Word lengths are optimized for efficient communication

Steven T. Piantadosi, Harry Tily, and Edward Gibson


1984
GEORGE ORWELL

Newspeak

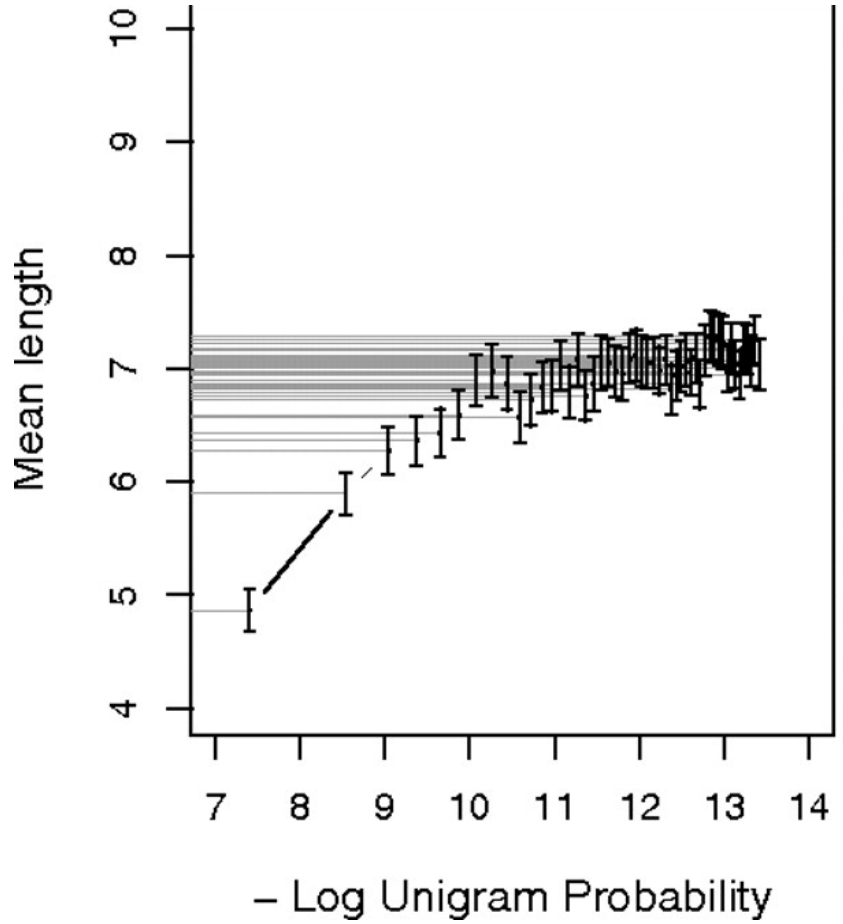
- steal, stole – steal, stealed
- good, bad – good, ungood
- warm, cold – warm, unwarm
- Adjectives: add –ful.
  - rapid – speedful
- Adverbs: add –wise.
  - quickly – speedwise
- Purpose: thought control, through streamlining and efficiency in language

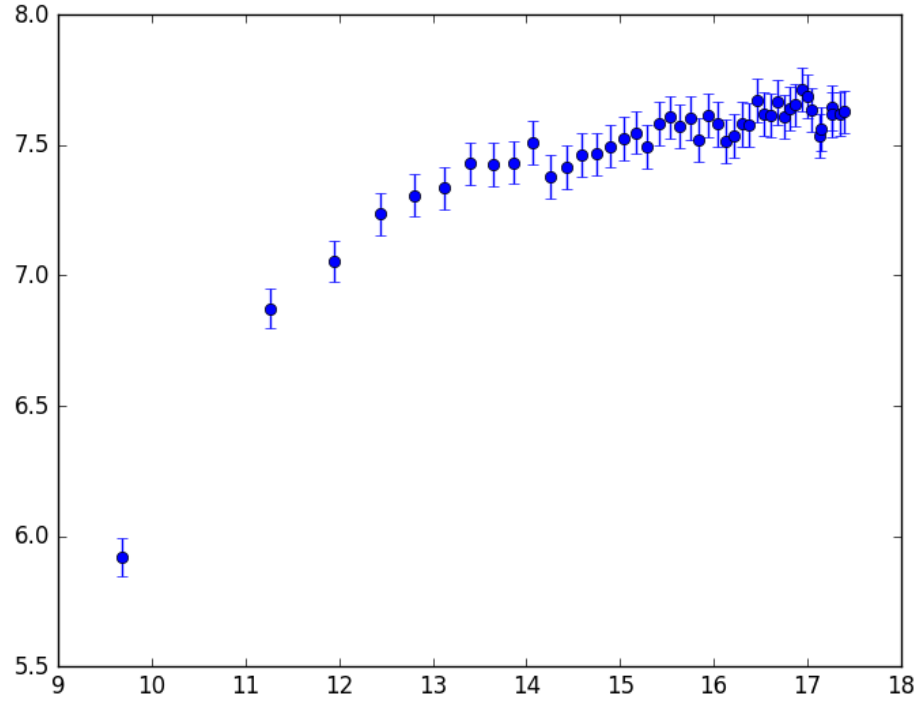
How efficient or optimized is human language?

