

# RETRODICTION AS DELAYED RECURRENCE: THE CASE OF ADJECTIVES IN ITALIAN AND ENGLISH

Raquel G. Alhama<sup>1,2</sup> Francesca Zermiani<sup>2,3</sup> Atiqah Khaliq<sup>2</sup>

<sup>1</sup>Tilburg University, Department of Cognitive Science & Artificial Intelligence

<sup>2</sup>Max Planck Institute for Psycholinguistics, Language Development Department

<sup>3</sup>University of Stuttgart, Department of Teaching and Learning with Intelligent Systems

## Introduction

- Sensitivity to statistical regularities allows for efficient lexical processing.
- Children with more efficient lexical processing are also those with faster vocabulary growth [2, 3].
- Typological variation on the statistical regularities of different languages should be either equally tracked during processing, or reflected in cross-linguistic differences in learning.
- We focus on modeling one such typological feature: in particular, word order of descriptive adjectives in English (which occur pre-nominally), and Italian (which appear mostly post-nominally, but also pre-nominally).

## Corpus Analyses

- We extracted child-directed speech transcriptions from all the English and Italian corpora in CHILDES [5] (age: 0-60 m.o.)
- For each adjective-noun or noun-adjective pair  $w_1w_2$  we computed its conditional probability as  $P(w_i|w_j) = \text{counts}(w_1w_2)/\text{counts}(ctx)$ , where  $i = 2, j = 1, ctx = w_1$  for forward conditional probabilities and  $i = 1, j = 2, ctx = w_2$  for backward conditional probabilities.

