Extremely costly intensifiers are stronger than quite costly ones.

Erin Bennett (erindb@stanford.edu)
Department of Psychology, 450 Serra Mall , Stanford, CA 94305

Noah Goodman (ngoodman@stanford.edu)
Department of Psychology, 450 Serra Mall , Stanford, CA 94305

Abstract

abstract

Keywords: intensifiers; degree adverbs; scalar adjectives; pragmatics; m-implicature

Introduction

i need to clean this up

Intensifiers, for example "extremely" or "very", are adverbs that modify scalar adjectives to increase their degree. We argue that the meanings of intensifiers are likely not just arbitrarily paired with their words, but that some intensifiers have stronger meanings as a result of the cost it takes to utter them.

The specific meanings that scalar adjectives take can be inferred pragmatically from context (cite). The specific meaning of an intensifier might be inferred from context in a similar way.

Longer and more suprising intensifiers are more costly to utter, and so in this paper we look at how suprisal and syllable length relate to the strengths of different intensifiers. We find that longer, more surprising intensifiers tend to have stronger meanings.

The fact that the surprisal of an intensifier might influence its strength means that as the frequency and therefore the surprisal of an intensifier changes over time in a dialect, the strength will also change. An interesting consequence is that new intensifiers might continually need to be created to replace old, faded ones. We show that intensifiers are quite malleable and that people can learn that within a new "version of English" a particular intensifier is used much more frequently than in standard English, and they consequently infer a weaker meaning for the overused word.

ch i think
nave ev-
ce for...

Experiment 1

To explore the hypothesis that the interpretations of intensifiers are a function of their cost, we first wanted to see whether two possible ways of measuring the cost of a word, inverse frequency (rarer words being more costly) and syllable length, were correlated with the meanings of intensifiers.

Method[1]

40 participants with US IP addresses participated in our Experiment 1 on Amazon's Mechanical Turk.

We asked participants to give us judgements of prices based on a person's description of an object that included an intensifier. There were three categories of objects (*laptop*, *watch*, and *coffee maker*) and 40 intensifiers (listen in Table 1). We chose intensifiers to have a wide range of frequencies and excluded intensifiers that are either more commonly used to signal affect than to signal degree (e.g. "depressingly expensive" might indicate a degree, but it definitely indicates affect) or are ambiguous between other parts of speech (e.g. "super" can be used as an intensifier, as in "super expensive", but it can also be used as an adjective, as in "super hero"). Each particpant gave price judgements for every intensifier-category pairing in randomized order, for a total of 120 price judgements. We chose the domain of price and used only the adjective "expensive" (Figure 1), because price gave a quantitative scale on which to measure the different intensifers and because we thought participants would have similar enough experience with the distributions over prices for these objects.

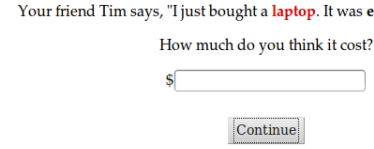

Figure 1: Screenshot from Experiment 1 target question.

Corpus Methods We collected length in syllables and frequency for every intensifier (Table 1) as measures of their cost. The frequency was collected from the Google Web 1T 5-grams database (Brants & Franz, 2006)[2]

Results and Discussion

If the meaning of an intensifier is stronger for higher cost intensifiers, we would expect to find that as surprisal and length in syllables increase, the prices participants give will also increase. We find that this is the case.

In a linear mixed effects regression with syllables and sur-

[1] The full experiment can be found at `http://web.stanford.edu/~erindb/degree-adverbs/experiments/exp5_2014-12-01/exp5.html`

[2] We also ran the same analyses on frequency information collected from the Google Books American Ngrams Corpus (Michel et al., 2011) as well, and found similar results.
In addition, we did the same using the bigram frequencies of "*[intensifer]* expensive" rather than the unigram frequencies of the intensifiers alone. These data were much more sparse. For bigrams, we found no significant effects of surprisal using the books database and a negative effect using the web database.

