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

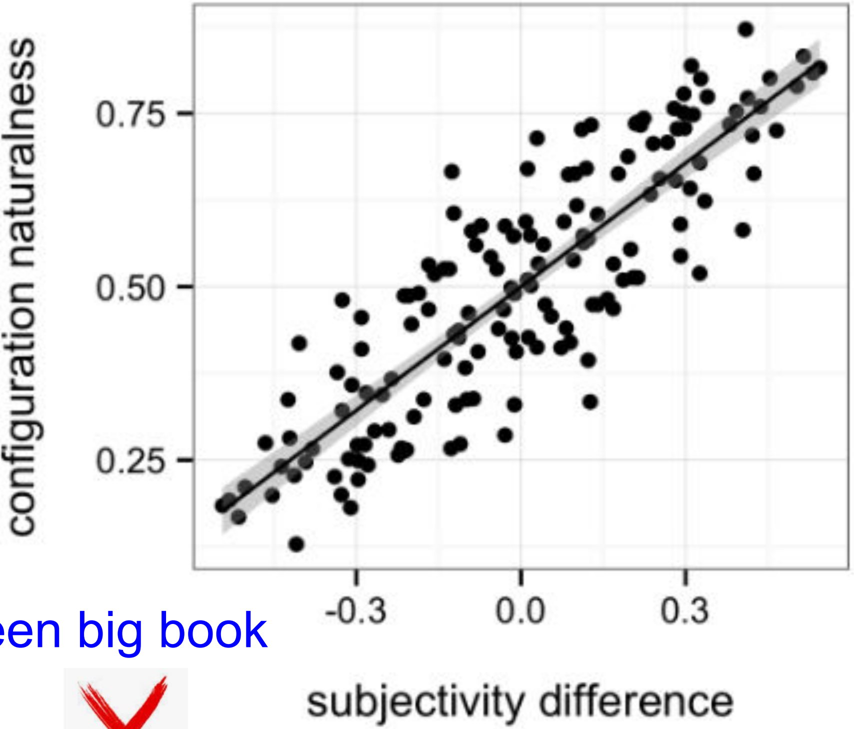
- **Inherentness** (Whorf 1945)
- **Specificity** (Sweet 1898, Ziff 1960)
- **Absoluteness** (Sproat & Shih 1991)
- **Concept-Formability** (Svenonius 2008)
- **Subjectivity** (Hetzron 1978)

Scontras et al. (2017):

Subjectivity captures all of these

The more **subjective** an adjective, the **farther** from the noun it occurs.

big green book



green big book

From Scontras et al. (2017)

a. xiao lü huanping
small green vase
‘the small green vase’ (Sproat & Shih, 1991, 566)

b. *lü xiao huanping

Mandarin Chinese

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

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

Mutual Information

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

Probability that
Noun occurs, given
the modifier **Adj**

Mutual Information

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

Probability that **Noun** occurs, given the modifier **Adj**

Frequency of **Noun**

Mutual Information

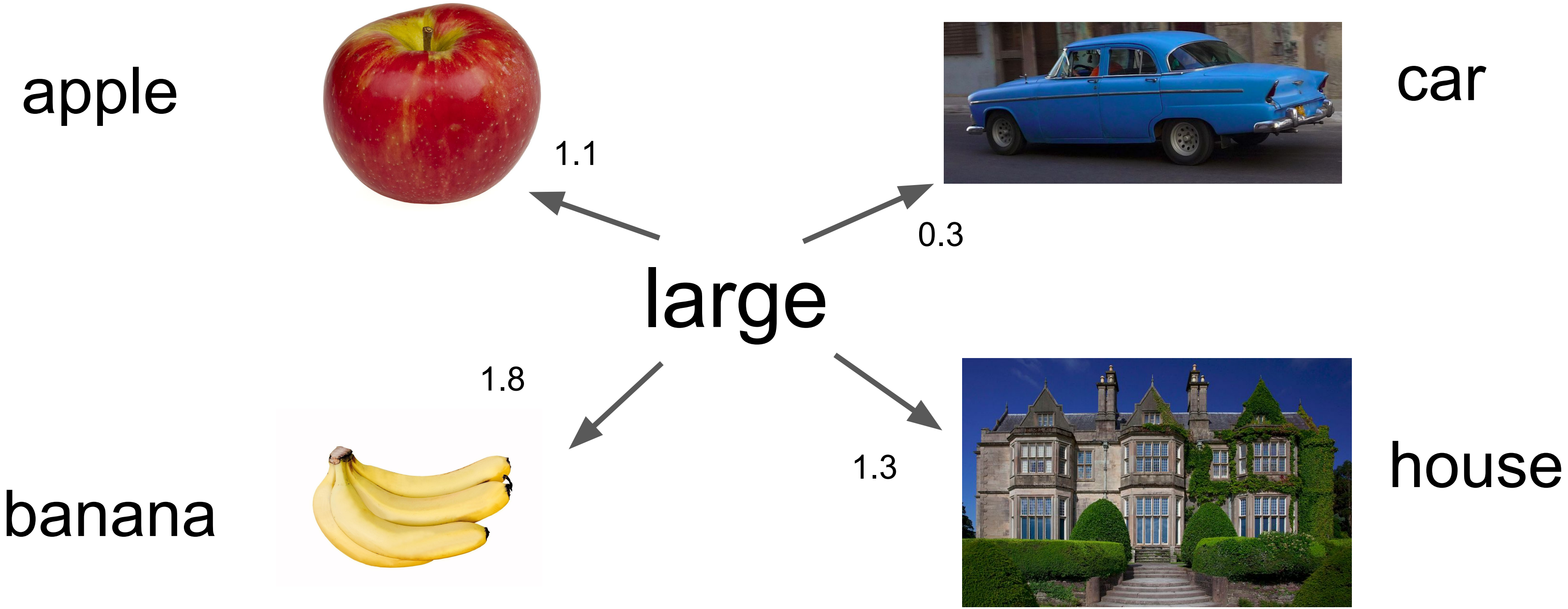
$$\text{PMI}(\text{Adj},\text{Noun}) = \log P(\text{Noun}|\text{Adj}) - \log P(\text{Noun})$$

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

Common measure of collocation (Manning and Schuetze 1999)

Mutual Information

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



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

Mutual Information

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

apple

car

3.5

-1.3

delicious

3.9

-1.8

banana

house

(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

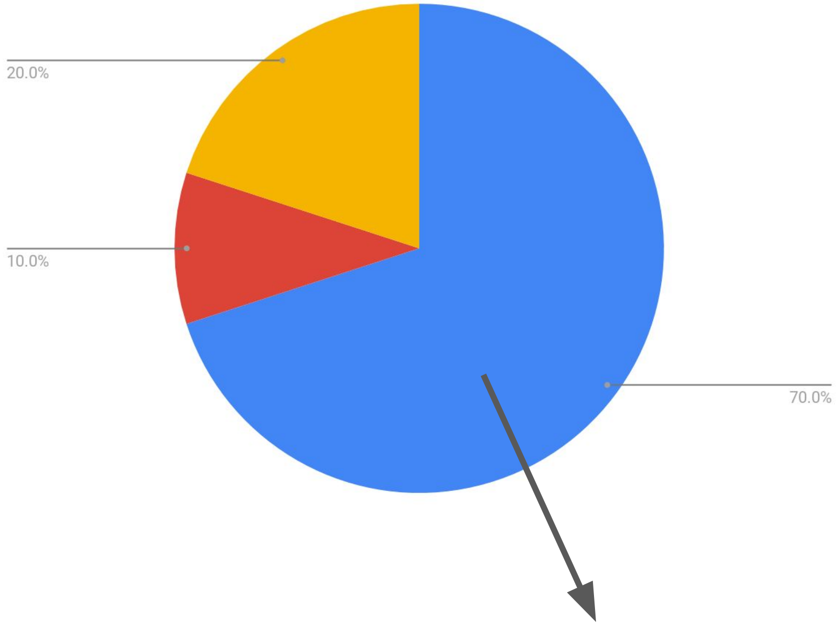
74 Million sentences

Corpus Study

BookCorpus:

11,038 English novels

74 Million sentences



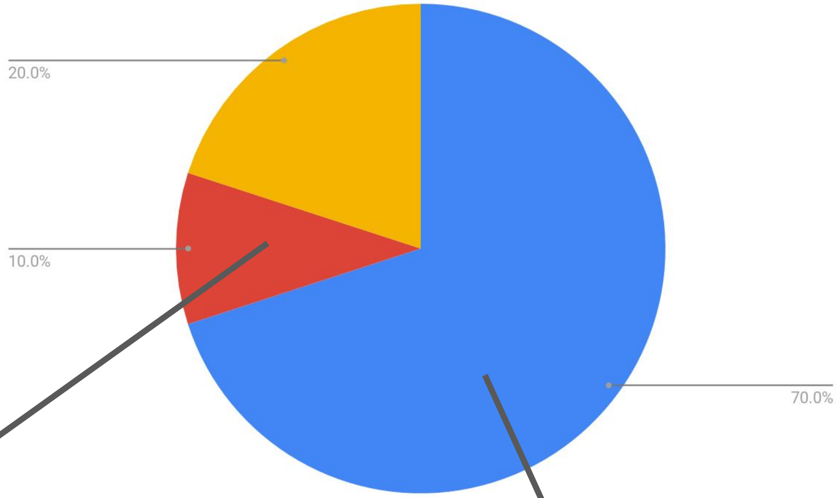
estimate MI, controlling for existing ordering preferences

Corpus Study

BookCorpus:

11,038 English novels

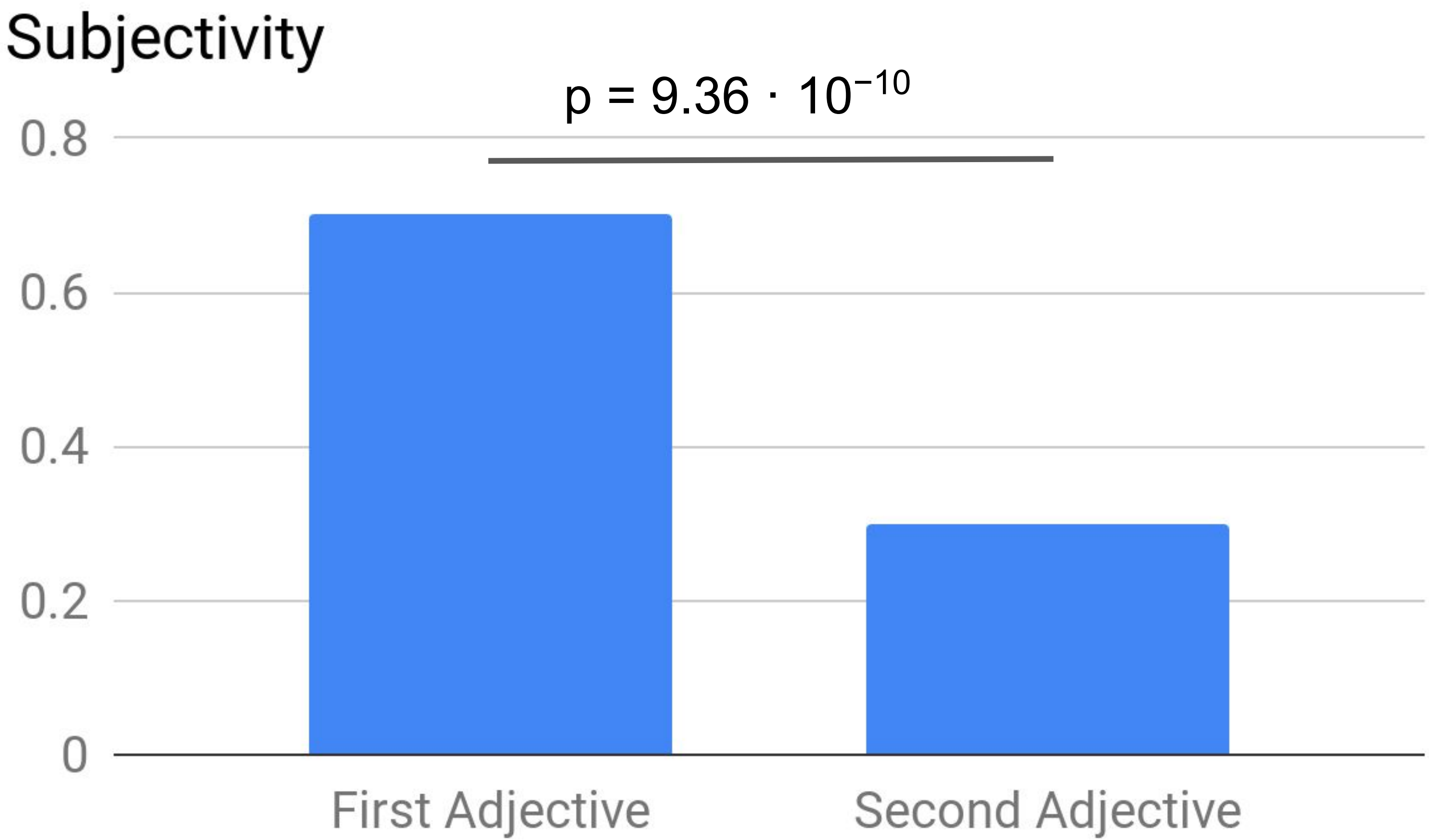
74 Million sentences

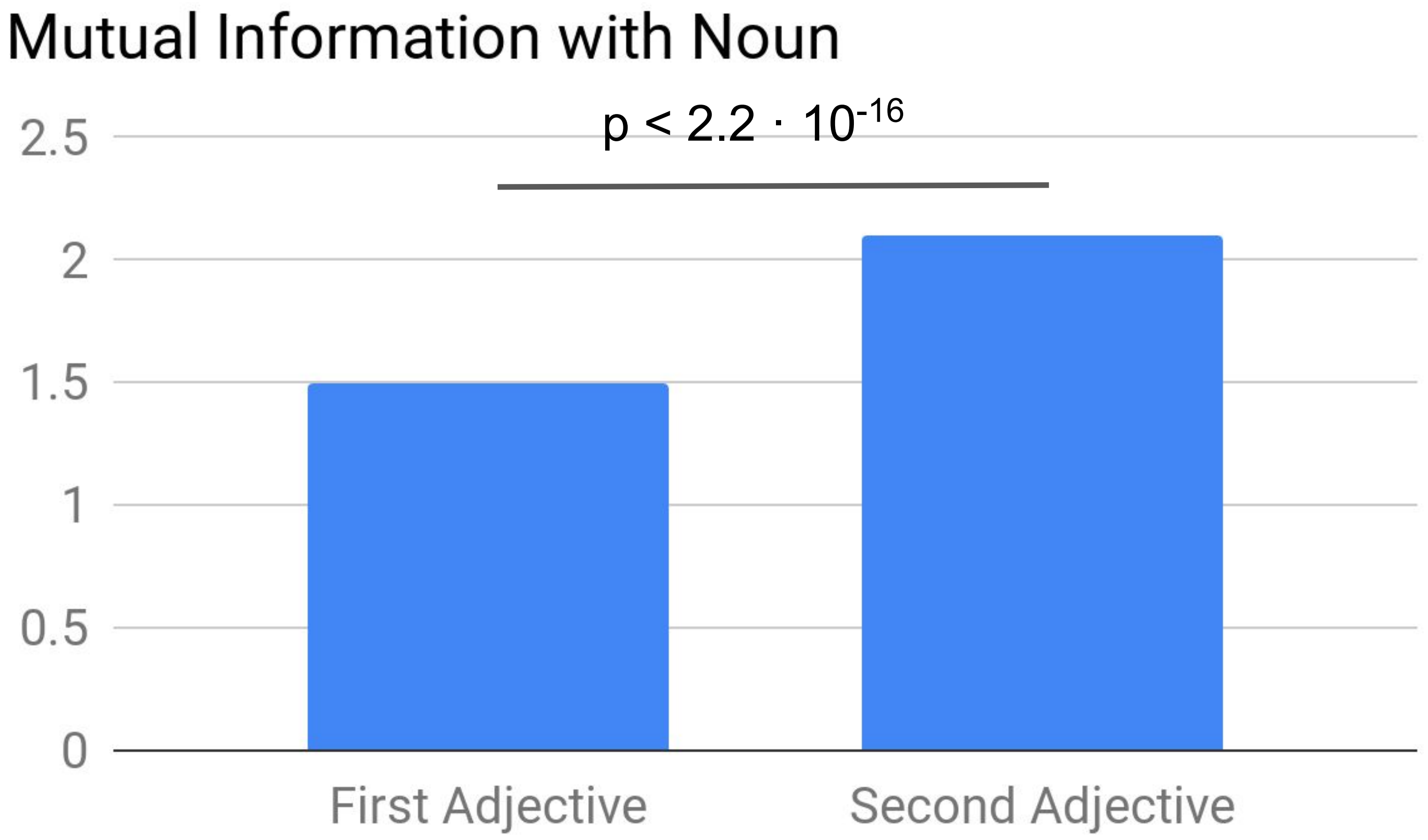


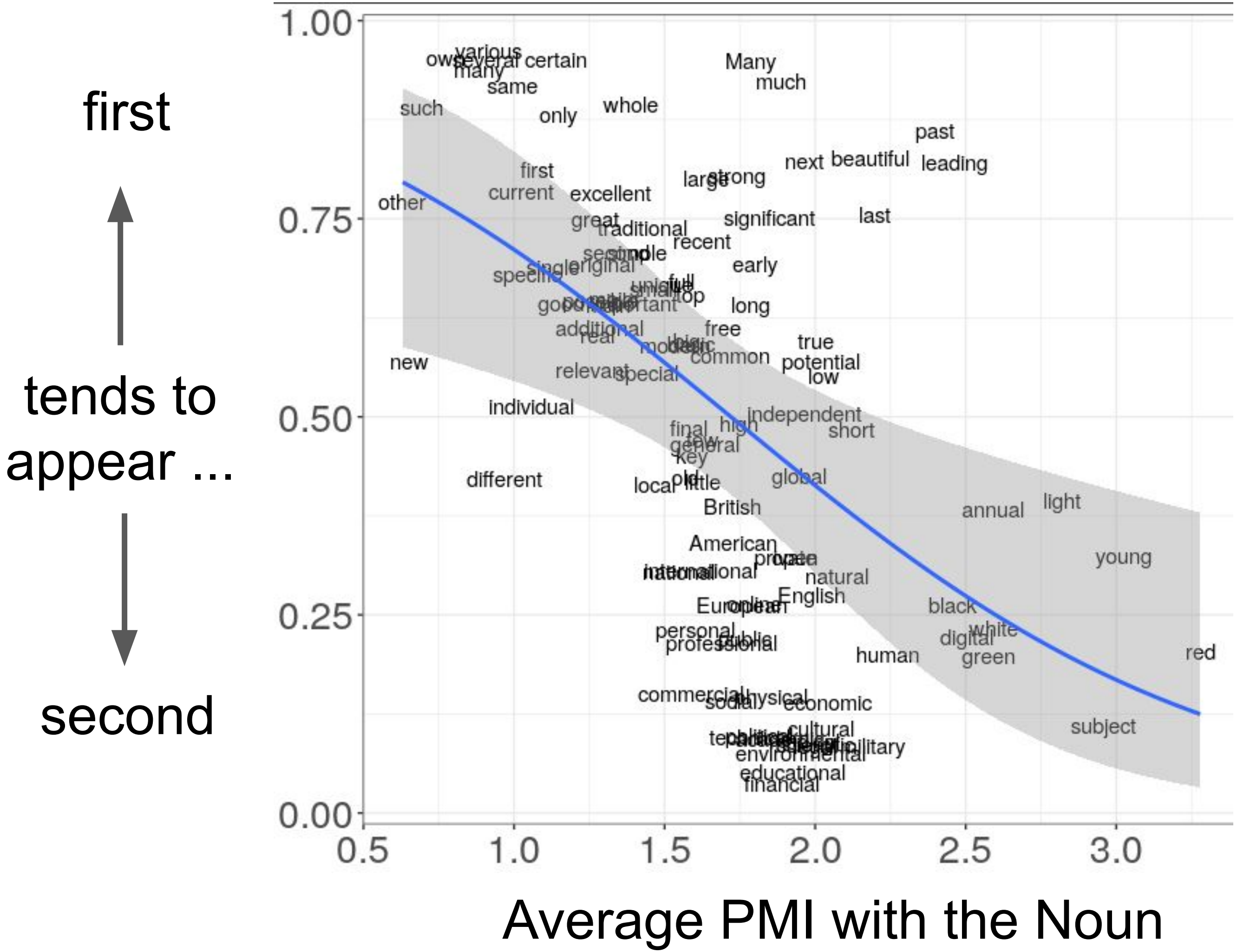
extract all occurrences of
“DET ADJ ADJ NOUN”

~ 4700 datapoints

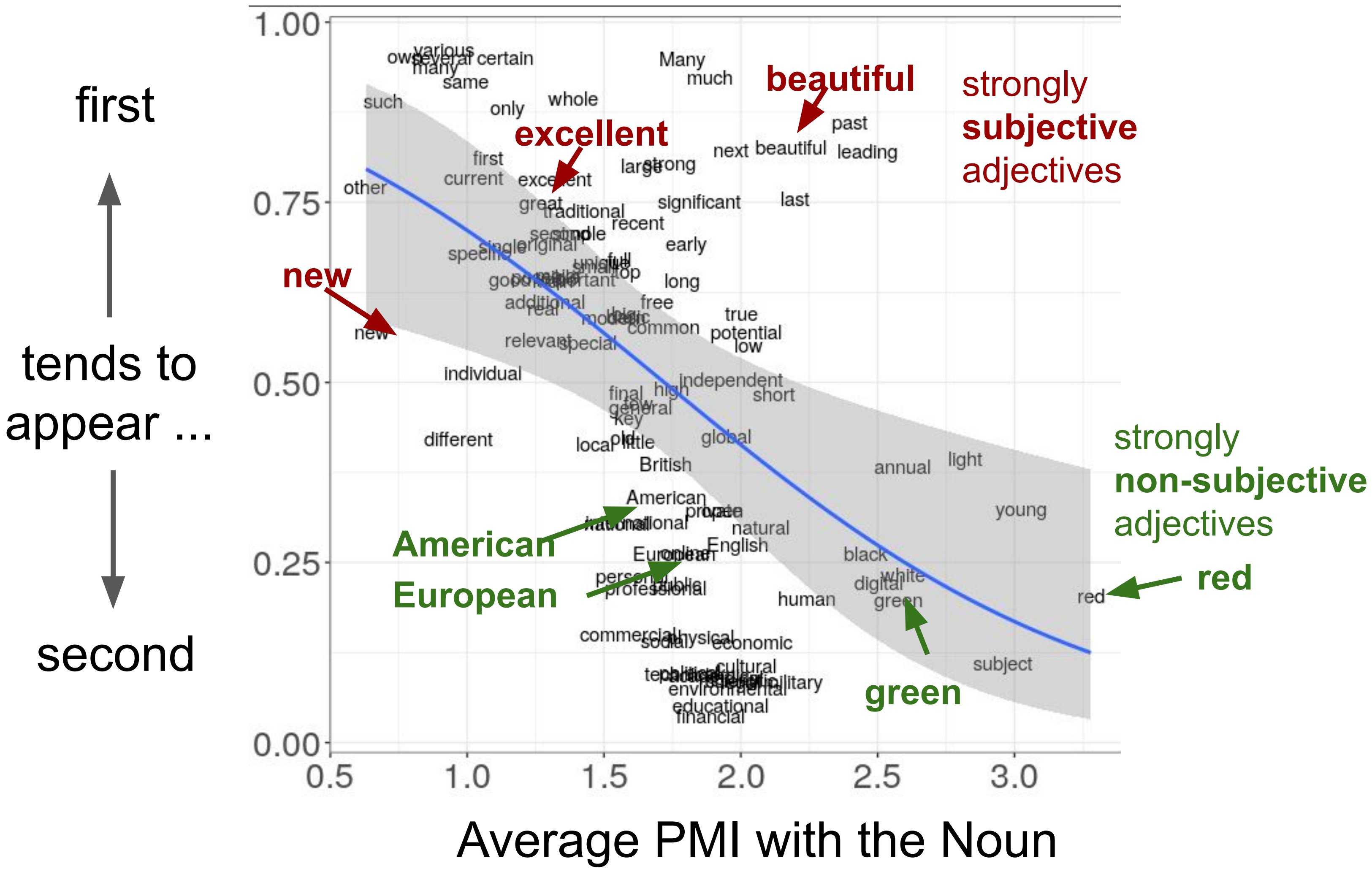
estimate MI, controlling for existing ordering preferences


Subjectivity
p = 9.36 · 10^−10
0.8
0.6
0.4
0.2
0
First Adjective
Second Adjective

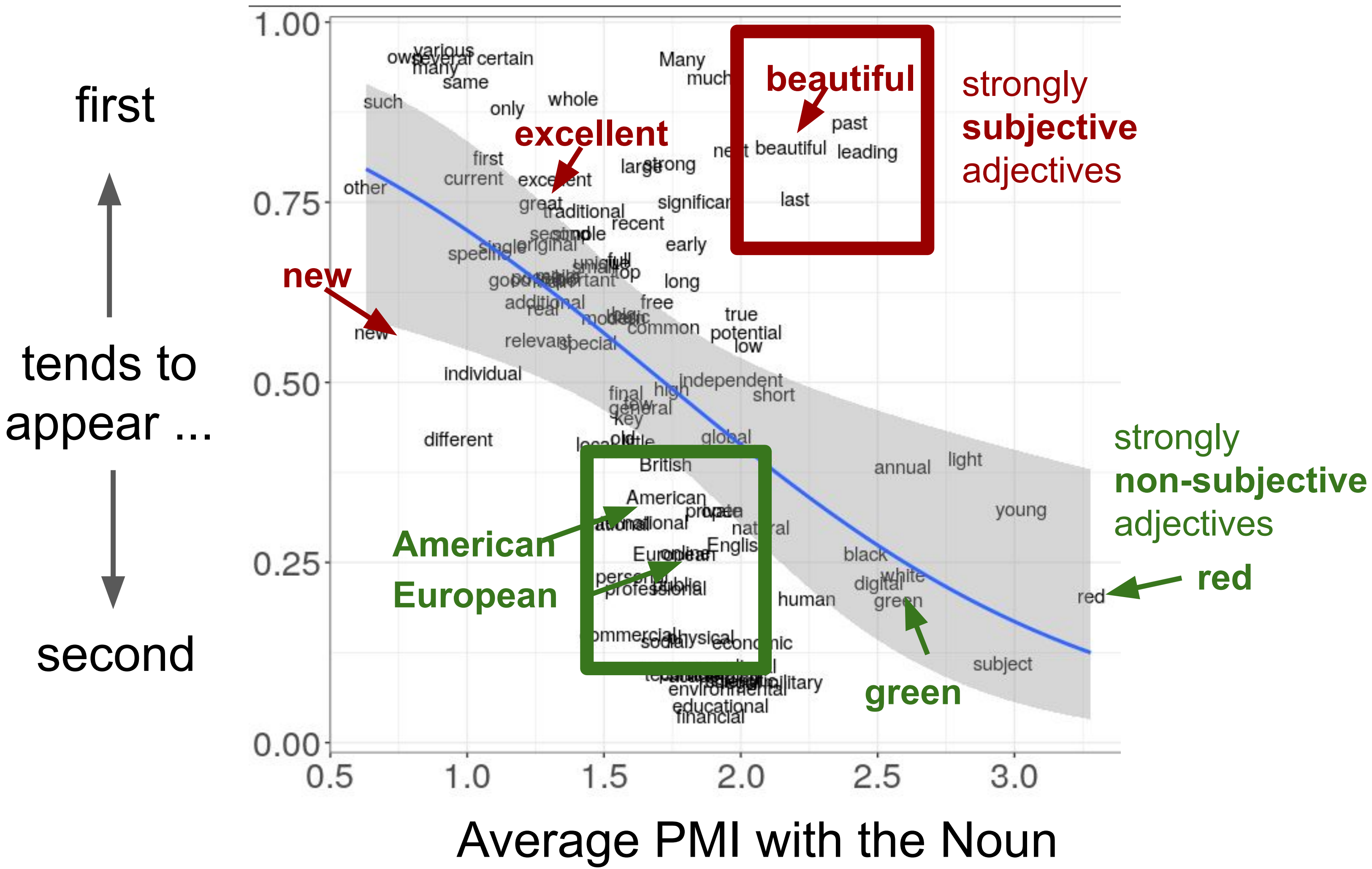

Mutual Information with Noun
p < 2.2 · 10-16
2.5
2
1.5
1
0.5
0
First Adjective
Second Adjective



Spearman’s rho: 0.60 ($p < 10^{-16}$)



first
tends to appear ...
second
1.00
0.75
0.50
0.25
0.00
0.5
1.0
1.5
2.0
2.5
3.0
Average PMI with the Noun
beautiful
excellent
new
strongly subjective adjectives
strongly non-subjective adjectives
American
European
red
green
various
certain
many
same
such
only
whole
Many
much
past
next
leading
first
current
other
significant
last
recent
early
long
free
true
potential
low
common
relevant
special
individual
independent
short
final
high
general
key
different
global
British
annual
light
young
natural
English
black
white
digital
human
subject
economic
cultural
environmental
educational
financial

