

# Some adaptive properties of the lexicon with respect to efficient communication

Harry Tily

joint work with Steve Piantadosi & Ted Gibson

Brain and Cognitive Sciences  
Massachusetts Institute of Technology

## Question

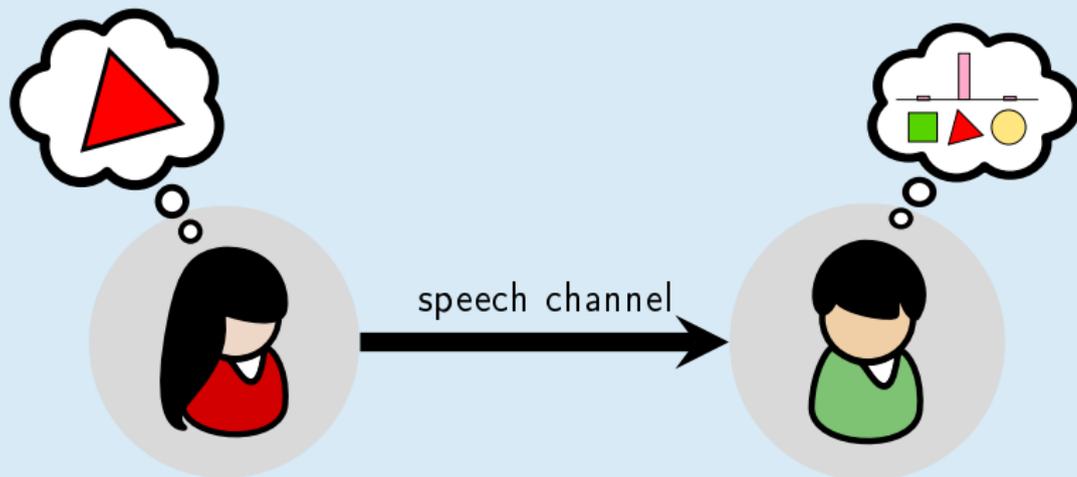
Is human language a “good solution” to the problem of communication?



*Is [language] well designed for [...] communication? I think that's the wrong question. [...] If you want to make sure that we never misunderstand one another, for that purpose language is not well designed, because you have such properties as ambiguity. If we want to have the property that the things that we usually would like to say come out short and simple, well, it probably doesn't have that property.*

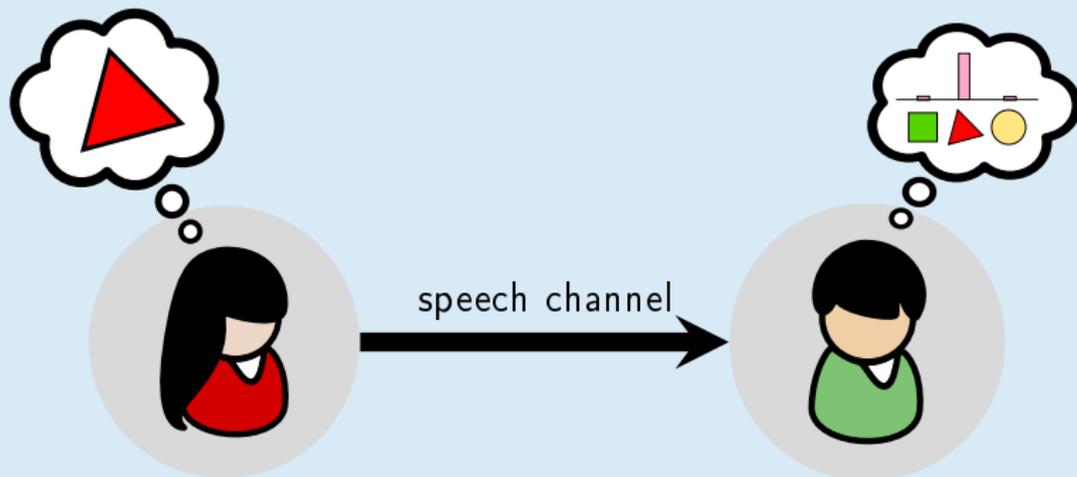
*(Chomsky, 2002:p107)*

# Communication



transmission is *effortful*, and the speech channel is *noisy*

# Communication



transmission is *effortful*, and the speech channel is *noisy*

# Rational Analysis

- rational analysis of cognition

(Newell, 1990; Anderson, 1991; Oaksford & Chater, 1998; cf Simon, 1957, 1992)

Chomsky's (2010) *Strong Minimalist Thesis*: human language is the *most efficient mapping possible* between meaning and sound

*Behavior cannot be predicted from optimality criteria without information about the strategies and knowledge agents possess or acquire. The study of the behaviour of an adaptive system is not a logical study of optimization but an empirical study of the side conditions that place limits on the approach to the optimum.* (Simon, 1992)

# Rational Analysis

- rational analysis of cognition

(Newell, 1990; Anderson, 1991; Oaksford & Chater, 1998; cf Simon, 1957, 1992)

Chomsky's (2010) *Strong Minimalist Thesis*: human language is the *most efficient mapping possible* between meaning and sound

*Behavior cannot be predicted from optimality criteria without information about the strategies and knowledge agents possess or acquire. The study of the behaviour of an adaptive system is not a logical study of optimization but an empirical study of the side conditions that place limits on the approach to the optimum.* (Simon, 1992)

# Rational Analysis

- rational analysis of cognition

(Newell, 1990; Anderson, 1991; Oaksford & Chater, 1998; cf Simon, 1957, 1992)

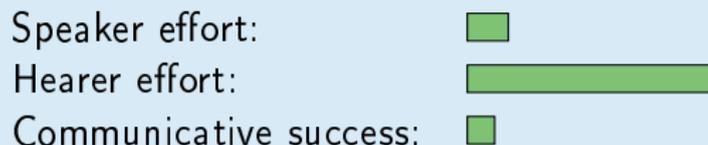
Chomsky's (2010) *Strong Minimalist Thesis*: human language is the *most efficient mapping possible* between meaning and sound

*Behavior cannot be predicted from optimality criteria without information about the strategies and knowledge agents possess or acquire. The study of the behaviour of an adaptive system is not a logical study of optimization but an empirical study of the side conditions that place limits on the approach to the optimum.* (Simon, 1992)

## Some logically possible languages

### A super-concise language (“telepathy”)

- all meanings are given the same sound



## Some logically possible languages

### A concise but taxing language

- each meaning given a subtly different single phone



## Some logically possible languages

A clear but tedious language (“legalese”)

- every meaning has a very distinct wordform

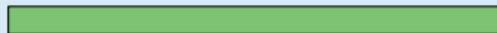
Speaker effort:



Hearer effort:



