


What Google Books Ngram Corpus reveals about adjective order in English



By Angie Garcia and Amanda
Robinson

Introduction

a green great dragon vs. a great green dragon

Introduction

Theoretical question:

How are adjectives ordered in English?

Talk Overview

1. Theoretical Background
2. Study
3. Conclusion

Theoretical Background

The rule of adjective ordering: **Subjectivity Hypothesis**

- Adjectives belong to different semantic classes.
- Different classes can be ranked on a scale in terms of their **subjectivity**.
- Adjectives that express the speaker's personal perspective are more subjective.
 - *Beautiful, great* are more subjective
 - *green* is less subjective

Theoretical Background

Subjectivity Hypothesis:

(1) **Evaluation** > **Color** > **Origin**

a **great**

green

Noun
dragon

a **beautiful**

English

rose

Study: Goal

GOAL: To test the predictions of the Subjectivity Hypothesis by analyzing how adjectives are distributed in Google Books Ngram Corpus.

- Google Books Ngram Corpus: largest collection of **5,195,769** digitized books spanning from 1800 - 2008.

Study: Design & Procedure

- Step 1: We created a dictionary of English adjectives (14,000 words), using WordNet as a source.
- Step 2: Downloaded 195 bigrams from Google Books Ngram Corpus.
- Step 3: Created a corpus of adjectival bigrams: 120,000+ unique bigrams.

For each step of this project we wrote a custom computer program.

Study: Design & Procedure

- We selected 326 individual adjectives that appeared both as first and second words in bigrams.
- We assigned each adjective to a semantic class.
 - Example
 - Green = Color
 - Great = Evaluation
- The final list consisted of 140 adjectives that were assigned to the same semantic class by both authors.

Study: Results

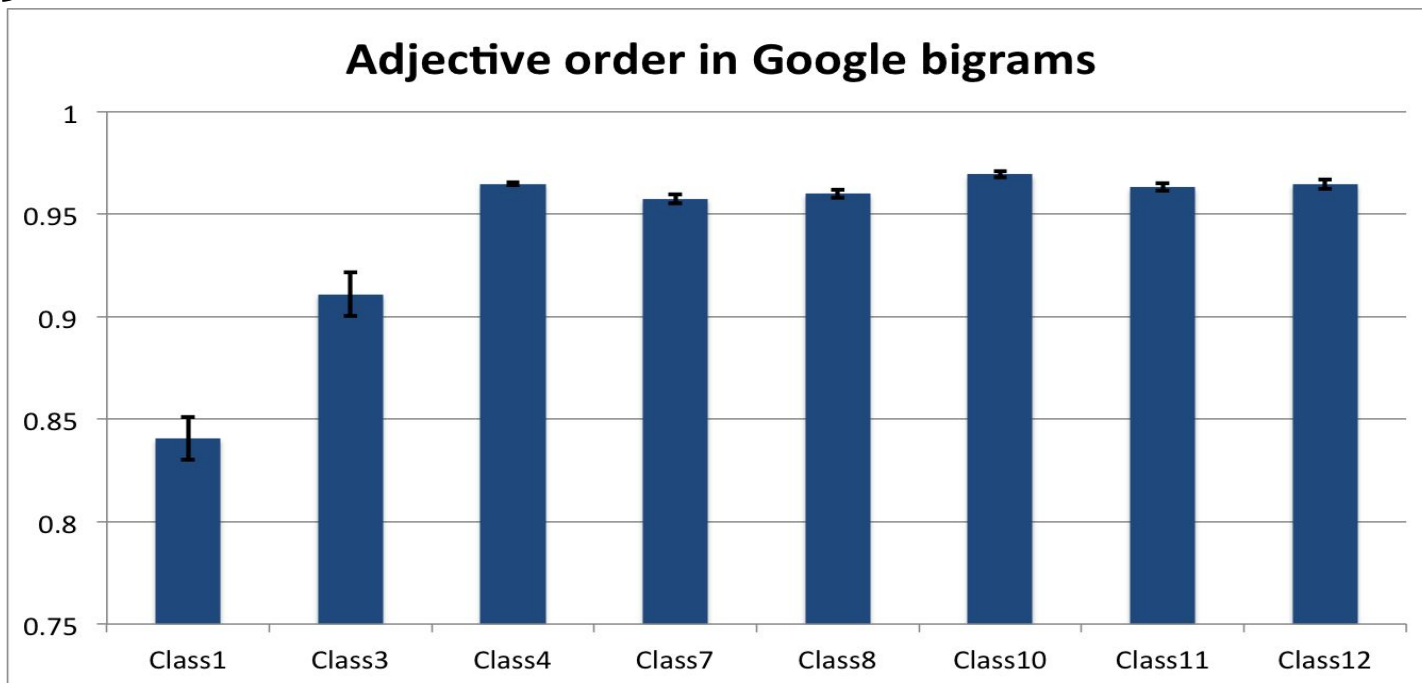


Figure 1. Adjective order in Google bigrams. The numbers on the y-axis represent the proportion of times adjectives are mentioned first in google bigrams. Error bars represent +/-1SE.

Study: Results

- The results are highly significant.
(Pearson's $r(138) = .5, p < .001$).
- Results support subjectivity-based approach to adjective ordering.

Conclusions

- **Theoretical contribution:** We evaluated the theories of adjective ordering in English, and specifically provided support for the Subjectivity Hypothesis.
- **Methodological contribution:** We propose a new method for evaluating the theoretical approach to adjective ordering using Google Books Ngram Corpus.

Conclusions

- Adjective order is an example of “hidden rules” in language: all speakers use this rule but cannot explain it.
- A better understanding of “hidden rules”, such as adjective order, provides Insights on how the human language faculty works.
- A better understanding of language takes us one step closer to understanding who we are as a species.

Thank you!

Acknowledgments

We are grateful to Ricardo Romero-Sanchez, Gregory Ruffin, Alex Lenarsky, Professor Anastasia Smirnova, and other members of the Experimental and Computational Linguistics Lab for their comments and feedback.

Reference

1. Hetzron, R. (1978). On the relative order of adjectives. In H. Sella (Ed.), *Language Universals*. Tübingen: Narr.
2. Martin, J. E. (1969). Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior*, 8, 697-704.
3. Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Pinker, S. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176-182.
4. Miller, G.A (1995). WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.
5. Scontras, G., Degen, J., & Goodman, N. D. (2016). Subjectivity predicts adjective ordering preferences. *Open Mind* (to appear).