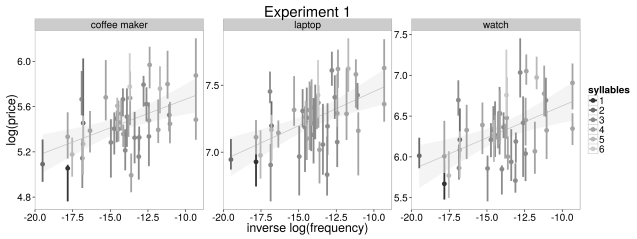

Figure 2: Results of Experiment 1. As surprisal and length in syllables increase, participants' free response prices increased.

prisal and their interaction as fixed effects[3] and random intercepts and slopes for syllables and surprisal for both participant and object, we found significant main effects of surprisal (estimate=0.054, p=0.012) and syllable length (estimate=0.093, p=0.0041) as well as a significant interaction (estimate=0.019, p=0.00018).

This suggests that intensifiers that are more surprising and longer (and therefore are more costly to utter) also have stronger meanings.

interpret interaction

make a big deal

Experiment 2

In Experiment 2, we replicated our finding from Experiment 1 using a different dependent measure which we expect to be more sensitive to small differences in meaning and an extension to other adjectival scales.

Method[4]

30 participants with US IP addresses participated in our Experiment 2 on Amazon's Mechanical Turk.

We divided the 40 intensifiers from Experiment 1 into four lists of 10 intensifiers each (Table 2). Each list was randomly paired with one of four adjectives ("old", "expensive", "beautiful", and "tall"). For each adjective-list pairing, participants were shown every combination of the 10 intensifiers and the one adjective on the left side of the screen. They were asked to move the adjective phrases from the left to the right side of the screen, reordering the phrases from the lowest to the highest degree (Figure 3). Each participant did four trials of this process, seeing all four lists and all four adjectives. The pairings between list and adjective were randomized between participants. The division of the intensifiers into lists of 10 was always constant, i.e. the same 10 intensifiers were always shown together.

[3]We centered both surprisal and syllable length by subtracting their means.

[4]The full experiment can be found at `http://web.stanford.edu/~erindb/degree-adverbs/experiments/exp4/exp4.html`

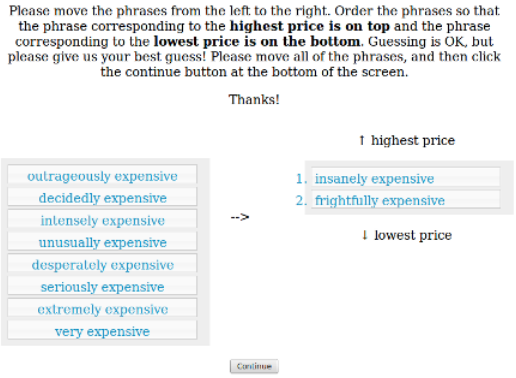

Figure 3: Screenshot from Experiment 2 target question.

Results and Discussion

turns out, breaking things up by adjective, there's no significant interaction between syllables and surprisal. and there is a slight difference between the surprisal slopes for the adjectives. "tall" and "expensive" are slightly steeper than for "beautiful" and "old" (Figure 4). The main effects of syllables and surprisal are robust and almost identical across the different adjectives (Figure 5).

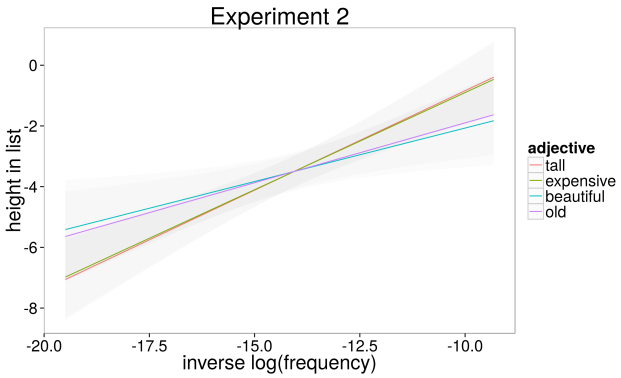

Figure 4: tall and expensive are steeper than beautiful and old

In a linear regression of ranking within the list as a function of surprisal, syllable length, and their interaction[5], we found significant main effects of surprisal (estimate=0.46, p<2e-16) and syllables (estimate=0.64, p=8.2e-11) as well as a significant interaction (estimate=0.069, p=0.046).