Communicative success:



# G. K. Zipf



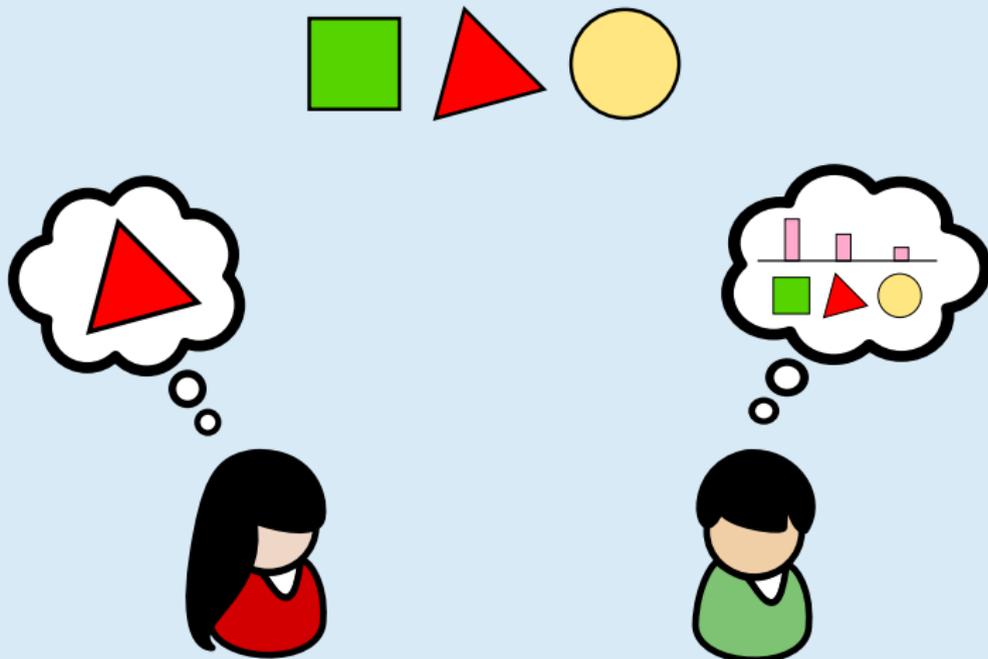
- Zipf's (1929; 1935) *Principle of Least Effort*
  - language can be expressed more concisely if the most frequent words are short
  - can't make all words equally short
    - but at least shorten those that get used most

# An information theoretic restatement

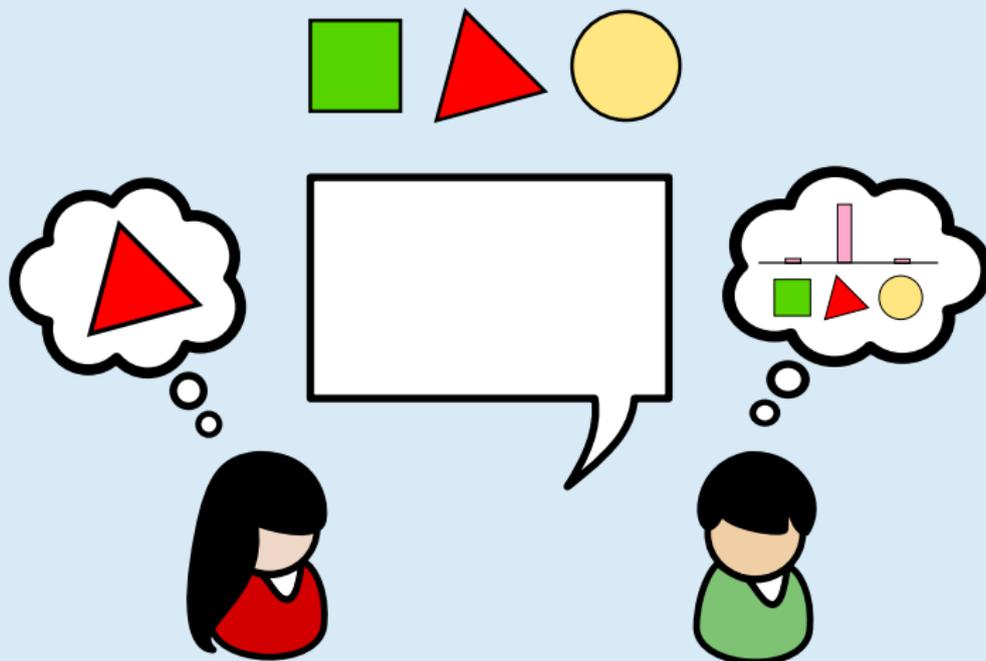
- Shannon's (1948)  
*Source Coding Theorem*
  - optimal code length for a meaning  $x$  is proportional to the *information* it carries,  
 $I(x) = -\log P(x)$
- and  $P(x) \approx$  frequency