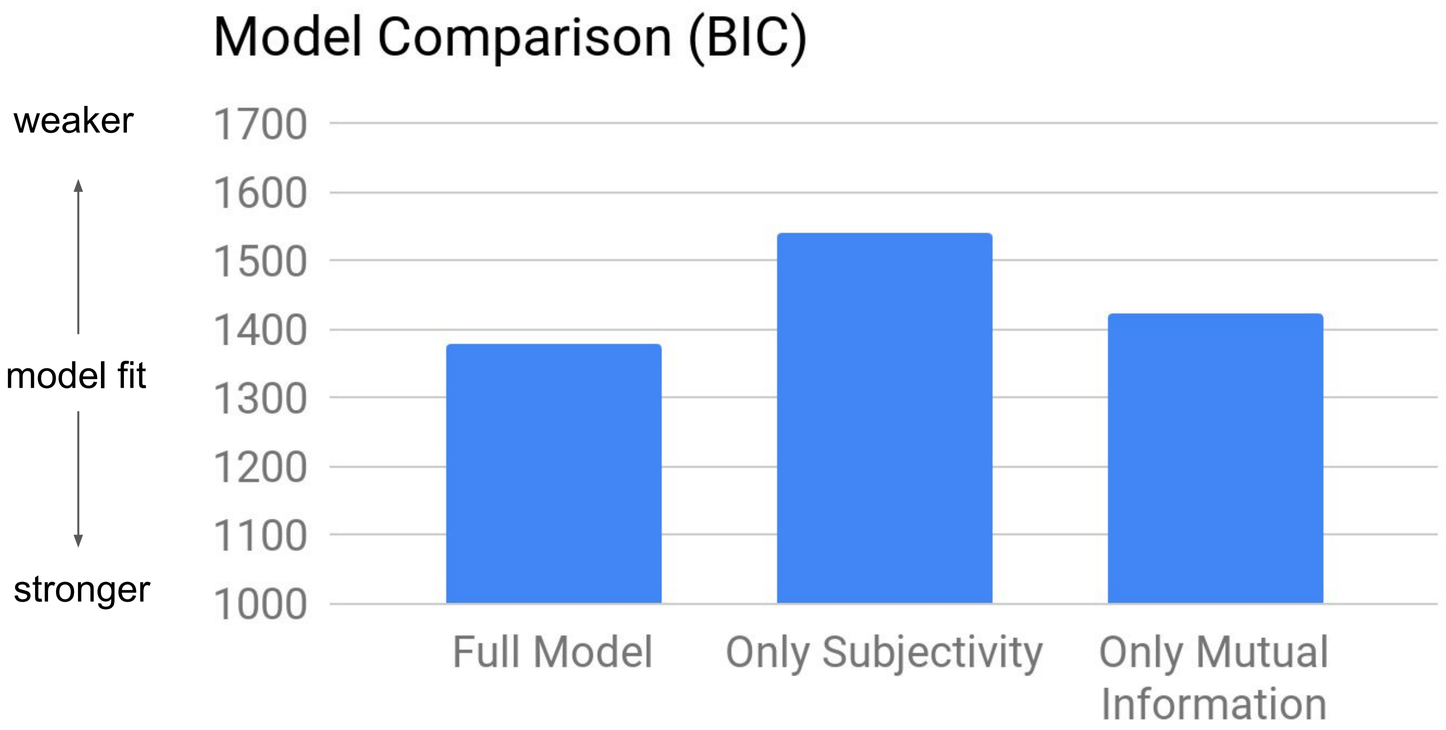


first
tends to appear ...
second
1.00
0.75
0.50
0.25
0.00
0.5
1.0
1.5
2.0
2.5
3.0
Average PMI with the Noun
beautiful
strongly
subjective
adjectives
excellent
new
strongly
non-subjective
adjectives
red
American
European
green
such
other
same
only
whole
certain
Many
much
past
leading
last
first
current
recent
early
long
free
true
potential
low
individual
different
independent
short
final
general
global
British
annual
light
young
black
human
subject
educational
financial

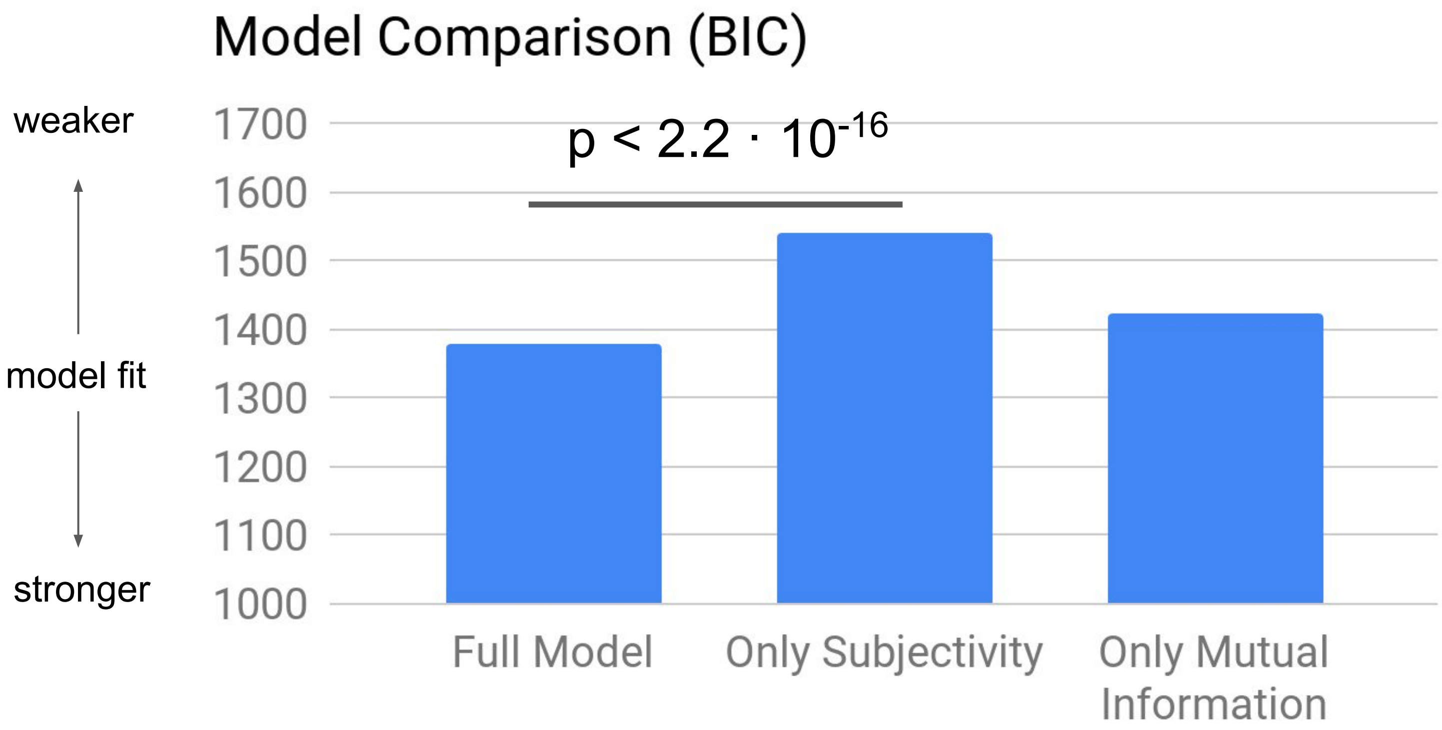
Relation of MI and Subjectivity

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

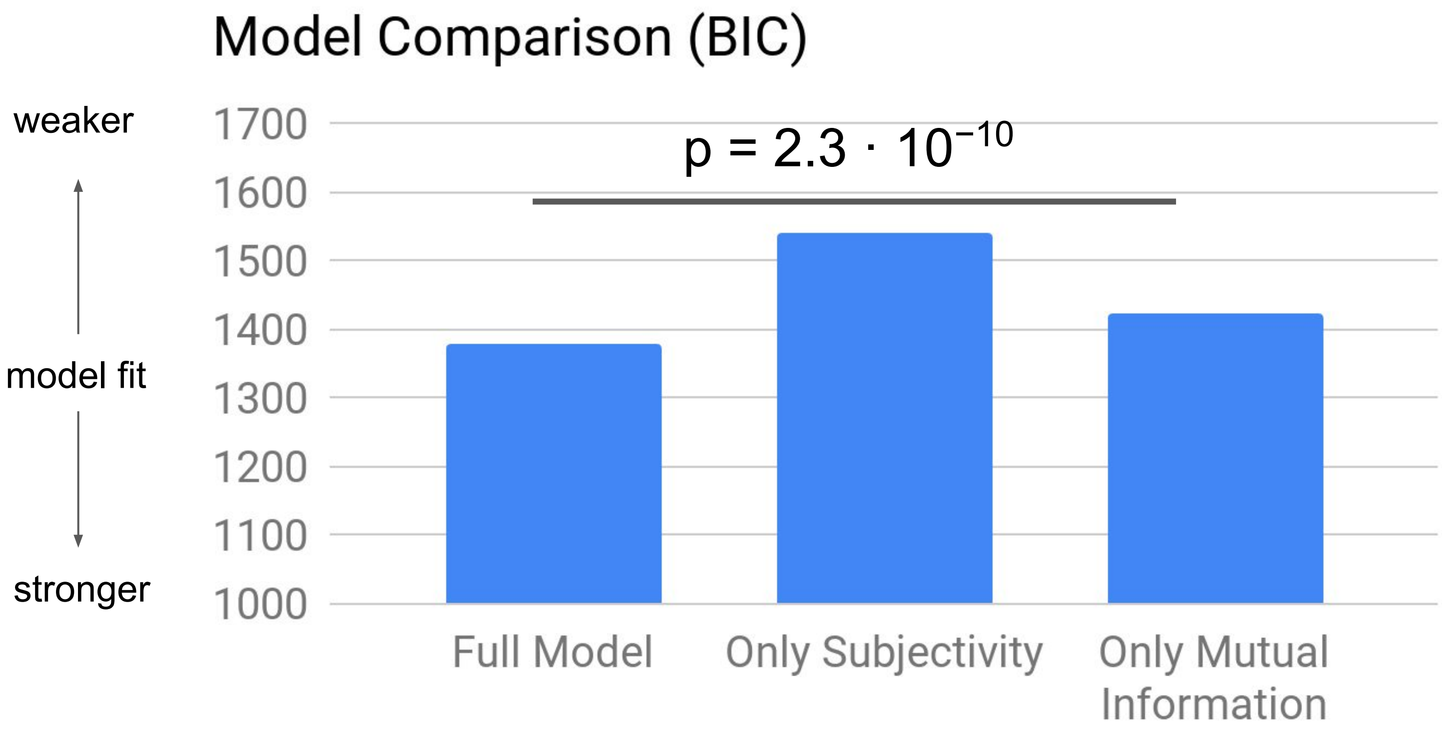
1. $PMI(A_1, N) - PMI(A_2, N)$
2. $Subj(A_1) - Subj(A_2)$



Subjectivity and Mutual Information independently impact ordering.



Subjectivity and Mutual Information independently impact ordering.



Subjectivity and Mutual Information independently impact ordering.

Mutual Information predicts **Noun-Specific** Effects:

new good luck

PMI	-3.1	4.1
Subjectivity	0.5	0.8

international young people

PMI	-3.0	3.9
Subjectivity	0.26	0.64

open curly braces

PMI	2.5	9.5
Subjectivity	0.38	0.40

Mutual Information predicts **Noun-Specific** Effects:

new good luck

PMI	-3.1	4.1
Subjectivity	0.5	0.8

international young people

PMI	-3.0	3.9
Subjectivity	0.26	0.64

open curly braces

PMI	2.5	9.5
Subjectivity	0.38	0.40

Mutual Information predicts **Noun-Specific** Effects:

new good luck

PMI	-3.1	4.1
Subjectivity	0.5	0.8

international young people

PMI	-3.0	3.9
Subjectivity	0.26	0.64

open curly braces

PMI	2.5	9.5
Subjectivity	0.38	0.40

Mutual Information predicts **Noun-Specific** Effects:

new good luck

PMI	-3.1	4.1
Subjectivity	0.5	0.8

international young people

PMI	-3.0	3.9
Subjectivity	0.26	0.64

open curly braces

PMI	2.5	9.5
Subjectivity	0.38	0.40

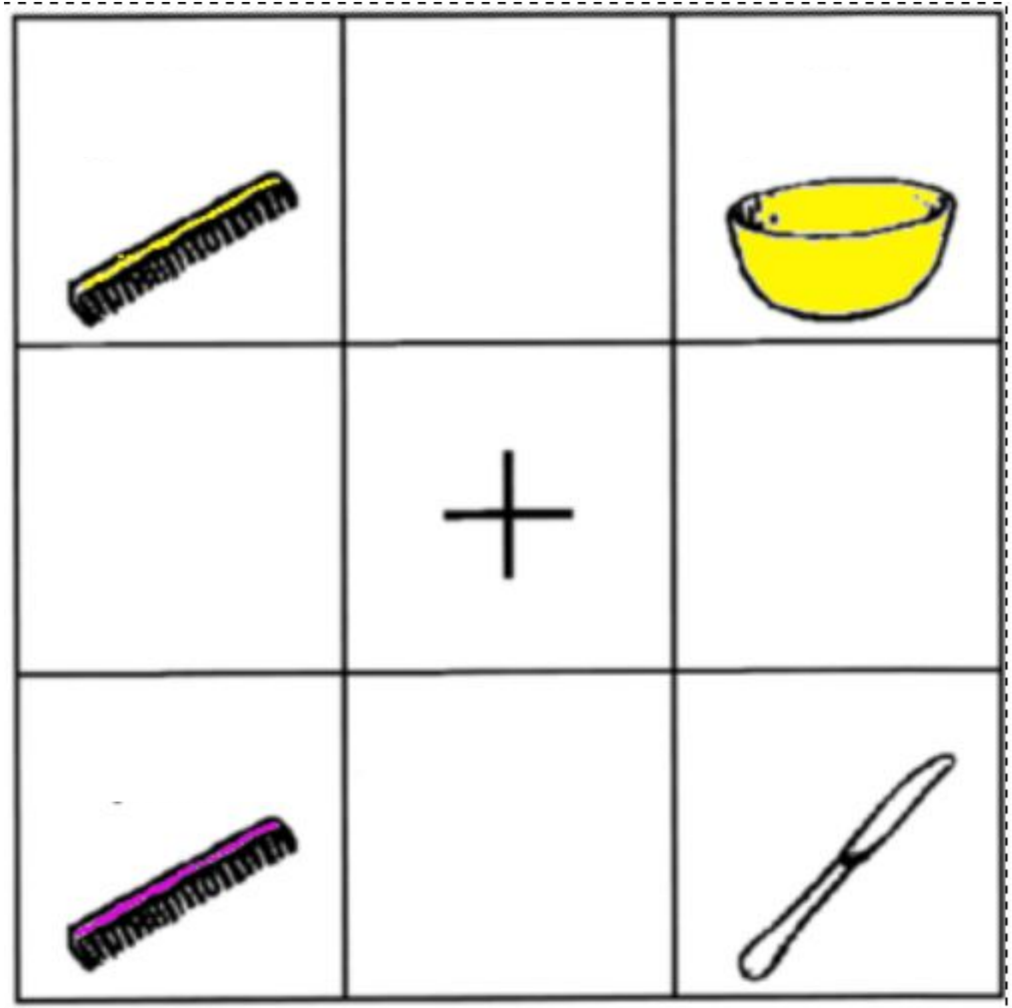
Goal:

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

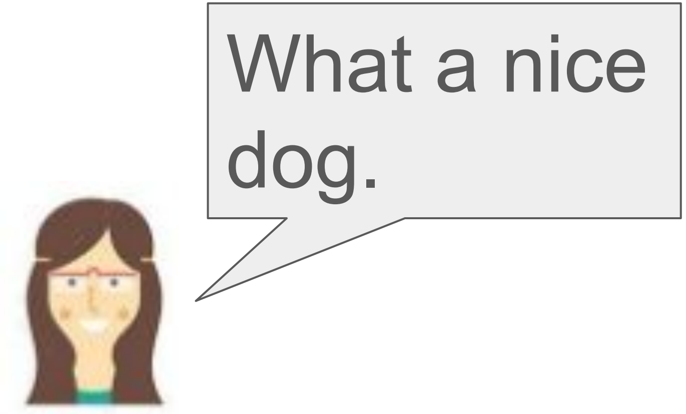
Adjectives can help **pick out** referents.

Adjectives can **describe** and **comment on** a referent.

The Use of Adjectives

Adjectives can help **pick out** referents.

Adjectives can **describe** and **comment on** a referent.



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!

from COCA (Davies, 2017)

Forrest looks at the massive crowd.

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

We loo

The to

Abrupt

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

Telling the red blood to stop flowing.

Look at the little boy!

from COCA (Davies, 2017)

Modeling Approach

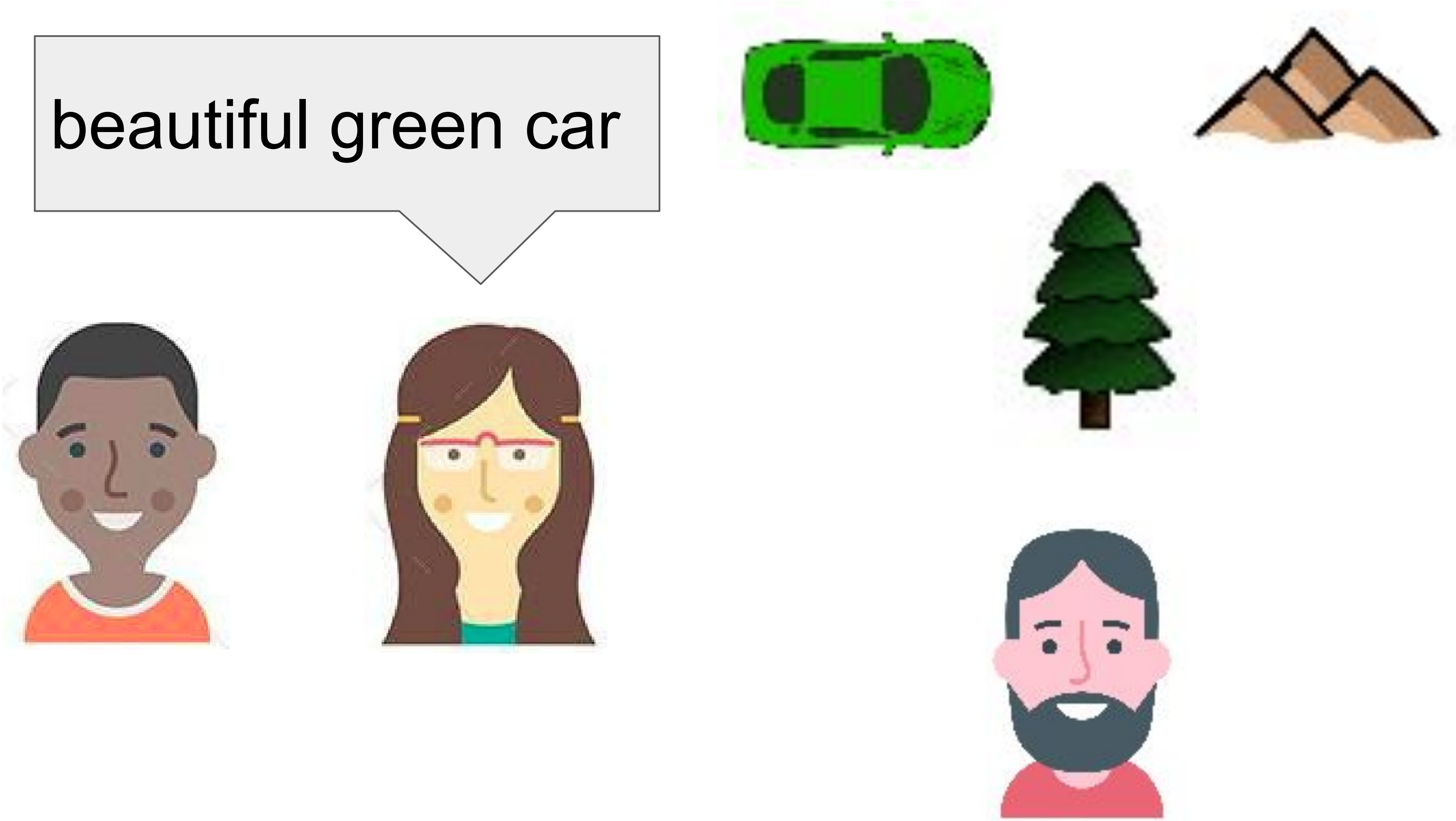
1. Formalize nonrestrictive use of adjectives
2. Define a rational Bayesian model of communication
3. Show how memory limitations lead effects of subjectivity and mutual information
4. Evaluate on Corpus Data

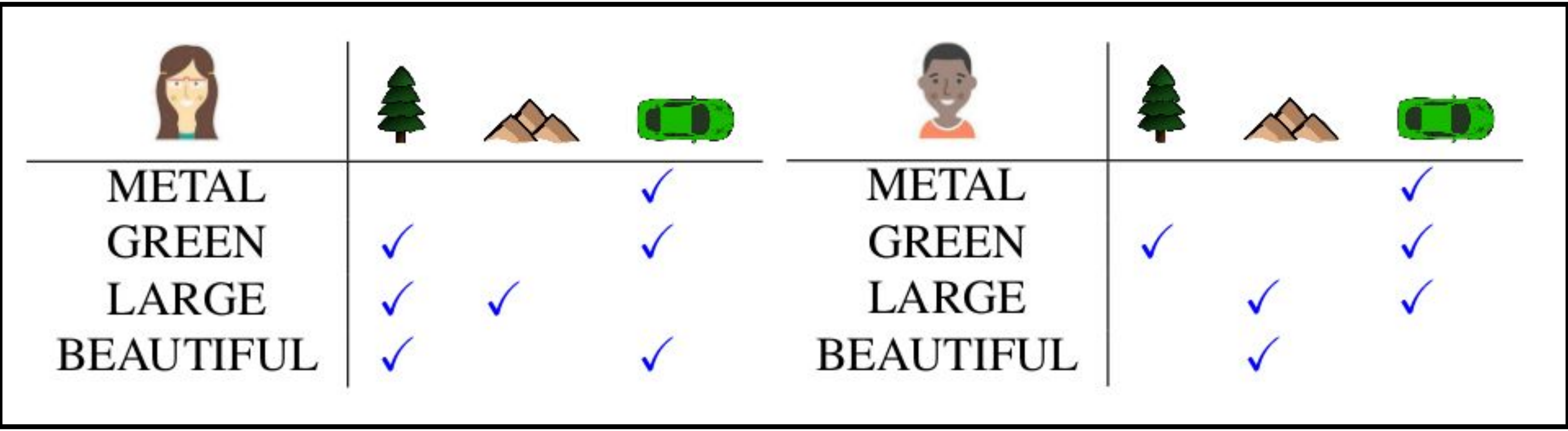
Modeling Approach

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

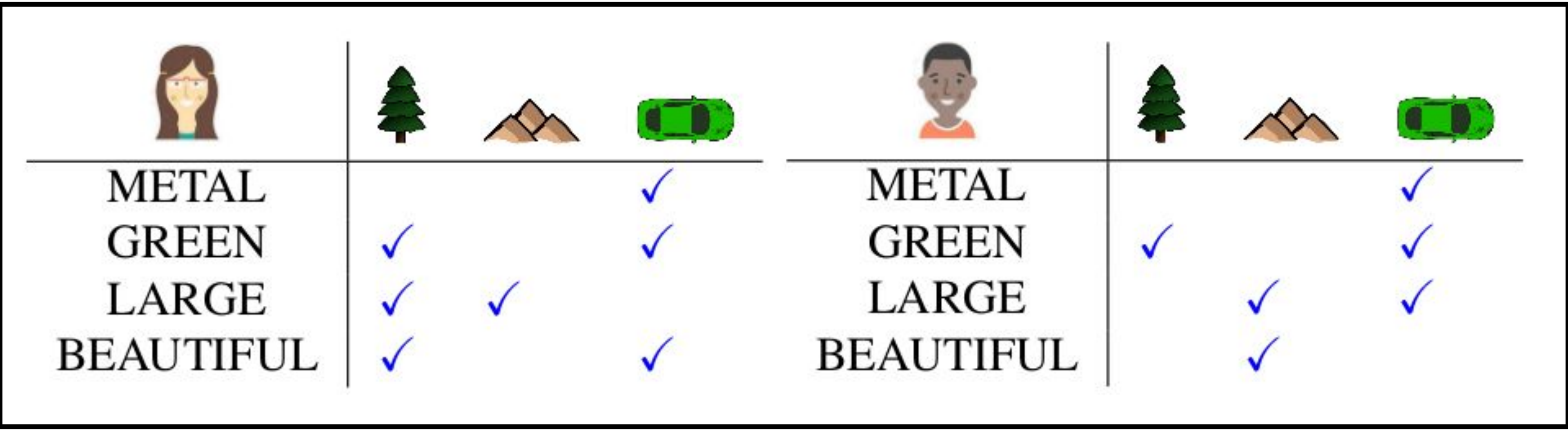




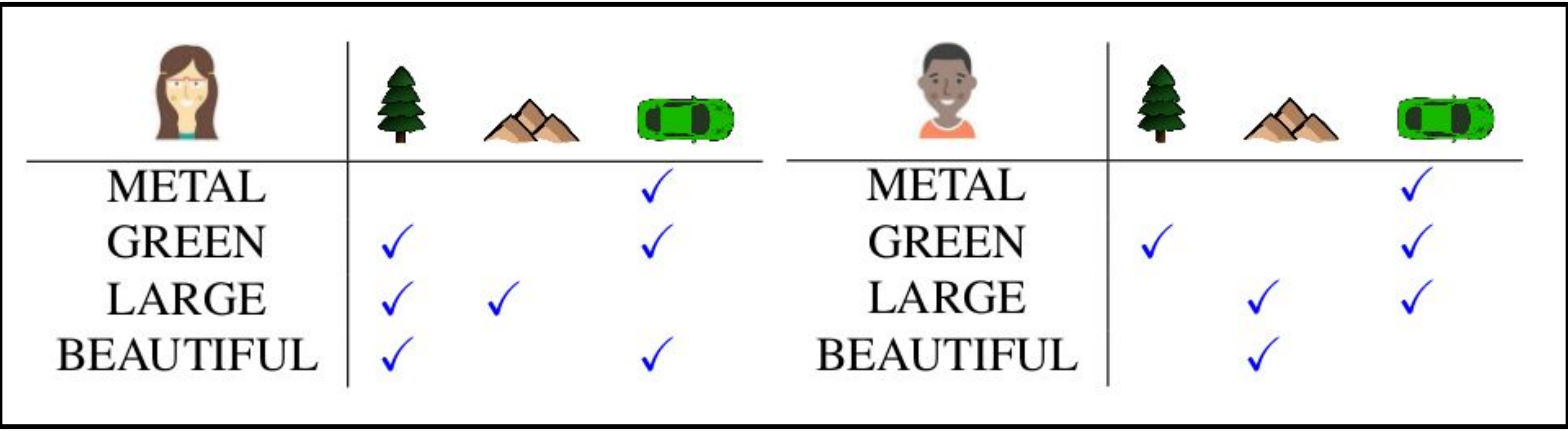


beautiful green car



World state = **Truth value assignment** to the cells in this table



Speakers mostly **agree** on **objective** judgments

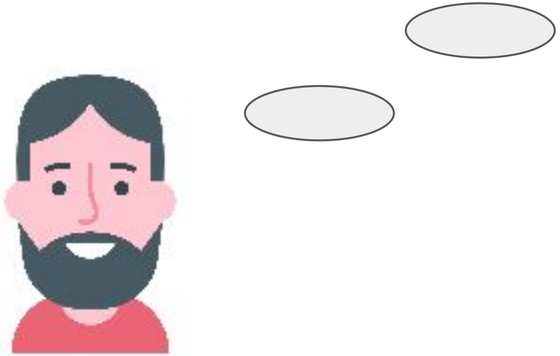


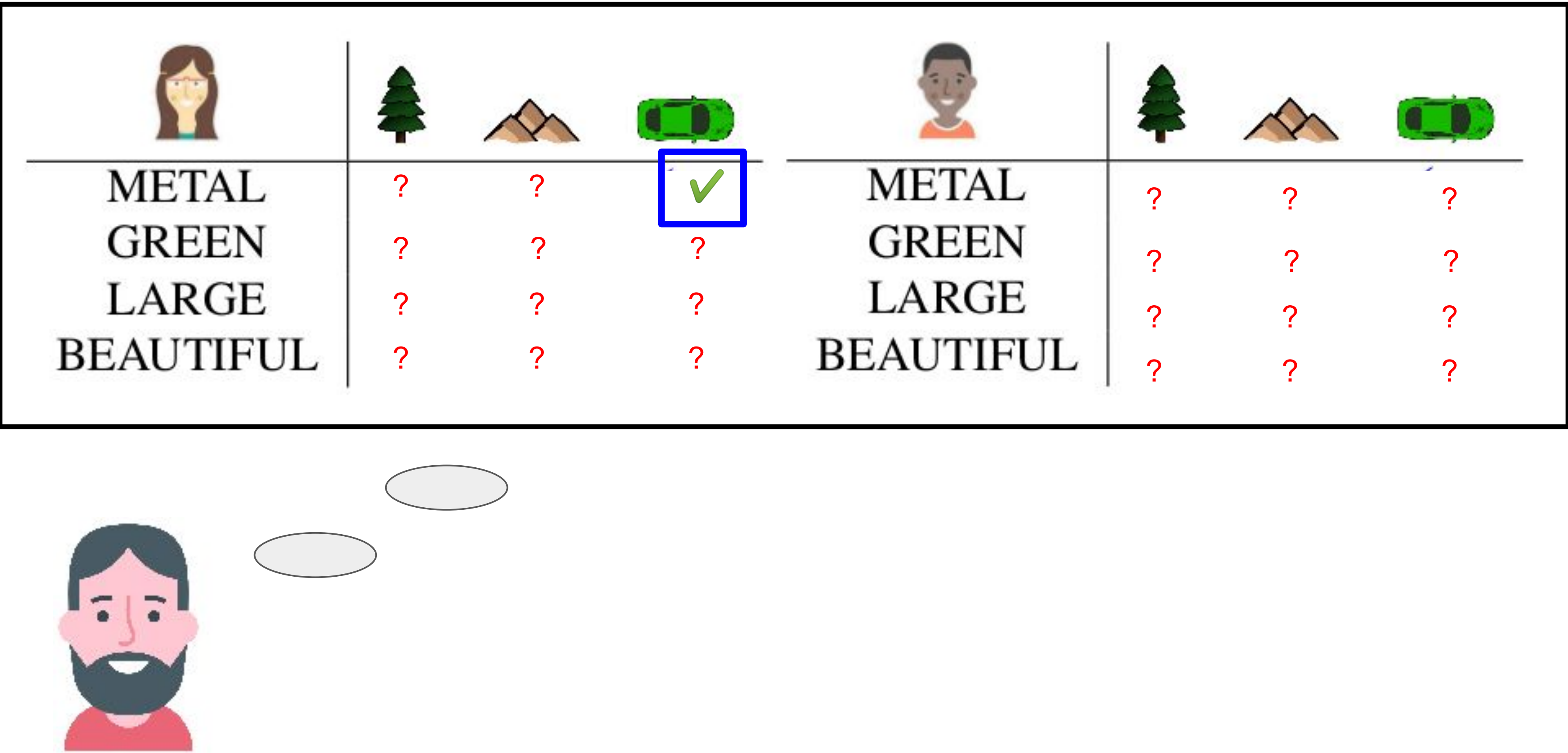
Speakers mostly **agree** on **objective** judgments

