An Information-Theoretic Explanation of Adjective Ordering Preferences

Michael Hahn
Department of Linguistics
Stanford University

Work with: Richard Futrell, Judith Degen, Dan Jurafsky, Noah Goodman

Theoretical and Experimental Approaches to Modification
Tübingen, January 2021
Adjective Ordering

the big blue table    VS    the blue big table
the beautiful old house    VS    the old beautiful house
the delicious boiling curry    VS    the boiling delicious curry
Adjective Ordering

the big blue table  VS  the blue big table
the beautiful old house  VS  the old beautiful house
the delicious boiling curry  VS  the boiling delicious curry
Adjective Ordering

- the big blue table VS the blue big table
- the beautiful old house VS the old beautiful house
- the delicious boiling curry VS the boiling delicious curry

Various generalizations have been offered

- **Inherentness** (Whorf 1945)
- **Specificity** (Sweet 1898, Ziff 1960)
- **Absoluteness** (Sproat & Shih 1991)
- **Concept-Formability** (Svenonius 2008)
- **Subjectivity** (Hetzron 1978)
Adjective Ordering

the **big** blue table VS the blue **big** table
the **beautiful** old house VS the old **beautiful** house
the **delicious** boiling curry VS the boiling **delicious** curry

Various generalizations have been offered

- **Inherentness** (Whorf 1945)
- **Specificity** (Sweet 1898, Ziff 1960)
- **Absoluteness** (Sproat & Shih 1991)
- **Concept-Formability** (Svenonius 2008)
- **Subjectivity** (Hetzron 1978)

Scontras et al. (2017):

**Subjectivity** captures all of these
The more **subjective** an adjective, the **farther** from the noun it occurs.

From Scontras et al. (2017)
a. xiao lü huanping
   small green vase
   ‘the small green vase’ (Sproat & Shih, 1991, 566)

b. *lü xiao huanping
Research Question:
Can adjective ordering be explained in terms of general principles of language use and processing?
Empirical Question:

Are factors other than subjectivity relevant?
Mutual Information

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

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

Probability that \textbf{Noun} occurs, given the modifier \textbf{Adj}
Mutual Information

PMI(Adj,Noun) = log P(Noun|Adj) - log P(Noun)

Probability that \textbf{Noun} occurs, given the modifier \textbf{Adj}

Frequency of \textbf{Noun}
Mutual Information

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

Quantifies degree to which words appear together more frequently than expected at chance

Common measure of collocation (Manning and Schuetze 1999)
Mutual Information

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

PMIs computed from COCA, https://www.english-corpora.org/coca/
Mutual Information

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

(PMIs computed from COCA, https://www.english-corpora.org/coca/)
Mutual Information

Hypothesis:
Adjectives with higher mutual information with the noun tend to come closer to the noun.
Corpus Study

BookCorpus:
11,038 English novels
74 Million sentences
Corpus Study

BookCorpus:
11,038 English novels
74 Million sentences

estimate MI, controlling for existing ordering preferences
Corpus Study

BookCorpus:
11,038 English novels
74 Million sentences

extract all occurrences of "DET ADJ ADJ NOUN"
~ 4700 datapoints

estimate MI, controlling for existing ordering preferences
Subjectivity

$\text{p} = 9.36 \cdot 10^{-10}$

- First Adjective
- Second Adjective
Mutual Information with Noun

\[ p < 2.2 \cdot 10^{-16} \]

- First Adjective
- Second Adjective
Average PMI with the Noun tends to appear...
Average PMI with the Noun

tends to appear ...

first

second

American European

new

excellent

beautiful

strongly subjective adjectives

strongly non-subjective adjectives

red

green
Average PMI with the Noun

first
tends to appear ...
second
Relation of MI and Subjectivity

Predict order of $A_1, A_2$ in logistic mixed-effects model from

1. $\text{PMI}(A_1, N) - \text{PMI}(A_2, N)$
2. $\text{Subj}(A_1) - \text{Subj}(A_2)$
Subjectivity and Mutual Information independently impact ordering.
Subjectivity and Mutual Information independently impact ordering.