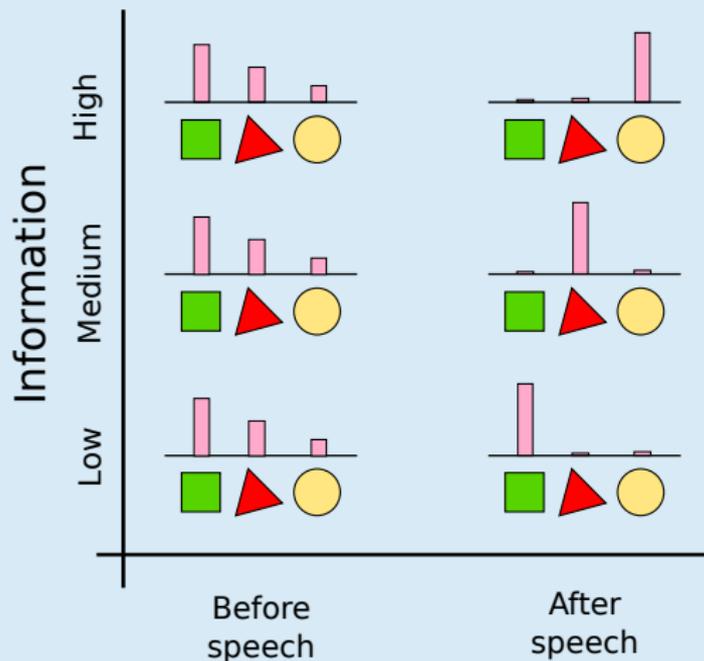
# Communication as belief update



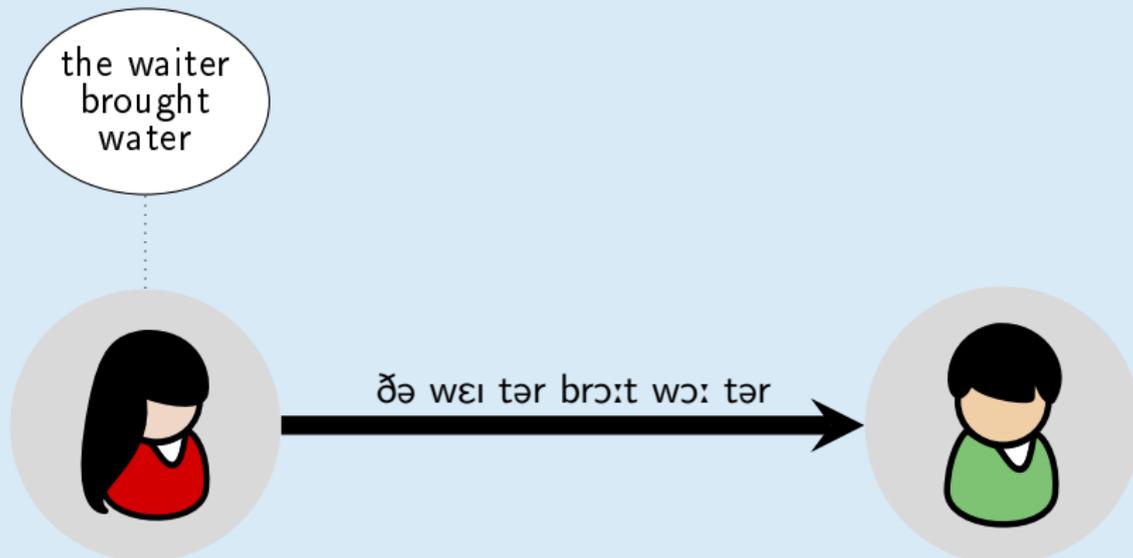
# Communication as belief update



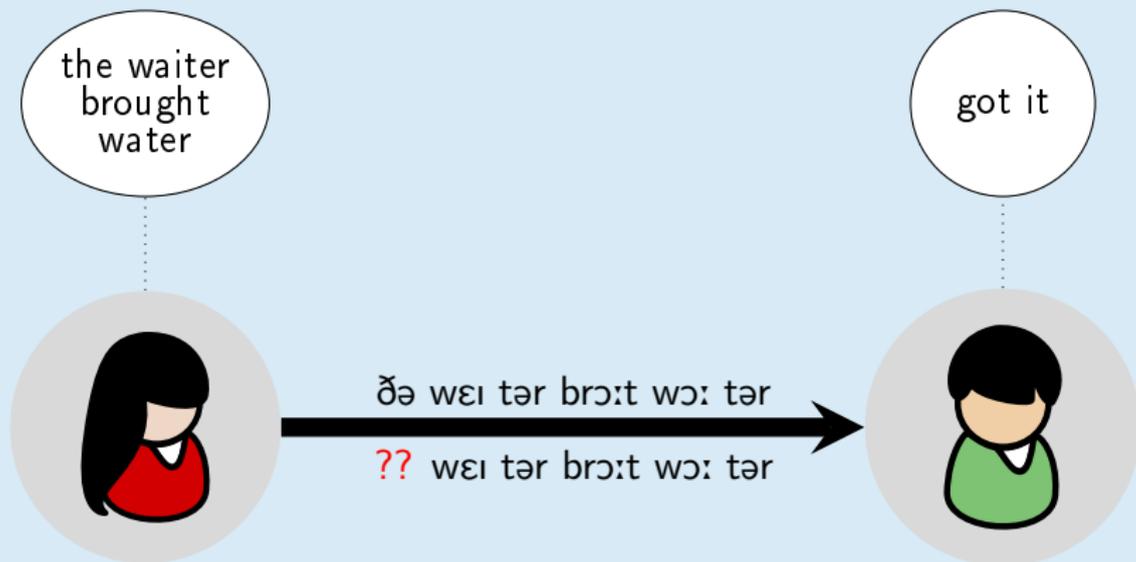
## Information in belief update



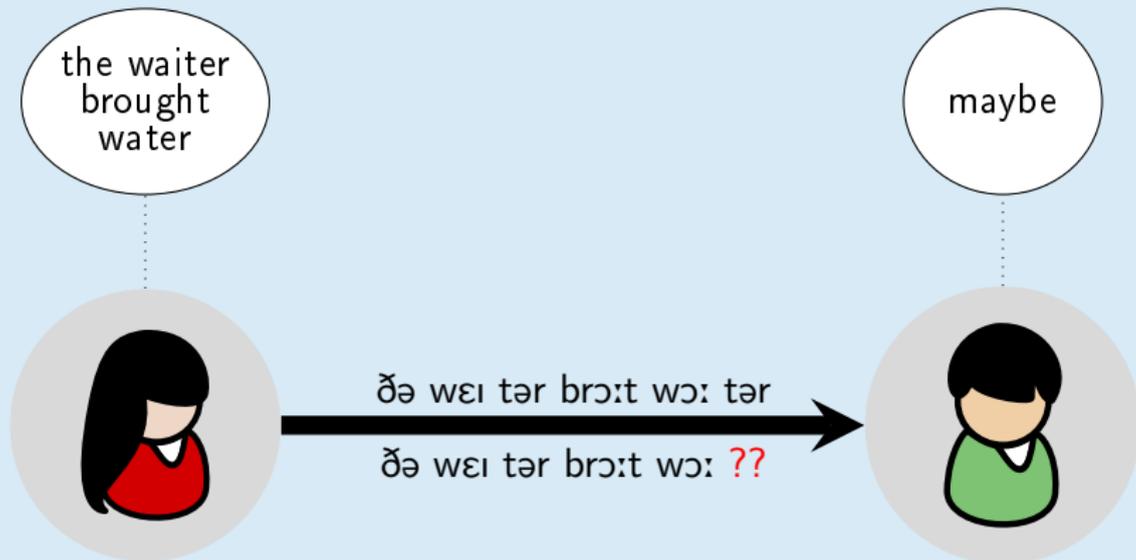
# The noisy channel



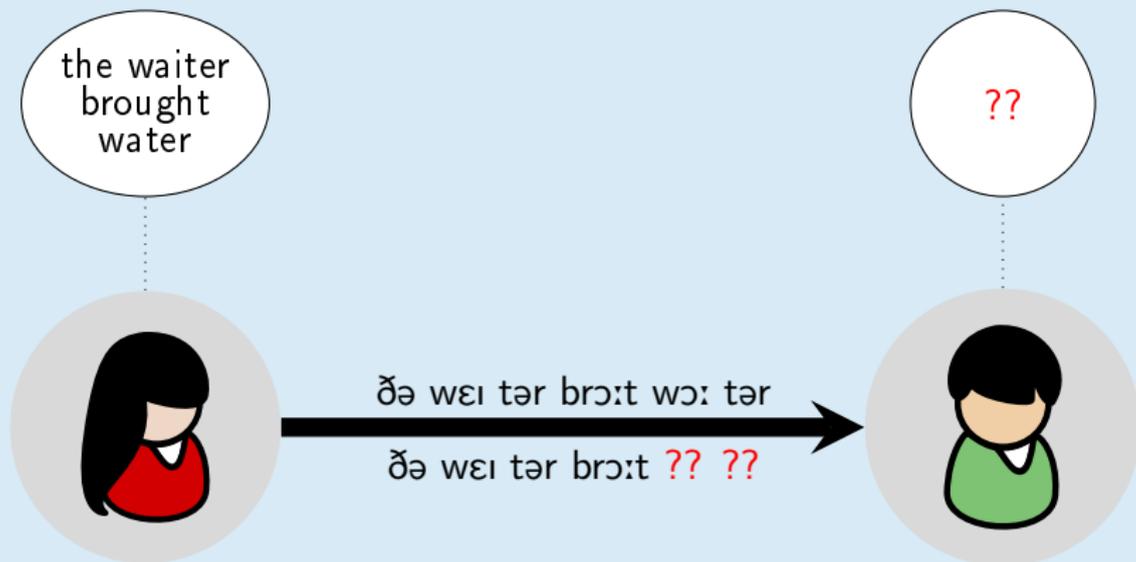