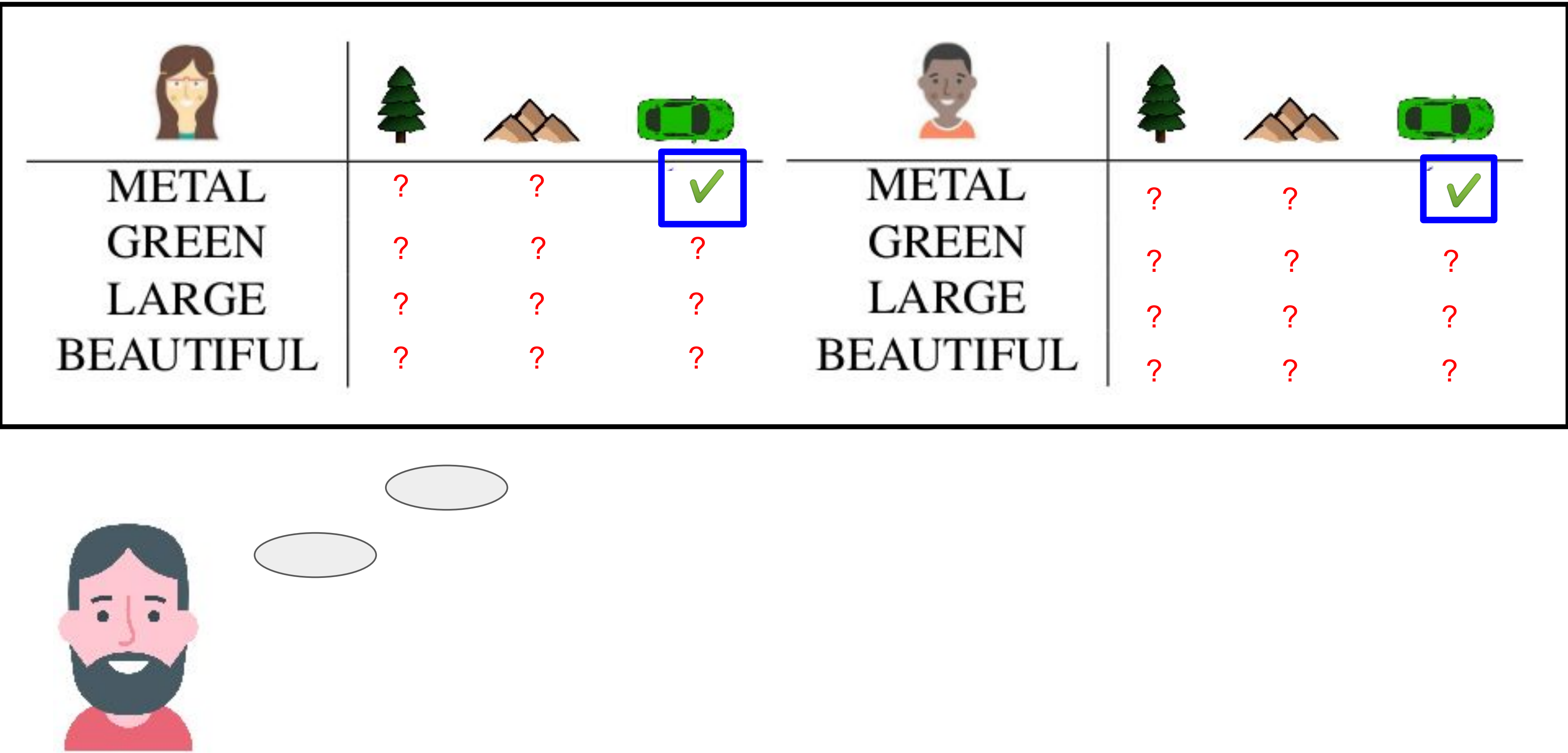
More **disagreement** for more **subjective** judgments

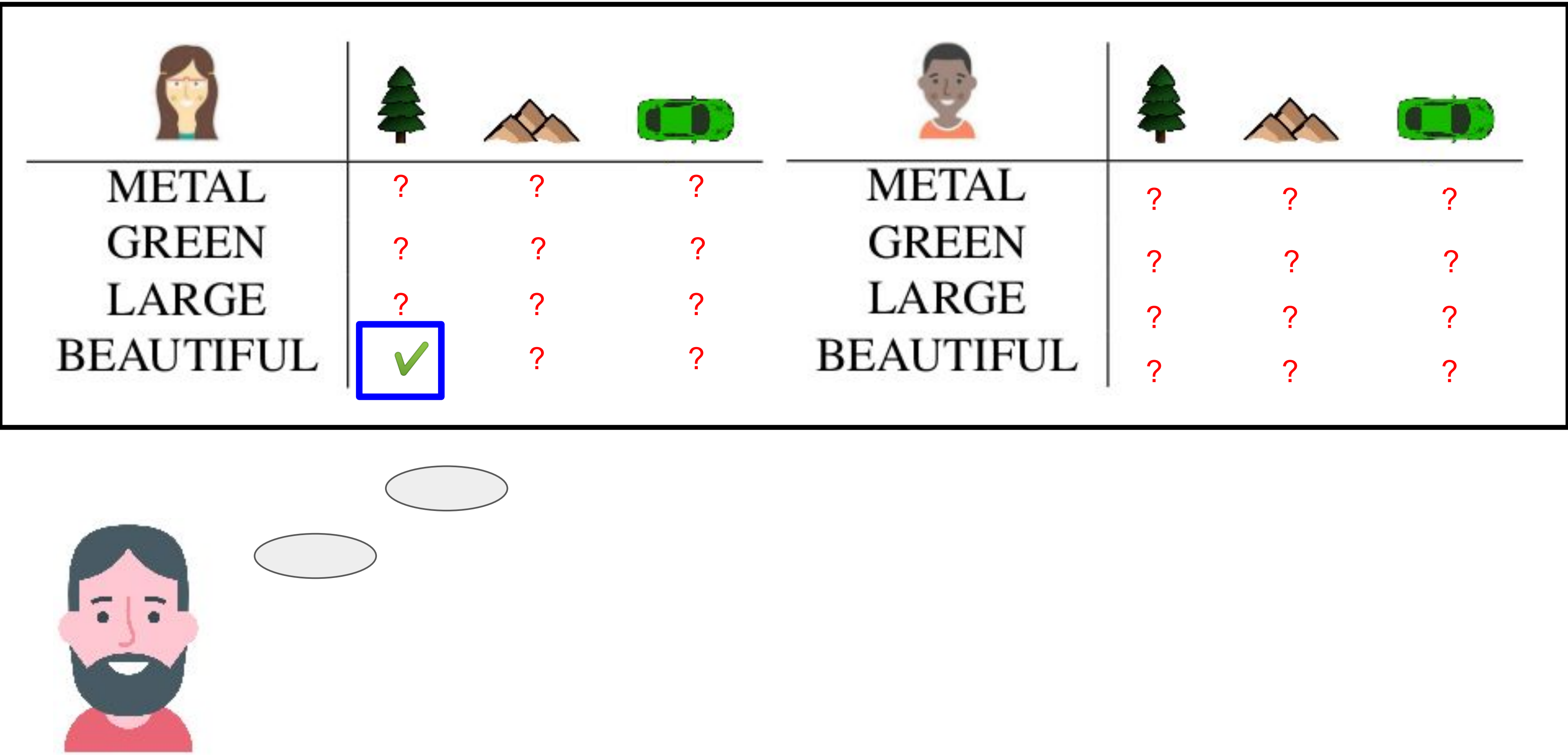
METAL	?	?	?
GREEN	?	?	?
LARGE	?	?	?
BEAUTIFUL	?	?	?

METAL	?	?	?
GREEN	?	?	?
LARGE	?	?	?
BEAUTIFUL	?	?	?

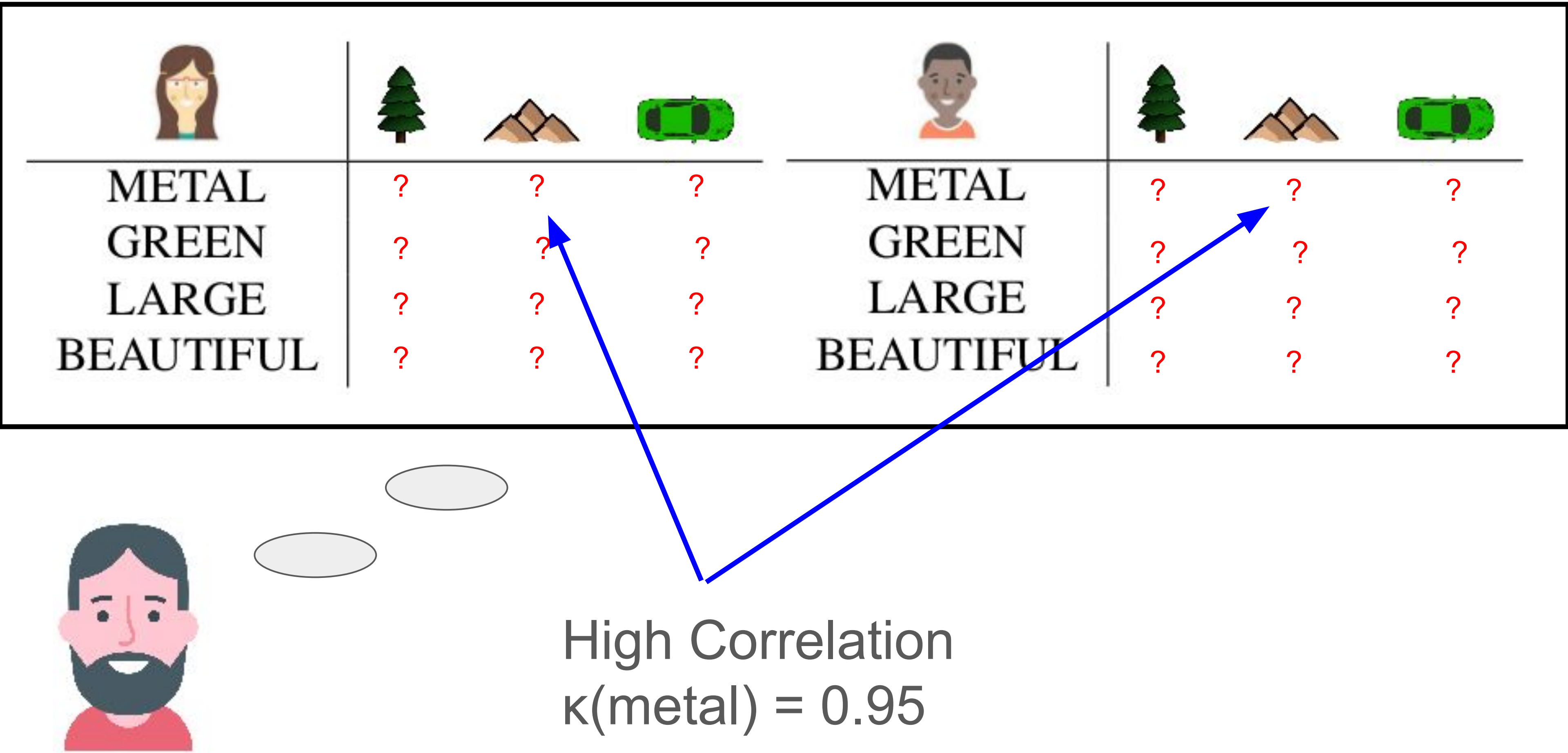



METAL
GREEN
LARGE
BEAUTIFUL
METAL
GREEN
LARGE
BEAUTIFUL

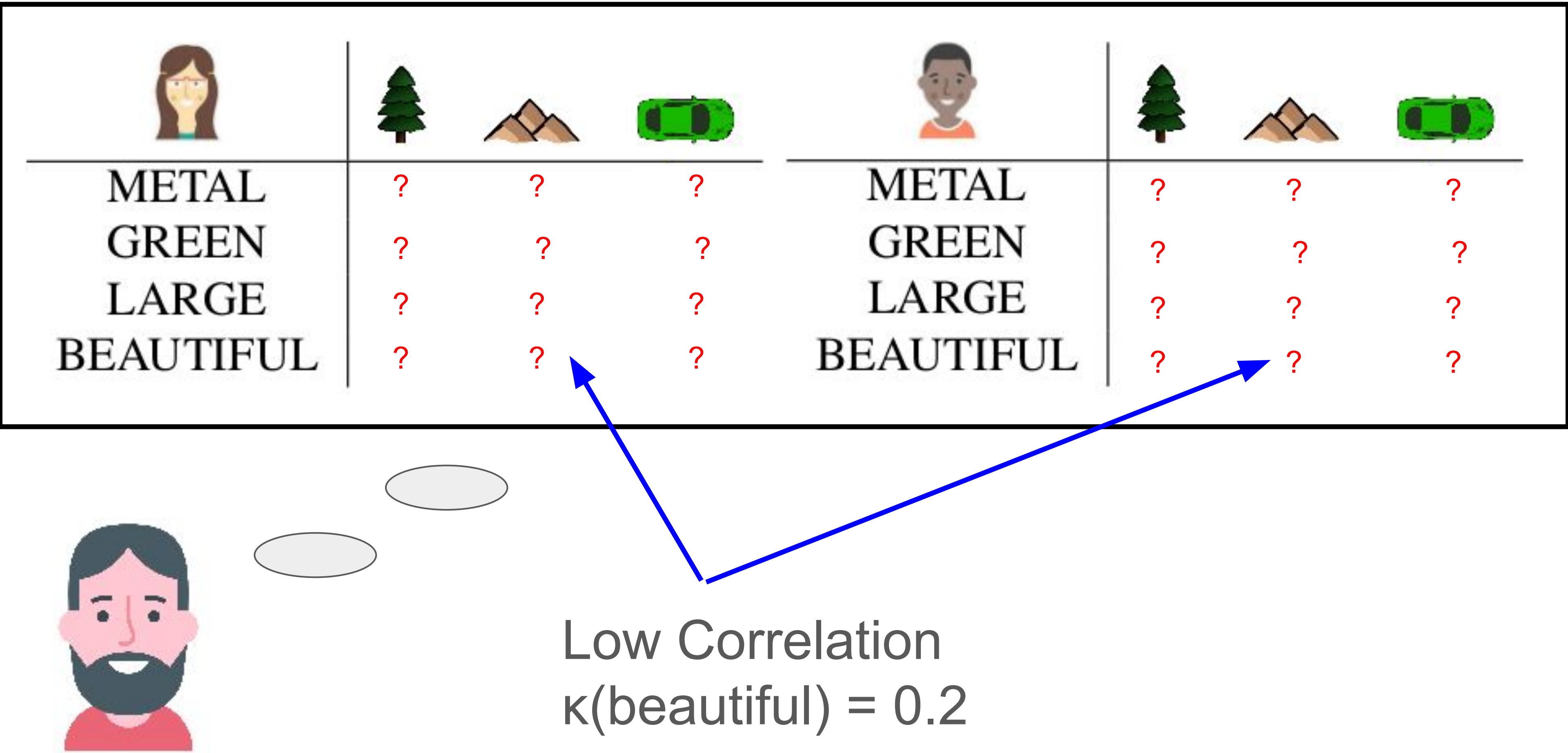

METAL
GREEN
LARGE
BEAUTIFUL
METAL
GREEN
LARGE
BEAUTIFUL

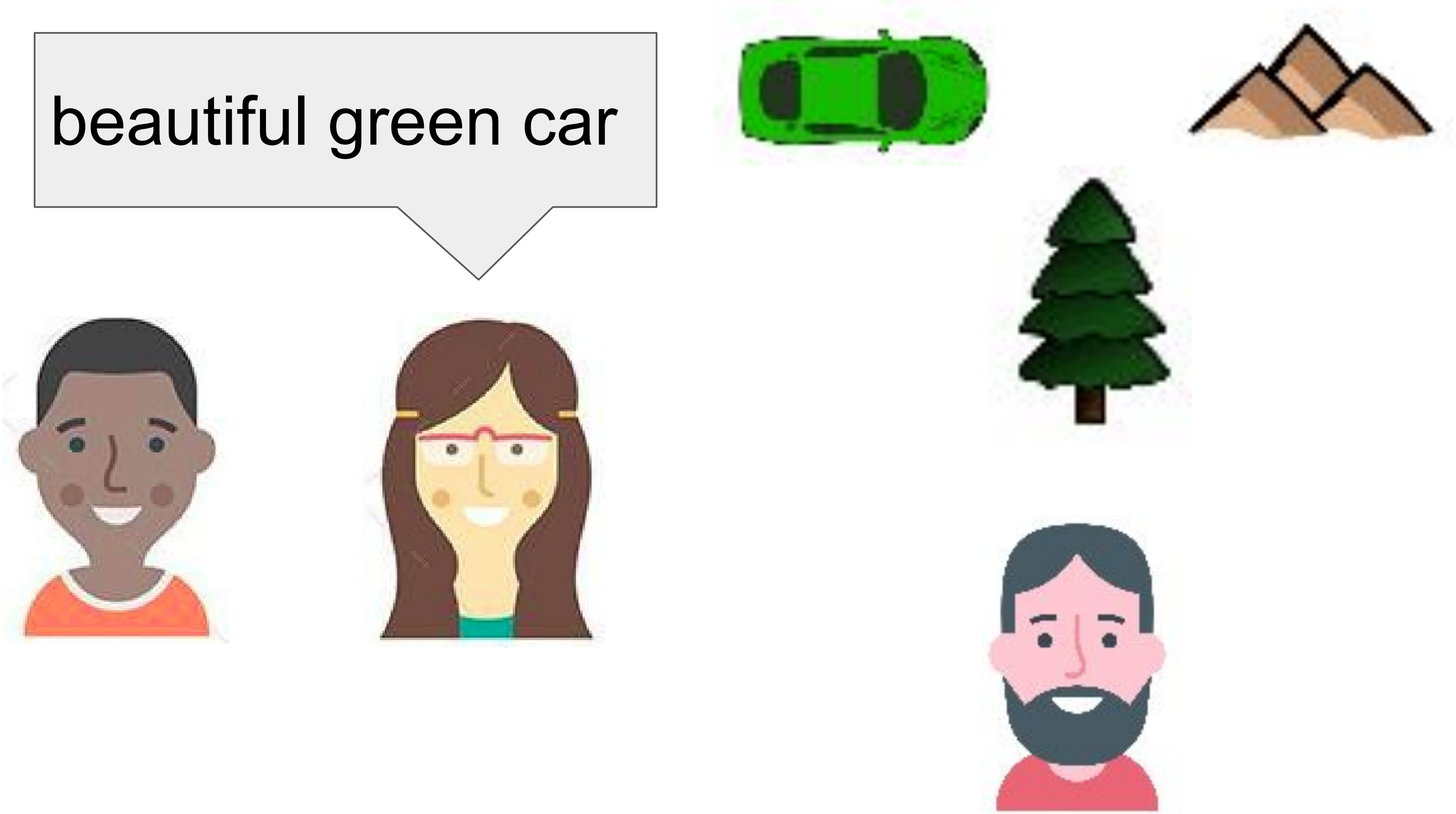

METAL ? ? ?
GREEN ? ? ?
LARGE ? ? ?
BEAUTIFUL ✓ ? ?
METAL ? ? ?
GREEN ? ? ?
LARGE ? ? ?
BEAUTIFUL ? ? ?

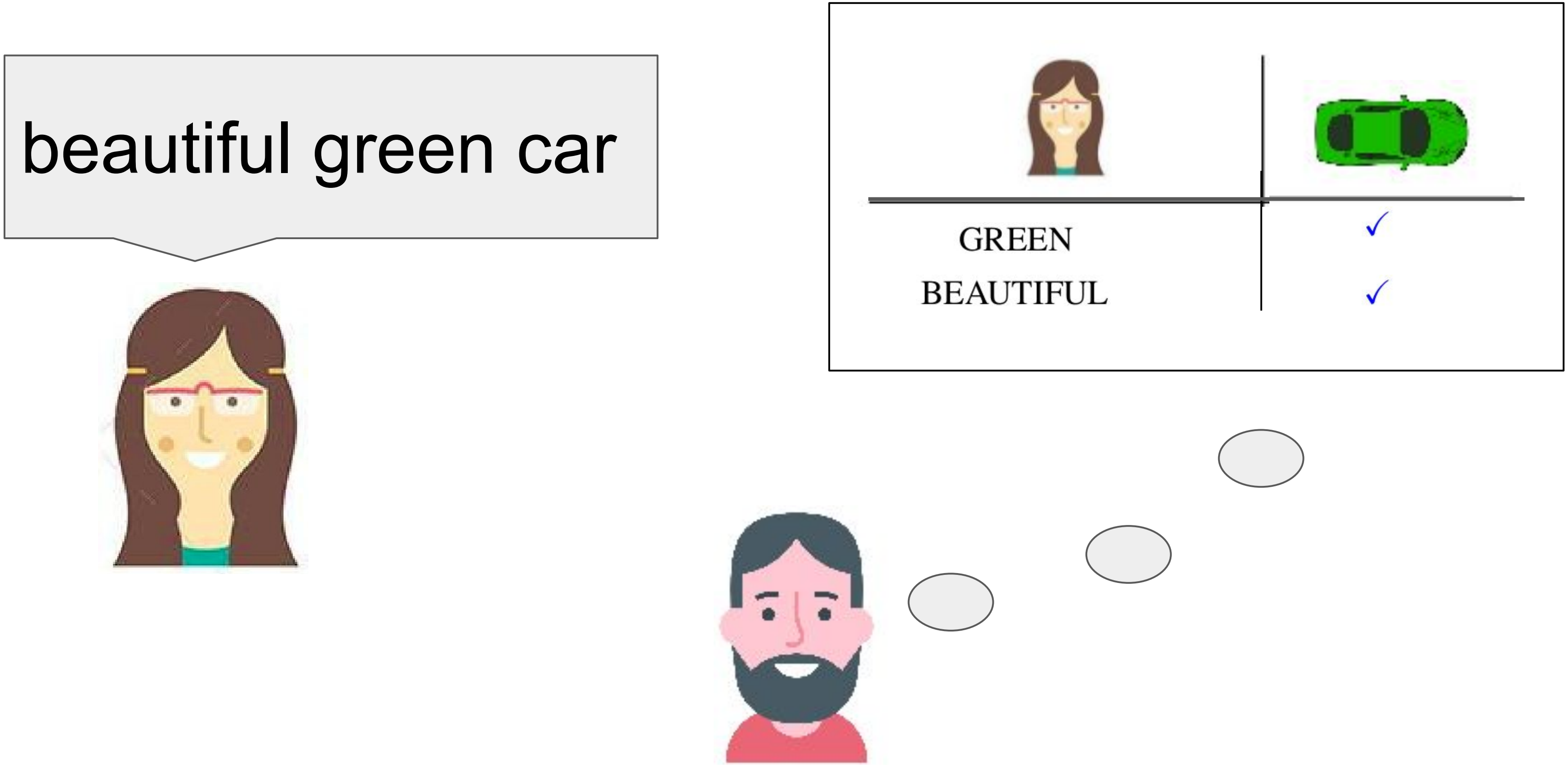

METAL
GREEN
LARGE
BEAUTIFUL
? ? ?
? ? ?
? ? ?
? ? ?
METAL
GREEN
LARGE
BEAUTIFUL
? ? ?
? ? ?
? ? ?
? ? ?

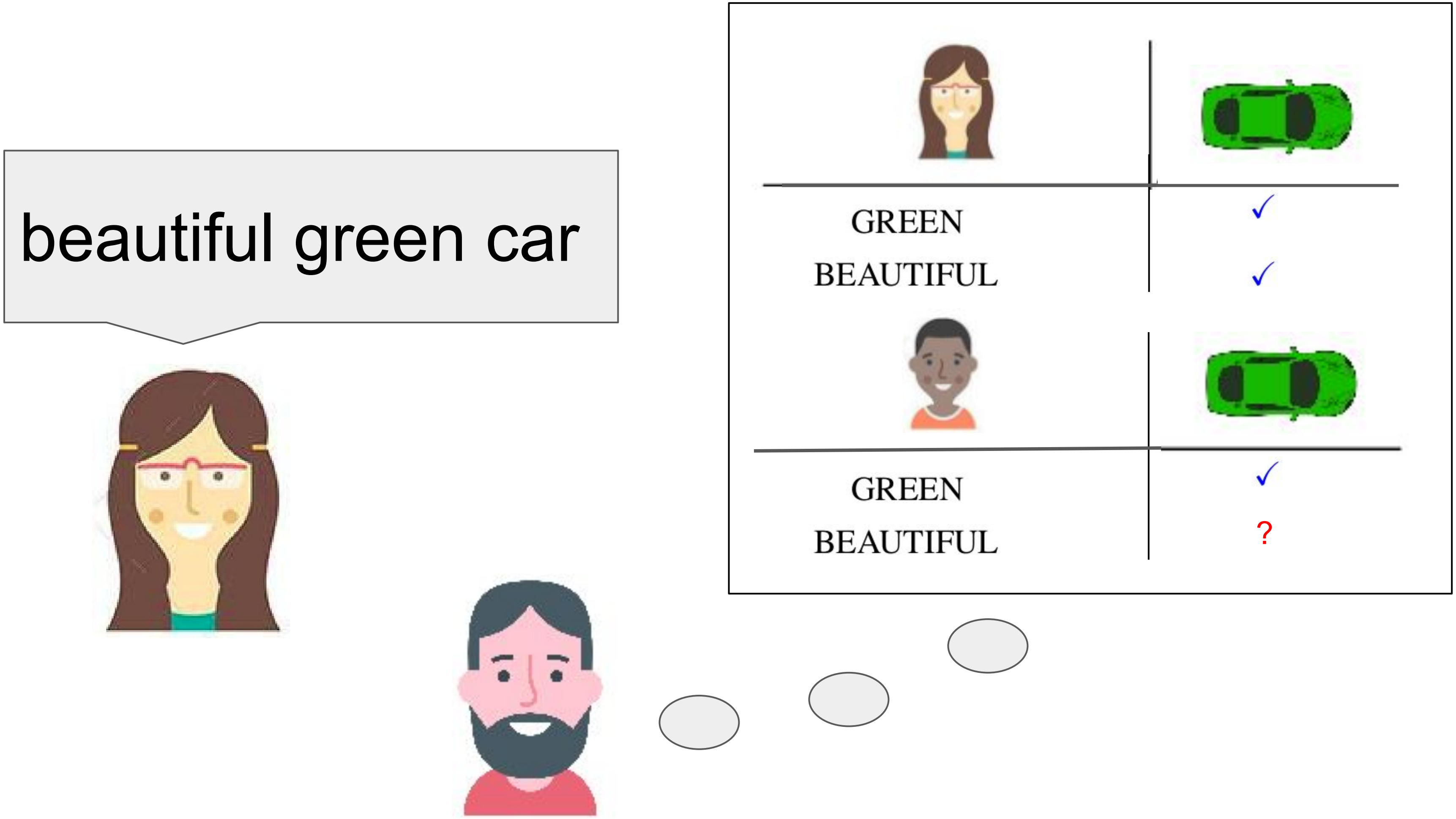


METAL
GREEN
LARGE
BEAUTIFUL
METAL
GREEN
LARGE
BEAUTIFUL
High Correlation
κ(metal) = 0.95


METAL ? ? ?
GREEN ? ? ?
LARGE ? ? ?
BEAUTIFUL ? ? ?
METAL ? ? ?
GREEN ? ? ?
LARGE ? ? ?
BEAUTIFUL ? ? ?
Low Correlation
κ(beautiful) = 0.2


beautiful green car


beautiful green car
GREEN ✓
BEAUTIFUL ✓


beautiful green car
GREEN
BEAUTIFUL
GREEN
BEAUTIFUL
?