We again found that participants assign stronger meanings to intensifiers with higher surprisals and syllable lengths.

The relationship between strength and suprisal could be explained by the stronger meanings themselves being more unusual and surprising and therefore corresponding to words that are used less frequently. However, this story would not

[5]We centered both surprisal and syllable length by subtracting their means.

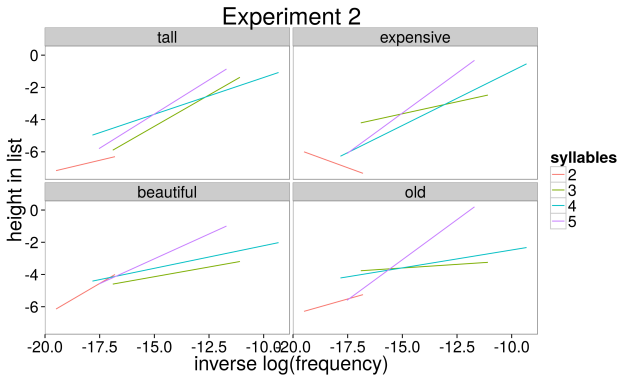

Figure 5: tall and expensive are steeper than beautiful and old

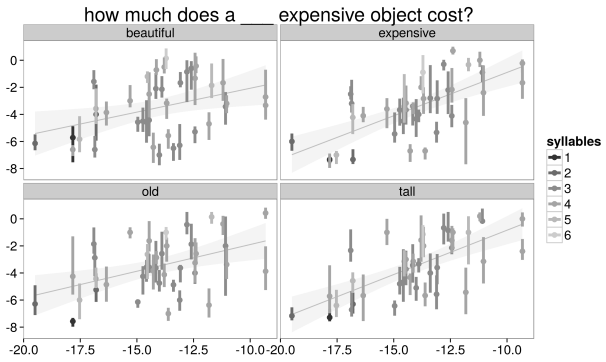

Figure 6: Results of Experiment 2. As surprisal and length in syllables increase, participants' rankings increased.

be able to explain why syllable length above and beyond surprisal would predict stronger meanings.

Experiment 3

In Experiment 3, we tested whether manipulating the surprisal associated with an intensifier can change the strength of an intensifier. If this were the case, it would provide more evidence for our hypothesis that the surprisal of a word causes its meaning to be stronger due to the added cost.

Method[6]

i need to clean this up

We looked at two short intensifiers of equal length: "truly" and "very". In our two conditions (varied between participants), each intensifier was either the target or the control. In a comic-style training story, the target intensifier was repeated 22 times by a speaker who lived "across the country" and had "a distinct way of speaking." The control intensifier was not used by the speaker in the story.

[6]The full experiment can be found at `http://web.stanford.edu/~erindb/degree-adverbs/experiments/exp8/exp8.html`

To guage whether participants had learned that the speaker's use of the target intensifier was unusually high, we asked participants how many times in the next 1000 words they thought the speaker would use the control word, the target word, and another word that had been frequently repeated.

To determine what strength participants inferred for the target and control intensifiers, we gave participants a final panel of the comic where the speaker described a new coffee maker he bought (Fig 7). We asked participants to guess the price of the coffee maker given different possible descriptions the speaker could have used.

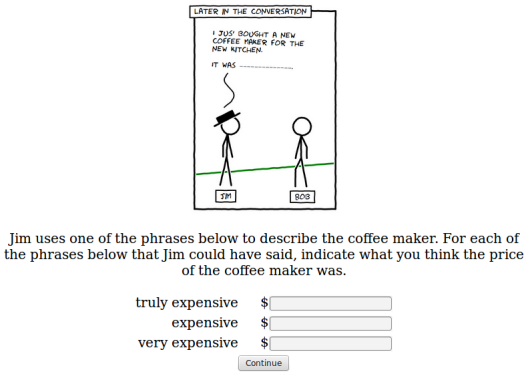

Figure 7: Screenshot from Experiment 3 target question.