# The noisy channel



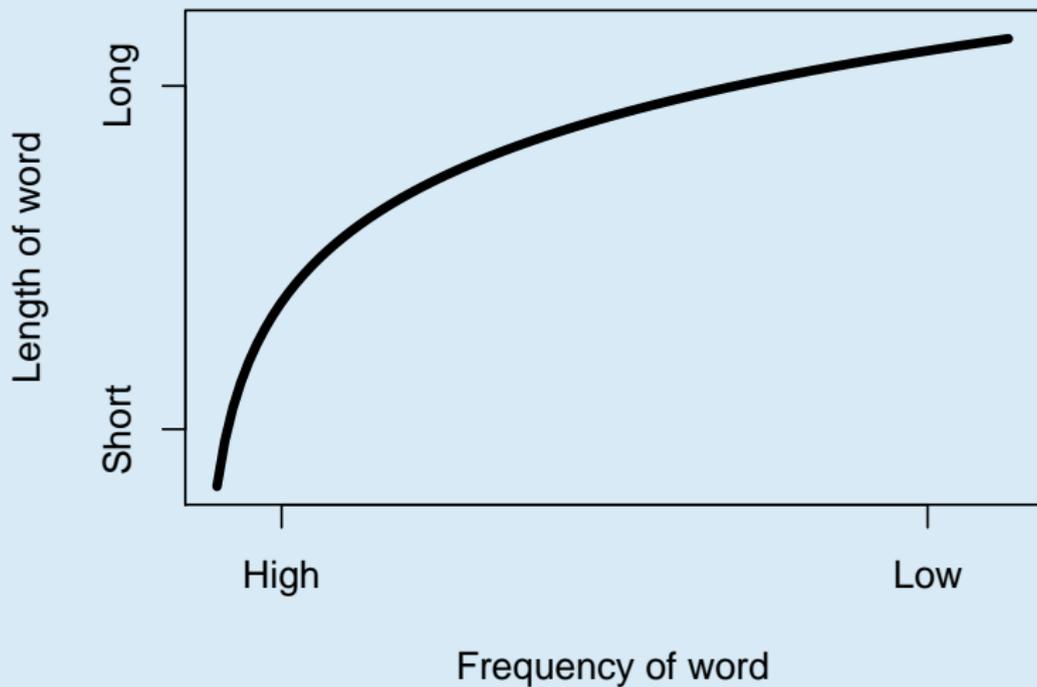
# The noisy channel



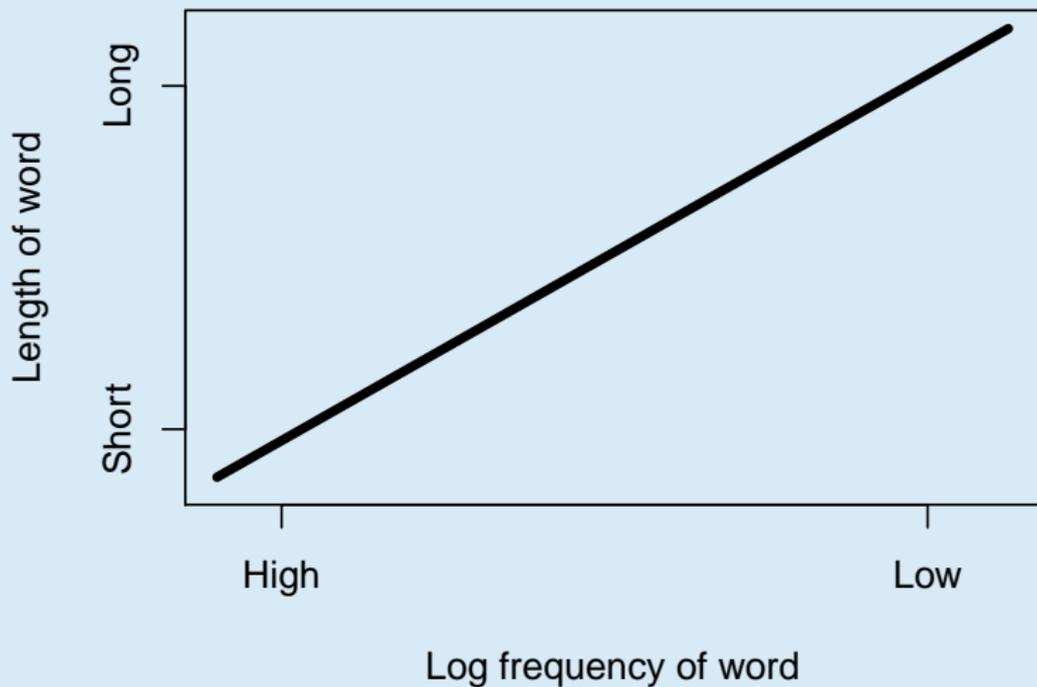
# The noisy channel



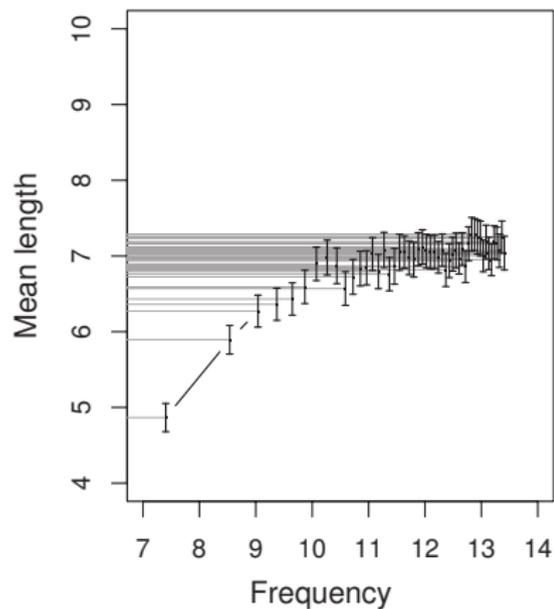
## Expected pattern



## Expected pattern



# English word lengths



$\log P(w)$

## Information and Frequency

if words are independent:

$$I(w) = -\log P(w)$$

if words are nonindependent:

$$I(w) = -\log P(w|\text{previous context})$$

## Amending Zipf's theory

Given that words are nonindependent,

- frequency-length correlation → pressure for *shortness*
- does *information* or *frequency* correlate better with length?
  - space out information evenly
  - don't overtax channel capacity

(see Jaeger, 2006; Levy & Jaeger 2007)

# Calculate the information a word carries on average

- Google database: 1 trillion words of web text
- estimate probability of each word given the two that precede it

$$\hat{I}(w) = \sum_c P(c|w) \cdot -\log P(w|c)$$

- information of  $w$  in context  $c$
- weighted by the frequency with which  $c$  is the context for  $w$

$$\approx -\frac{1}{N} \sum_{i=1}^N \log P(w|c_i)$$

where  $N$  is the total count of  $w$  in the corpus,  $c_i$  is the context for the  $i$ th instance of  $w$

# Calculate the information a word carries on average

- Google database: 1 trillion words of web text
- estimate probability of each word given the two that precede it

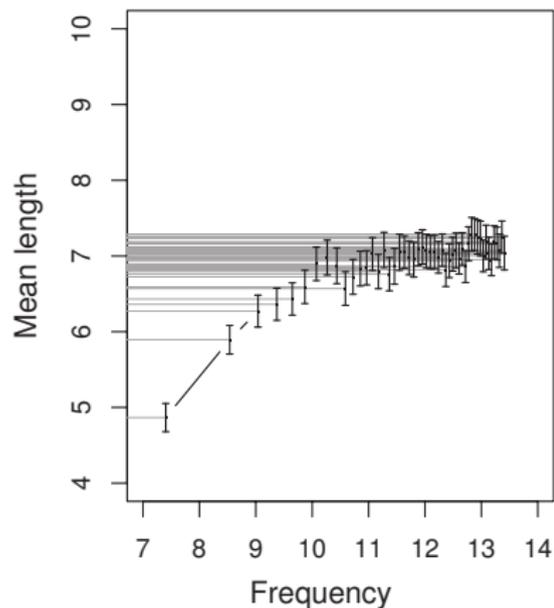
$$\hat{I}(w) = \sum_c P(c|w) \cdot -\log P(w|c)$$

- information of  $w$  in context  $c$
- weighted by the frequency with which  $c$  is the context for  $w$

$$\approx -\frac{1}{N} \sum_{i=1}^N \log P(w|c_i)$$

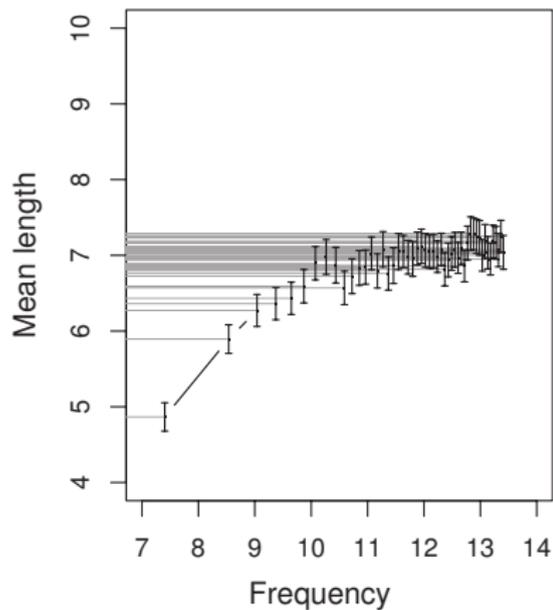
where  $N$  is the total count of  $w$  in the corpus,  $c_i$  is the context for the  $i$ th instance of  $w$