Model Comparison (BIC)

![Bar chart showing model comparison with BIC values: Full Model, Only Subjectivity, Only Mutual Information. The p-value is p < 2.2 \cdot 10^{-16}.]
Subjectivity and Mutual Information independently impact ordering.
### Mutual Information predicts Noun-Specific Effects:

<table>
<thead>
<tr>
<th></th>
<th>new good luck</th>
<th>international young people</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PMI</strong></td>
<td>-3.1</td>
<td>-3.0</td>
</tr>
<tr>
<td><strong>Subjectivity</strong></td>
<td>0.5</td>
<td>0.26</td>
</tr>
</tbody>
</table>

**Open curly braces**

<table>
<thead>
<tr>
<th></th>
<th>open curly braces</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PMI</strong></td>
<td>2.5</td>
</tr>
<tr>
<td><strong>Subjectivity</strong></td>
<td>0.38</td>
</tr>
</tbody>
</table>
**Mutual Information** predicts **Noun-Specific** Effects:

<table>
<thead>
<tr>
<th></th>
<th>PMI</th>
<th>PMI</th>
<th>Subjectivity</th>
<th>Subjectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>new good luck</td>
<td>-3.1</td>
<td>4.1</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>international young people</td>
<td>-3.0</td>
<td>3.9</td>
<td>0.26</td>
<td>0.64</td>
</tr>
<tr>
<td>open curly braces</td>
<td>2.5</td>
<td>9.5</td>
<td>0.38</td>
<td>0.40</td>
</tr>
</tbody>
</table>
**Mutual Information** predicts **Noun-Specific** Effects:

<table>
<thead>
<tr>
<th>new good luck</th>
<th>international young people</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMI</td>
<td>PMI</td>
</tr>
<tr>
<td>-3.1</td>
<td>-3.0</td>
</tr>
<tr>
<td>4.1</td>
<td>3.9</td>
</tr>
<tr>
<td>Subjectivity</td>
<td>Subjectivity</td>
</tr>
<tr>
<td>0.5</td>
<td>0.26</td>
</tr>
<tr>
<td>0.8</td>
<td>0.64</td>
</tr>
</tbody>
</table>

**open curly braces**

| PMI                    | Subjectivity               |
| 2.5                    | 0.38                       |
| 9.5                    | 0.40                       |
**Mutual Information** predicts **Noun-Specific** Effects:

<table>
<thead>
<tr>
<th></th>
<th>new good luck</th>
<th>international young people</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PMI</strong></td>
<td>-3.1  4.1</td>
<td>-3.0  3.9</td>
</tr>
<tr>
<td><strong>Subjectivity</strong></td>
<td>0.5  0.8</td>
<td>0.26  0.64</td>
</tr>
</tbody>
</table>

open curly braces

<table>
<thead>
<tr>
<th></th>
<th>PMI  2.5  9.5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subjectivity</strong></td>
<td>0.38  0.40</td>
</tr>
</tbody>
</table>
Goal:

Provide a model of adjective use that explains effects of subjectivity and mutual information.
The Use of Adjectives

Adjectives can help **pick out** referents.
Click on the yellow comb.

Sedivy, Chambers, and Tanenhaus (1999)
The Use of Adjectives

Adjectives can help *pick out* referents.

Adjectives can *describe* and *comment on* a referent.
In the cool evening she started taking drives, leaving her old mother at home.

Adjectives can describe and comment on a referent.

Adjectives can help pick out referents.

Adjectives can describe and comment on a referent.

What a nice dog.

Does not help pick out a referent.

Speaker comments on referent.
Forrest looks at the massive crowd.

I see the door to the house open..., and in the yellow light I see Kate.

We look at the little animal faces, and we know they need a home.

The toes of animals tapped on the metal roof in the dark.

Abruptly, the beautiful face softened.

Telling the red blood to stop flowing.

Look at the little boy!
Forrest looks at the **massive** crowd.

I see the door to the house open..., and in the **yellow** light I see Kate.

The toes of animals tapped on the **metal** roof in the dark.

Telling the **red** blood to stop flowing.

Abruptly, the **beautiful** face softened.

Look at the **little** boy!