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)

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

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

Listener Model

Listener performs Bayesian reasoning to **infer world state**.

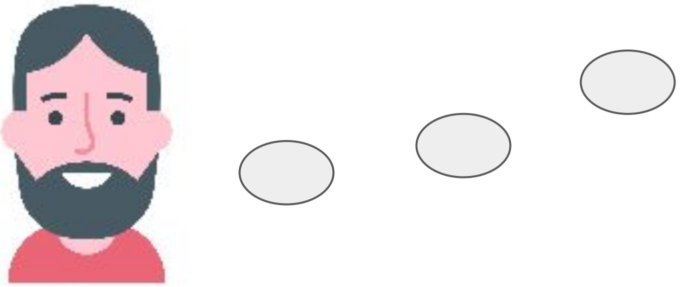
$$P_{\text{listener}}(w|u) \propto P_{\text{prior}}(w)\ \delta_{\text{u is true for speaker in w}}$$

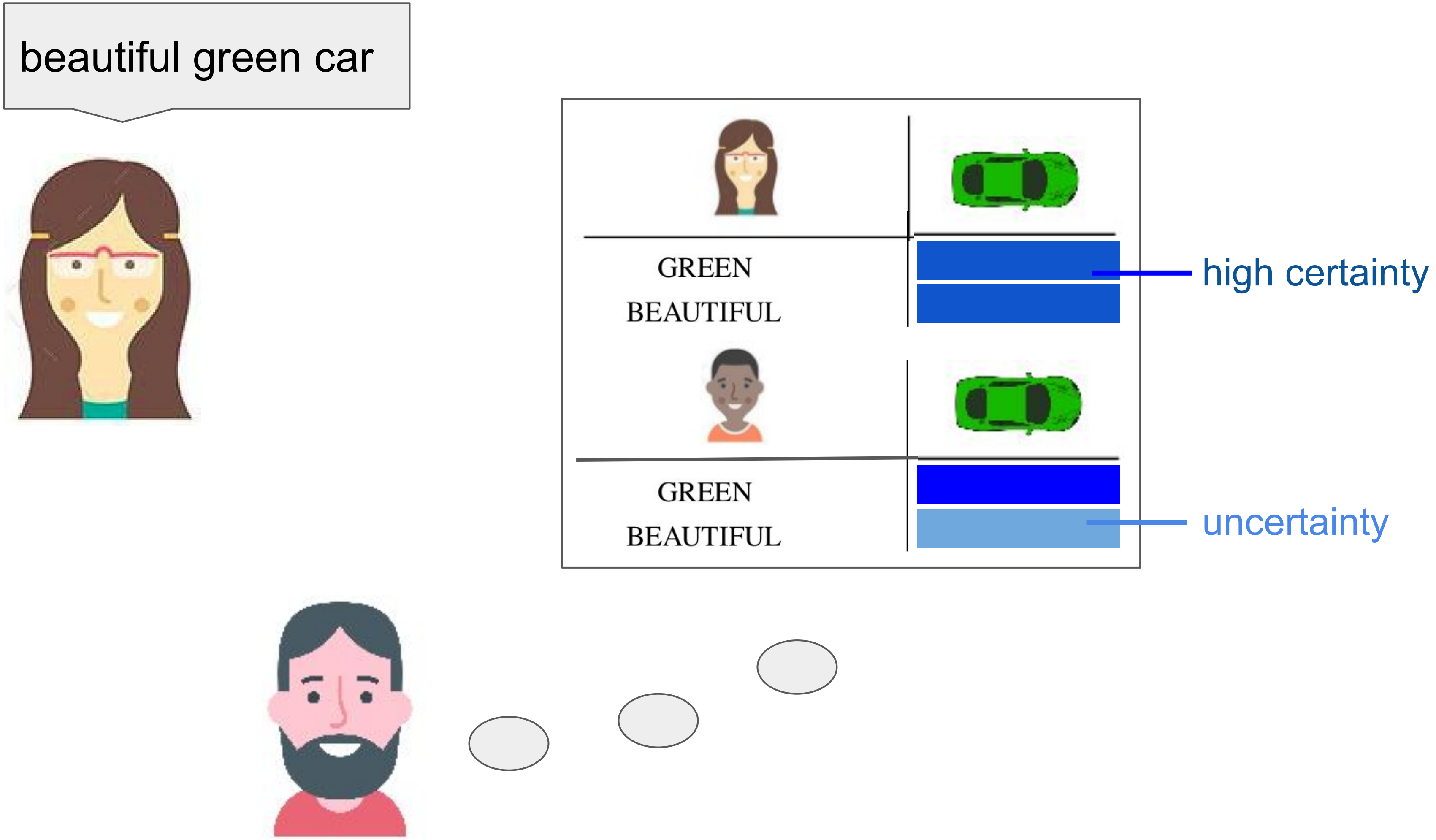
state of
the world

utterance
received

	
GREEN	✓
BEAUTIFUL	✓
	
GREEN	✓
BEAUTIFUL	??

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




beautiful green car
GREEN
BEAUTIFUL
GREEN
BEAUTIFUL
high certainty
uncertainty

Rational Communication: Speakers and Listeners

Formalize model in the framework of **Bayesian pragmatics**

(Franke 2008; Frank and Goodman, 2012)

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

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

Speaker Model

Speaker chooses utterance to optimize utility

(Franke 2008; Frank and Goodman 2012).

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

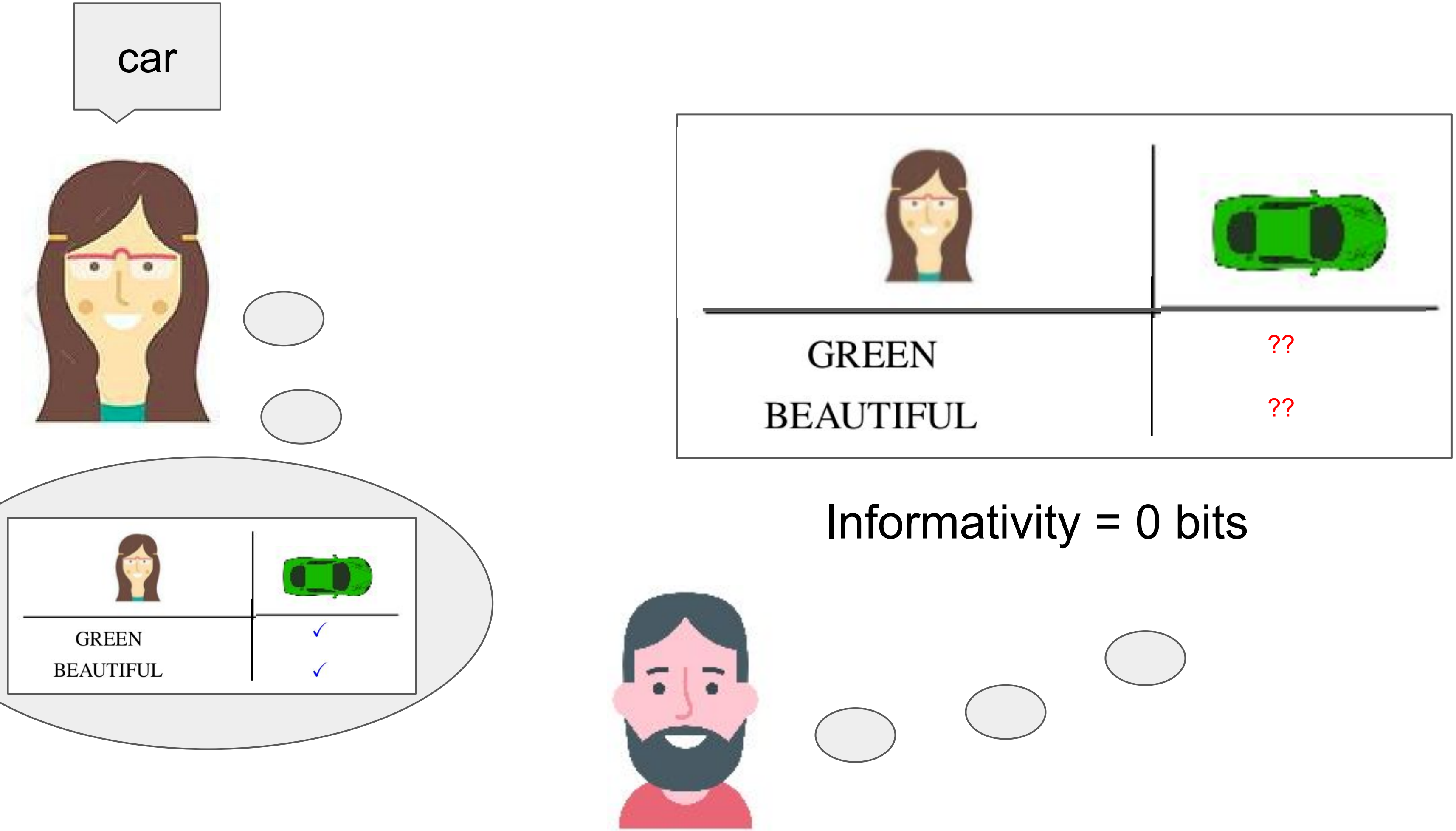
Informativity of utterance `u’

Cost of utterance

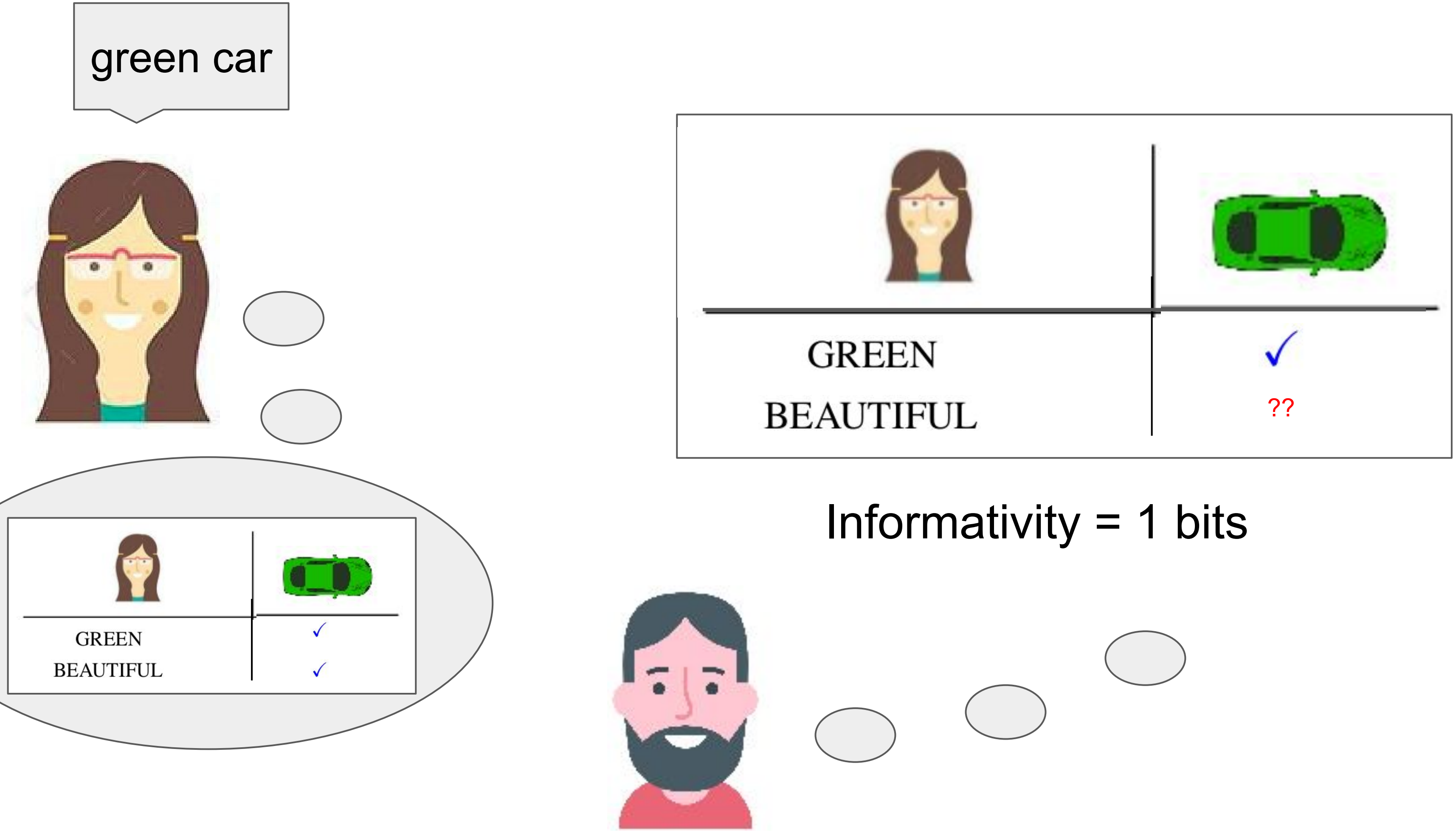
Speaker Model

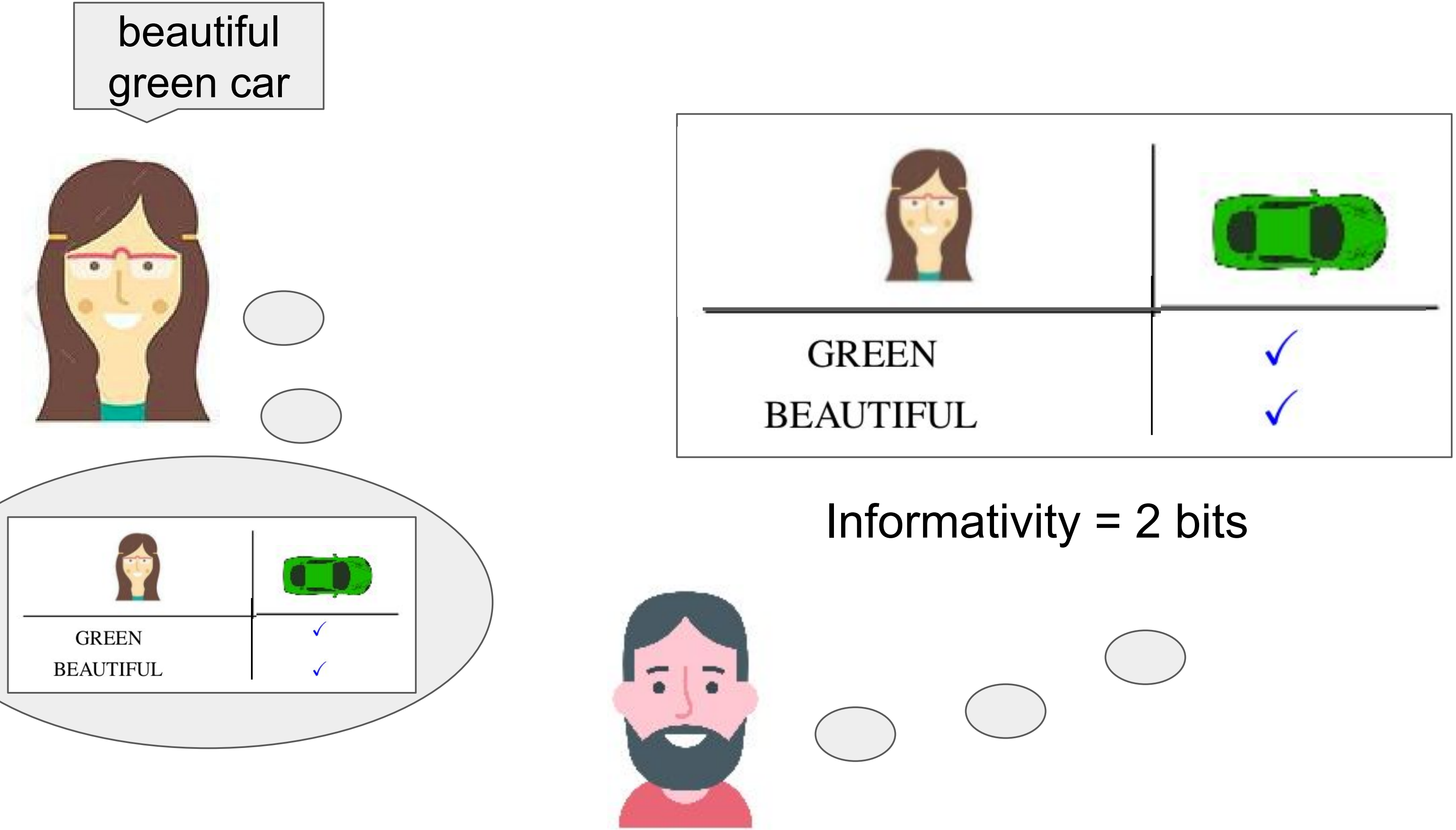
$$P_{speaker}(u) \propto \exp(\alpha \cdot \mathbf{I(u)} - \beta \cdot C(u))$$

Typically: Reduction in the listener's uncertainty about the world state, measured in bits (e.g., Frank and Goodman, 2012; Goodman and Stuhlmueller, 2013).



car
GREEN
BEAUTIFUL
✓
✓
GREEN
BEAUTIFUL
??
??
Informativity = 0 bits


green car
GREEN
BEAUTIFUL
✓
✓
GREEN
BEAUTIFUL
✓
??
Informativity = 1 bits


beautiful green car
GREEN
BEAUTIFUL
GREEN
BEAUTIFUL
Informativity = 2 bits

Speaker Model

Speaker chooses utterance to optimize utility

(Franke 2008; Frank and Goodman 2012).

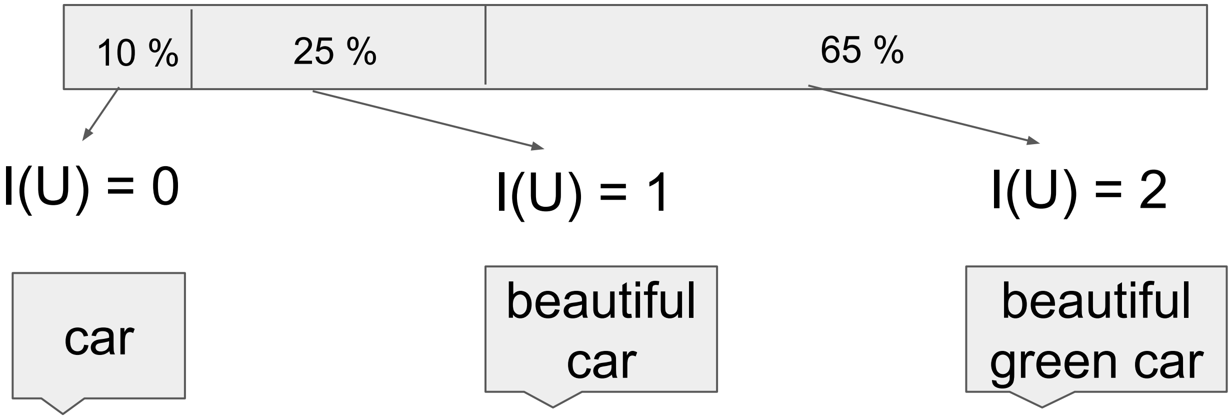
$$P_{speaker}(u) \propto \exp(\alpha \cdot \mathbf{I(u)} - \beta \cdot \mathbf{C(u)})$$

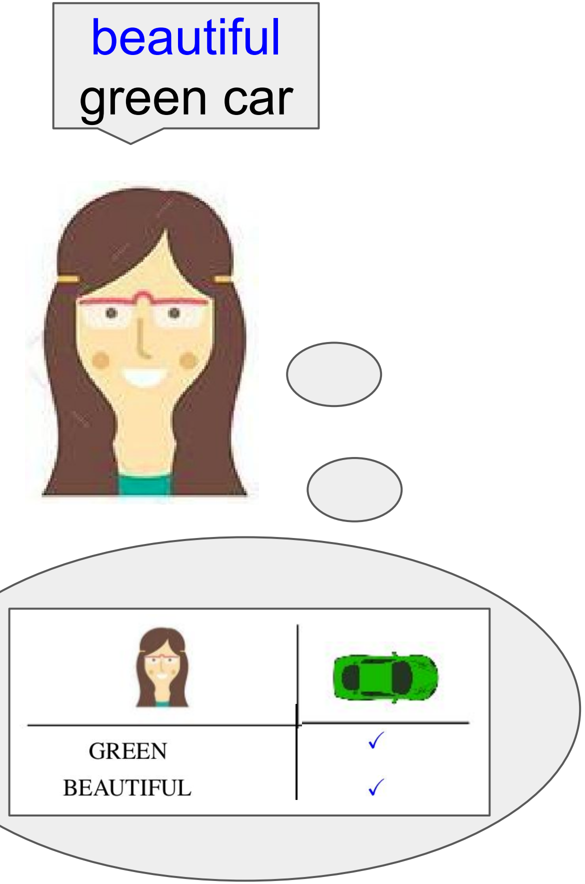
Informativity of

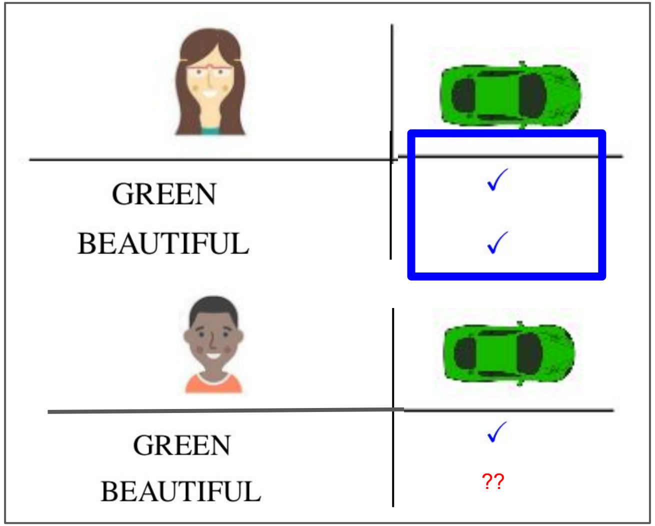
utterance `u’

Speaker Model

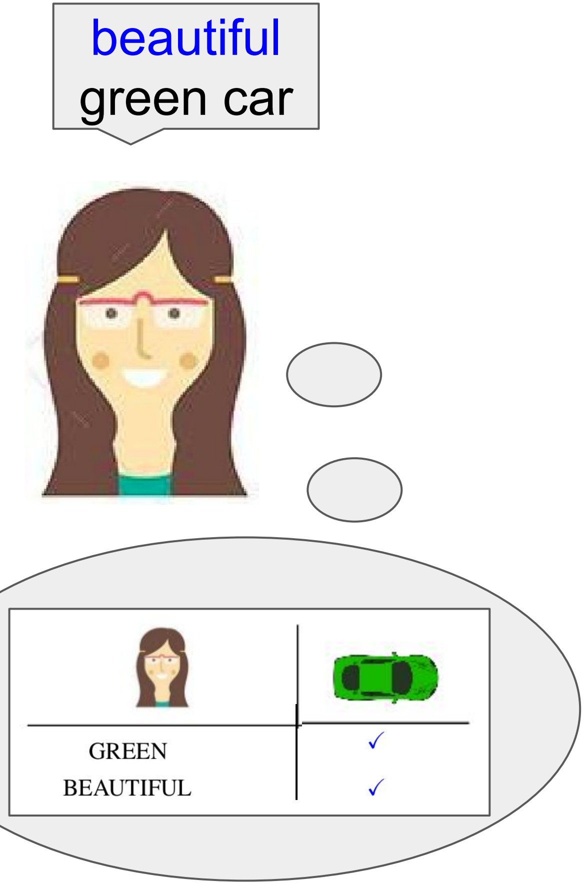
$$P_{speaker}(u) \propto \exp(\alpha \cdot \mathbf{I(u)})$$

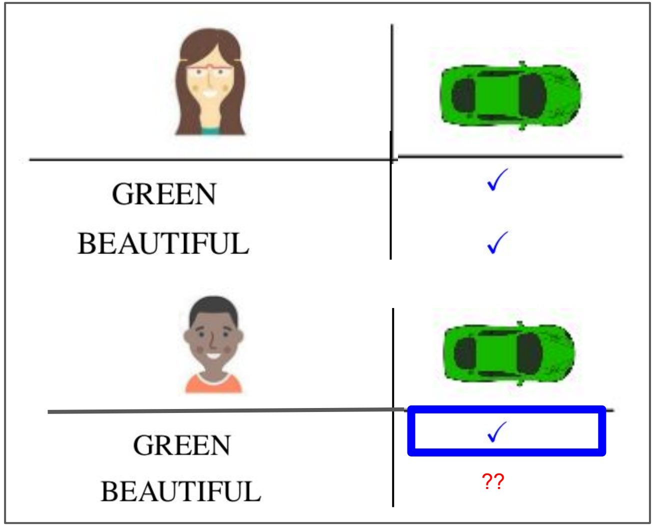






Informativity about  = 2 bits



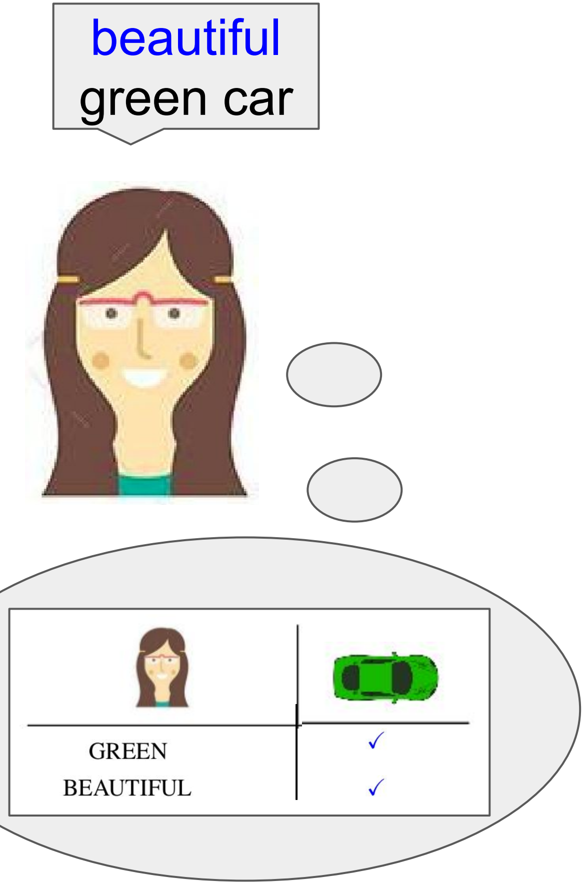


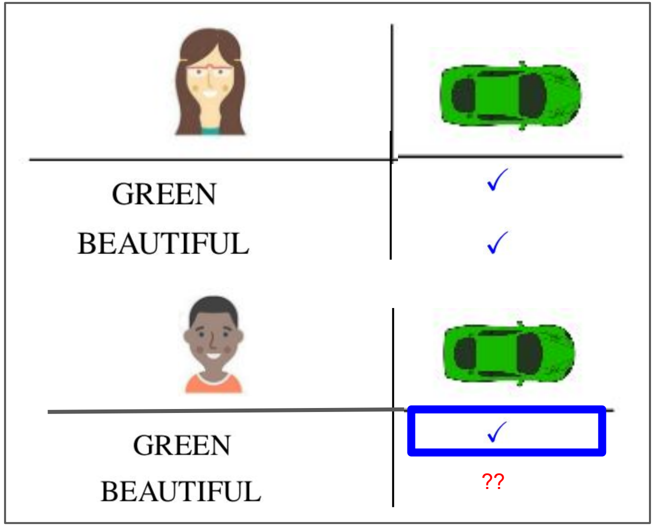
Informativity about  = 2 bits

+

Informativity about  = 1 bits

= I(U)





Informativity about  = 2 bits

+

Informativity about  = 1 bits

= I(U)

Cooperative speakers communicate knowledge that generalizes to other people.