# English word lengths

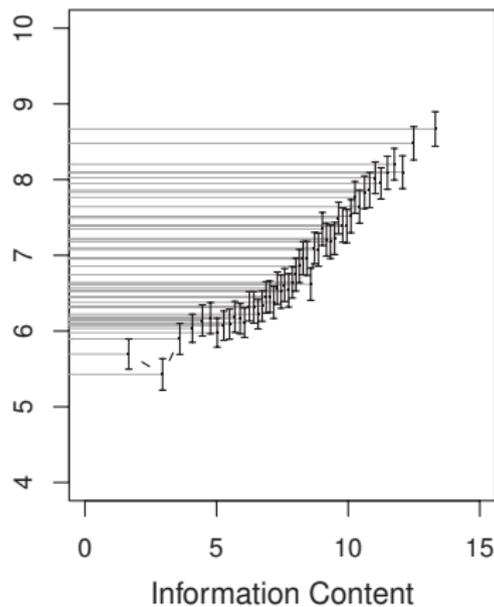


$\log P(w)$

## English word lengths

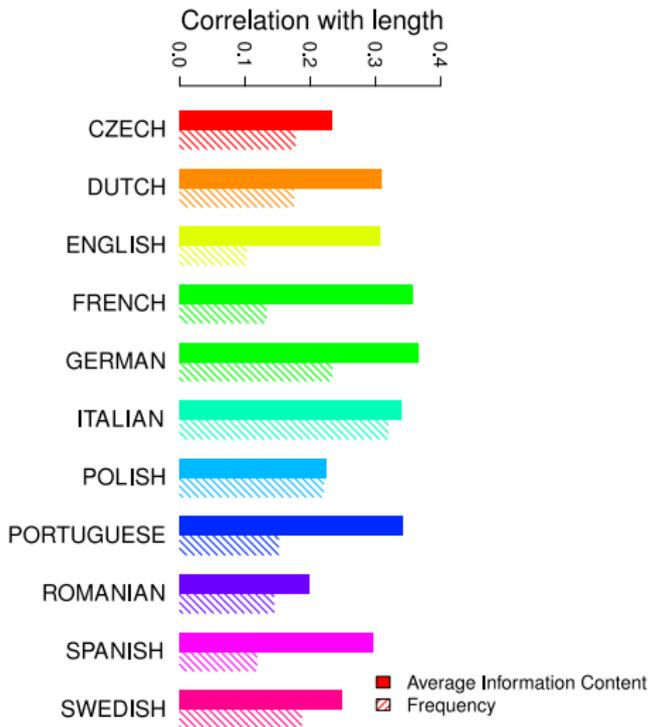


$$\log P(w)$$

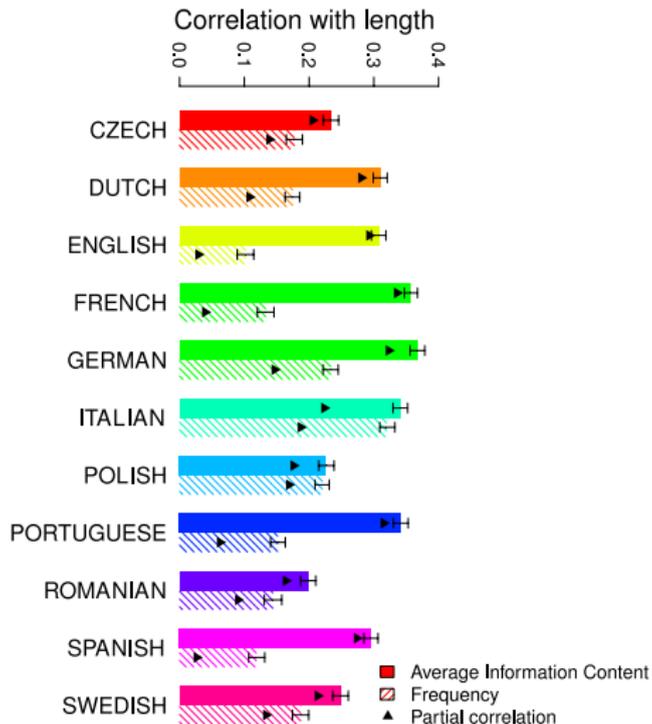


$$\sum_c P(c|w) \log P(w|c)$$

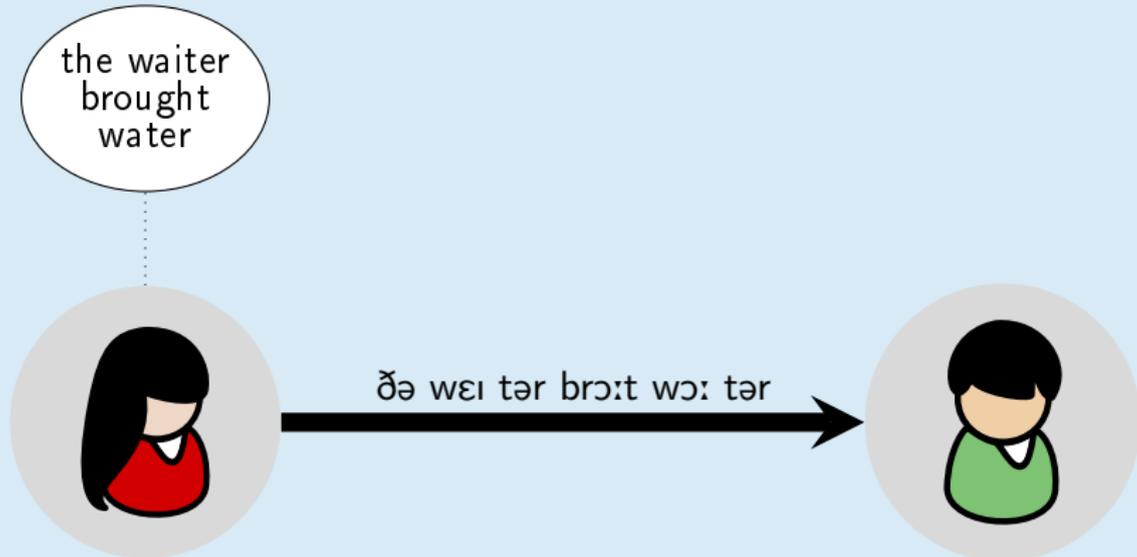
# Cross-linguistically



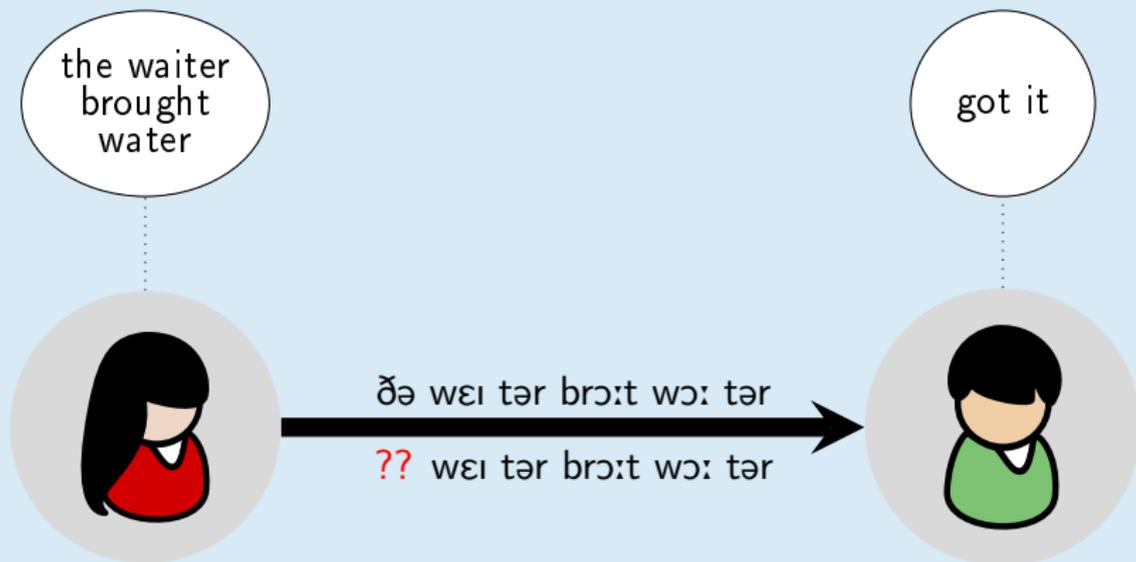
# Cross-linguistically



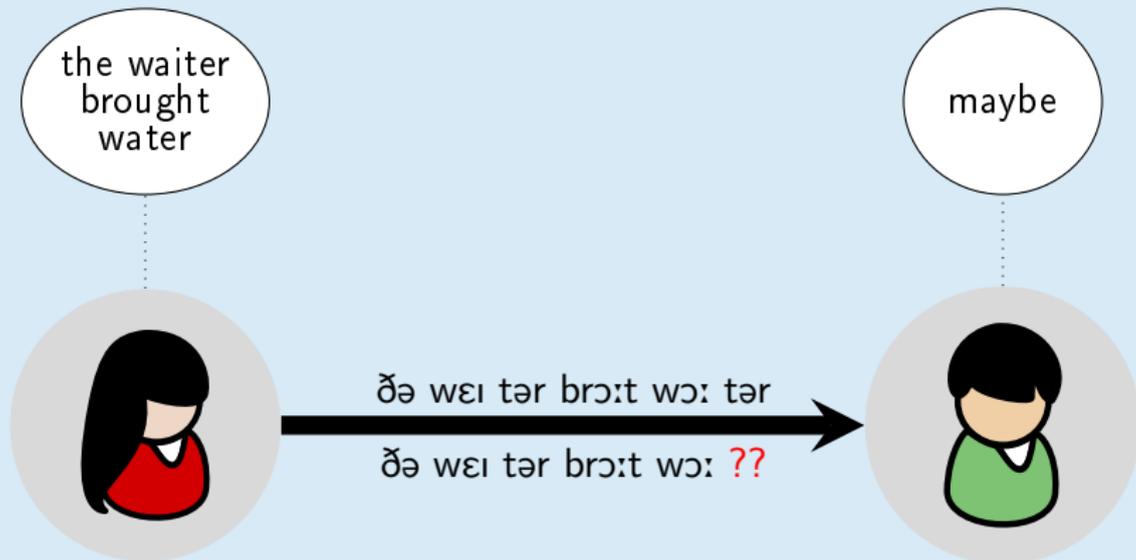
## Back to the noisy channel



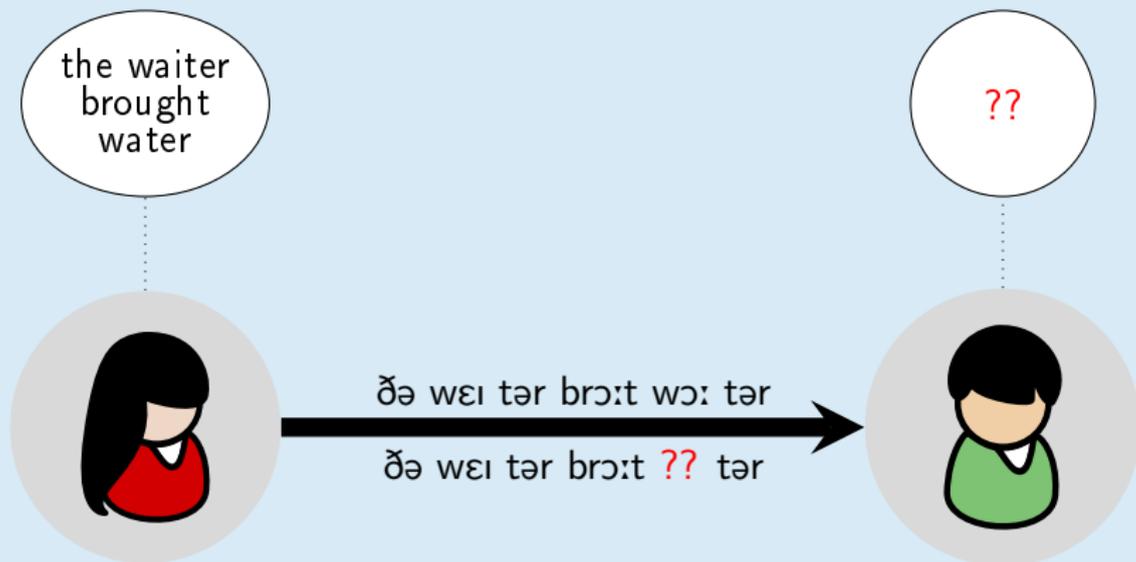
## Back to the noisy channel



## Back to the noisy channel



## Back to the noisy channel



# Lexical stress

Calculate the information transmitted by each syllable in each word

