



Mutual Information Impacts Adjective Ordering Across Languages

Michael Hahn (Stanford), Richard Futrell (MIT), Judith Degen (Stanford), Dan Jurafsky (Stanford), Noah Goodman (Stanford)
mhahn2@stanford.edu



Adjective Ordering

Across languages, adjectives show preferences for some orderings over others.

“the large wooden table” > “the wooden large table”

“the beautiful old house” > “the old beautiful table”

Generalizations include adjectives’

- **Inherentness** (Whorf 1945) and **Specificity** to the noun (Sweet 1898, Ziff 1960)
- **Absoluteness** (Sproat & Shih 1991) and **Subjectivity** (Hetzron 1978)

but there is no consensus on **explanation**.

Mutual Information

If humans predict words based on limited context representations due to memory limitations, words that have **high mutual information (MI)** should appear **closer together** to minimize prediction error (Qian & Jaeger, 2013; Futrell & Levy 2017).

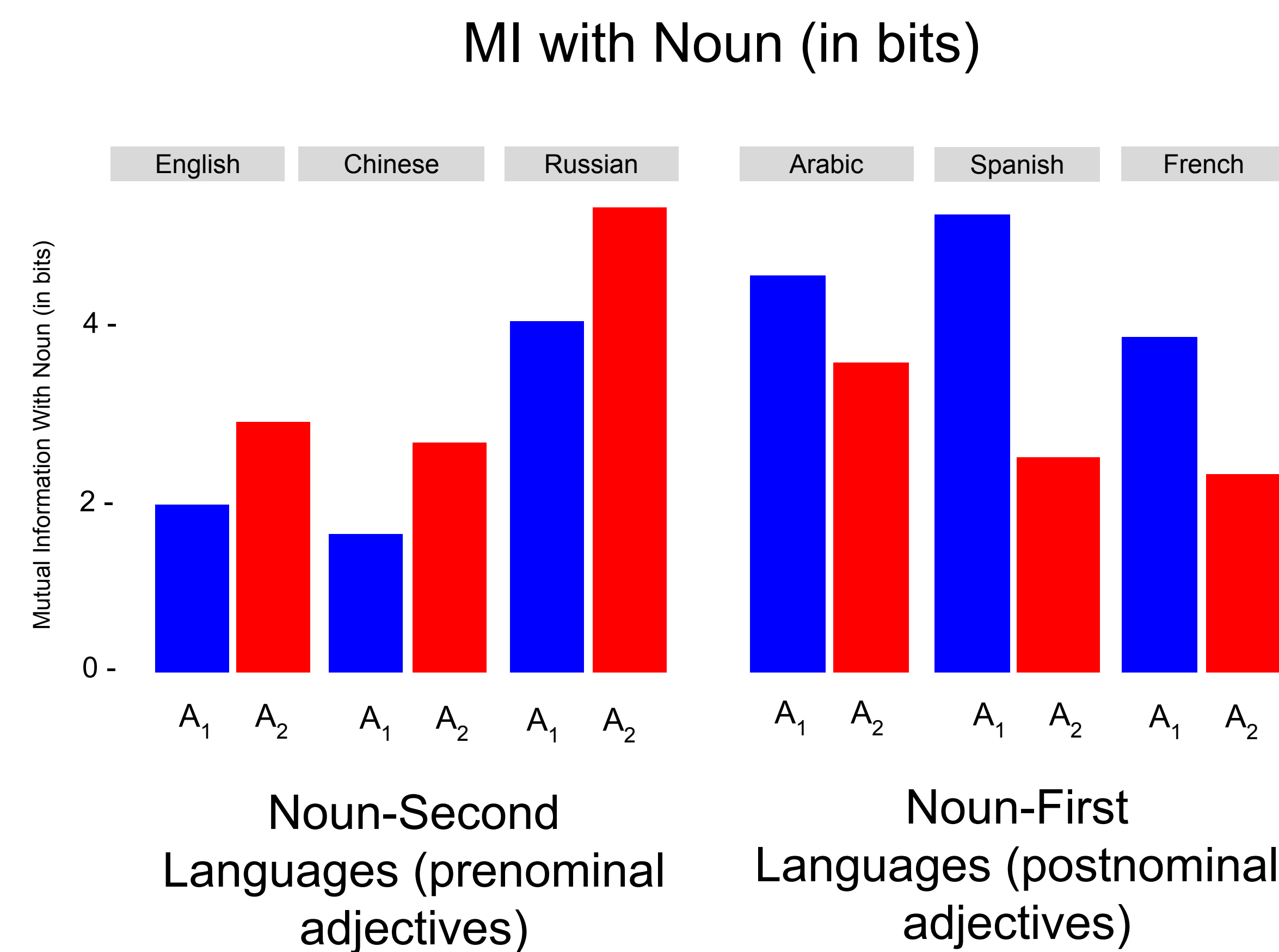
$$MI(\text{Adj}, \text{Noun}) = \log P(\text{Noun}|\text{Adj}) - \log P(\text{Noun})$$

Measures **collocation**, plausible measure of **specificity**.

Prediction: Adjectives with higher MI with noun are closer to it

Experiment I

Hypothesis: Higher MI ~ Closer to Noun



Across languages, adjectives with **higher MI** with noun occur **closer** to noun.

