2: Incidental Memorization ("Flower, animal or tool")

- Familiarize participants with 30 pseudo-words they must classify
- *Surprise test*: 'Have you heard this word'? 15 targets + 15 foils
- [Memorization ~ Channel Capacity]



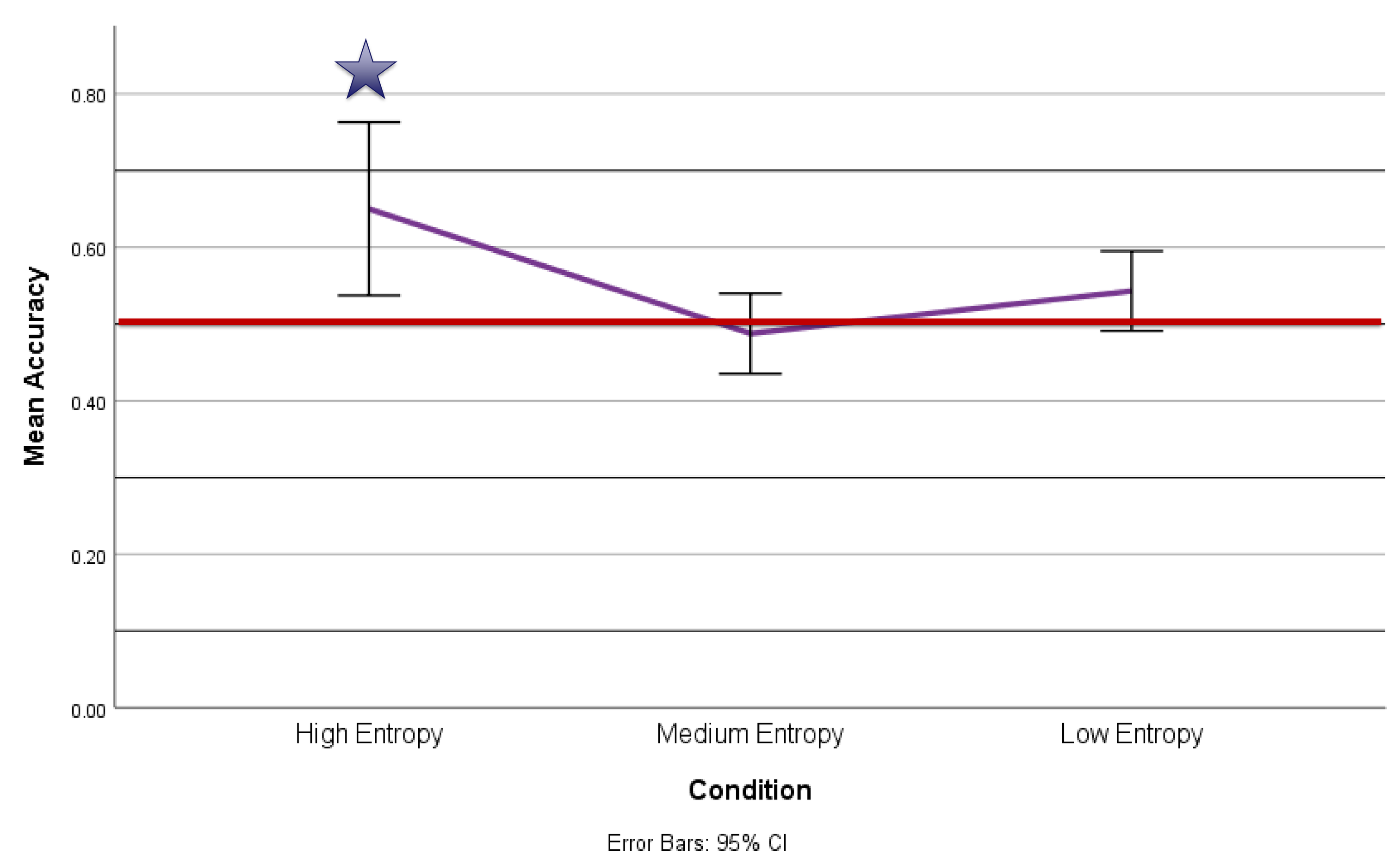
Test 3: Word Recall (Only Low Entropy) ("Did you hear this word in the first phase?")

- 28 items: 12 target a, X and b words and 16 foils
- [Does generalization of a_b rely on encoding a/b but not X?]

Results

- **Test 1:** Generalized Linear Mixed Model (Accurate/Inaccurate response to each test item) - > Significant effect of **Condition** ($F(2, 909) = 5.441$, $p = .004$)
- Post-hoc Comparisons: **High Entropy** > **Medium** ($p = .001$), **High** > **Low** ($p = .024$); **Medium** = **Low** ($p = .238$)
- **High Entropy Condition** – significantly above chance learning ($p = .019$)
- **Test 2:** No correlation with Incidental Memorization (item-specific encoding)
- **Test 3** (Low Entropy): No correlation between Accuracy in Dependency-Learning and Accuracy in Word Recall (a/b, or X, or both)*

*Using d' instead of Accuracy as a dependent variable, a Linear Mixed Model found a/b Recall to be a significant positive predictor of Dependency-Learning ($p < .05$) and X Recall a near-significant negative predictor ($p = .095$)



Discussion

- High **Entropy** promotes better **generalization** of Non-Adjacent Dependencies even when X set size is kept constant
- Entropy does not linearly predict performance:
  - Is there a **threshold** of entropy that shifts the balance from item-specific learning to generalization?
  - Or a **U-shape** whereby Low Entropy facilitates the (item-specific) encoding of *a_b* frames?
- No correlation with item-specific encoding
- Future research: **How can we measure the effect of Channel Capacity?**

Conclusion

It is Entropy, and not mere set size, that drives Non-Adjacent Dependency-Learning!

Selected References

Aslin, R.N., and Newport, E. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, 21, 170–176.

Gomez, R. L. (2002). Variability and detection of invariant structure. Psychological Science, 13(5), 431-436.

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

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

This work was supported by the Netherlands Organization for Scientific Research (NWO), through grant number 322-75-009/1116 awarded to Silvia Radulescu.