- “How much does hearing just the syllable /wɔː/ help you identify the word?”

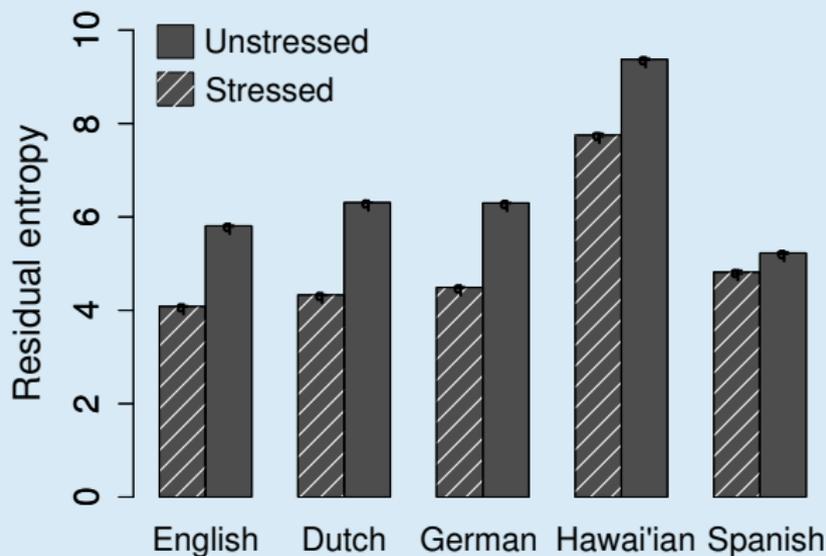
## Definition of residual entropy

- log probability (information) of word  $w$  given that it contains the syllable  $s$

$$H(w|s \in w) = \sum_{w:s \in w} [P(w|s \in w) \cdot -\log P(w|s \in w)]$$

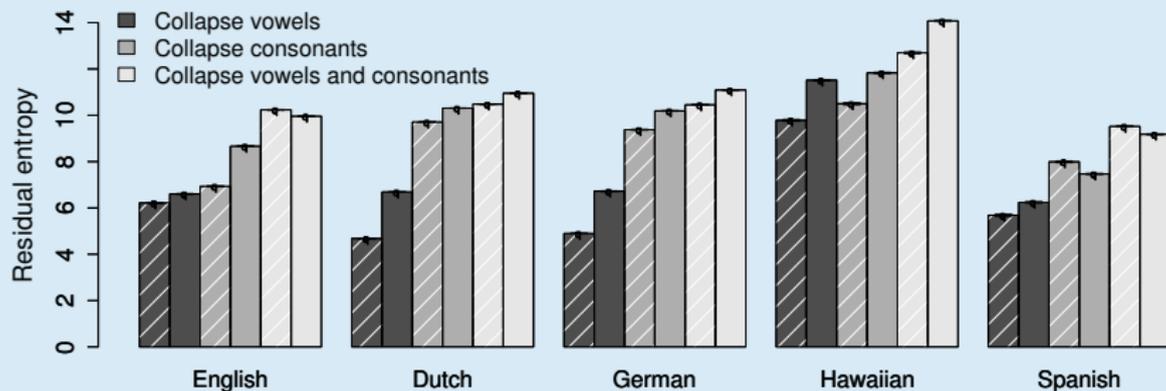
- weighted by that same probability

## Uncertainty left after hearing one syllable



# Is this only a vowel reduction effect?

English unstressed vowels tend to reduce to schwa ( $\text{ə}$ )