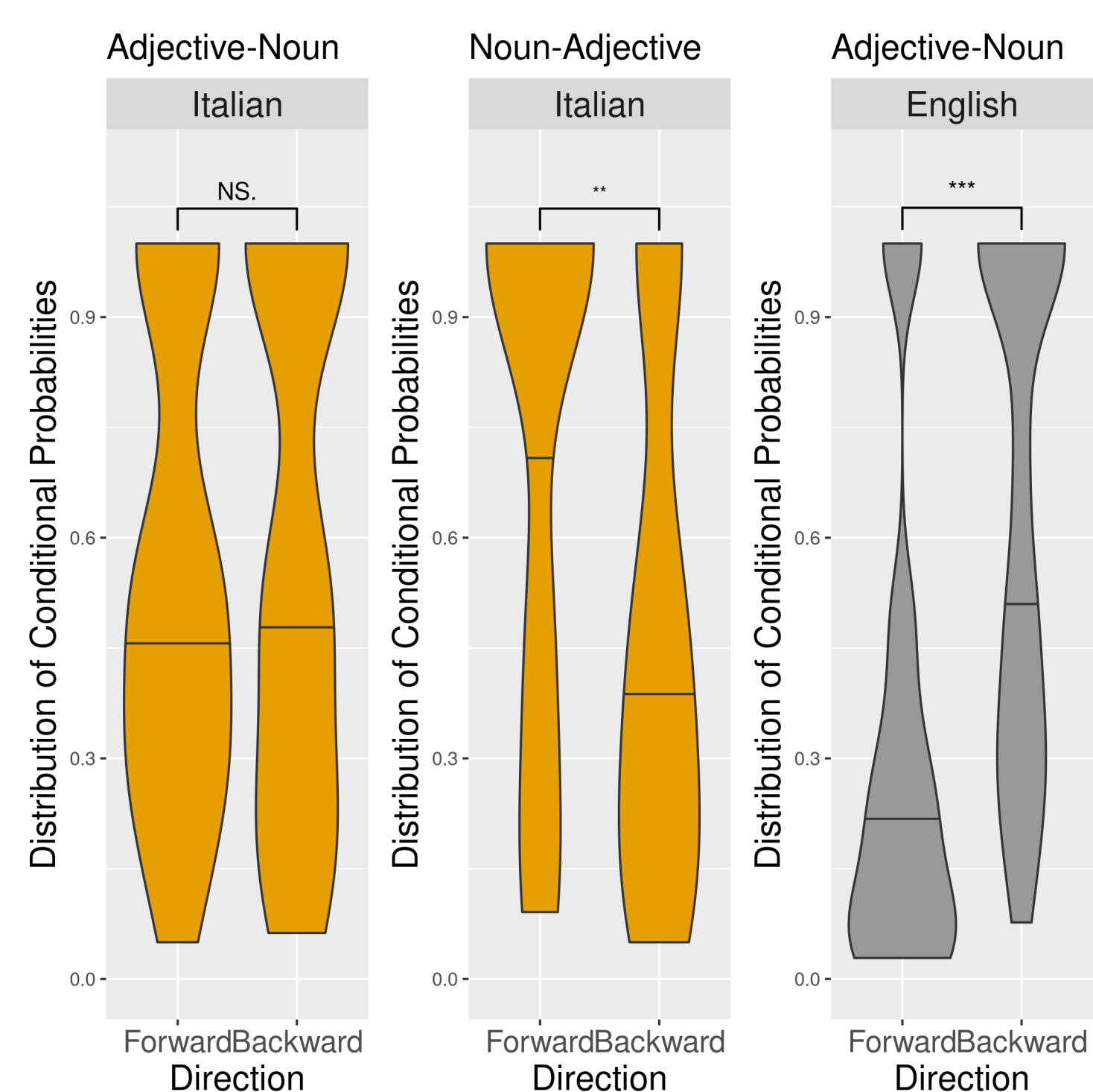


Fig. 1: Distribution of probabilities.

- Forward conditional probabilities are significantly more reliable for adjectives occurring in the Italian canonical noun-adjective ordering ( $p < 0.01$ ). The opposite is the case for English, in which predicting backwards is significantly more reliable ( $p < 0.001$ ).
- We thus confirmed that the adjective order in English and Italian was reflected in the conditional probabilities between adjectives and nouns.

## The Acquisition of Adjectives

- Efficient processing correlates with faster vocabulary growth. Thus, if there is a difference in processing forward and backward dependencies, it should be reflected in a cross-linguistic difference in vocabulary acquisition.
- We analyzed data collected with the MacArthur-Bates Communicative Development Inventory forms (CDIs; Wordbank [4]) to estimate the age of acquisition (AoA) of adjectives and nouns.

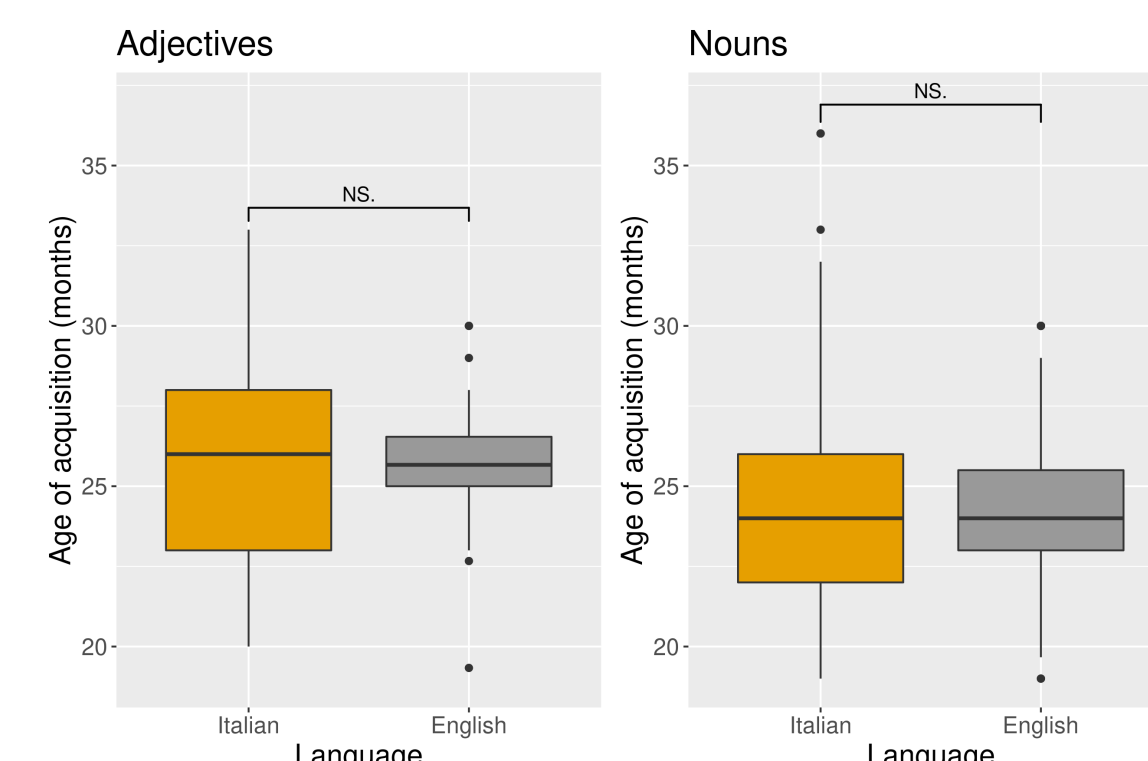


Fig. 2: Age of Acquisition (AoA).

- There is no significant difference between the AoA of adjectives and nouns in each language. Thus children learning Italian must be employing their forward predictive skills, while children learning English need to draw upon their capacity to retrodict.

## Do RNNs Retrodict?

- We trained an RNN [1] on the child-directed data.