---

Model will be centered around speakers communicating **descriptions** and **attitudes**.

from COCA (Davies, 2017)
Modeling Approach

1. Formalize nonrestrictive use of adjectives
2. Define a rational Bayesian model of communication
3. Show how memory limitations lead effects of subjectivity and mutual information
4. Evaluate on Corpus Data
Modeling Approach

1. Formalize nonrestrictive use of adjectives
2. Define a rational Bayesian model of communication
3. Show how memory limitations lead effects of subjectivity and mutual information
4. Evaluate on Corpus Data
beautiful green car
**World state**  =  **Truth value assignment** to the cells in this table

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

![Image showing the truth value assignment to the cells in the table.](image-url)
Speakers mostly **agree** on **objective** judgments
Speakers mostly agree on objective judgments

More disagreement for more subjective judgments
<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>METAL</td>
<td>GREEN</td>
<td>LARGE</td>
<td>BEAUTIFUL</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>☑</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METAL</td>
<td>GREEN</td>
<td>LARGE</td>
<td>BEAUTIFUL</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>METAL</td>
<td>GREEN</td>
<td>LARGE</td>
<td>BEAUTIFUL</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>✔</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>METAL</td>
<td>GREEN</td>
<td>LARGE</td>
<td>BEAUTIFUL</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>METAL</td>
<td>GREEN</td>
<td>LARGE</td>
<td>BEAUTIFUL</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-----------</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

High Correlation

$\kappa(\text{metal}) = 0.95$
<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>METAL</th>
<th>GREEN</th>
<th>LARGE</th>
<th>BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

Low Correlation
\[ \kappa(\text{beautiful}) = 0.2 \]
beautiful green car
beautiful green car
beautiful green car
Modeling Approach

1. Formalize nonrestrictive use of adjectives
2. Define a rational Bayesian model of communication
3. Show how memory limitations lead effects of subjectivity and mutual information
4. Evaluate on Corpus Data
Rational Communication: Speakers and Listeners

Formalize model in the framework of Bayesian pragmatics (Franke 2008; Frank and Goodman, 2012)

\[ P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w) \delta_{u \text{ is true for speaker in } w} \]

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]
Listener Model

Listener performs Bayesian reasoning to infer world state.

\[ P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w) \delta \]

- state of the world
- utterance received
- \( u \) is true for speaker in \( w \)
beautiful green car

\[ P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w) \delta_u \text{ is true for speaker in } w \]
<table>
<thead>
<tr>
<th>GREEN BEAUTIFUL</th>
<th>GREEN BEAUTIFUL</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
</tbody>
</table>

- **beautiful green car**
- **high certainty**
- **uncertainty**
Rational Communication: Speakers and Listeners

Formalize model in the framework of Bayesian pragmatics
(Franke 2008; Frank and Goodman, 2012)

\[ P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w) \delta_u \text{ is true for speaker in } w \]

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]
Speaker Model

Speaker chooses utterance to optimize utility (Franke 2008; Frank and Goodman 2012).

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

- **Informativity** of utterance `u`
- **Cost** of utterance
Speaker Model

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

Typically: Reduction in the listener’s uncertainty about the world state, measured in bits (e.g., Frank and Goodman, 2012; Goodman and Stuhlmueller, 2013).
Informativity = 0 bits
green car

Informativity = 1 bits
beautiful green car

Informativity = 2 bits
Speaker Model

Speaker chooses utterance to optimize utility (Franke 2008; Frank and Goodman 2012).

\[
P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u))
\]

**Informativity** of
utterance `u'
Speaker Model

$$P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u))$$

<table>
<thead>
<tr>
<th>I(U) = 0</th>
<th>I(U) = 1</th>
<th>I(U) = 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>25%</td>
<td>65%</td>
</tr>
</tbody>
</table>

- car
- beautiful car
- beautiful green car
beautiful
green car

Informativity about = 2 bits
beautiful green car

Informativity about green = 2 bits
Informativity about beautiful = 1 bits

\[ I(U) = I \text{green} + I \text{beautiful} \]
Informativity about \( \text{GREEN BEAUTIFUL} \) = 1 bits

Informativity about \( \text{GREEN BEAUTIFUL} \) = 2 bits

\[ \text{Informativity about} \quad \text{GREEN BEAUTIFUL} = \text{I(U)} \]

Cooperative speakers communicate knowledge that generalizes to other people.
Speaker Model: Cost

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]

Cost of the utterance
Speaker Model: Cost

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]

C(u) = − log P(u)

Surprisal of the utterance
(c.f. Bennett & Goodman, 2018; Peloquin et al, 2019)
Speaker Model: Cost

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]

\[ C(u) = -\log P(u) \]

We will assume no prior preference:

\[ P(A_1 A_2 N) = P(A_2 A_1 N) \]
Rational Communication: Speakers and Listeners

Formalize model in the framework of Bayesian pragmatics (Franke 2008; Frank and Goodman, 2012)

\[ P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w) \delta_{\text{u is true for speaker in } w} \]