Results and Discussion

i need to clean this up

We found that participants did learn that the speaker's use of a word was much higher when it was a target than when it was a control (Fig 8). In a linear regression with word type (target or control) as a fixed effect and random intercepts for word and participant, word type was a significant predictor of frequency (estimate=34.06, p=0.0405).

In addition, when participants believed the speaker's use of a word was much higher, they believed the meaning the speaker intended to convey with the word was lower (Fig 9). The difference between "*(truly|very)* expensive" and "expensive" was greater for the target word than for the control word. In a linear gregression with word type as a fixed effect and random intercepts for word and participant, word type was a significant predictor of difference score (estimate=-31.39, p=0.0226).

In a linear regression with word type (target or control) as a fixed effect and random intercepts for word and participant, word type was a significant predictor of frequency (estimate=34.06, p=0.0405).

Individuals' estimates of frequency correlate with their estimates of price (Fig 10, Pearson's r=0.396, p=0.0169).

Discussion

conclusion

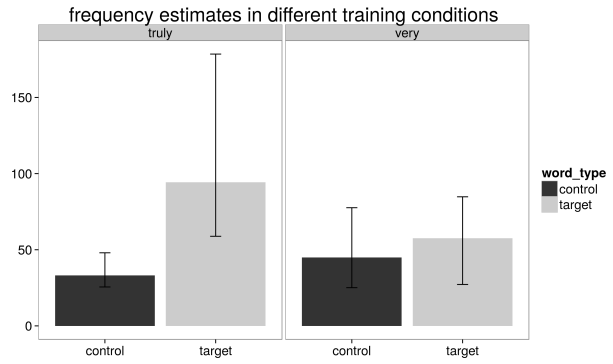

Figure 8: Results of Experiment 3. Intensifier is given a higher frequency estimate when it is target than when it is control, showing that participants learned a new frequency for that intensifier from the training.

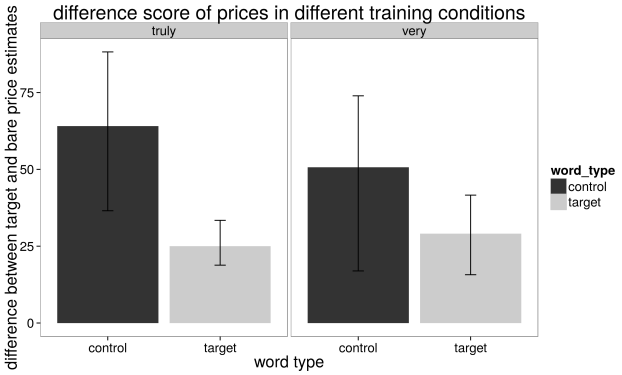

Figure 9: Results of Experiment 3. Price estimate for intensifier is lower after the intensifier is repeated (target condition), showing that overuse within a dialect results in a less strong meaning.

Acknowledgments

References

Brants, T., & Franz, A. (2006). *Web 1T 5-gram Version 1.* Philadelphia: Linguistic Data Consortium.

Lewis, M., Sugarman, E., & Frank, M. C. (2014). The structure of the lexicon reflects principles of communication. In *Proceedings of the 36th annual meeting of the cognitive science society.*

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... others (2011). Quantitative analysis of culture using millions of digitized books. *science*, *331*(6014), 176–182.

