Extremely costly intensifiers are stronger than quite costly ones

Erin Bennett (erindb@stanford.edu), **Noah D. Goodman** (ngoodman@stanford.edu)
Department of Psychology, Stanford University.

Abstract

We show how the wide range in strengths of intensifying degree adverbs (e.g. *very* and *extremely*) could be explained by pragmatic inference based on differing cost, rather than differing semantics. This predicts a linear relationship between the meaning of intensifiers and their length and log-frequency. We test this prediction in two studies, using two different dependent measures, finding that higher cost does predict stronger meanings. In two additional studies with the same dependent measures, we confirm that the relationship between length and meaning is present even in novel words. We discuss the implications for adverbial meaning and the more general question of how extensive non-arbitrary form-meaning association may be in language.

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

Introduction

How do different words get their meanings? For instance, why is an "extremely good paper" better than a "quite good paper"? The traditional answer (de Saussure, 1916) is that different meanings have been arbitrarily and conventionally assigned to the different word forms. This view has been challenged by a number of examples in which word meaning appears to be non-arbitrarily related to properties of the word. In some cases, the phonetic form of a word is systematically related to its meaning, for example rounded vowels and voiced consonants tend to refer to round objects (Köhler, 1947; Ramachandran & Hubbard, 2001; Holland & Wertheimer, 1964; Davis, 1961). In other cases, orthographic form is diagnostic of meaning, for example, speakers of Hebrew who have never seen Chinese characters are nonetheless above chance at matching them to their corresponding Hebrew words (Koriat & Levy, 1979). Similarly, the length of words predicts aspects of their meanings: across languages longer words refer to more complex meanings (Lewis, Sugarman, & Frank, 2014). In this paper, we explore adjectival intensifiers[1], like *extremely* and *quite*, as a case study in which to empirically explore the relationship of meaning to factors like word form and distribution of usage. Intensifiers form a good case study both because they are amenable to simple quantitative measures of meaning (such as the numeric extent to which they shift the interpretation of a scalar adjective) and because theoretical considerations, which we lay out shortly, suggest a relationship between their meaning and their usage cost (e.g., due to frequency and length).

In the next section, we discuss a possible semantics for intensifiers, building off of previous work on scalar adjectives. We show how pragmatic effects could explain much of the variation in the meanings of intensifiers. this predicts that the meanings of intensifiers are influenced by their form (in length) and their distribution (frequency) of usage. The impact of word length is reminiscent of the results of Lewis et al. (2014), who studied noun categories. While word frequency is known to have major effects on sentence processing (Levy, 2008, e.g.), the prediction that frequency should affect meaning is more novel.

We confirm, in our first two experiments, that English intensifiers in adjective phrases are indeed interpreted as much higher degrees (e.g. in the case of *expensive*, higher prices) for both longer and less frequent intensifiers. This holds in quantitative judgments of meaning and in forced comparisons, and across a number of adjectival dimensions. In our second two experiments, we replicate this finding, and extend it to novel intensifiers, showing that length is a significant predictor of the strength of an intensifiers meaning. We conclude with a discussion of different interpretations of these phenomena and future directions.

The semantics of intensifying degree adverbs

Our paper focuses on intensifying degree adverbs applied to scalar adjectives[2]. Scalar adjectives have been described as having a threshold semantics (Kennedy, 2007), where, for example, *expensive* means "having a price greater than θ" and θ is a semantic variable inferred from context (e.g., \$100). Above the threshold degree θ, the adjective is true of an object, and below, the adjective is false. Lassiter and Goodman (2013) give a formal probabilistic Rational Speech Acts (RSA) (Frank & Goodman, 2012; Goodman & Stuhlmüller, 2013) model of how this threshold might be established by pragmatic inference that takes into account statistical background knowledge (such as the distribution of prices for watches), which we extend in the Appendix.

We explore the idea that an adjective phrase with an intensifying degree adverb derives much of its meaning from a M(arkedness)-impicature (Levinson, 2000): more marked (costly to utter) versions of an adjective phrase will be inter-

[1] Intensifiers are adverbs that modify scalar adjectives to increase the degree. The word "intensifier" is often used to denote the full range of degree adverbs, be they "amplifiers", or "downtoners" (Quirk, Greenbaum, Leech, & Svartvik, 1985). The "intensifiers" we are looking at in this paper are, according to this typology, "amplifiers" because they increase (rather than decrease) the threshold associated with a gradable predicate. This typology also distinguishes between two different kinds of amplifiers: those that increase an adjective maximally (e.g. *completely* and *utterly*) and those that merely increase (e.g. *greatly* and *terribly*). We do not make this distinction. The word "intensifier" is sometimes used for a completely different linguistic phenomenon, where a reflexive is used for emphasis, e.g. "The king himself gave the command," which we do not analyze in this paper.