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]
Rational Communication: Speakers and Listeners

Formalize model in the framework of Bayesian pragmatics
(Franke 2008; Frank and Goodman, 2012)

$$P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w) \delta_{u \text{ is true for speaker in } w}$$

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

Informativity about

Informativity about
Rational Communication: Speakers and Listeners

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]

So far, no ordering preferences are predicted!

big green tree \quad \text{green big tree}

Identical Informativity and Cost
Rational Communication: Speakers and Listeners

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]

Proposal:

Memory limitations in processing break symmetry.
Memory Limitations

Firmly established as factor in language understanding

Classical example: Long dependencies harder to process
(e.g., Gibson, 1998; McElree, 2000; Lewis & Vasishth, 2005; Bartek et al., 2011; Nicenboim, 2015)
Memory Limitations: Formal Model  (Futrell and Levy, 2017)
Memory Limitations: Formal Model (Futrell and Levy, 2017)
Memory Limitations: Formal Model  (Futrell and Levy, 2017)
Memory Limitations: Formal Model (Futrell and Levy, 2017)

Assumption 1:

Previous words in the input may be lost from memory stochastically.
Memory Limitations: Formal Model (Futrell and Levy, 2017)

Assumption 1:
Previous words in the input may be lost from memory stochastically
Assumption 1:
Previous words in the input may be lost from memory stochastically
Memory Limitations: Formal Model (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** in the sequence.
Listener Model with Memory Loss

big

big
Listener Model with Memory Loss

big green

big green
Listener Model with Memory Loss

big green tree

BEAUTIFUL
BIG
GREEN
Listener Model with Memory Loss

?? green tree

big green tree
Listener Model with Memory Loss

Rational listener marginalizes over possible completions (Futrell & Levy, 2017)
Listener Model with Memory Loss

<table>
<thead>
<tr>
<th>BEAUTIFUL</th>
<th>BIG</th>
<th>GREEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>big green tree</td>
<td>beautiful green tree</td>
<td>....</td>
</tr>
</tbody>
</table>

big green tree
Listener Model with Memory Loss

??? big tree

green big tree
Listener Model with Memory Loss

green big tree
beautiful big tree
....

BEAUTIFUL
BIG
GREEN

green big tree
<table>
<thead>
<tr>
<th>BEAUTIFUL</th>
<th>BIG</th>
<th>GREEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BEAUTIFUL</th>
<th>BIG</th>
<th>GREEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Listener able to generalize better across persons?
Prediction:

Assuming forgetful listener, placing **subjective** adjective **first** has **higher expected informativity** under the model.
Memory Loss in the Cost
Memory Loss in the Cost

$A_1 - \log P(A_1)$
Memory Loss in the Cost

\[ A_1 \]
\[ A_1 \quad A_2 \quad - \log P(A_1) \quad - \log P(A_2|A_1) \]
Memory Loss in the Cost

\[ -\log P(A_1) \]
\[ -\log P(A_2|A_1) \]
\[ -\log P(N|?? A_2) \]

Will be smaller if \( \text{PMI}(N, A_2) \) is larger!
Our Proposed Model

Rational communication with Bayesian inference

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]
Our Proposed Model

Rational communication with Bayesian inference

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]

including reasoning about multiple speakers
Our Proposed Model

Rational communication with Bayesian inference

\[
P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u))
\]

including reasoning about multiple speakers and incremental, rational processing under memory limitations.
Evaluation

Task: Predict adjective order in corpus data
Evaluation

**Task:** Predict adjective order in corpus data

Model Parameters:

- $\kappa(A) = 1 - \text{subjectivity}(A)$

![Chart showing model parameters and example values]

$\kappa(\text{big}) = 0.2$  
$\kappa(\text{metal}) = 0.85$
Evaluation

**Task:** Predict adjective order in corpus data

**Model Parameters:**

- $\kappa(A) = 1 - \text{subjectivity}(A)$
- MI: from corpus analysis
Evaluation

**Task:** Predict adjective order in corpus data

**Model Parameters:**

- $\kappa(A) = 1 - \text{subjectivity}(A)$
- **MI:** from corpus analysis
- Other parameters inferred using Bayesian Data Analysis in Pyro (http://pyro.ai/)

$$P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u))$$
**Evaluation**

**Task:** Predict adjective order in corpus data

**Model Parameters:**
- $\kappa(A) = 1 - \text{subjectivity}(A)$
- **MI:** from corpus analysis
- Other parameters inferred using Bayesian Data Analysis in Pyro (http://pyro.ai/)

