

Expressives in Composition

A computational approach

Joshua Bowles

February 12, 2010

1 What are Expressives and Why Should we Care?

Expressive words are lexical items that contain some emotive content. They do not, however, impact the truth or falsity of a statement (usually called propositional content).

Expressives are relevant to us because we typically want to present our research through writing that is both neutral and negotiates its emotional investment. Through using expressive terms we reveal bias by revealing emotional investment. This is certainly one way to argue against using expressive terms in scientific writing. But there is, I argue, an even stronger reason: as we will see in the examples below, expressive terms do not appear to be compatible with making precise statements. Use of expressive content seems to mask precision and encourage vague reference to the world and data.

2 Some Examples and Data

2.1 Examples

- (1) Expressive words used in this study
 - a. like, damn, soo, wow, crap, super, hell, awesome, totally, really, literally, idiot, stupid, dumb, amazing
- (2) Common expressives
 - a. swears, exclamatives, intensives
- (3) The cheetah can run **so** fast.
 - a. The cheetah can run **so** 60 miles per hour.
- (4) Pluto is **really** far away.
 - a. Pluto is **really** 5,913,520,000 kilometers away.

- (5) Toyota gas-pedal problems are **freakin'** rare.
 a. Toyota gas-pedal problems are **freakin'** 0.00001%.

Notice also that here we are addressing a problem that has come up before: vague language. We replace the terms *fast*, *far*, *rare* with measurable quantities. Notice what happens to the expressives – in some cases, problems (3-a) and (4-a), they appear to act like an appeal to/for credibility.

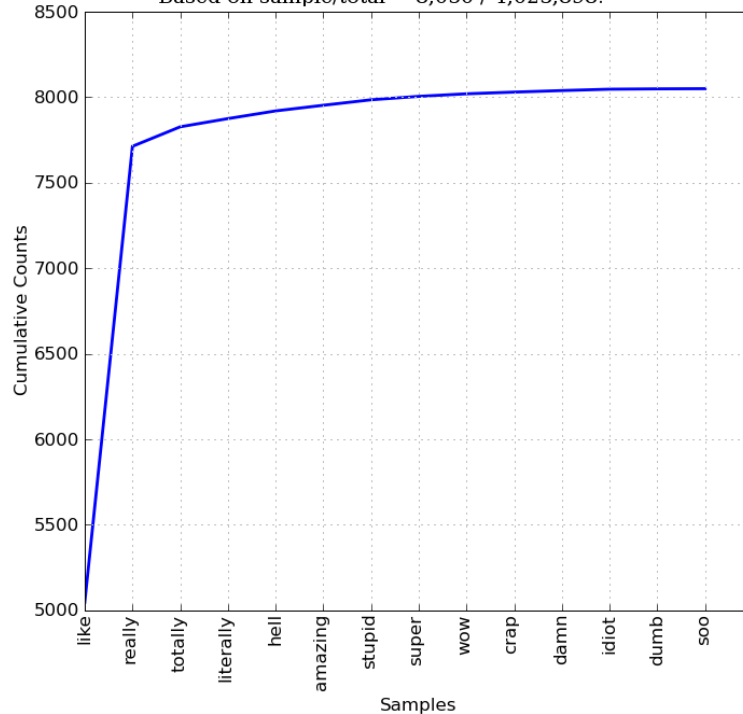
2.1.1 Description of Method

Frequency counts of the words in example (1-a) are made in two different collections of texts. One collection (or corpus) contains lectures by academics and runs a total of about 4 million words; the other corpus is based on collections of emails, blogs, chatrooms, and reviews (i.e., movie, product, restaurant reviews) and runs about 1 million. Cumulative counts can be seen in the following graphs.

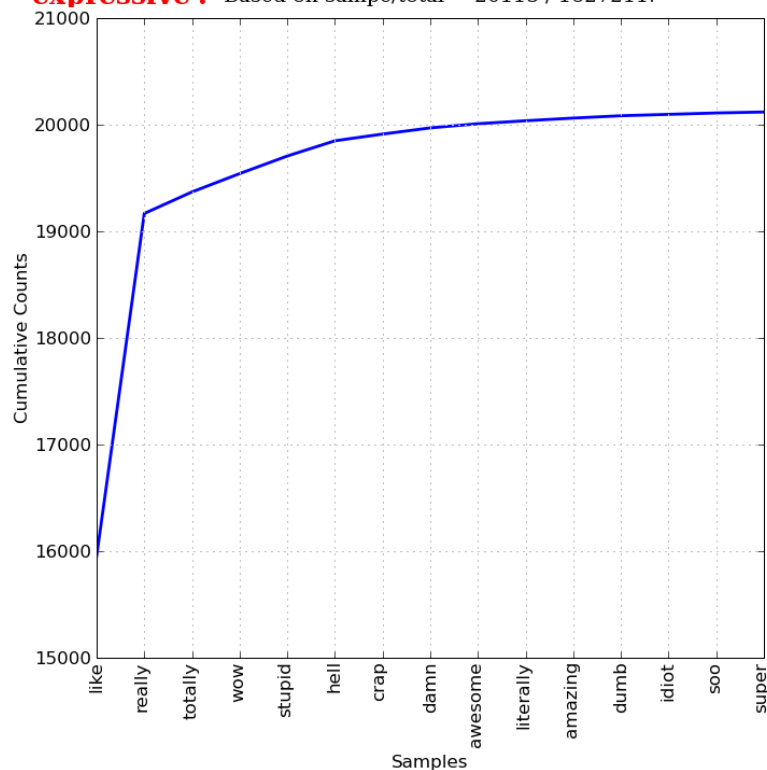
2.2 Cumulative Count Graphs

Academic corpus: 0.2% of words are 'expressive'