[2] Some of these intensifiers can also apply to verbal and nominal predicates, and different restrictions apply for different intensifiers, e.g. *I truly like carrots* is an acceptable utterance, whereas *I very like carrots* is not. See Bolinger (1972) for a discussion.

preted as implicating higher values. This hypothesis predicts the relative cost of one intensifier over another would affect the interpretation of a resulting adjective phrase.

A possible model of this effect is in the Appendix, where we extend Lassiter and Goodman (2013)'s adjectives model by assigning a seperate threshold to each intensified (or bare) adjective phrase. As in previous RSA models that include utterances with similar semantics but different costs (Bergen, Goodman, & Levy, 2012), we find an M-implicature, such that more costly intensifiers result in stronger adjective phrases.

Background

Previous researchers have proposed that adjective phrases modified by intensifiers have the same semantics as unmodified adjective phrases, except with new, higher thresholds (Kennedy & McNally, 2005; Klein, 1980; Wheeler, 1972). That is, some threshold, inferred from context, exists above which objects are *expensive* and below which they are not, and the intensifier *very* determines a new, higher threshold for *very expensive*. They suggest that the intensified thresholds are determined by first collecting the set of objects in the comparison class for which the bare adjective is true, and then using that as the comparison class to infer a new threshold, i.e. *very expensive laptop* means "expensive for an expensive laptop". This analysis results in the expected intensification of adjectives ("expensive for an expensive laptop" has a higher threshold for being true than simply "expensive for a laptop") and is appropriately sensitive to different domains (e.g. the absolute difference in price between thresholds for *expensive* and *very expensive* is much higher in the context of "That space station is very expensive," than in the context of "That coffee is very expensive."). However, this account does not, in and of itself, distinguish between the graded strengths of different intensifiers, for example, *very expensive* and *phenomenally expensive*.

Intuition suggests that different intensifiers do have different strengths (e.g. *outrageously* seems stronger than *quite*), and we provide further evidence of this in our experiments, where participants interperet and compare different intensifiers. It could be that the degree of strength of different intensifiers is conventionally specified by the lexicon. But the semantics must then specify how these entries affect the very flexible threshold of the relevant adjective. In addition, the multitude of intensifiers (Bolinger, 1972) and their apparent productivity[3] suggest a more parsimonious solution would be welcome. That is, having a lexically determined meaning for each different intensifier might overlook the similarity among words of this class.

M-implicature proposal

We propose instead that each time a scalar adjective is used, in each phrase, it introduces a free threshold variable (that is, a new token threshold is inferred for every time the lexical entry of the adjective is accessed). Further we propose that intensifiers contribute *nothing* to the literal, compositional semantics[4]. This implies that different adjectival phrases (e.g. "very expensive watch" and "extremely expensive watch") have equivalent meanings, though with thresholds that will be separately assigned based on context. *However*, the intensifiers do affect the production cost of the corresponding sentences, and it is this cost difference that results in meaning differences.

a high-level sketch of how this M-implicature works.

A possible model of this proposal is detailed in the Appendix.

Factors affecting utterance cost

We have identified the intensifier's cost as a potentially critical factor of its interpreted meaning. To connect this prediction to empirical facts, we still must specify (at least a subset of) the factors we expect to impact cost. The most natural notion of cost is the effort a speaker incurs to produce an utterance. This could include cognitive effort to access lexical items from memory, articulatory effort to produce the sound forms, and other such direct costs. Speakers might also seek to minimize comprehension cost for their listeners, resulting in other contributions to cost. For the purposes of this paper, we restrict to the most straightforward contributors to production cost and use proxies that are straightforward to quantify: length (longer utterances are more costly) [5] and frequency (rarer intensifiers are harder to access and therefore more costly). In a number of different tasks, lexical frequency affects difficulty in an approximately logarithmic way. For instance word recognition time (McCusker, 1977) and reading time in context (Smith & Levy, 2013) are both logarithmic in frequency. We thus use the log-frequency (whose negative is also called *surprisal*) as the quantitative contribution to cost.

We thus predict a contribution of longer and higher surprisal intensifiers to the meaning. This leaves open the the relative importance of length and surprisal, and potential interactions (as well as other factors that might enter into cost), which can be explored via regression models.

Correlation experiments: utterance cost and strength

The proposal detailed above predicts an association between measures of cost and strength of interpretations. In our first two experiments, we tested whether our possible measures of cost can in fact predict the interpreted degree along a scalar adjective's scale.

[3] For example, *altitidinously expensive* is not in common usage, but one can easily interpret *altitidinously* as a novel intensifier.

[4] We take this strong view for rhetorical purposes. It is highly likely that some intensifiers have other aspects of meaning.

[5] We measure length in number of syllables, although length in characters (which might be a relevant source of utterance cost in a written format, as our experiments were in) has similar predictive power to syllable length in all of our analyses.

Experiment 1

In Experiment 1, we test the qualitative prediction that as cost of an utterance increases, so will the interpreted meaning of the intensifier. We tested this prediction by eliciting free response price estimates from people and determining whether these prices are correlated with our independent measures of utterance cost.