Speaker Model: Cost

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

Cost of the utterance

Speaker Model: Cost

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

$$\mathbf{C(u) = -\log P(u)}$$

Surprisal of the utterance

(cf. Bennett & Goodman, 2018; Peloquin et al 2019)

Speaker Model: Cost

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

$$\mathbf{C(u) = -\log P(u)}$$

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)

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

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

Informativity about 

Informativity about 

Rational Communication: Speakers and Listeners

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

So far, no ordering preferences are predicted!

big green tree

green big tree

Identical Informativity and Cost

Rational Communication: Speakers and Listeners

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

Proposal:

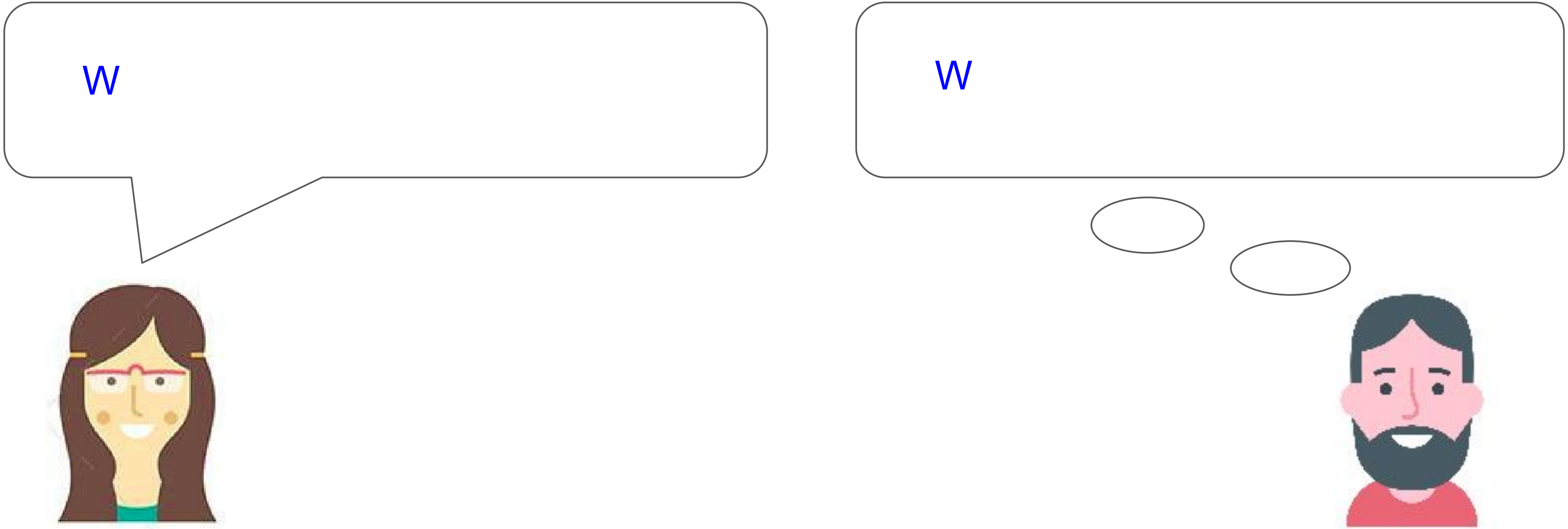
Memory limitations in processing **break symmetry**.

Memory Limitations

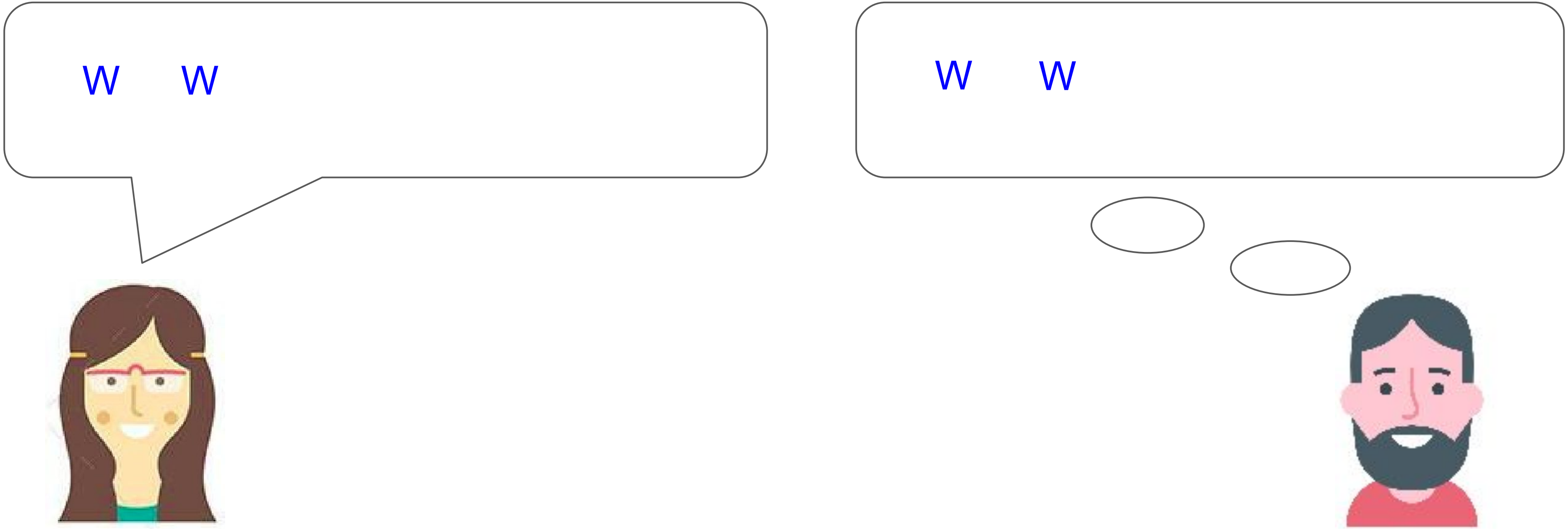
Firmly established as factor in language understanding

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)

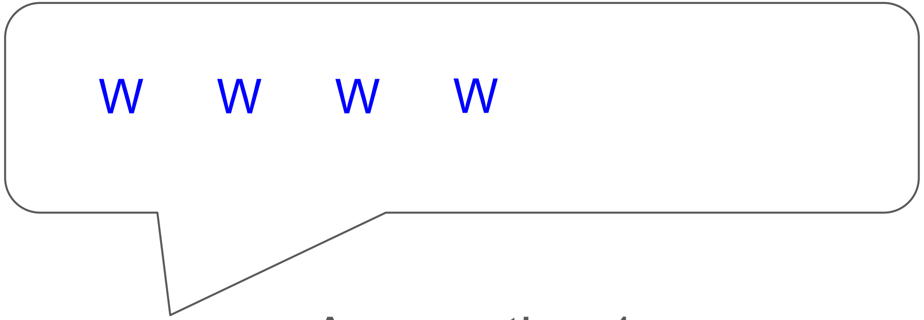


Memory Limitations: Formal Model (Futrell and Levy, 2017)



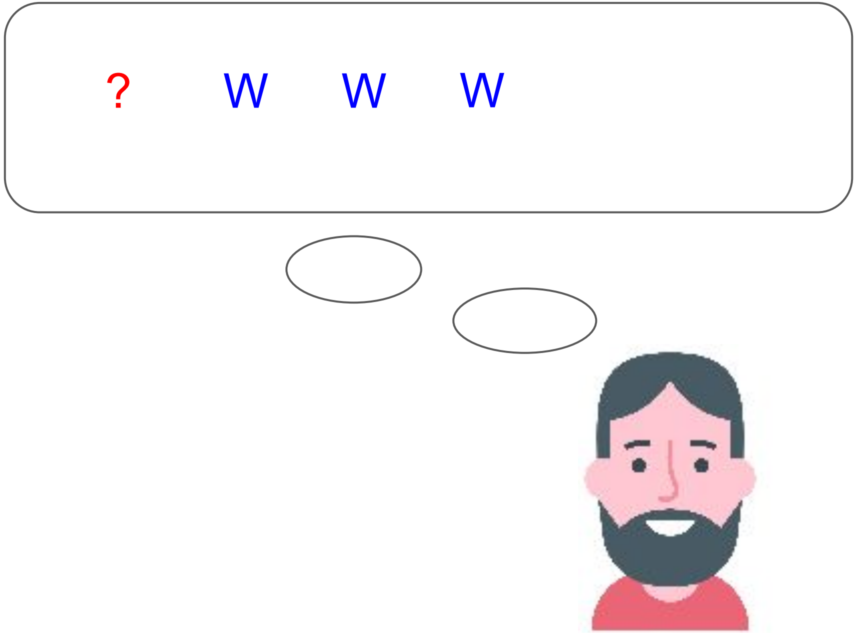
Memory Limitations: Formal Model (Futrell and Levy, 2017)

Memory Limitations: Formal Model (Futrell and Levy, 2017)

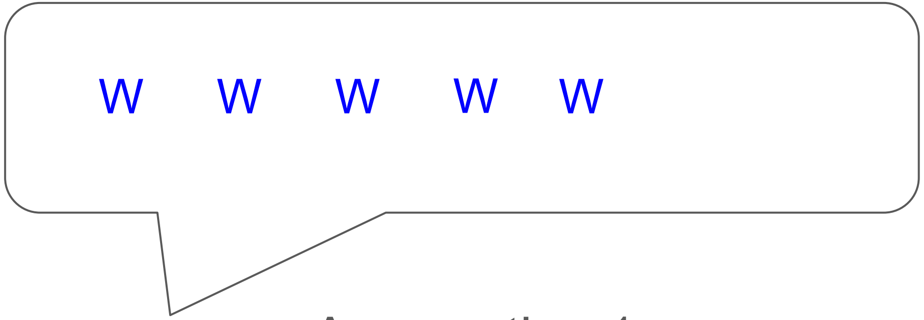


Assumption 1:

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

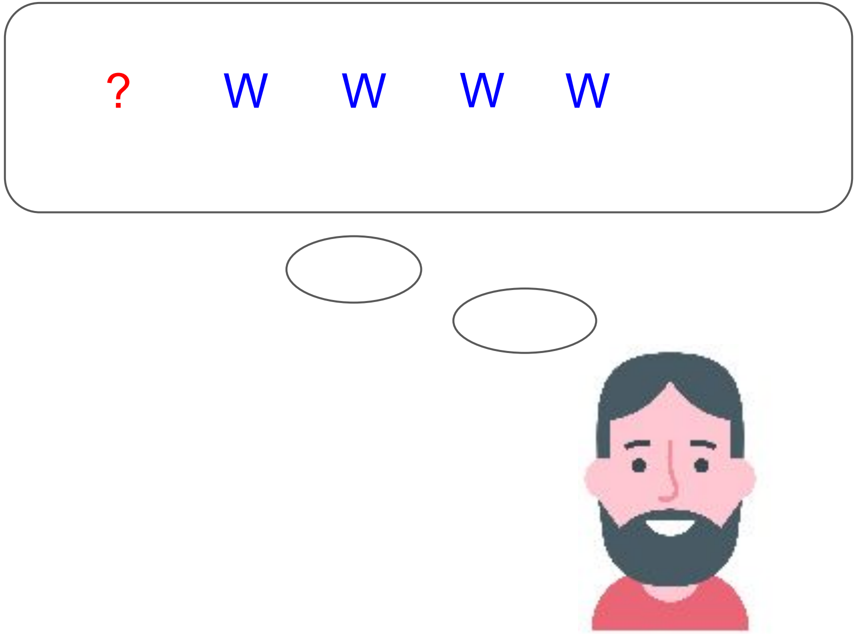


Memory Limitations: Formal Model (Futrell and Levy, 2017)

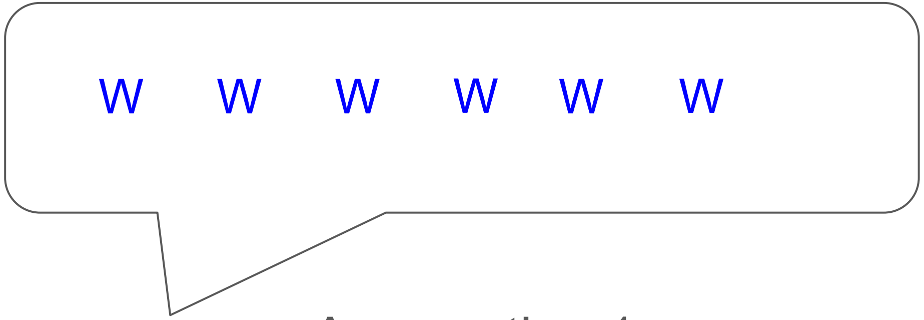


Assumption 1:

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

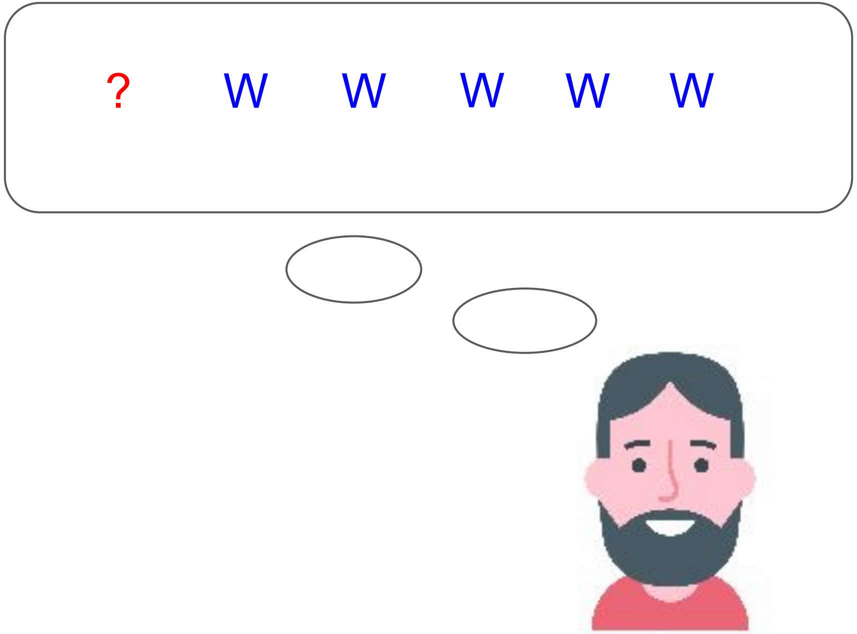


Memory Limitations: Formal Model (Futrell and Levy, 2017)

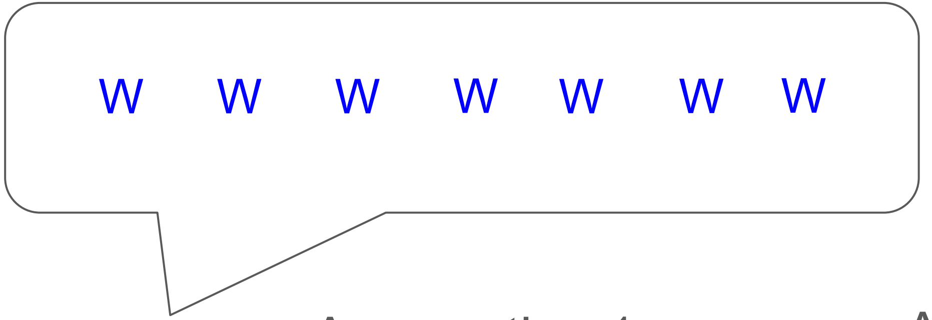


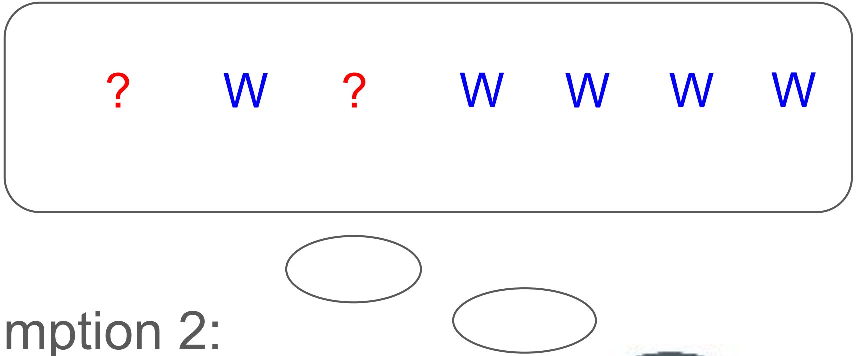
Assumption 1:

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



Memory Limitations: Formal Model (Futrell and Levy, 2017)





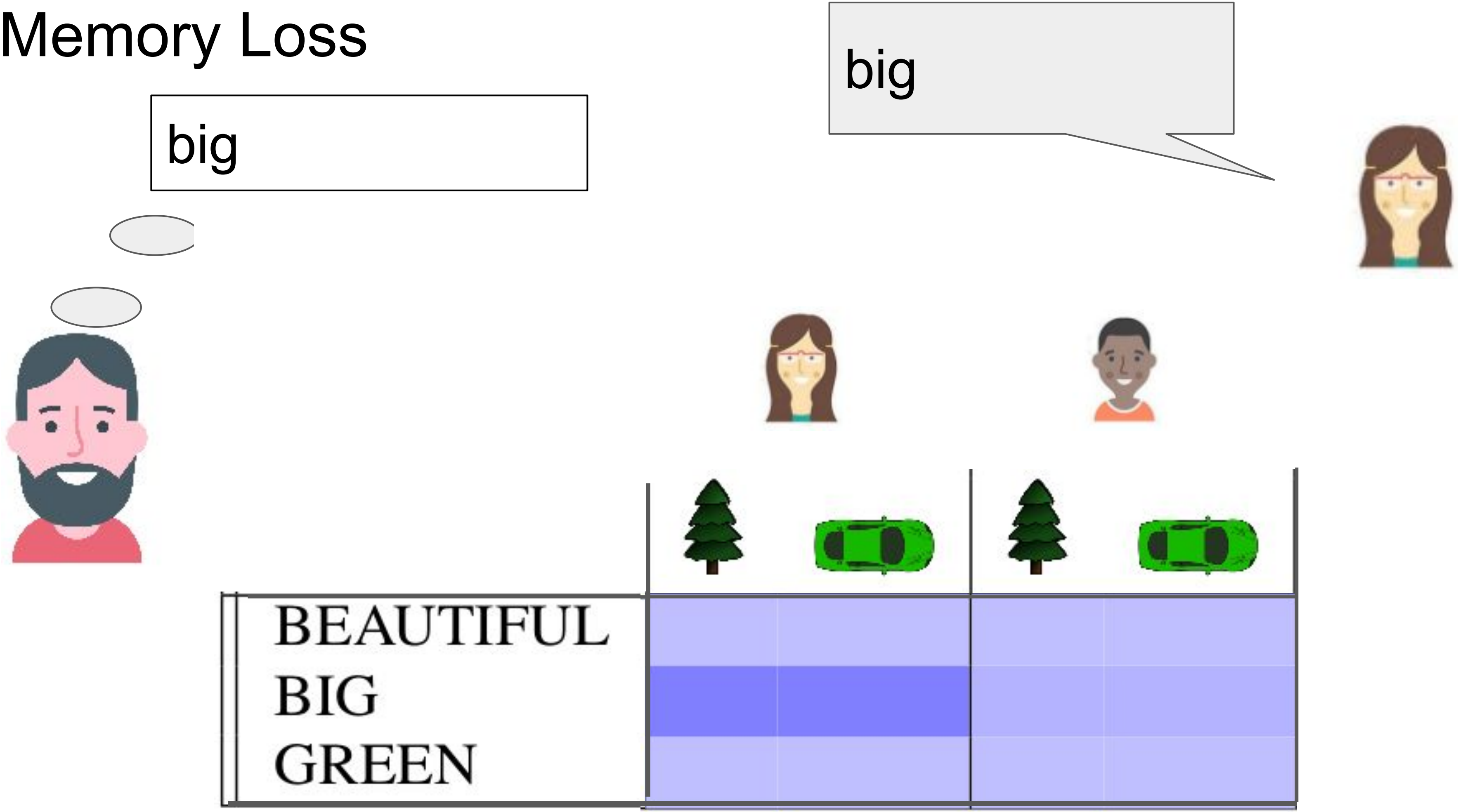
Assumption 1:

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

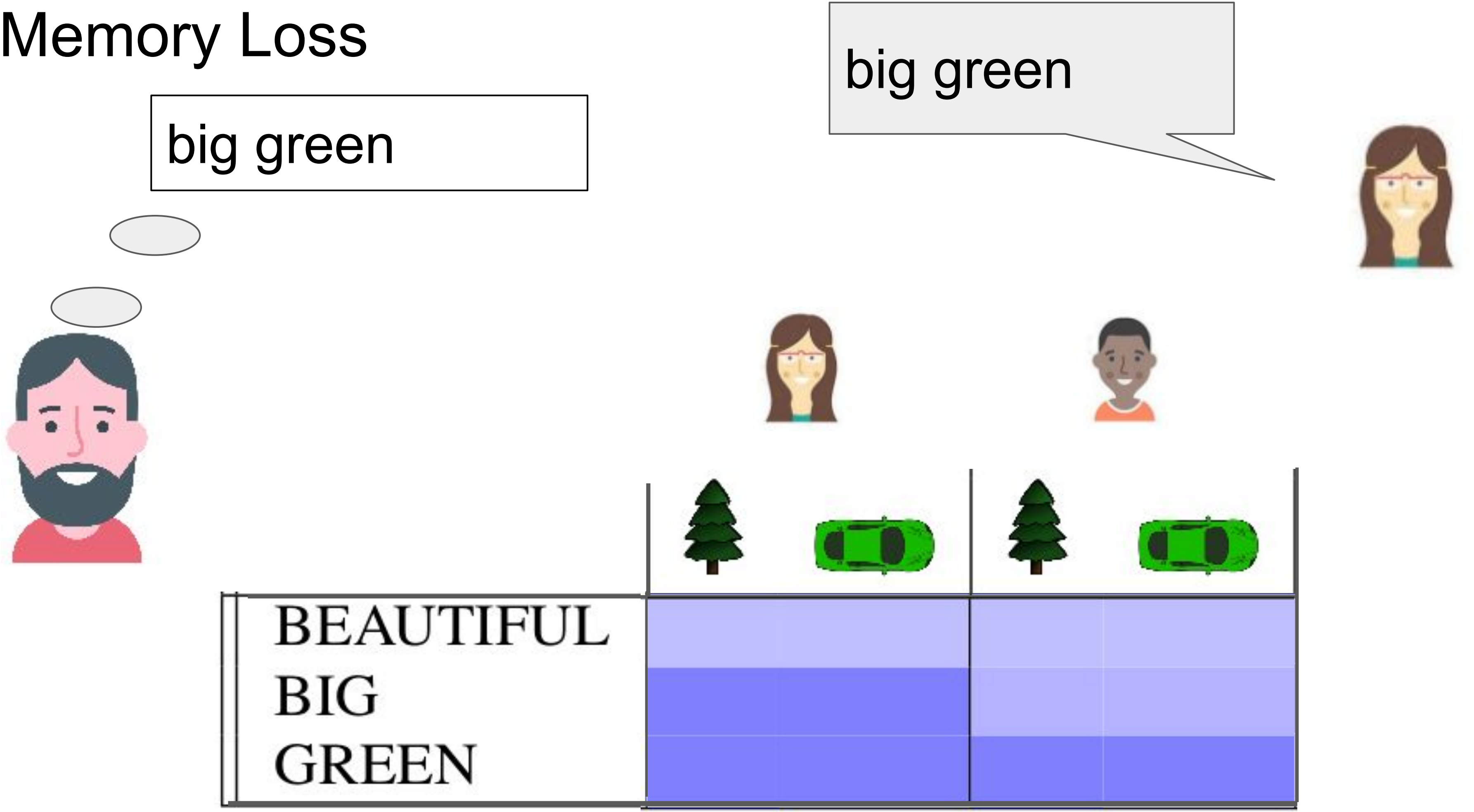
Assumption 2:

Probability of loss **increases** as one goes **further back** in the sequence.