**Evaluation Datasets**
Set from corpus analysis (~ 4,700 examples)
Evaluation

Task: Predict adjective order in corpus data

Model Parameters:
- \( \kappa(A) = 1 - \text{subjectivity}(A) \)
- MI: from corpus analysis
- Other parameters inferred using Bayesian Data Analysis in Pyro (http://pyro.ai/)

Evaluation Datasets
- Unseen data set (~ 10,000 examples)
- Set from corpus analysis (~ 4,700 examples)
Classification accuracy:
- \( \phi \) fixed
  - 93.7% on set from corpus analysis
  - 93.1% on unseen data (~10,000 examples)
- \( \phi \) depends on adjective:
  - 97.3% on set from corpus analysis
  - 96.2% on unseen data

Future research:
- Compare inferred values for \( \phi(A) \) to elicited probabilities
- Elicit priors depending on both adjective and object

Model

- Subjectivity+MI
  - Logistic Regression

Graph showing accuracy for original and unseen data for two models: Subjectivity+MI Logistic Regression and Model.
Languages with Postnominal Adjectives

l-kitaabu l-’axdaru š-šayiru
the-book the-green the-small
‘the little green book’ (Fassi Fehri, 1999, 107)

Standard Arabic
Languages with Postnominal Adjectives

Subjectivity-based ordering reported for

- Arabic (Kachakeche & Scontras, 2020)
- Tagalog (Samonte & Scontras, 2019)

Similarly for many other languages (Dixon, 1982; Hetzron, 1978; Sproat & Shih, 1991).
Languages with Postnominal Adjectives

tree green big

tree big green
Languages with Postnominal Adjectives

- Tree green
- Tree big

<table>
<thead>
<tr>
<th>Beautiful</th>
<th>BIG</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beautiful</th>
<th>BIG</th>
<th>Green</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Two images of the same scene, one with the words "tree green" and the other with the words "tree big"
Languages with Postnominal Adjectives

Listener able to generalize better across persons

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big

Languages with Postnominal Adjectives

- tree green
- tree big
Languages with Postnominal Adjectives

\[
\begin{align*}
N & \quad - \log P(N) \\
N & \quad A_1 & \quad - \log P(A_1 | N) \\
?? & \quad A_1 & \quad A_2 & \quad - \log P(A_2 | ?? A_1)
\end{align*}
\]
Languages with Postnominal Adjectives

Will be lower when PMI($A_1$, $N$) is higher.
Languages with Postnominal Adjectives

Mutual Information With Noun (in bits)

<table>
<thead>
<tr>
<th>Language</th>
<th>A1</th>
<th>A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Arabic</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Discussion

*Subjectivity* and *MI* independently impact adjective ordering.
Discussion

**Subjectivity** and **MI** independently impact adjective ordering.

Provided model of adjective ordering integrating standard **Bayesian reasoning**

\[ P_{\text{speaker}}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u)) \]
Discussion

Subjectivity and MI independently impact adjective ordering.

Provided model of adjective ordering integrating standard Bayesian reasoning with incremental processing under memory limitations.

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

Assumption 2: Probability of loss increases as one goes further back in the sequence.
Discussion

Subjectivity and MI independently impact adjective ordering.

Provided model of adjective ordering integrating standard Bayesian reasoning with incremental processing under memory limitations, achieving 96% accuracy on corpus data.
Discussion

Subjectivity and MI independently impact adjective ordering.

Provided model of adjective ordering integrating standard Bayesian reasoning with incremental processing under memory limitations, achieving 96% accuracy on corpus data.

Suggests that adjective ordering can be explained by general principles of human communication and language processing.
Discussion

**Subjectivity** and MI independently impact adjective ordering.

Provided model of adjective ordering integrating standard **Bayesian reasoning** with **incremental processing** under **memory limitations**, achieving **96% accuracy** on corpus data.

Suggests that adjective ordering can be explained by general principles of **human communication** and **language processing**.

Subjective material tends to appear at **periphery** of phrases and clauses (Traugott, 2010).

**Future Research:** Test our model on other types of subjective content.

blue big book

blue big book

Reference resolution failed!