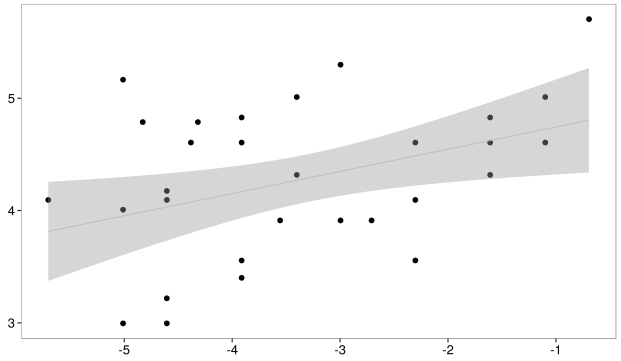

Figure 10: Results of Experiment 3. As participants' frequency estimates increase, their price estimates decrease.

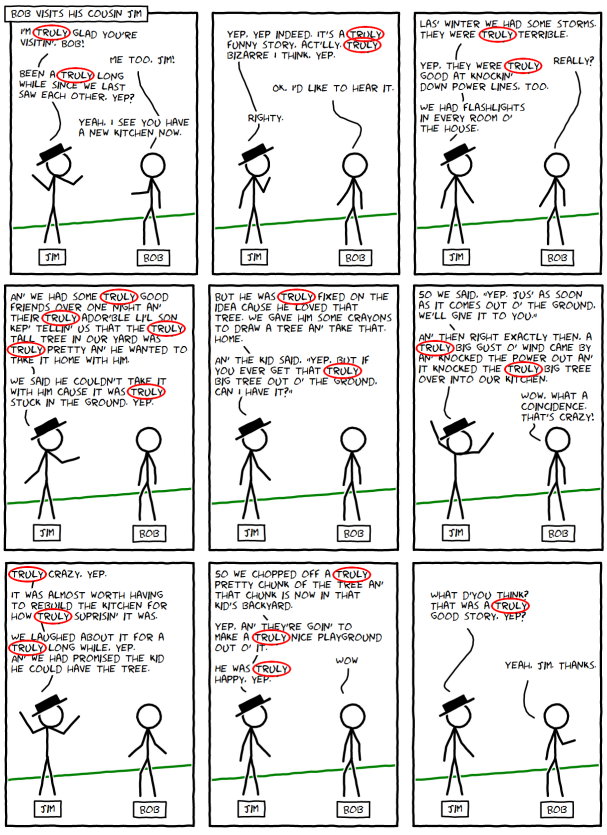

Figure 11: Full comic for Experiment 3, target intensifier "truly" is repeated 22 times, control target "very" is not used.

Table 1: Intensifiers from Experiment 1, number of occurences in Google Web 1T 5grams corpus, and number of syllables.

ngram	frequency	syllables
surpassingly	11156	4
colossally	11167	4
terrifically	62292	4
frightfully	65389	3
astoundingly	73041	4
phenomenally	120769	5
uncommonly	135747	4
outrageously	240010	4
fantastically	250989	4
mightily	252135	3
supremely	296134	3
insanely	359644	3
strikingly	480417	3
acutely	493931	3
awfully	651519	3
decidedly	817806	4
excessively	877280	4
extraordinarily	900456	6
exceedingly	977435	4
intensely	1084765	3
markedly	1213704	3
amazingly	1384225	4
radically	1414254	3
unusually	1583939	4
remarkably	1902493	4
terribly	1906059	3
exceptionally	2054231	5
desperately	2139968	3
utterly	2507480	3
notably	3141835	3
incredibly	4416030	4
seriously	12570333	4
truly	19778608	2
significantly	19939125	5
totally	20950052	3
extremely	21862963	3
particularly	41066217	5
quite	55269390	1
especially	55397873	4
very	292897993	2

Table 2: Intensifier Lists from Experiment 2: Rankings.

List A	List B	List C	List D
surpassingly	colossally	terrifically	frightfully
astoundingly	phenomenally	uncommonly	outrageously
fantastically	mightily	supremely	insanely
strikingly	acutely	awfully	decidedly
excessively	extraordinarily	exceedingly	intensely
markedly	amazingly	radically	unusually
remarkably	terribly	exceptionally	desperately
utterly	notably	incredibly	seriously
truly	significantly	totally	extremely
particularly	quite	especially	very