Based on sample/total = 8,050 / 4,023,898.



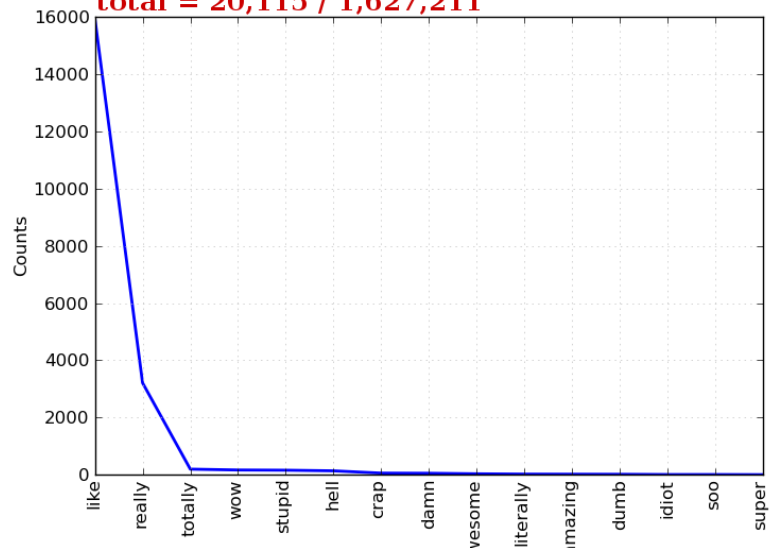
Chat, email, blog, reviews corpus: 1.2% are 'expressive'. Based on $\text{sampe}/\text{total} = 20115 / 1627211$.



2.3 Individual Word Count Graphs

What is interesting in these graphs is that the word *like* is the most frequent word in both corpora; this surprised me. There are also a number of problems with this very short study; I will name one here. The academic corpus consists of sampling from one pre-existing corpus which is composed entirely of academic lectures. This is problematic because we can assume with some confidence that one is more likely to use expressives in discussion versus writing—even in academic lectures. This means the percentage of 'expressive' words may actually be lower than 0.2%. Generally, the chat corpus has a better sampling, consisting of three pre-existing corpora I put together.

Chat, email, blog, reviews corpus (sample/
total = 20,115 / 1,627,211



Academic corpus: individual count (sample/
total = 8,050 / 4,023,898 = 0.2%)

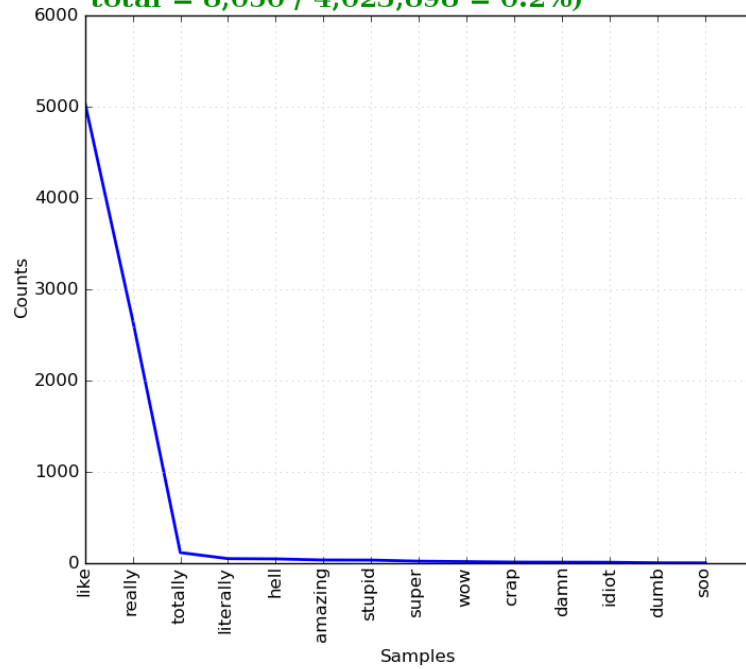


Table 1: Individual counts for 'expressives'

Word	Academic	Chat
like	5047	17010
damn	9	64
soo	1	13
wow	15	224
crap	10	63
super	20	10
hell	45	162
totally	114	206
really	2666	3305
literally	48	28
idiot	8	13
stupid	32	166
dumb	2	22
amazing	33	25
awesome	0	39