Their model:
Grounded in reference resolution
Predicts that conjunction weakens/eliminates the effect
(Rosales & Scontras, 2019; Scontras et al., 2020)

Our model:
Grounded in nonrestrictive usage
Centered around incremental processing aiming to be compatible with experimental evidence on processing
Accounts for MI effect in addition to Subjectivity effect
Mutual Information beyond Adjective Order
Mutual Information in Adverb Order

frankly > fortunately > allegedly > probably > once/then > perhaps > wisely > usually > already > no longer > always > completely > well

(Cinque 1999, p. 34)
\[ \rho = 0.60 \]

Average Mutual Information between Adverb and Verb

Average (Log) Distance between Adverb and Verb

Adverbs appearing at least 20K times

- furthermore
- however
- actively
- closely
- strongly
- always
\[ \rho = 0.60 \]

Average Mutual Information between Adverb and Verb

Average (Log) Distance between Adverb and Verb

Adverbs appearing at least 20K times

clause-oriented adverbs

VP-oriented adverbs

always

furthermore

however

furthermore

however

actively

closely

strongly
Predict order of pairs \( \text{Adverb}_1 \text{ Adverb}_2 \) in corpus using logistic regression from

1. **Mutual Information:** \( \text{pmi(Adverb}_1, \text{Verb)} - \text{pmi(Adverb}_2, \text{Verb)} \)
2. **Ranks** of adverbs in the hierarchy

    frankly > fortunately > allegedly > probably > once/then > perhaps > wisely > usually > already > no longer > always > completely > well

    (Cinque 1999, p. 34)
Model Comparison (BIC)

- Weakness
- Model Fit
- Strength

- Full Model
- Only Mutual Information
- Only Rank

$p < 2 \cdot 10^{-16}$
Model Comparison (BIC)

Full Model  Only Mutual Information  Only Rank

No Improvement
Mutual Information beyond Adjective Ordering

Mutual Information between words at given distance (controlling for redundancy with intervening words)

Computed from: English treebanks in Universal Dependencies (Nivre et al., 2017)

(Hahn, Degen, Futrell, in press)
(Hahn, Degen, Futrell, in press)
(Hahn, Degen, Futrell, in press)
specify random but internally consistent word order patterns e.g.
- SOV
- Noun-Adjective
- Genitive-Noun
- ...

(Hahn, Degen, Futrell, in press)
(Hahn, Degen, Futrell, in press)
(Hahn, Degen, Futrell, in press)
Mutual Information beyond Adjective Ordering

Mutual Information between words at given distance (controlling for redundancy with intervening words)

Real English
Median of Random Grammars

computed from: English treebanks in Universal Dependencies (Nivre et al, 2017)

(Hahn, Degen, Futrell, in press)
Mutual Information beyond Adjective Ordering

**Mutual Information**
between words at given distance

**Information Locality**:
Real orders place high-MI word pairs close together.
(Futrell and Levy, 2017)

Computed from:
- English treebanks in Universal Dependencies (Nivre et al., 2017)
- Median of random grammars (Hahn, Degen, Futrell, in press)

Distance Mutual Information between words at given distance

(Hahn, Degen, Futrell, in press)
Mutual Information beyond Adjective Ordering

Mutual Information between words at given distance

Information Locality:
Real orders place high-MI word pairs close together.
(Futrell and Levy, 2017)

Minimizes surprisal cost under memory limitations
(Hahn, Degen, Futrell, in press)
Mutual Information between words at given distance.

Distance

Real Orders

Median of Random Grammars

(Hahn, Degen, Futrell, in press)
Conclusion

**Subjectivity** and **MI** independently impact adjective ordering.

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

![Graph showing model comparison (BIC)](image)

- apple: 3.5
- banana: 3.9
- car: 1.3
- house: 1.8
Conclusion

Subjectivity and MI independently impact adjective ordering.

Proposed model of adjective ordering integrating Bayesian reasoning with incremental processing under memory limitations.
Conclusion

Subjectivity and MI independently impact adjective ordering.

Proposed model of adjective ordering integrating Bayesian reasoning with incremental processing under memory limitations.

Suggest that adjective ordering can be explained by general principles of human communication and language processing.
Conclusion

Subjectivity and MI independently impact adjective ordering.

Proposed model of adjective ordering integrating Bayesian reasoning with incremental processing under memory limitations.

Suggest that adjective ordering can be explained by general principles of human communication and language processing.

Mutual Information predicts order in language more generally.
Thank you!