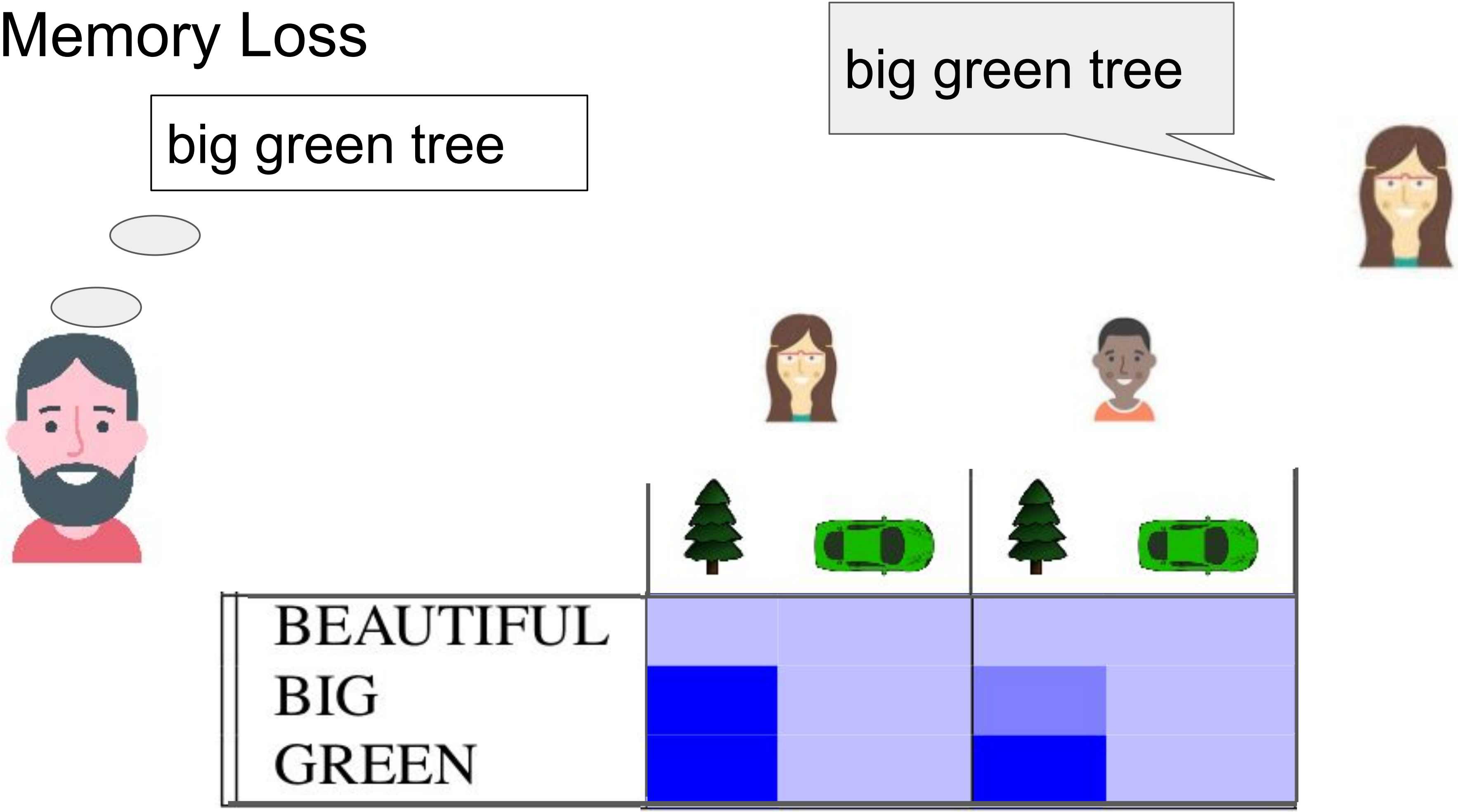
Listener Model with Memory Loss



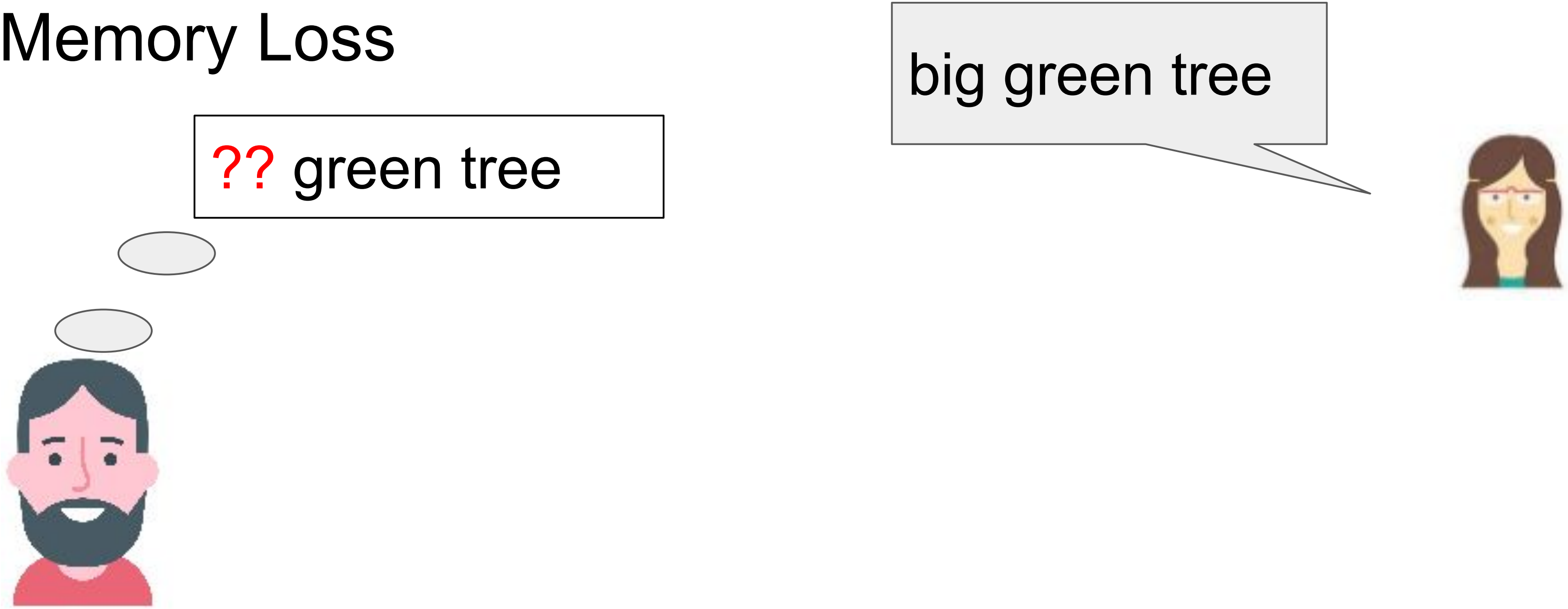
Listener Model with Memory Loss



Listener Model with Memory Loss



Listener Model with Memory Loss



Listener Model with Memory Loss

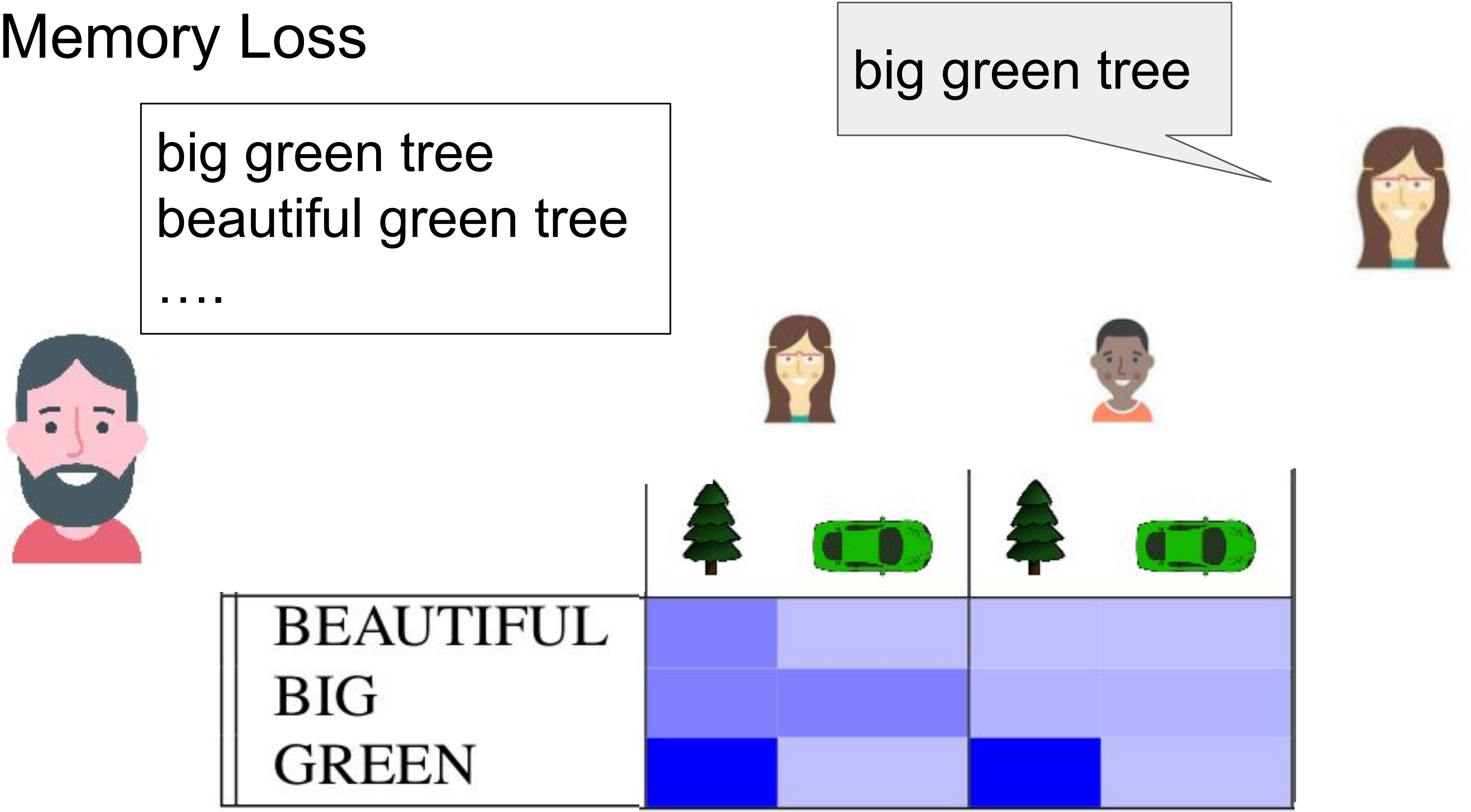
?? green tree

big green tree
beautiful green tree
ugly green tree
....

big green tree

Rational listener marginalizes over **possible completions** (Futrell & Levy, 2017)

Listener Model with Memory Loss



Listener Model with Memory Loss

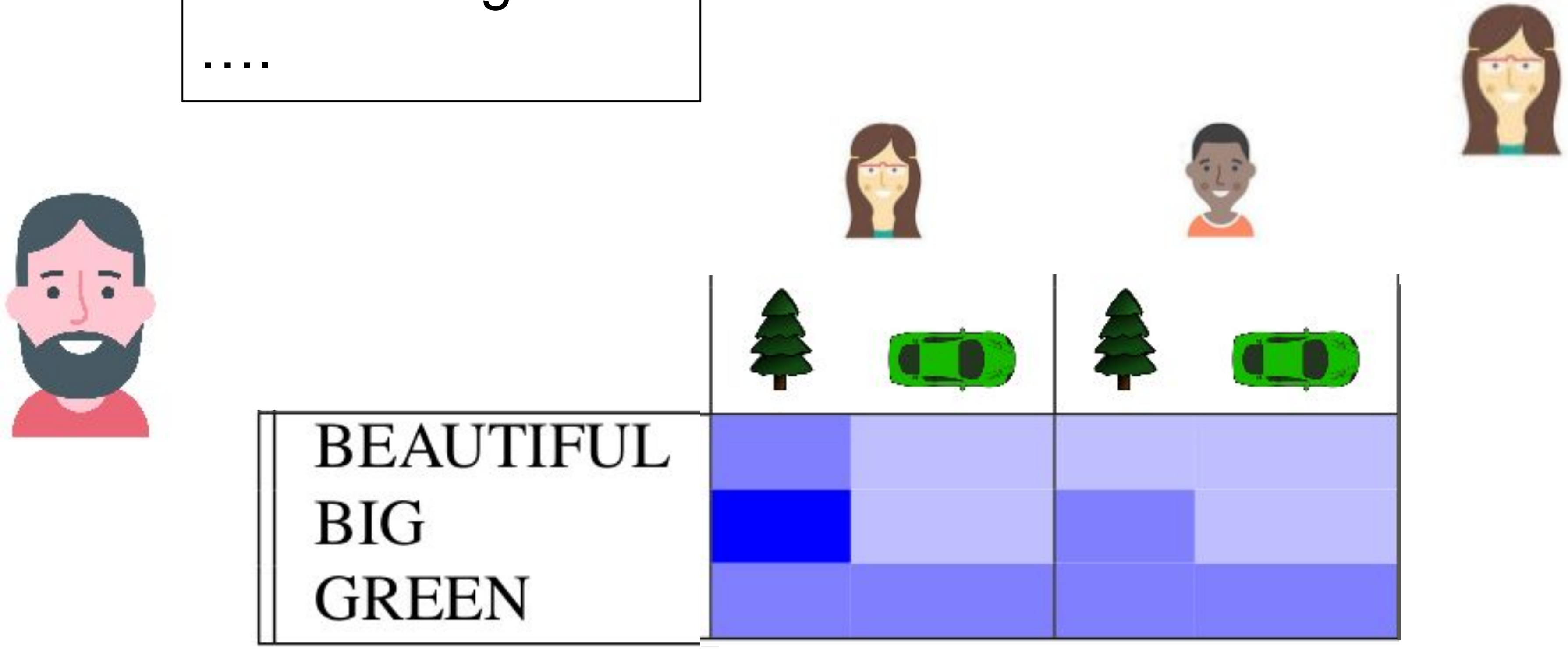
??? big tree

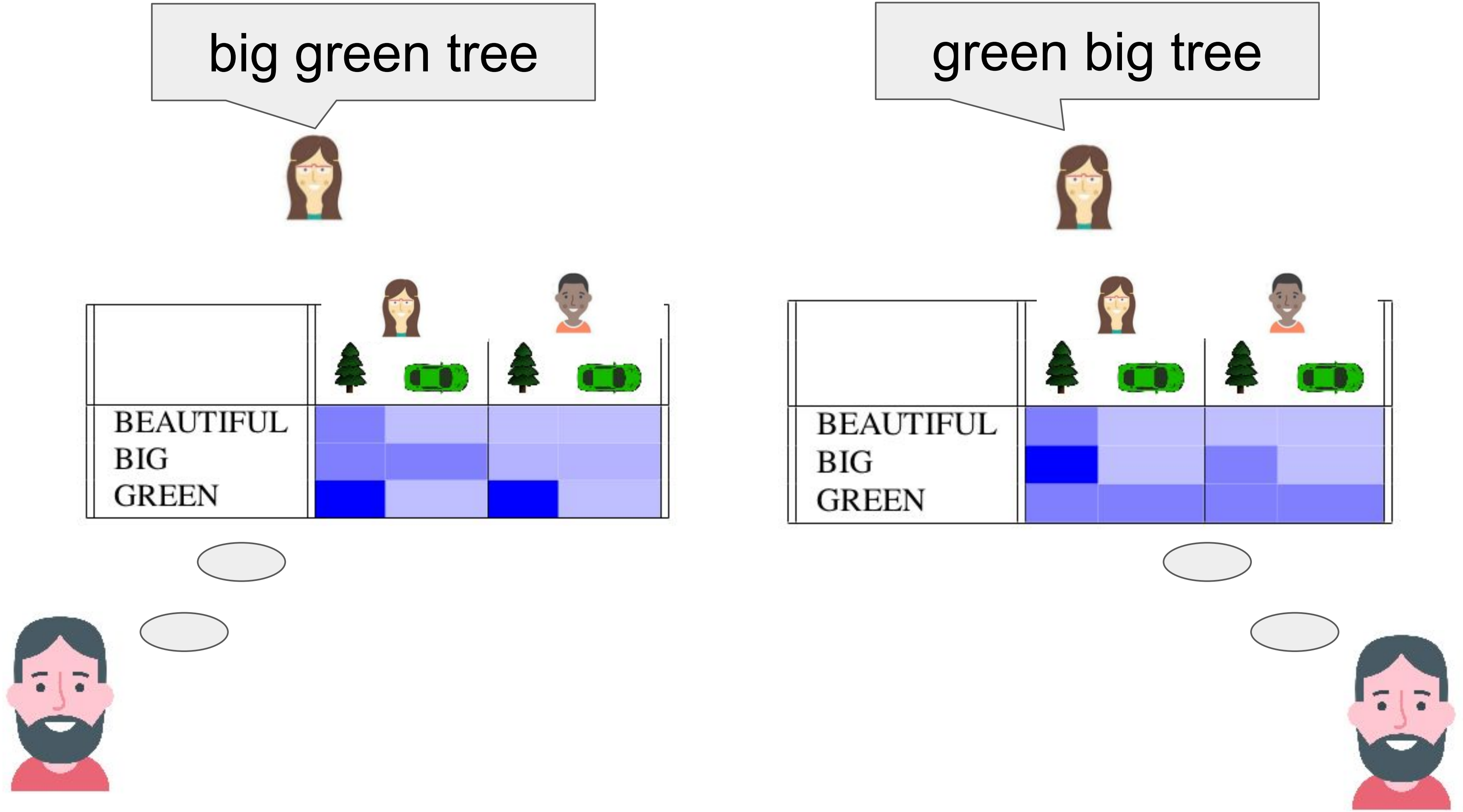
green big tree

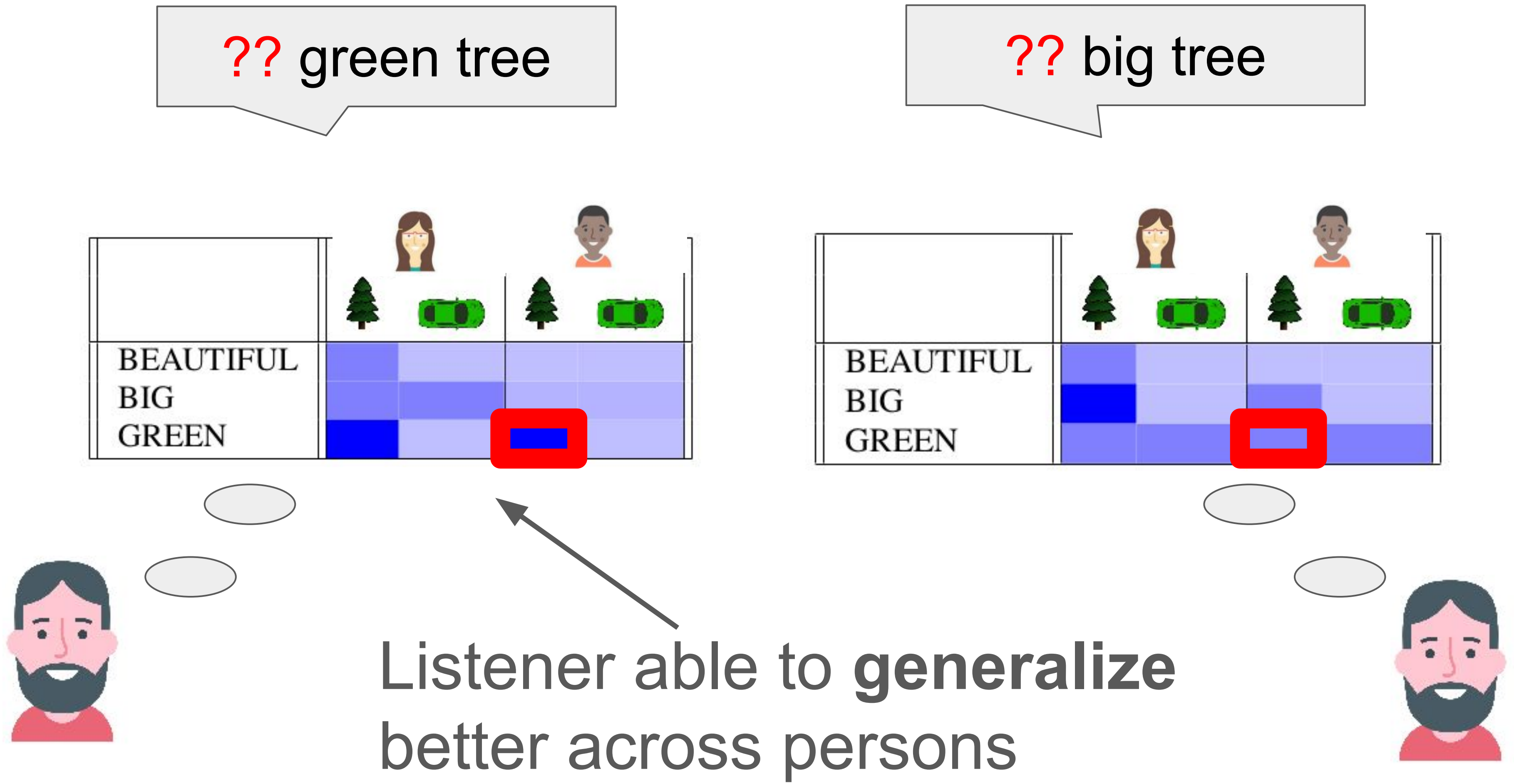
Listener Model with Memory Loss

green big tree
beautiful big tree
....

green big tree




big green tree
green big tree
BEAUTIFUL
BIG
GREEN
BEAUTIFUL
BIG
GREEN



?? green tree
?? big tree
BEAUTIFUL
BIG
GREEN
BEAUTIFUL
BIG
GREEN
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 $\qquad$ $-\log P(A_1)$

$A_1 \quad A_2$ $\qquad$ $-\log P(A_2|A_1)$

Memory Loss in the Cost

A_1			$- \log P(A_1)$
A_1	A_2		$- \log P(A_2 \vert A_1)$
??	A_2	N	$- \log P(N \vert ??\ A_2)$

Will be smaller if $PMI(N, A_2)$ is larger!

Our Proposed Model

Rational communication with Bayesian inference

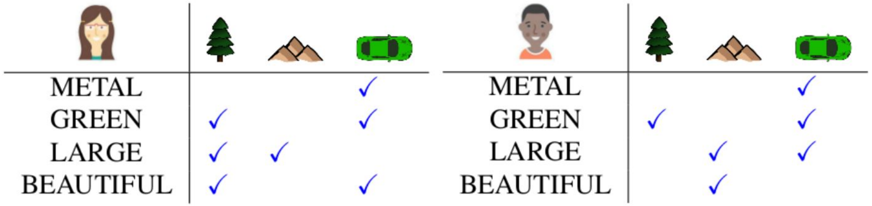
$$P_{speaker}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u))$$

Our Proposed Model

Rational communication with Bayesian inference

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

including reasoning about multiple speakers



METAL			✓
GREEN	✓		✓
LARGE	✓	✓	
BEAUTIFUL	✓		✓

METAL			✓
GREEN	✓		✓
LARGE		✓	✓
BEAUTIFUL		✓	

Our Proposed Model

Rational communication with Bayesian inference

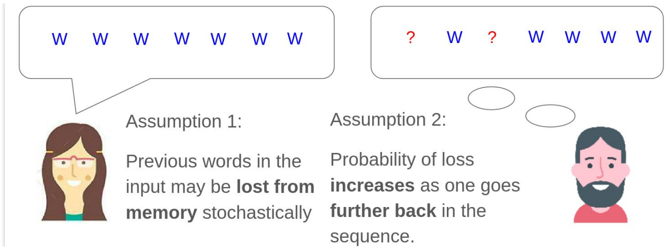
$$P_{speaker}(u) \propto \exp(\alpha \cdot I(u) - \beta \cdot C(u))$$

including reasoning about multiple speakers

METAL			✓
GREEN	✓		✓
LARGE	✓	✓	
BEAUTIFUL	✓		✓

METAL			✓
GREEN	✓		✓
LARGE		✓	✓
BEAUTIFUL		✓	

and incremental, rational processing under memory limitations.



Evaluation

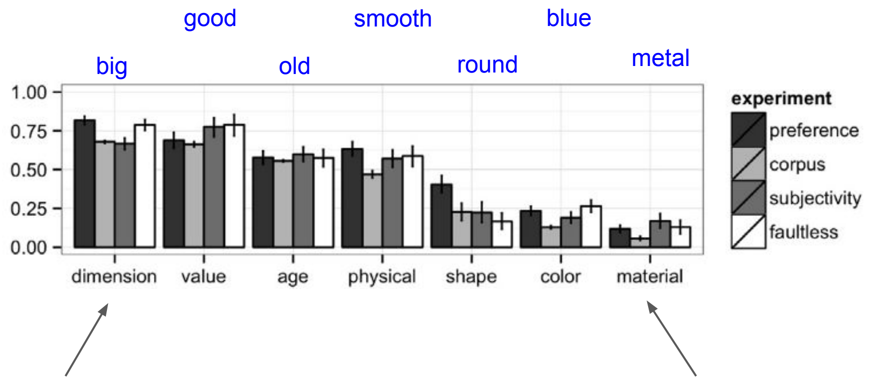
Task: Predict adjective order in corpus data

Evaluation

Task: Predict adjective order in corpus data

Model Parameters:

- **κ(A) = 1 - subjectivity(A)**



κ(big) = 0.2

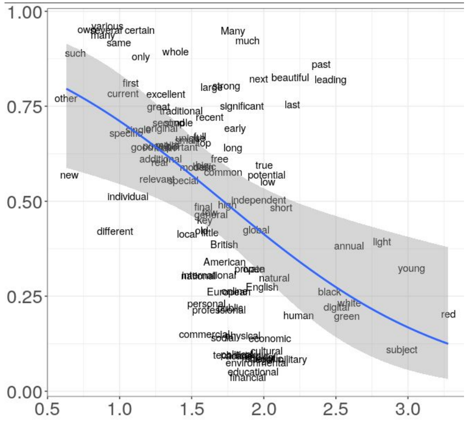
κ(metal)= 0.85

Evaluation

Task: Predict adjective order in corpus data

Model Parameters:

- **κ(A) = 1 - subjectivity(A)**
- **MI:** from corpus analysis



Evaluation

Task: Predict adjective order in corpus data

Model Parameters:

- **κ(A) = 1 - subjectivity(A)**
- **MI:** from corpus analysis
- Other parameters inferred using Bayesian Data Analysis in Pyro (http://pyro.ai/)

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

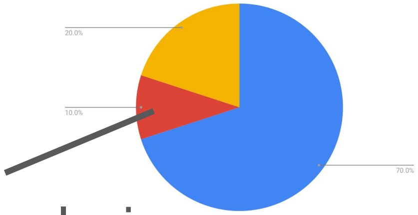
Evaluation

Task: Predict adjective order in corpus data

Model Parameters:

- **κ(A) = 1 - 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

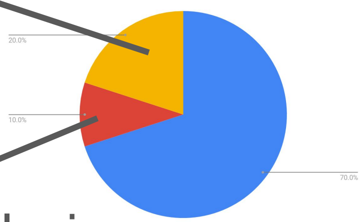
Task: Predict adjective order in corpus data

Model Parameters:

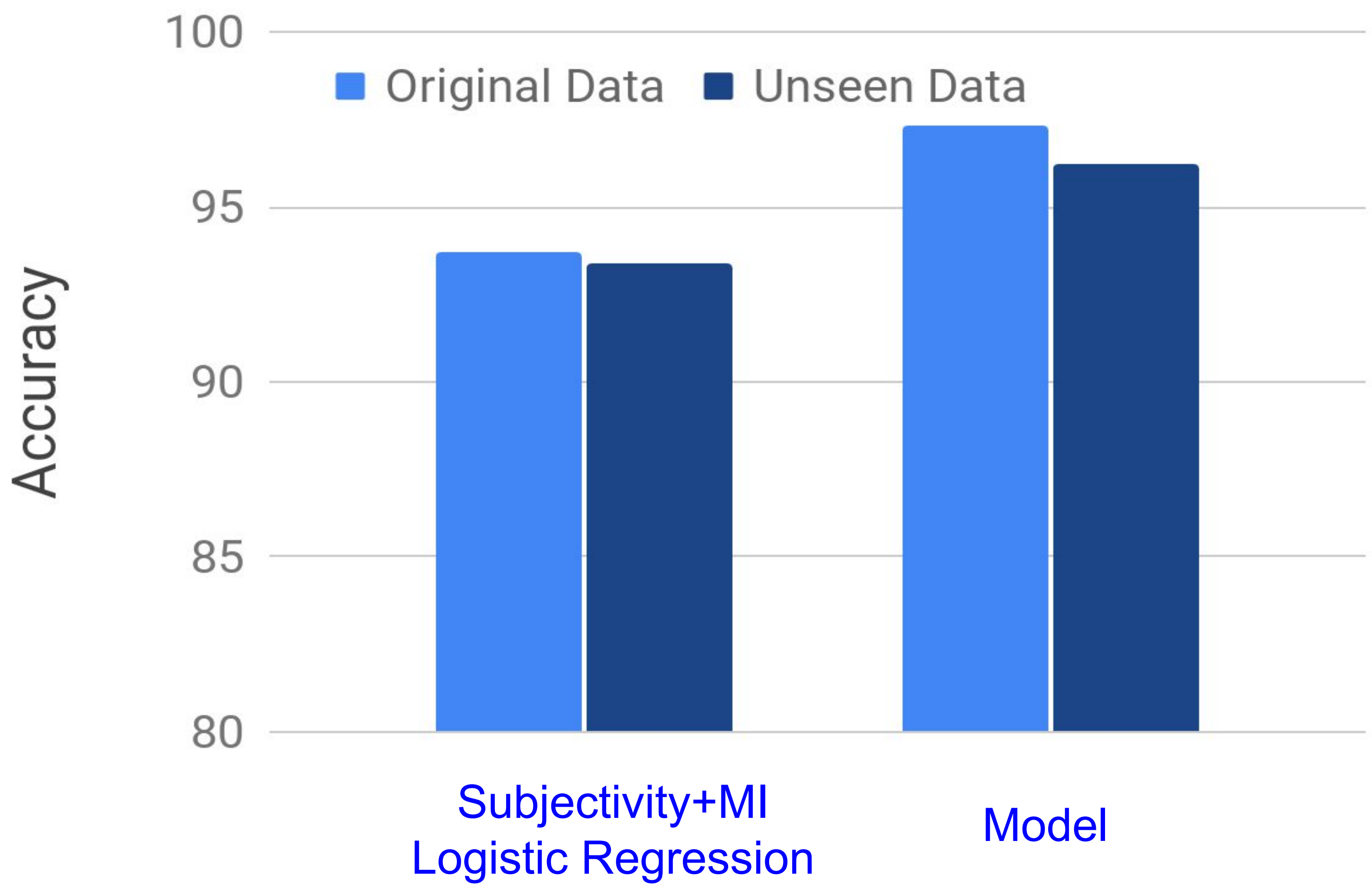
- **κ(A) = 1 - subjectivity(A)**
- **MI:** from corpus analysis
- Other parameters inferred using Bayesian Data Analysis in Pyro (http://pyro.ai/)

Evaluation Datasets

Unseen data set
(~ 10,000 examples)



Set from corpus analysis
(~ 4,700 examples)


100
95
90
85
80
Accuracy
Original Data
Unseen Data
Subjectivity+MI
Logistic Regression
Model

Languages with Postnominal Adjectives

l-kitaabu l-’axḍaru ṣ-ṣaγiiru
the-book the-green the-small

‘the little green book’ (Fassi Fehri, 1999, 107)