Word lengths are optimized for efficient communication

- Starting point: a theory by George Kinsley Zipf
- *Not* Zipf's law: frequency of a word is inversely proportional to its frequency rank
- Separate theory by Zipf: the **length** of a word is inversely related to its **frequency**
- Why?
  - Zipf's principle of least effort: less effort if common words are shorter
- Is it true?

Word Frequency vs. Word Length





If not frequency, then what?

- Authors' hypothesis: word length is based on *information content*

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

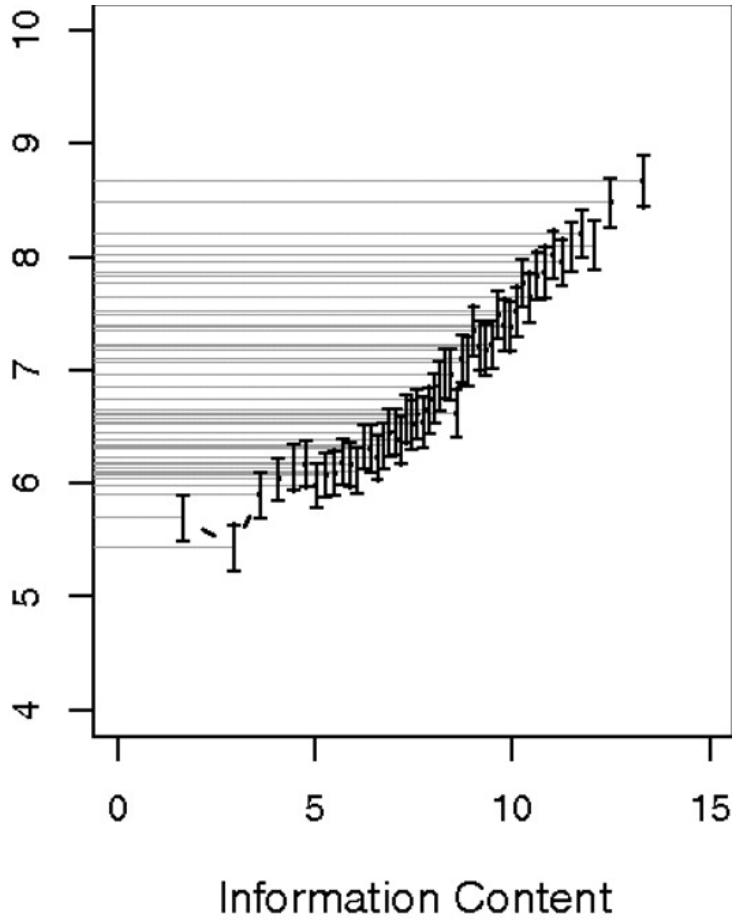
- where P(W = w | C = c_i) is the probability of the word given context *I*
- In this paper, context is the preceding *n* words

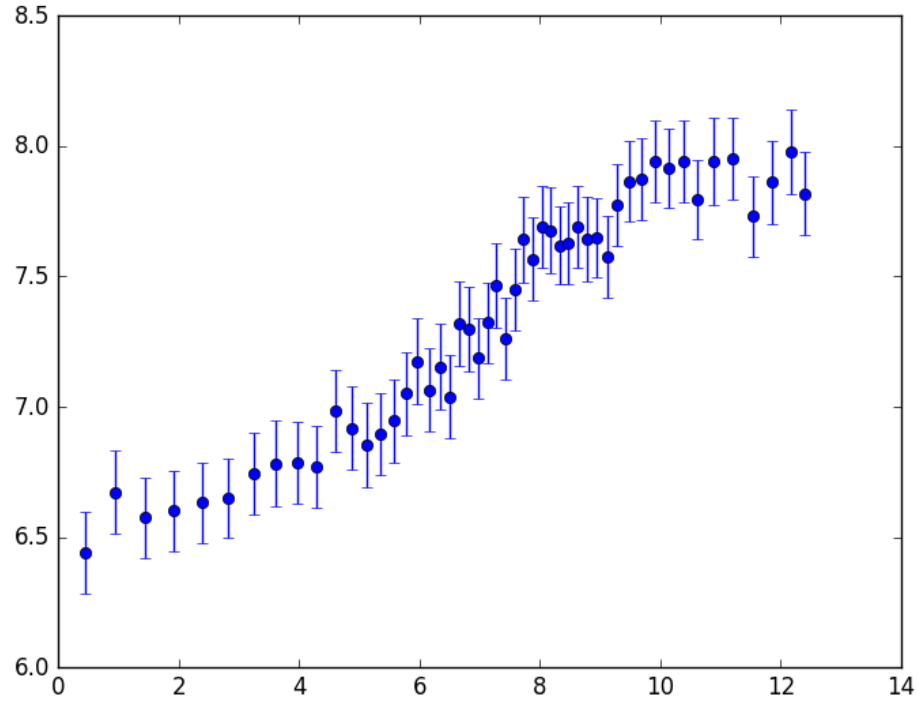
Example

- the cute, fuzzy _____
- **puppy**: higher probability, lower negative log probability, less surprisal, *less informative*
- **scorpion:** lower probability, higher negative log probability, more surprisal, *more informative*

