An Information-Theoretic Explanation of Adjective Ordering Preferences

Michael Hahn (Stanford), Judith Degen (Stanford), Noah Goodman (Stanford), Dan Jurafsky (Stanford), Richard Futrell (MIT)

Background

Adjectives are subject to ordering preferences:

"*the big blue plate*" > "*the blue big plate*"

"the beautiful old house" > *"the old beautiful house"*

"the delicious boiling curry" > *"the boiling delicious curry"*

Generalizations include `Specificity', `Inherentness', `Concept-Formability', `Subjectivity'

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

Can this be explained in terms of **general principles** of **language use** and **processing**?

Corpus Analysis: Subjectivity and MI

Previous hypotheses 'specificity', 'inherentness'. We formalize using Pointwise Mutual Information:

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

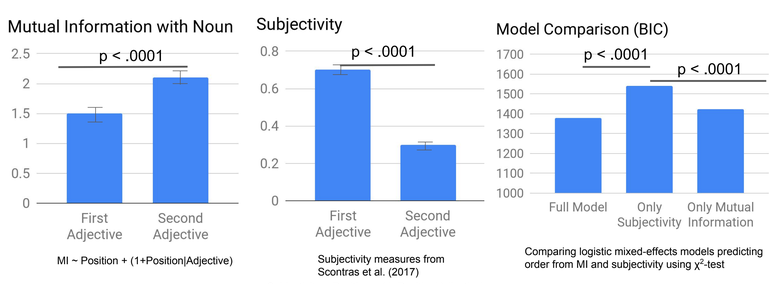
This is a common measure of collocation and a plausible quantitative measure of specificity.

Prediction: Adjectives with higher mutual information with the noun come closer to the noun.

We test this on BookCorpus and Gigaword.

Results: Prediction is confirmed

Subjectivity and Mutual Information independently impact ordering.



The Use of Adjectives

Goal: Provide a model of adjective use that **explains** effects of **subjectivity** and mutual **information**.

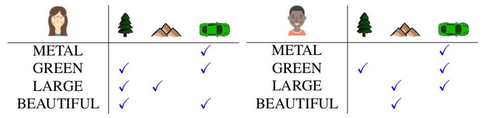
Adjectives can:

1. **pick out** a referent (Restrictive)
2. **describe** and **comment on** a referent (Non-Restrictive)

We focus on **non-restrictive** use.

RSA Model of Adjective Use

In a typical **world state**:

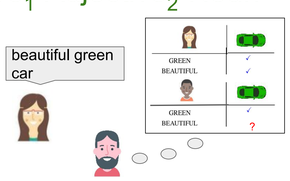


- Speakers mostly agree on objective judgments (METAL, GREEN)
- More disagreement for more subjective judgments (LARGE, BEAUTIFUL)

Formally, adjectives have different inter-speaker correlations under prior distribution over possible worlds.

Speaker utters sequence Adjective$_1$-Adjective$_2$-Noun

Listener performs Bayesian Inference about judgments of speaker and other people.



$$P_{listener}(w|u) \propto P_{prior}(w|u)\ \delta_{\text{utterance } u \text{ is true of the speaker in the world } w}$$

Speaker chooses utterance by trading off informativity and cost of utterance:

$$P_{speaker}(u) \propto \exp(\alpha \cdot (U(u) - \beta \cdot C(u)))$$

Informativity of utterance `u':

$$U(u) := -\text{KL}(P_{speaker}||P_{listener}(\cdot|u)) = \sum_w P_{speaker}(w) \log \frac{P_{listener}(w|u)}{P_{speaker}(w)}$$

Cost of producing `u':

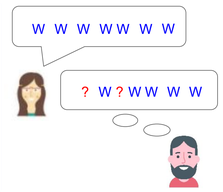
$$C(A_1A_2N) = -\log P(A_1A_2N)$$

Memory Loss Breaks Symmetry

Formal model by Futrell and Levy, 2017:

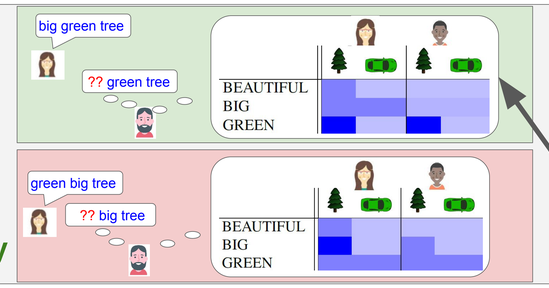
Assumption 1: Previous words in the input may be **lost from memory** stochastically

Assumption 2: Probability of loss **increases** as one goes **further back**.



Consequence: First adjective less likely to be integrated with noun by the listener.

Placing **subjective** adjective **first** has **higher informativity**



Listener learns more about third person→ higher informativity

Placing **high-MI** adjective **second** has **lower cost**

$$C(A_1A_2N) - C(A_2A_1N) = \lambda \cdot (\text{PMI}(A_1,N) - \text{PMI}(A_2,N))$$

Cost Difference between Orderings | Loss Rate | Difference in Mutual Information

Testing against Corpus Data

Predict adjective order in corpus data

Model Parameters:

- Inter-speaker correlation = 1 - subjectivity
- Other parameters inferred using Bayesian Data Analysis

Evaluation Datasets

1. Set from corpus analysis (~ 4,700 examples)
2. Unseen data set (~ 10,000 examples)

Conclusion

- Subjectivity and MI impact adjective ordering.
- Both effects can be explained by integrating Bayesian reasoning with memory limitations.
- Adjective ordering can be explained by general principles of human communication and language processing.

References: **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. **Futrell, Richard, and Roger Levy (2017).** Noisy-context surprisal as a human sentence processing cost model. *Proceedings of EACL.*