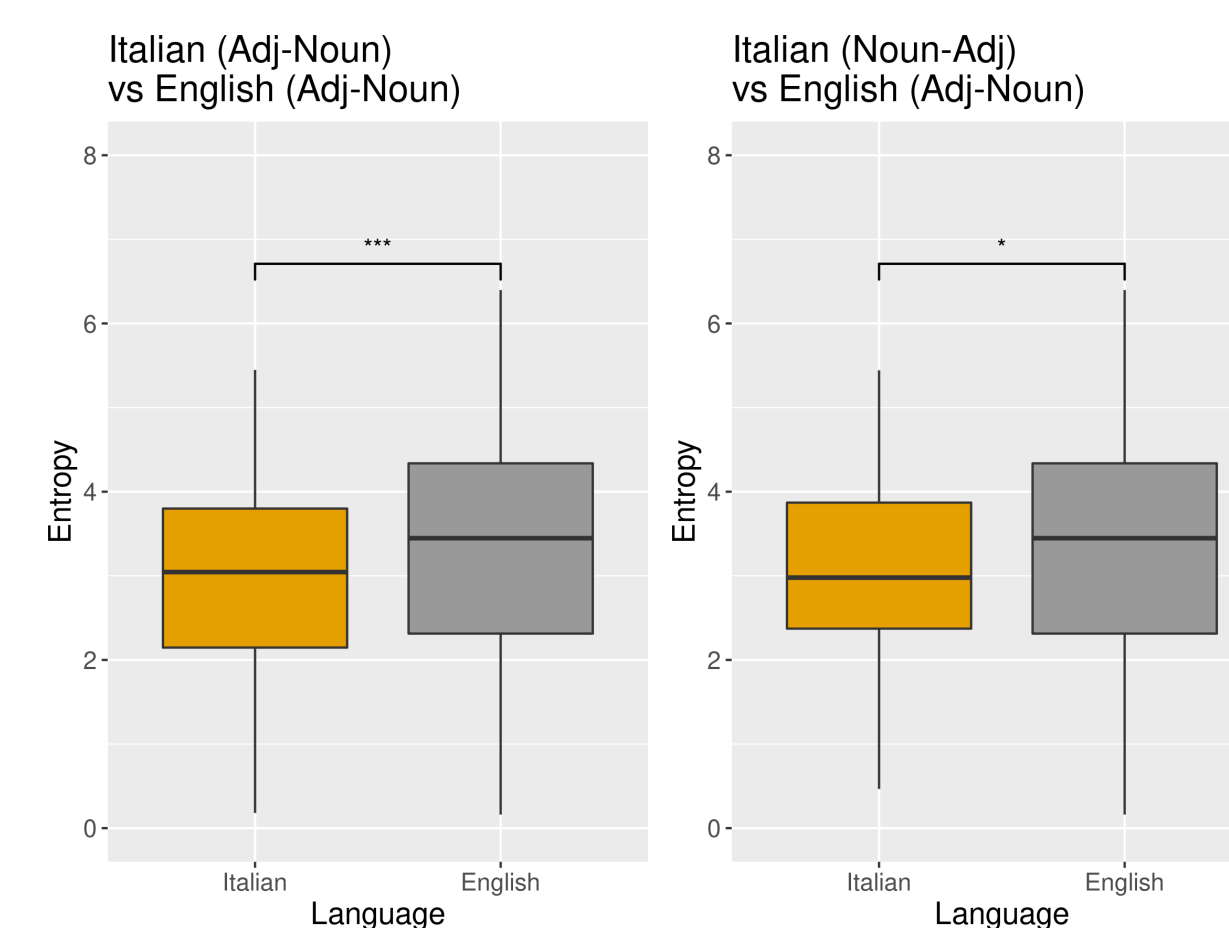


Fig. 3: Entropy of the RNN, after the first word in each adjective-noun or noun-adjective pair.

- The RNN is significantly less successful in learning English than Italian, as it cannot track backward dependencies (which are more prevalent in English than in Italian).

## Retrodiction as Delayed Prediction

- A strictly forward model like the standard RNN cannot account for learning backward dependencies.
- We explore an alternative account of retrodiction that functions as delayed prediction, based on the model presented in Turek et al. [6], known as Delayed Recurrent Neural Network (dRNN).
- In the dRNN, the weight update is performed at time  $t + d$ , where  $d$  is the pre-defined 'delay'. This entails that  $d$  extra words have been processed by the network before the error is backpropagated. We set a delay of one word.