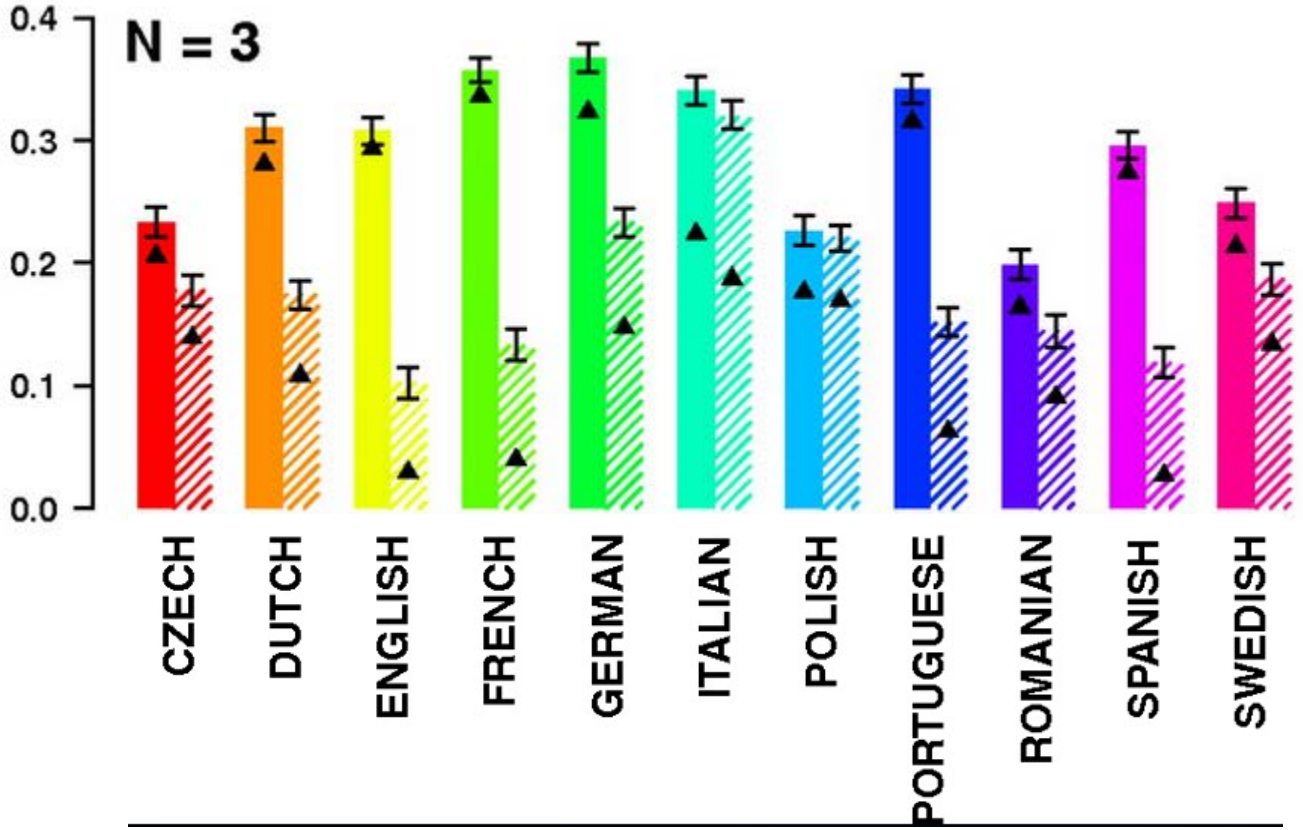
$$-\frac{1}{N}\sum_{i=1}^{N}\log P(W=w|C=c_i)$$

Informativeness vs. Word Length





Better correlation – across languages



Why?

- The principle of uniform density: a tendency, when creating language, to keep the number of bits per unit of time constant
  - Aylett M, Turk A (2004) The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. Lang Speech 47:31–56.
- Let the length, in characters, of a word be a proxy for how long it takes to produce it
- Longer words spread their higher information content across its length
- Shorter words have low information content
- Using short and long words together results in a constant information rate

Frequency and Information Content Related

- A word may be frequent *because* it has low information content
  - Zipf's theory isn't *entirely* wrong
- Authors use partial correlation to separate contributions of *frequency* and *information* to word length
- Conclusion:
  - Words tend to be shorter when they are less informative
  - Information content -> word length
  - *is better than*
  - Information content -> frequency -> word length

Conclusions

- Zipf’s theory *updated,* not necessarily disproven
- Orwellian language-engineering is a bad idea
  - “really bad” is more efficient than “double plus ungood”

Discussion

1. All languages? Bias in dataset for European languages. 5 Romance, 4 Germanic, 2 Slavic.
2. Can languages be artificially optimized for efficient communication? Esperanto; a “common language” a la Lazebnik
3. Can artificial languages be optimized for efficient communication? C, Java, Python
4. How did this optimization come about? Evolution analogy: what were the selective forces, how to characterize fitness?