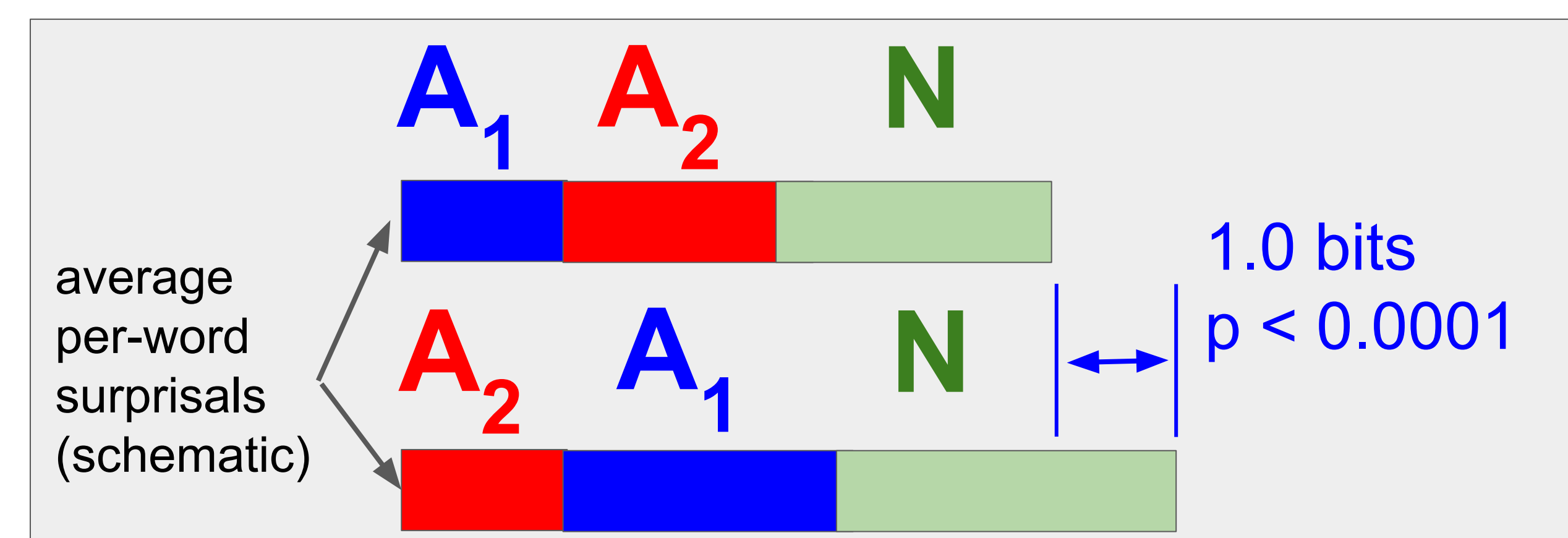
Experiment II

Question: Do differences in A_1, A_2 Mutual Information with the noun affect **ease of incremental prediction** for powerful statistical models of language?

Prediction: Lower aggregate surprisal on real ordering than reverse orders

Use **recurrent neural network** models. Order preferences were eliminated from training data.

We compare **surprisal** on real and reverse orderings:



For statistical model of language, real orderings **maximize ease of incremental prediction**.

Experiment III

Motivation: Scontras et al. (2017) show: The more **subjective** an adjective, the **farther** away from the noun it occurs.

Question: Do MI and subjectivity contribute to ordering preferences independently?

Model predicting whether A_1A_2N is from corpus or inverted

	β (SD)	News, RNN	News, Bigrams	Novels, RNN	Novels, Bigrams
MI A_1 with Context	0.12 (0.01)***	0.20 (0.04)***	0.16 (0.03)***	0.28 (0.05)***	
MI A_2 with Noun	0.08 (0.01)***	0.36 (0.03)***	0.08 (0.02)**	0.29 (0.03)***	
Subjectivity A_1	3.14 (1.15)**	7.57 (1.62)***	7.93 (1.42)***	8.40 (1.40)***	
Subjectivity A_2	-3.44 (1.20)**	-7.35 (1.58)***	-8.26 (1.36)***	-8.55 (1.34)***	
BIC	-6	15	26	30	
Deviance	14.6	34	43	49	
Pr(>Chisq)	0.0006879	5.01e-8	3.29e-10	2.84e-11	

Ratings for ‘How subjective is the adjective ...?’ from Scontras et al. (2017)

Model Comparison with MI-only model: **MI+Subjective is better than MI alone**

Corpora: English Gigaword (~120M sentences) and BookCorpus (~70M sentences)

At least in English, **mutual information** and **subjectivity** independently impact ordering.

Conclusion: Across languages, mutual information with noun impacts adjective ordering, as expected under assumptions of memory limitations and pressure to minimize prediction error.

References: Futrell, Richard, and Roger Levy. “Noisy-context surprisal as a human sentence processing cost model.” *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. 2017. **Gildea, D. and Jaeger, T. F.** (2015). Human languages order information efficiently. *arXiv, abs/1510.02823*. **Qian, T. and Jaeger, T. F.** (2012). Cue effectiveness in communicatively efficient discourse production. *Cognitive Science*, 36:1312–1336. **Scontras, G., Degen, J., & Goodman, N. D.** (2017). Subjectivity Predicts Adjective Ordering Preferences. *Open Mind*, 1(1), 53–66. **Sproat, R., & Shih, C.** (1991). The cross-linguistic distribution of adjective ordering restrictions. In C. Georgopoulos & R. Ishihara (Eds.), *Interdisciplinary approaches to language*. (pp. 565–593). Dordrecht, Netherlands: Kluwer Academic.