Standard Arabic

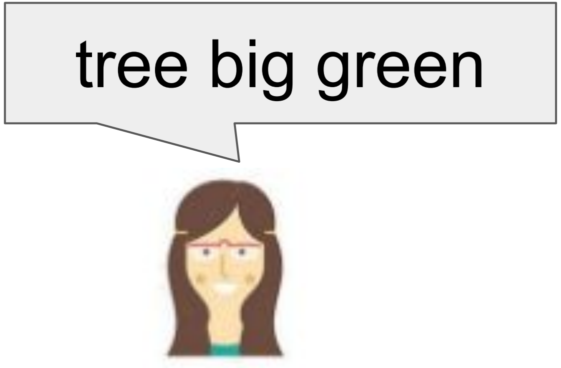
Languages with Postnominal Adjectives

Subjectivity-based ordering reported for

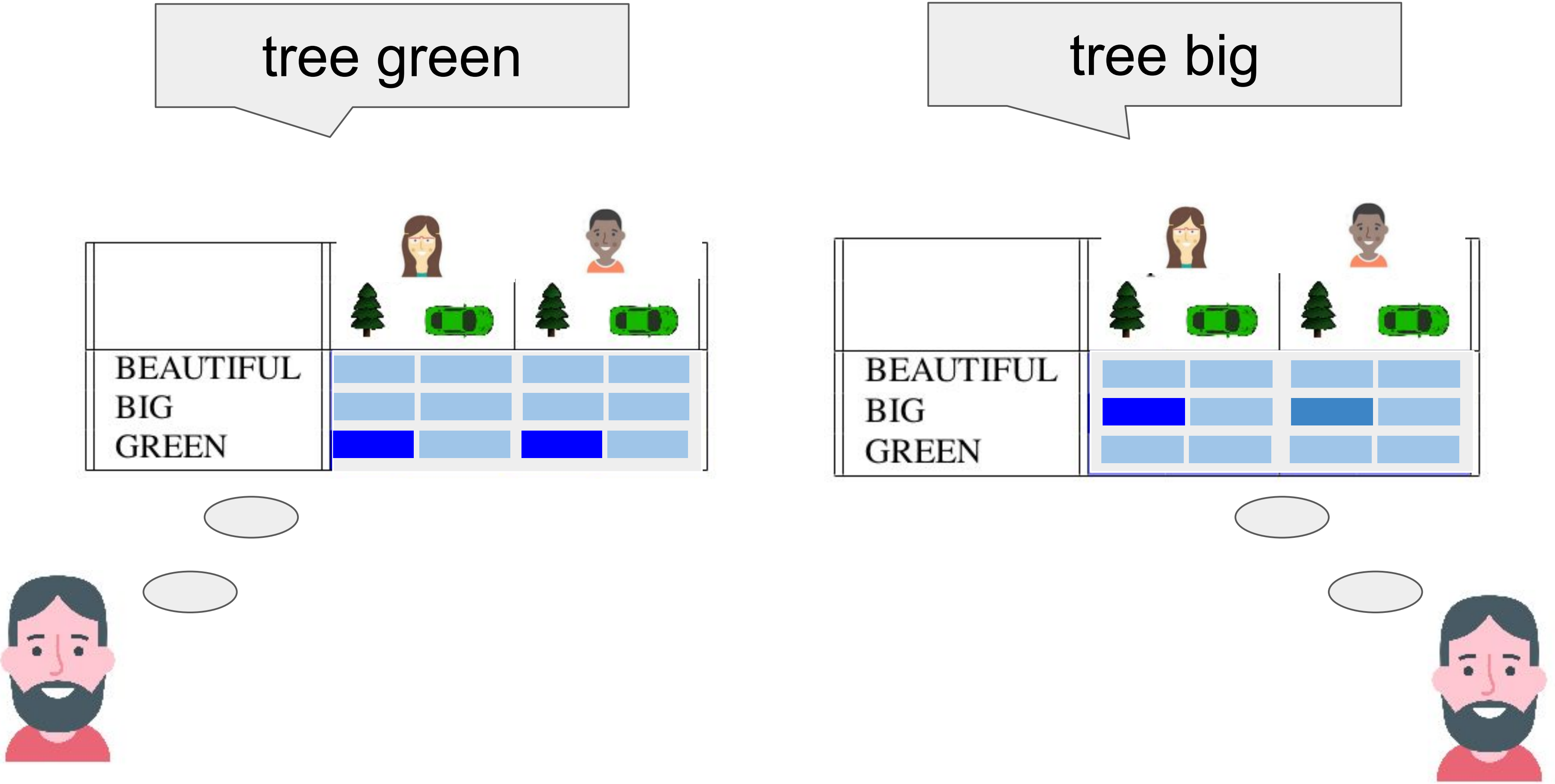
- Arabic (Kachakeche & Scontras, 2020)
- Tagalog (Samonte & Scontras, 2019)

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

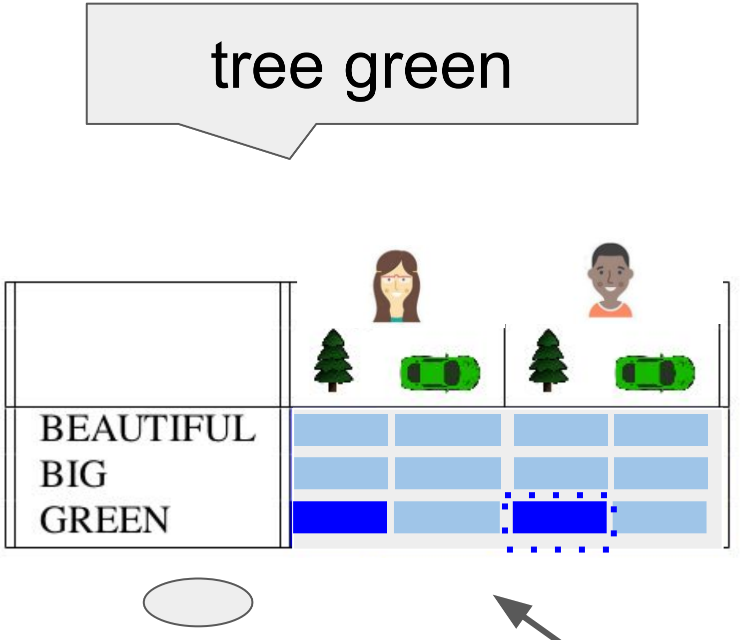
Languages with Postnominal Adjectives

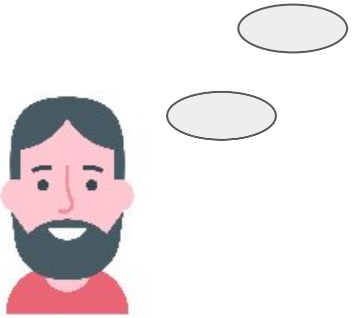


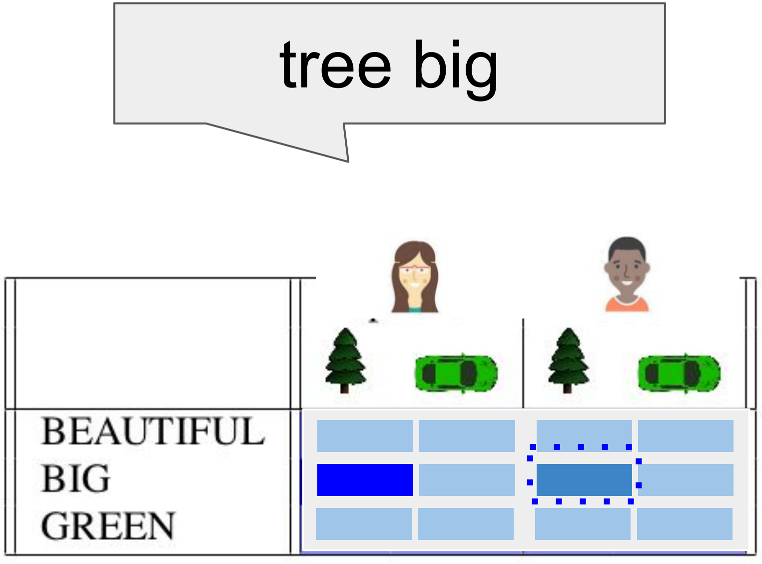
Languages with Postnominal Adjectives

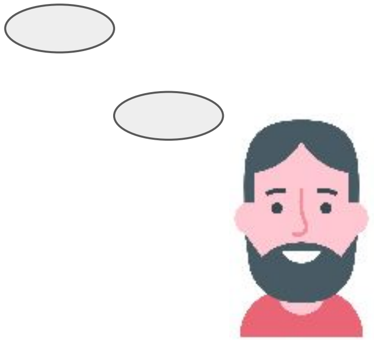


Languages with Postnominal Adjectives









Listener able to generalize better across persons

Languages with Postnominal Adjectives

N			$-\log P(N)$
N	A_1		$-\log P(A_1 \mid N)$
??	A_1	A_2	$-\log P(A_2 \mid ??\ A_1)$

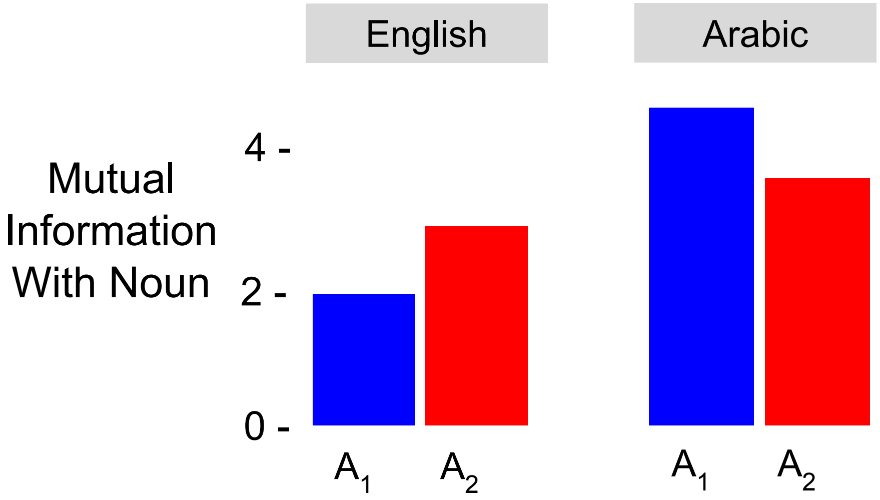
Languages with Postnominal Adjectives

N			$-\log P(N)$
N	A_1		$\boxed{-\log P(A_1 \mid N)}$
??	A_1	A_2	$-\log P(A_2 \mid ??\ A_1)$

Will be lower when
PMI(A_1, N) is higher.

Languages with Postnominal Adjectives

MI with Noun (in bits)



Discussion

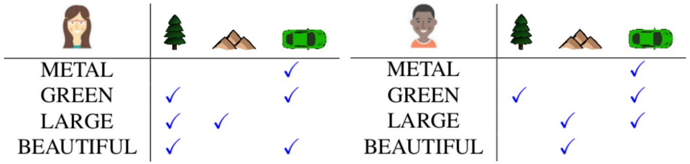
Subjectivity and **MI** independently impact adjective ordering.

Discussion

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

Provided model of adjective ordering integrating standard
Bayesian reasoning

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



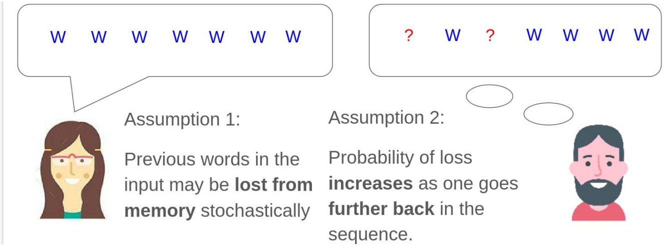
METAL			✓
GREEN	✓		✓
LARGE	✓	✓	
BEAUTIFUL	✓		✓

METAL			✓
GREEN	✓		✓
LARGE		✓	✓
BEAUTIFUL		✓	

Discussion

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

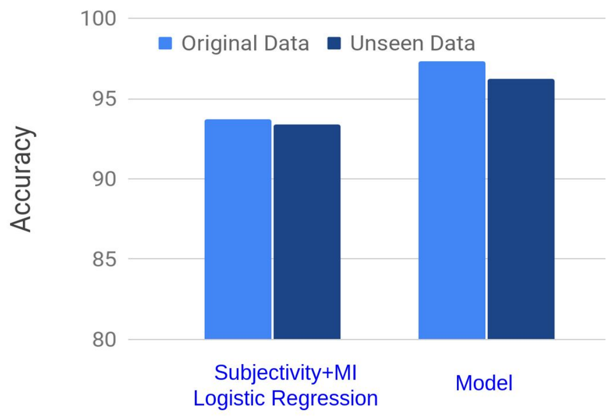
Provided model of adjective ordering integrating standard **Bayesian reasoning** with **incremental processing** under **memory limitations**



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.

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

blue big book

[[blue]] ([[big]] [[book]])

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

blue big book

[[blue]] ([[big]] [[book]])

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

blue big book

[[blue]] ([[big]] [[book]])

Reference resolution **failed**!

Related Account: Simonič 2018, Franke et al., 2019; Scontras et al., 2019.

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

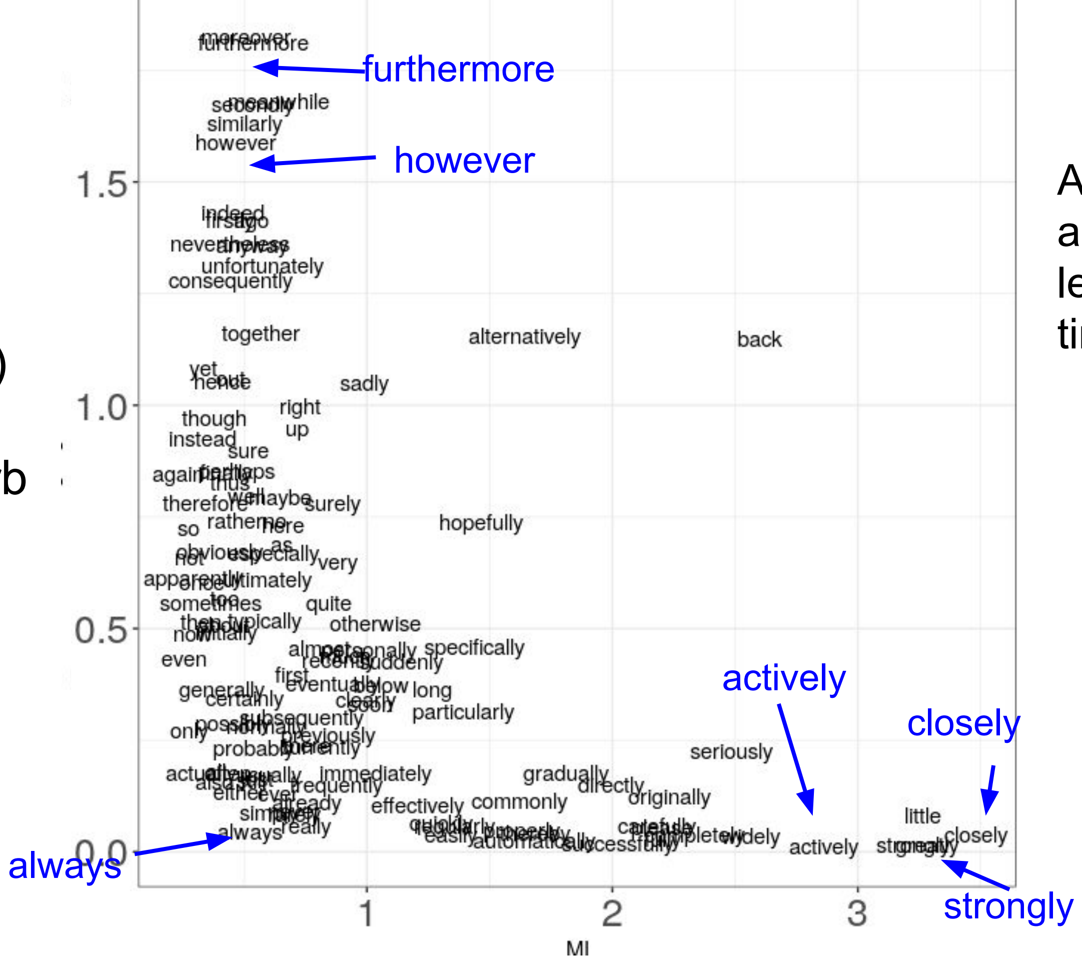
Mutual Information in Adverb Order

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

(Cinque 1999, p. 34)

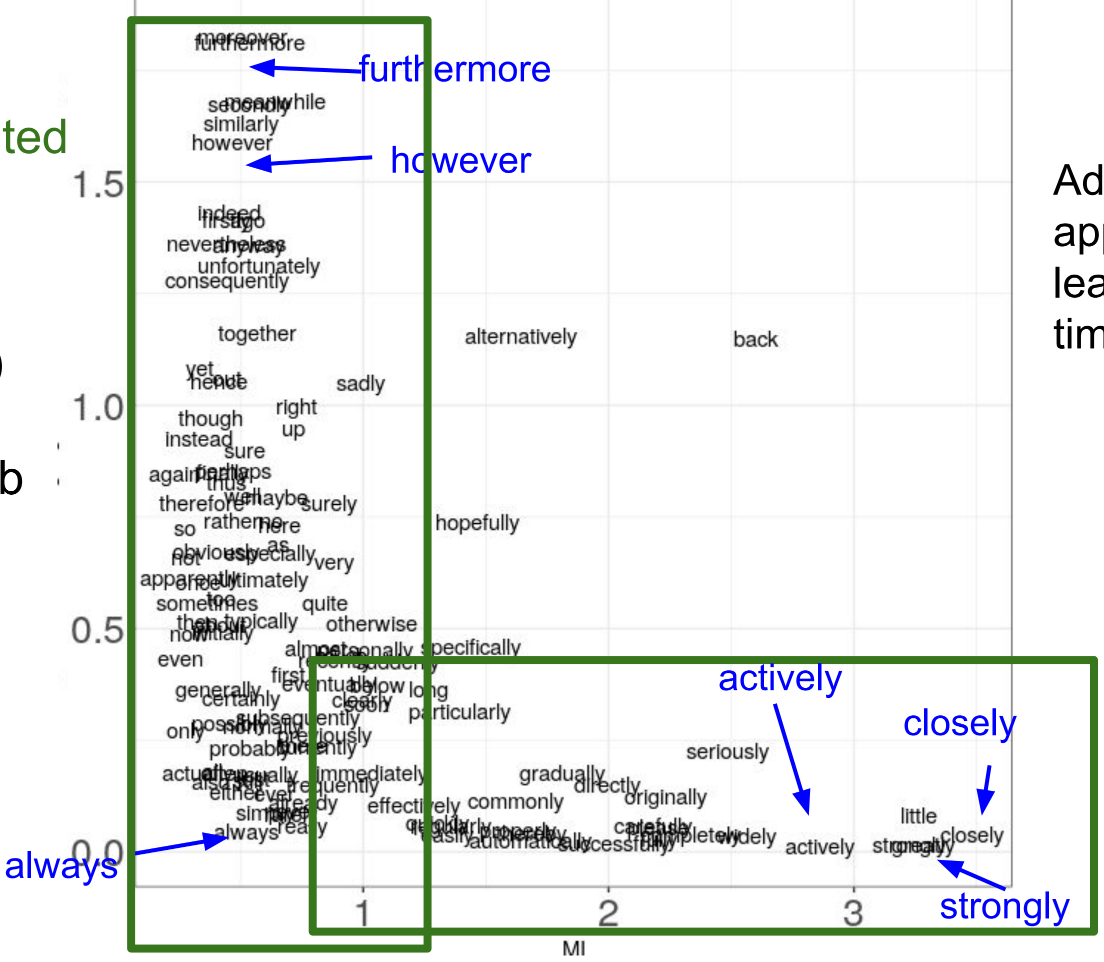
Adverbs appearing at least 20K times

Average (Log) Distance between Adverb and Verb



Average Mutual Information between Adverb and Verb clause-oriented adverbs

Average (Log) Distance between Adverb and Verb



Adverbs appearing at least 20K times

VP-oriented adverbs

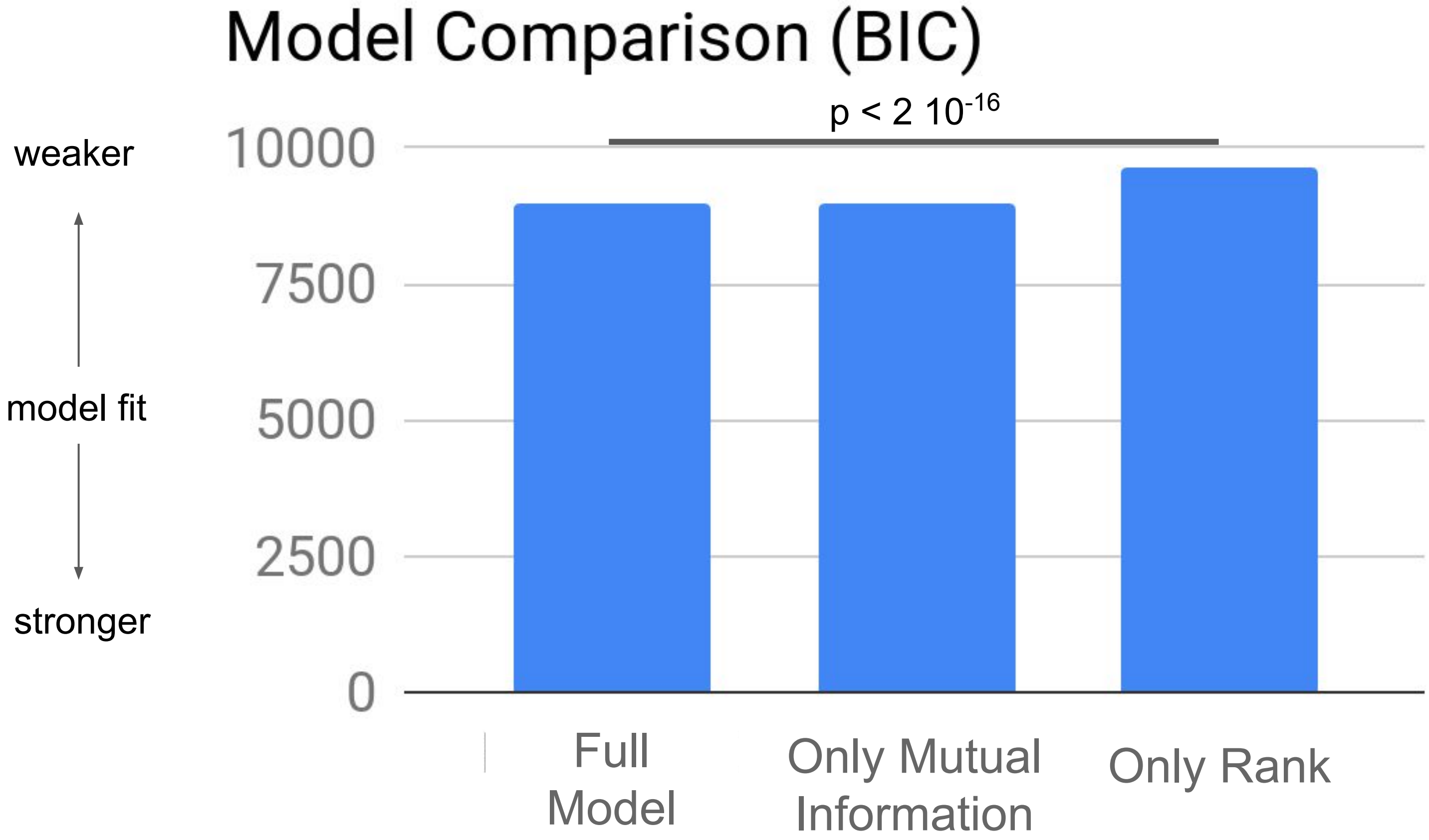
Average Mutual Information between Adverb and Verb

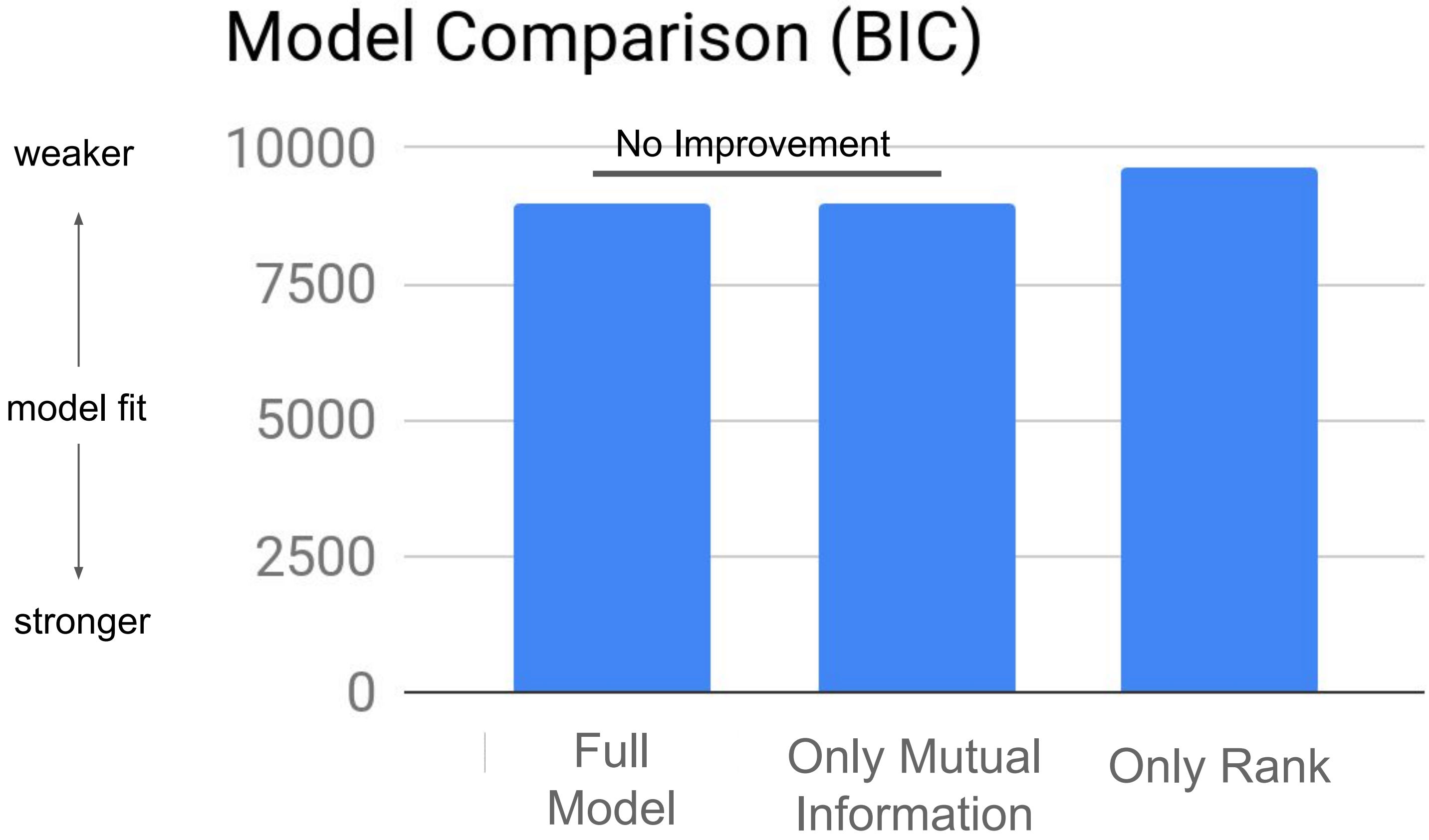
Predict order of pairs Adverb$_1$ Adverb$_2$ in corpus using logistic regression from

1. Mutual Information: pmi(Adverb1, Verb) - pmi(Adverb2, Verb)
2. Ranks of adverbs in the hierarchy

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

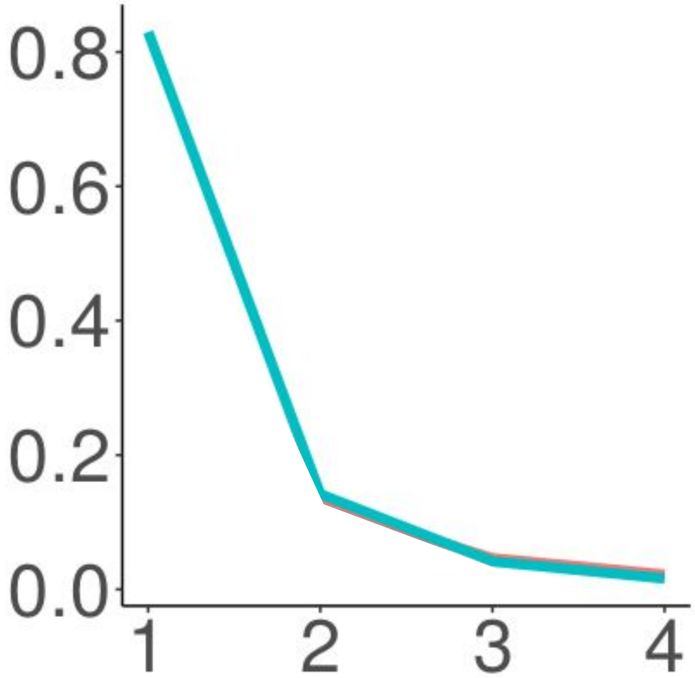
(Cinque 1999, p. 34)


Model Comparison (BIC)
p < 2 10-16
weaker
model fit
stronger
10000
7500
5000
2500
0
Full Model
Only Mutual Information
Only Rank


Model Comparison (BIC)
weaker
model fit
stronger
No Improvement
10000
7500
5000
2500
0
Full Model
Only Mutual Information
Only Rank

Mutual Information beyond Adjective Ordering

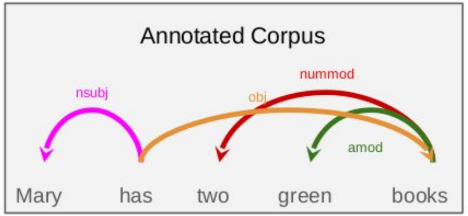
Mutual Information
between words at given distance
(controlling for redundancy with intervening words)