# Ambiguity

## bank

- noun*
1. The land alongside or sloping down to a river or lake
  2. A financial establishment that invests money deposited by customers

If context is informative about meaning, then words from an efficient code will look ambiguous out of context

# Ambiguity

## bank

- noun*
1. The land alongside or sloping down to a river or lake
  2. A financial establishment that invests money deposited by customers

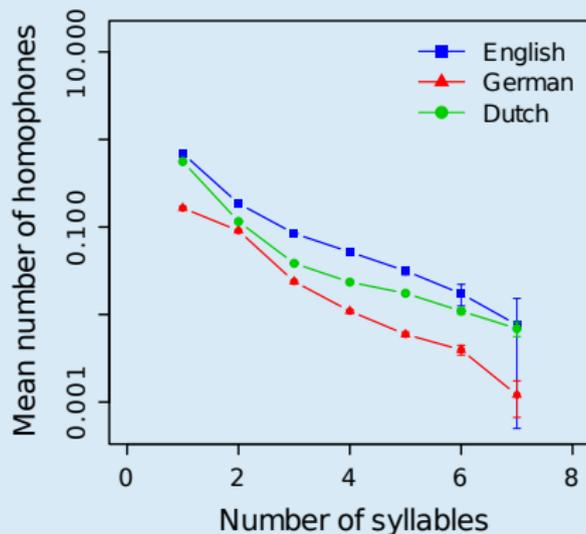
If context is informative about meaning, then words from an efficient code will look ambiguous out of context

# Ambiguity

## Hypothesis

Languages should re-use (make ambiguous) their most efficient words

# Homophony and length



All languages  $p < 0.001$  (quasi-Poisson regression)

# Referring expression choice

## Pronouns

## Descriptions

him

an employee

she

the individual concerned

it

our country

theirs

the corporation's products

them

hot-dipped galvanized steel products

# Information and efficient language

## Hypothesis

Language users choose referring expressions that are *appropriate to comprehenders' uncertainty*

- pronouns should be favored in *low uncertainty* situations

# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

Mr. Stone thought...



# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

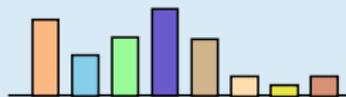
Mr. Stone thought the discipline was unfair;...



# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

Mr. Stone thought the discipline was unfair; he believed that...



# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

Mr. Stone thought the discipline was unfair; he believed that his manager wanted to get rid of...



# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

Mr. Stone thought the discipline was unfair; he believed that his manager wanted to get rid of him for...



# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

Mr. Stone thought the discipline was unfair; he believed that his manager wanted to get rid of him for personal reasons.

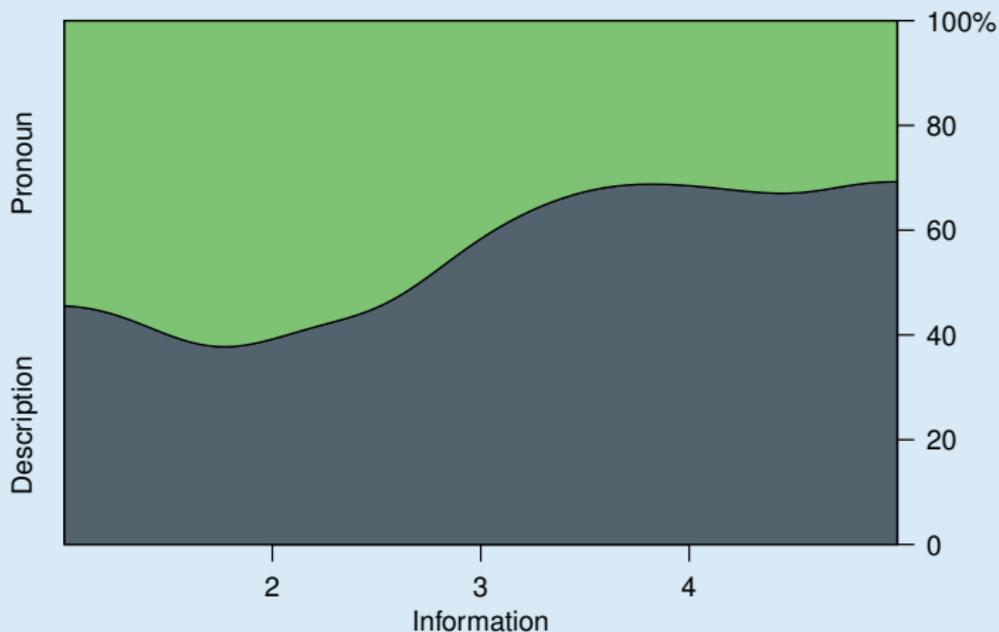
# The guessing game

Bob Stone stewed over a letter from his manager putting him on probation for insubordination.

Mr. Stone thought the discipline was unfair; he believed that his manager wanted to get rid of him for personal reasons.

- 82 newspaper articles
- total of 2211 NPs
- 50 participants see each

# Expression type as a function of information



Pronoun use decreases with information ( $p = .02$  by mixed effect regression)

## Regression model of expression choice

Does predictability influence expression choice *even controlling for other factors?*

- multinomial logit regression
- predictors
  - NP number & sentence number in text
  - number of referents mentioned so far
  - number of mentions to this referent so far
  - distance since last mention
  - grammatical function
  - coref with subject, object of previous clause
  - previous mention expression type
  - screen area of correct response (random click probability)
  - **information** ( $p = .02$ )

# Summary

- Language capitalizes on hearer's ability to infer meaning
  - saving the shortest words for predictable meanings
  - putting less information in unstressed syllables
  - re-using more efficient wordforms
  - allowing more or less detailed expressions of the same meaning
- This is a *rational strategy!*
  - minimize utterance length & complexity
  - while bounding miscommunication rate