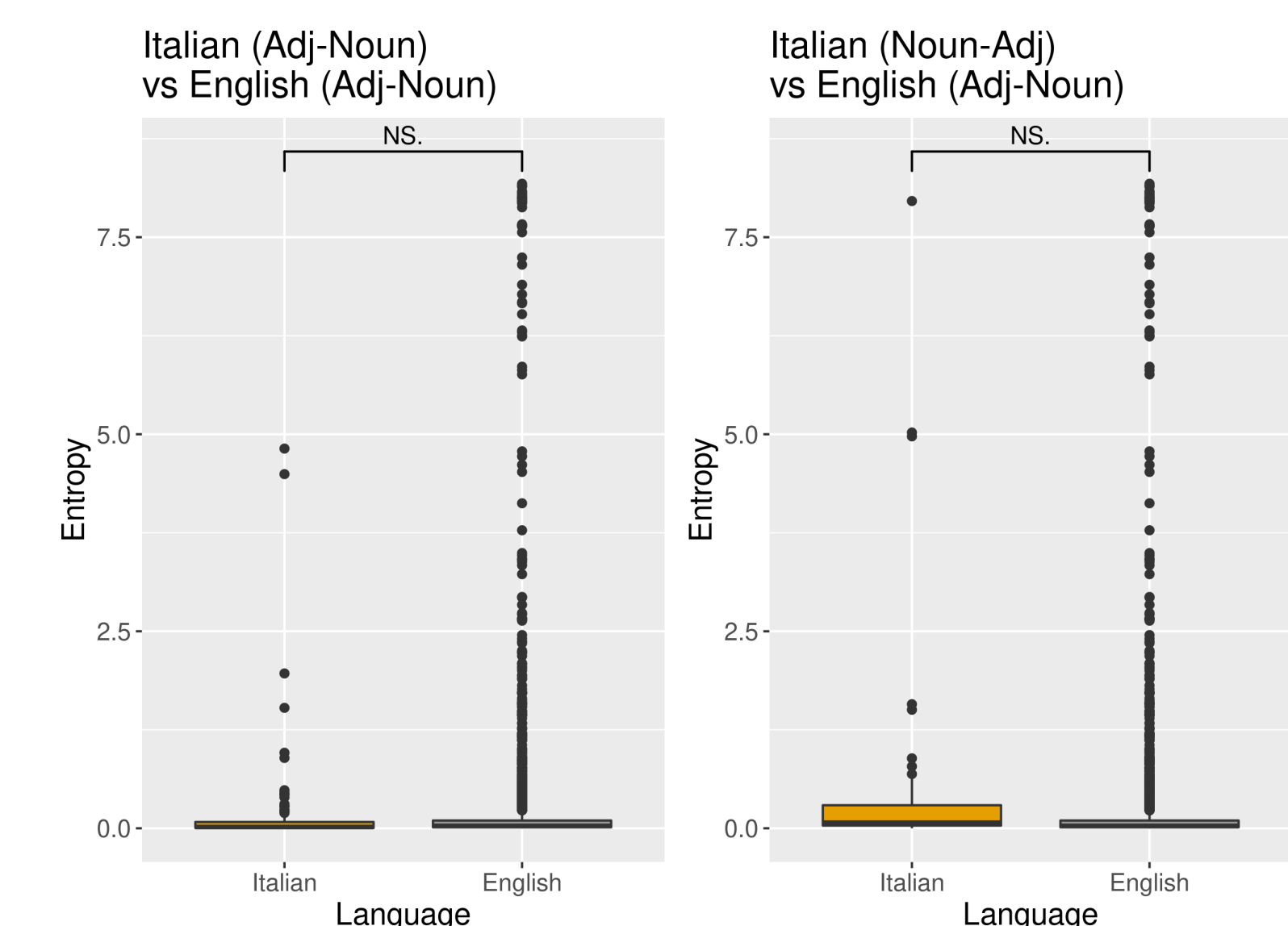


Fig. 4: Entropy of the dRNN, after the first word in each adjective-noun or noun-adjective pair.

- There are no significant differences between these languages, suggesting that this model can account for learning adjective constructions in both.

## Conclusions

Our work suggests that a full account of human processing and learning needs to address typological influences on distributional information, which require tracking of both forward and backward statistical dependencies. While we cannot account for these with standard RNN models, the dRNN can capture both forward and backward dependencies, offering a possible explanation for how humans are able to predict but also retrodict.

## References

- [1] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.
- [2] Anne Fernald and Virginia A Marchman. Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child development*, 83(1):203–222, 2012.
- [3] Anne Fernald, Amy Perfors, and Virginia A Marchman. Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, 42(1):98, 2006.
- [4] Michael C Frank, Mika Braginsky, Daniel Yurovsky, and Virginia A Marchman. Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, 44(3):677–694, 2017.
- [5] Brian MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Mahwah, NJ, 3 edition, 2000.
- [6] Javier S. Turek, Shailee Jain, Vy Vo, Mihai Capota, Alexander G. Huth, and Theodore L. Willke. Approximating stacked and bidirectional recurrent architectures with the delayed recurrent neural network. 2019.