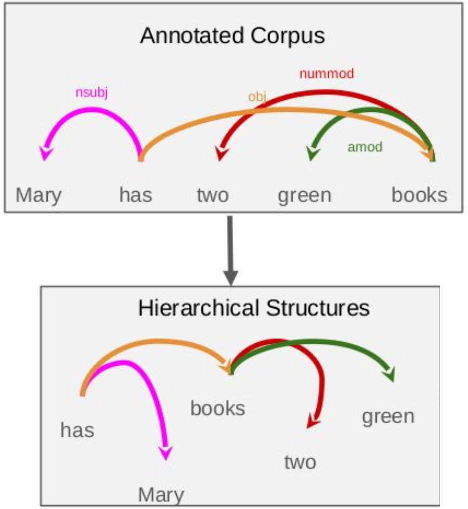
Distance

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

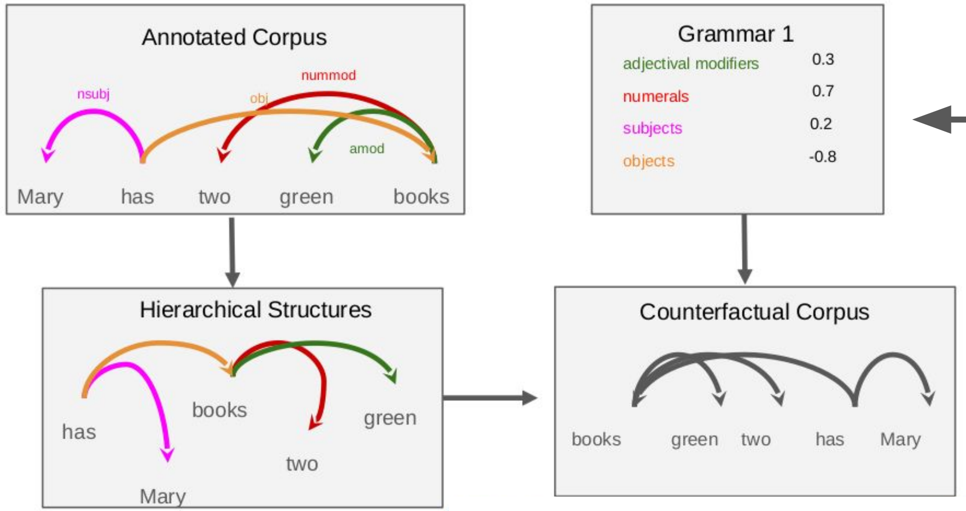
(Hahn, Degen, Futrell, in press)



(Hahn, Degen, Futrell, in press)



(Hahn, Degen, Futrell, in press)

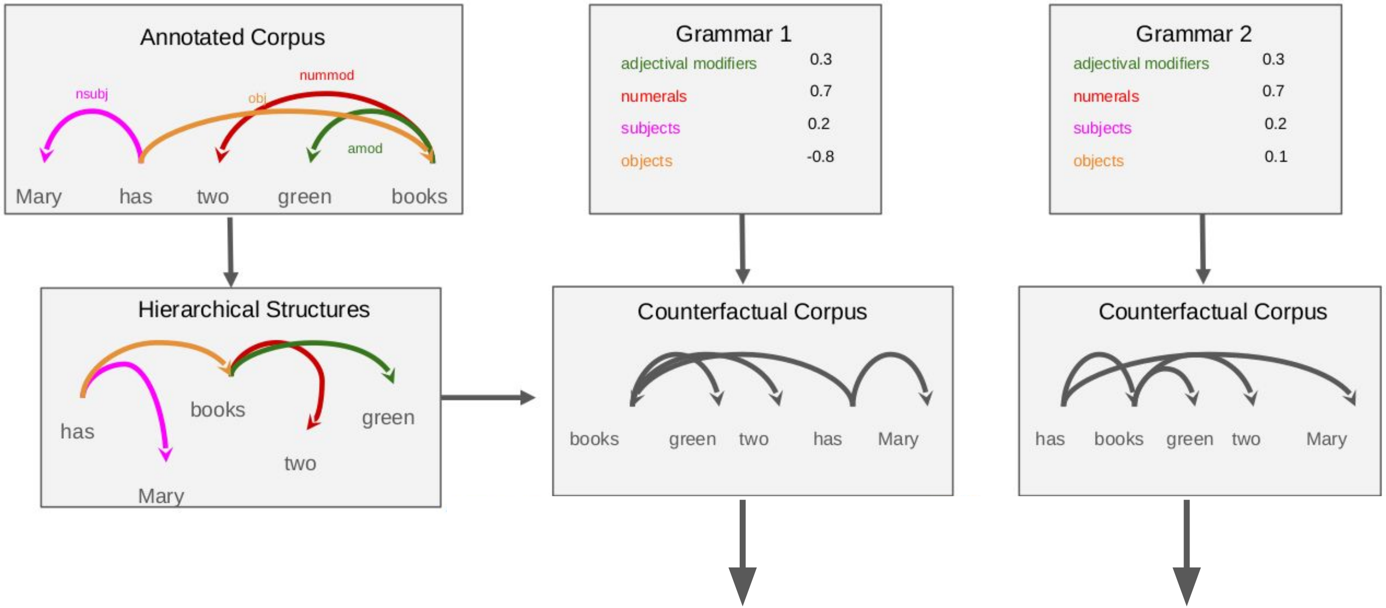


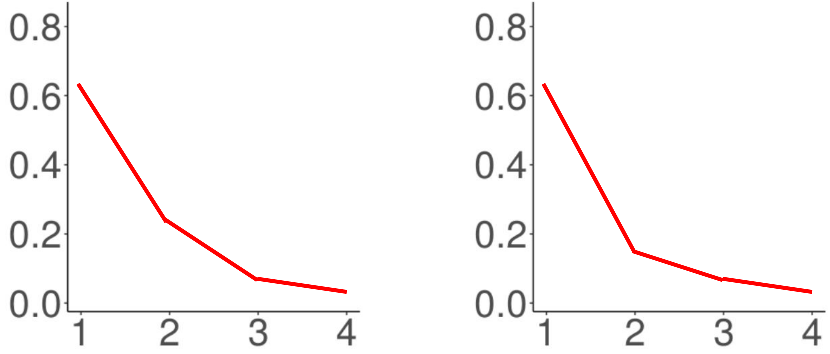
specify **random but internally consistent** word order patterns

e.g.

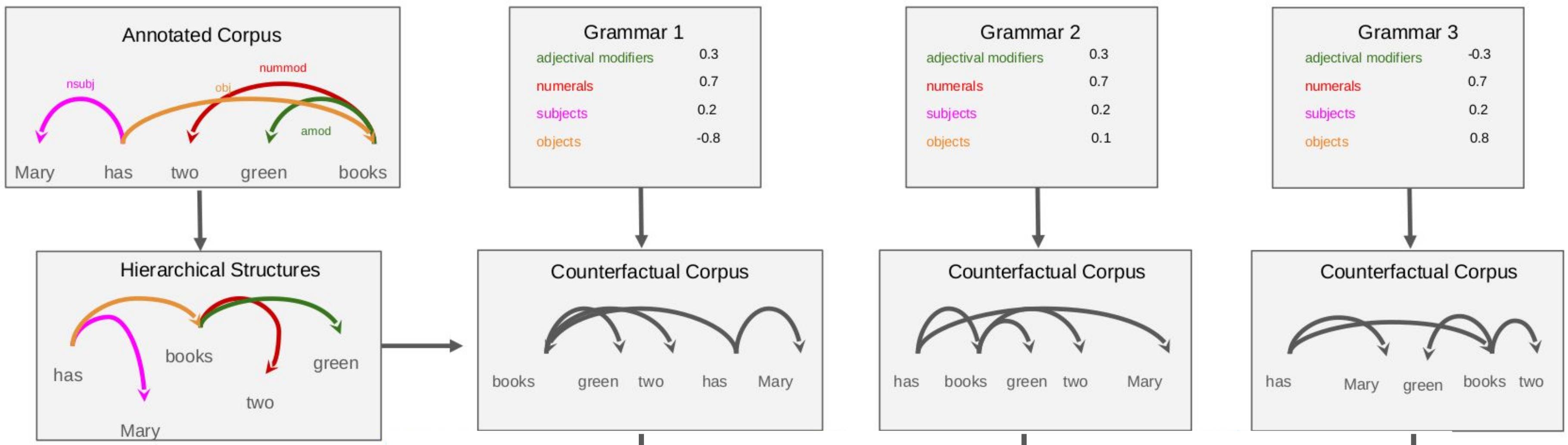
- SOV
- Noun-Adjective
- Genitive-Noun
- ...

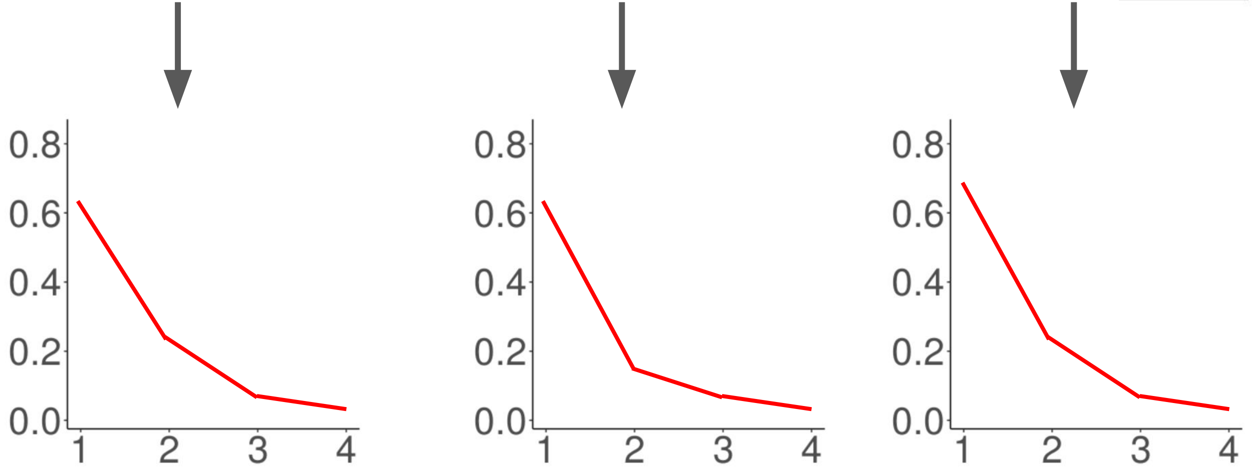
(Hahn, Degen, Futrell, in press)





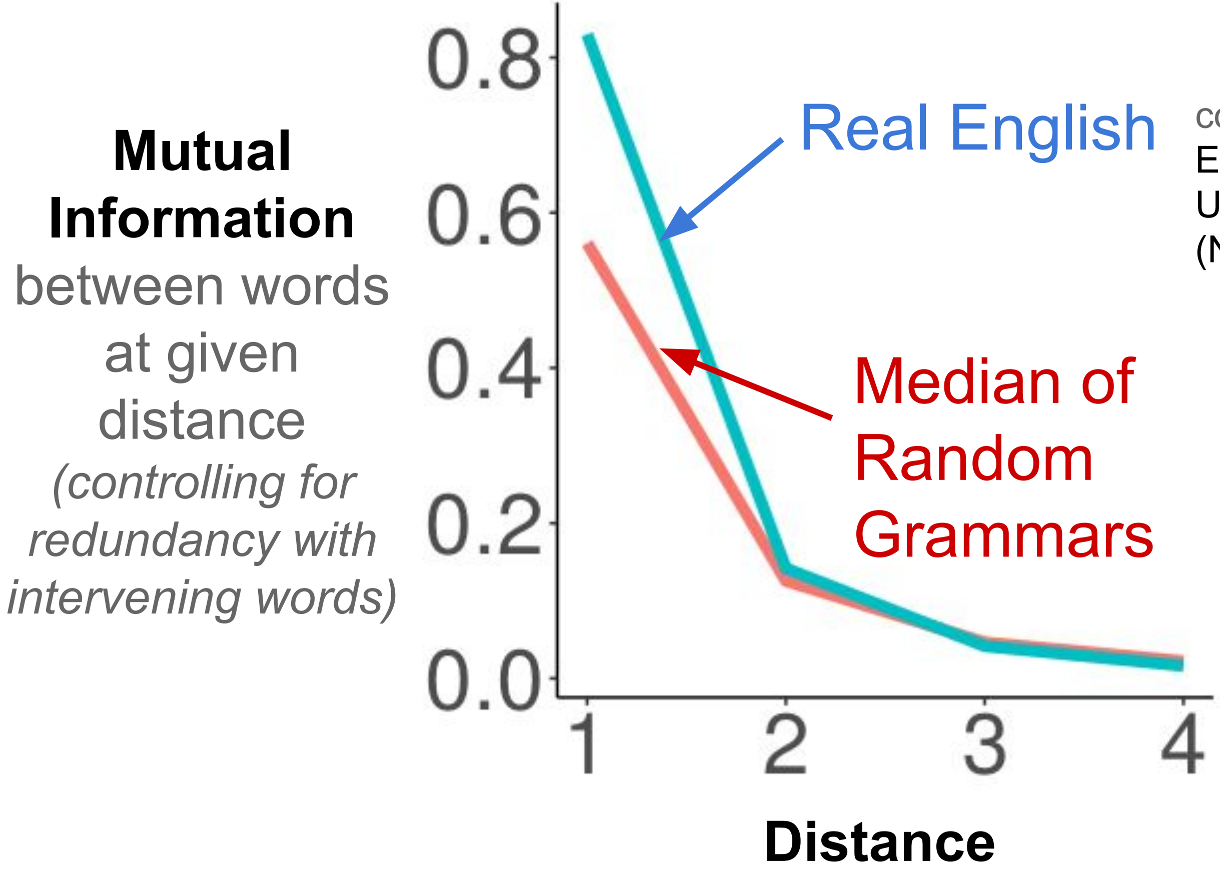
(Hahn, Degen, Futrell, in press)





(Hahn, Degen, Futrell, in press)

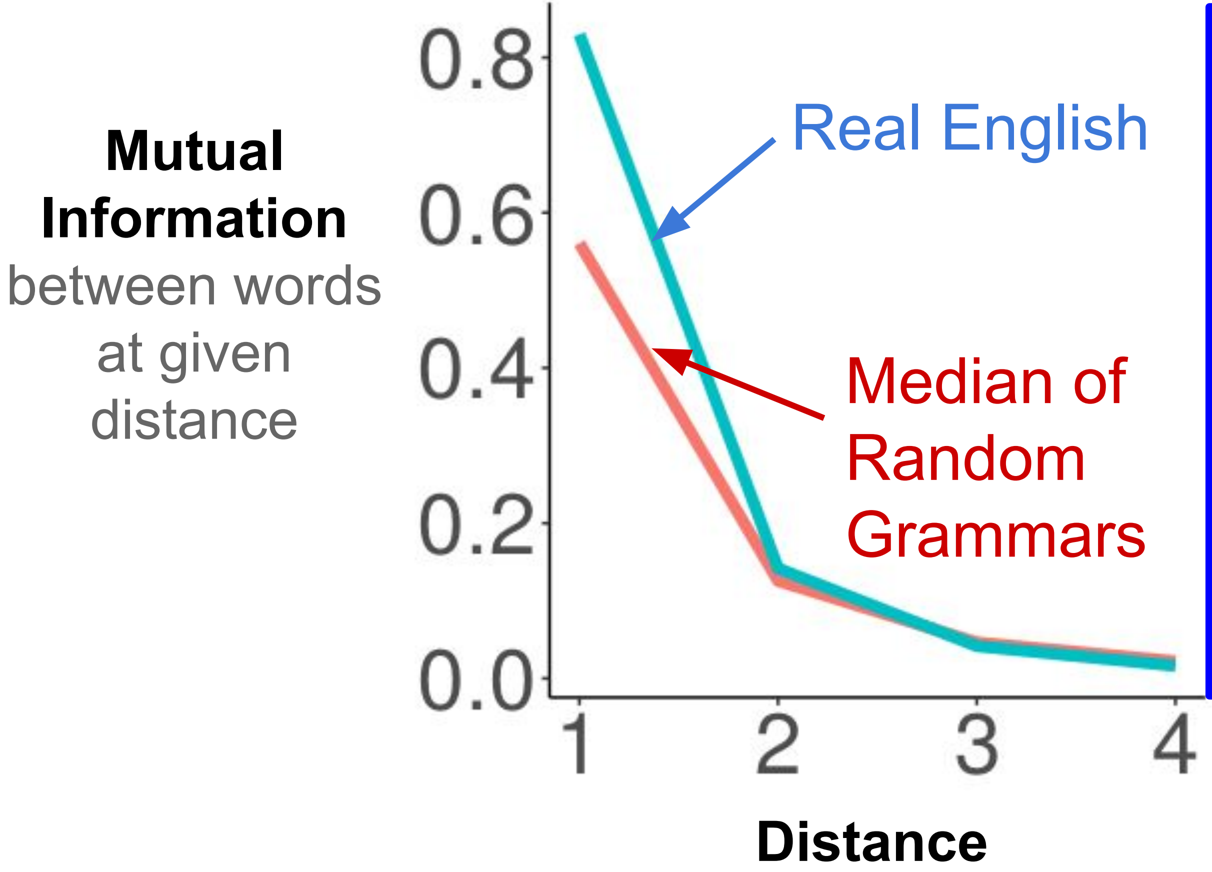
Mutual Information beyond Adjective Ordering



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

*(Hahn, Degen, Futrell,
in press)*

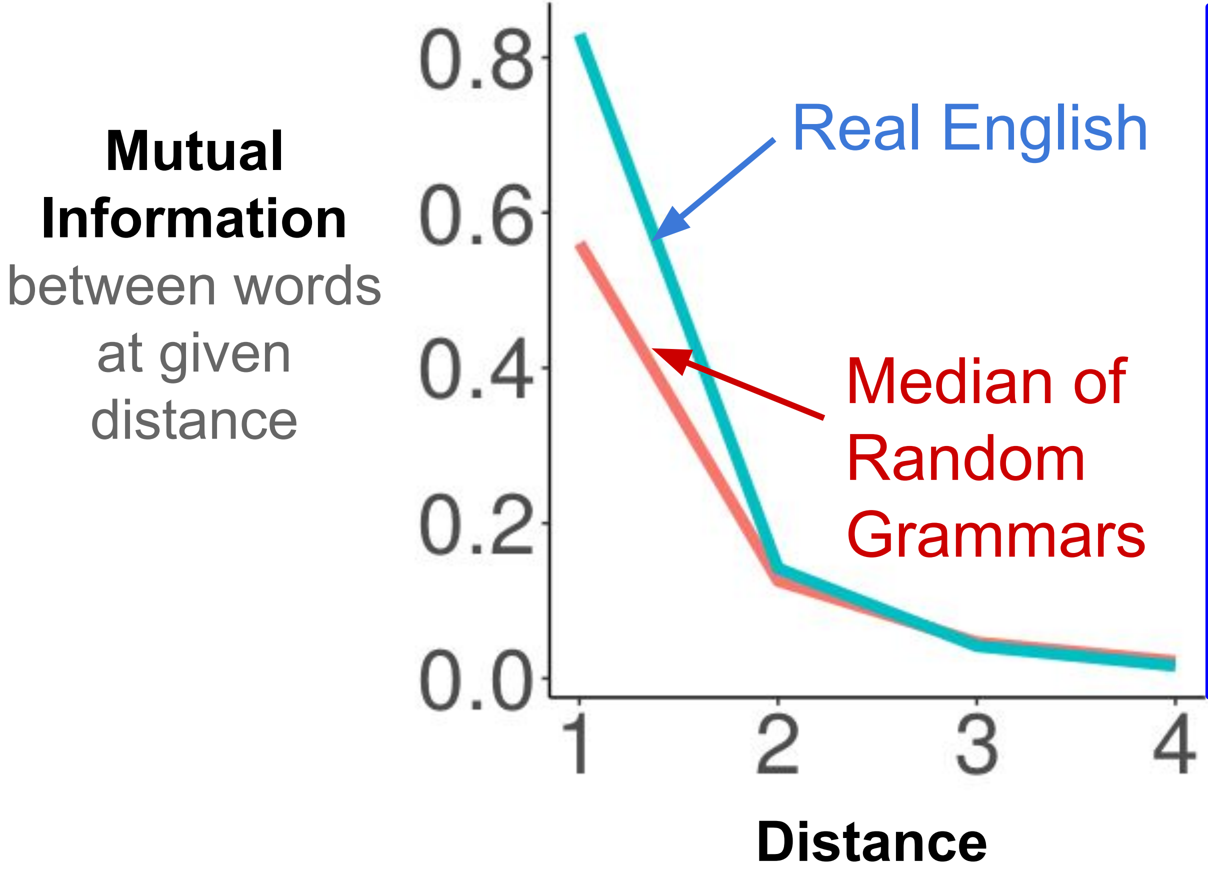
Mutual Information beyond Adjective Ordering



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

(Hahn, Degen, Futrell, in press)

Mutual Information beyond Adjective Ordering

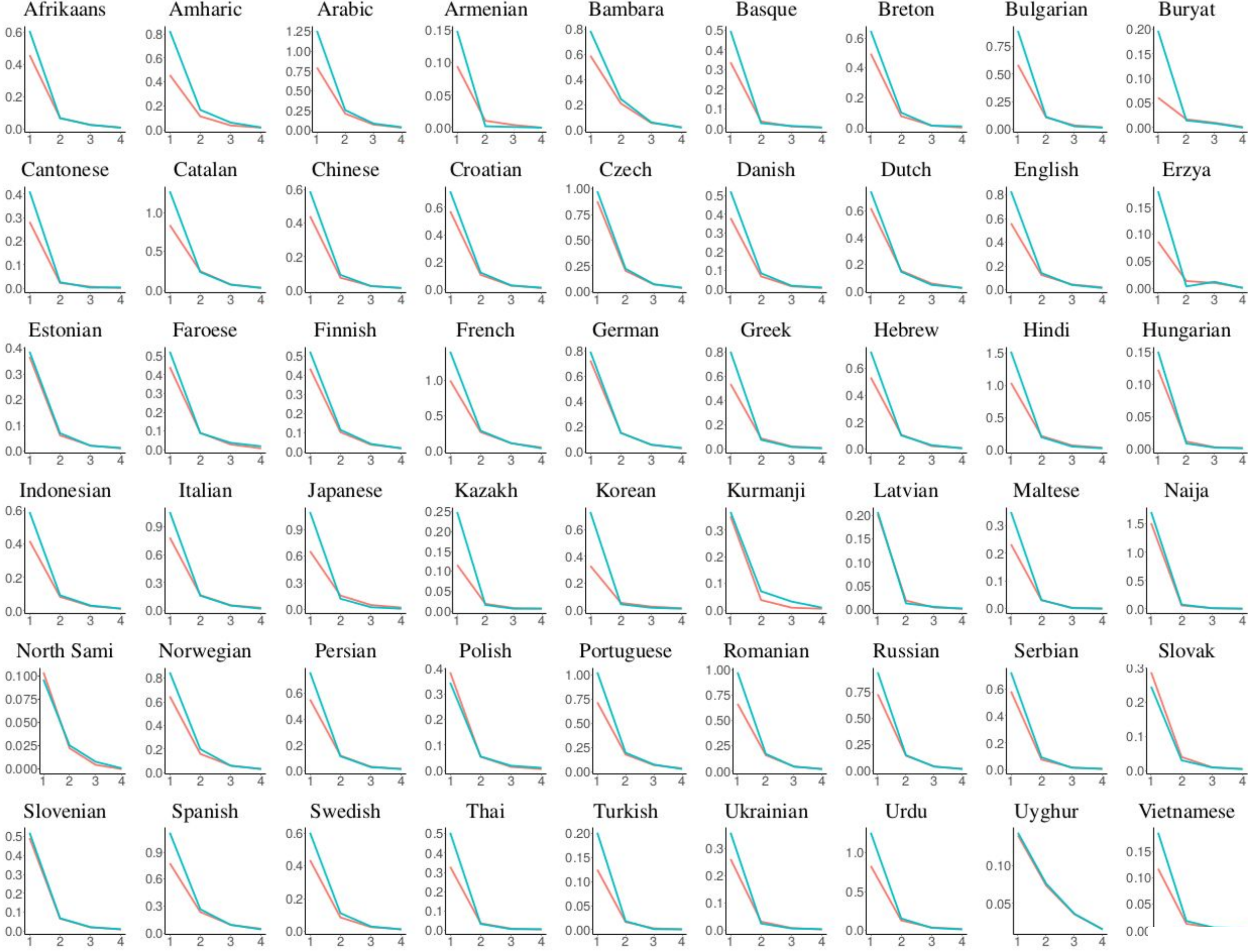


Information Locality:
Real orders place high-MI word pairs close together.
(Futrell and Levy, 2017)

Minimizes **surprisal cost** under **memory limitations** (Hahn, Degen, Futrell, in press)

(Hahn, Degen, Futrell, in press)

Mutual Information between words at given distance



Distance

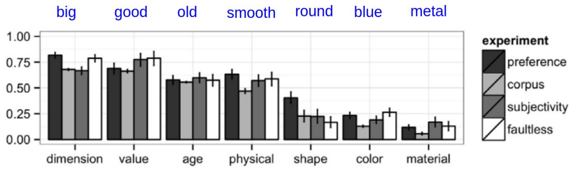
Real Orders

Median of Random Grammars

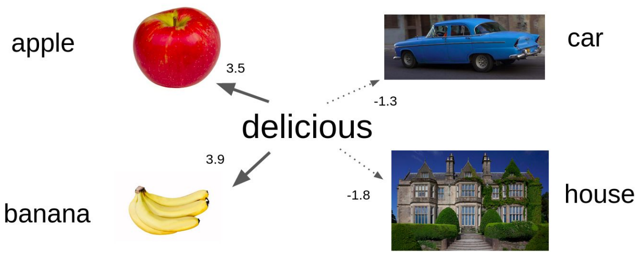
(Hahn, Degen, Futrell, in press)

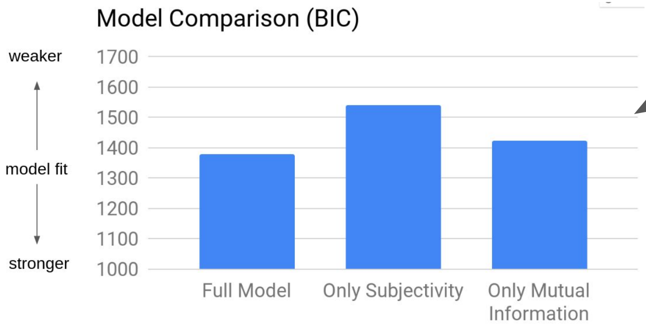
Conclusion

Subjectivity and MI independently impact adjective ordering.



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

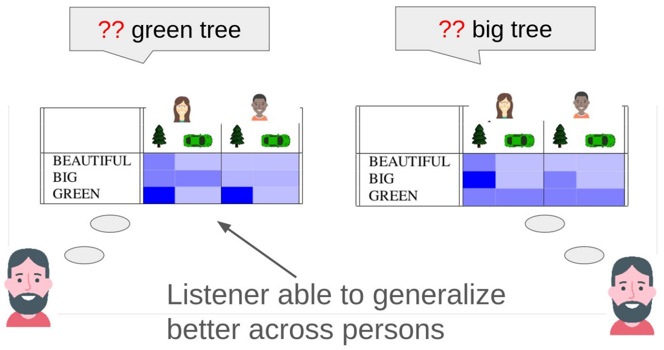


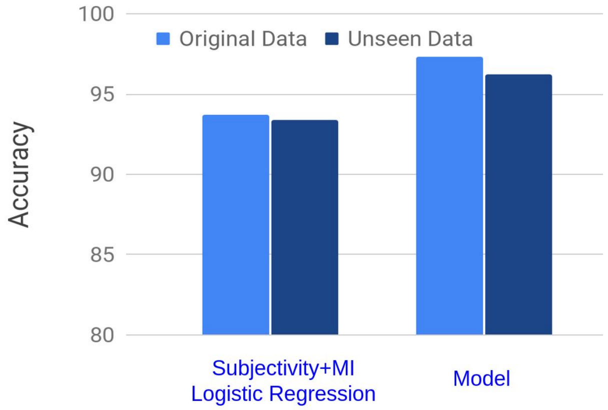


Conclusion

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.

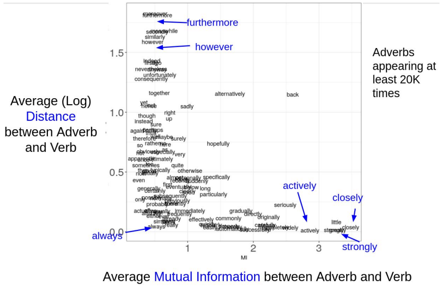
Conclusion

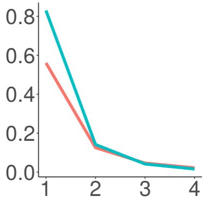
Subjectivity and MI independently impact adjective ordering.

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!