Method[6] 40 participants with US IP addresses were recruited through Amazon's Mechanical Turk and paid $0.40 for their participation. 1 participants was excluded from the analysis for admitting that they did not think they followed the instructions in a post-expeirment survey.

We asked participants to estimate the prices of different objects based on different descriptions of those objects. The descriptions included intensifiers paired with the adjective *expensive* (Figure 1). There were three categories of objects (*laptop*, *watch*, and *coffee maker*) and 40 intensifiers (see Table 1). We chose intensifiers that 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 mainly 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 judgments for every intensifier-category pairing in a randomized order (different for different participants), for a total of 120 price judgments per participant. We chose the domain of price and used only the adjective *expensive* because price constitutes a quantitative scale with standard units (dollars for our US participants) on which to measure the different intensifers.

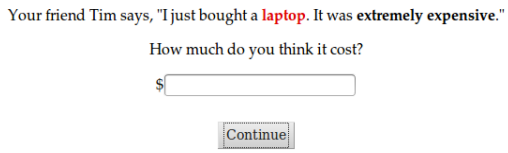

Figure 1: Screenshot from Experiment 1 target question.

Corpus Methods Table 1 shows word frequency and length in syllables for the intensifiers used in the experiment. The frequencies were collected from the Google Web 1T 5-grams database (Brants & Franz, 2006)[7] In the analysis below we use word length and word surprisal (negative log-frequency) as proxies for a word's cost, as motivated above. The syllable lengths of our intensifiers and the surprisals were correlated, but not strongly so (r = 0.27).

[6]The full experiment can be found at `http://cocolab.stanford.edu/cogsci2015/intensifiers/Experiment1`

[7] We also ran the same analyses on frequency information collected from the Google Books American Ngrams Corpus (Michel et al., 2011) and found similar results.

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

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

We ran a linear mixed effects regression with centered fixed effects of syllables, surprisal, and their interaction, and random intercepts and slopes for syllables and surprisal for both participant and object. We used the logarithm of participants' price estimates as the dependent variable, because of evidence that people's representation of numbers, including prices, is logarithmic (Dehaene, 2003, e.g.)[8].

Our results are shown in Figure 2. Both measures of cost play a role in predicting participants' price estimates. We found a significant main effects of surprisal ($\beta = 0.0536, SE = 0.00902, t(3) = 5.94, p < 0.05$) such that

[8]I.e. the perceptual distance between two prices the same dollar amount apart is more for small numbers (e.g. $3 and $6) and less for large numbers (e.g. $1,543 and $1,546).

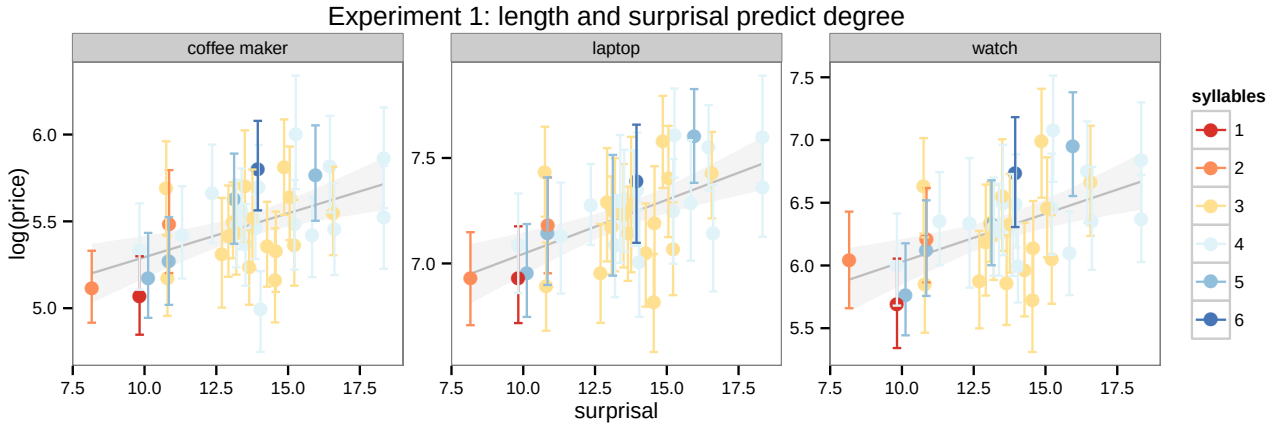

Figure 2: Results of Experiment 1. As surprisal and length in syllables increase, participants' free response prices increase.

less frequent words tend to be associated with higher price estimates. We also found a significant main effect of syllable length ($\beta = 0.0900, SE = 0.0189, t(4) = 4.76, p < 0.05$), above and beyond surprisal, such that longer words predict stronger meanings. We also found a significant interaction ($\beta = 0.0196, SE = 0.00520, t(3.5) = 3.77, p < 0.0005$) between surprisal and syllable length, wich may indicate that the relationship between the two predictors of cost is not simply additive, and that having multiple sources of communicative cost (i.e. length and surprisal) might increase the implicature even more.

Thus intensifiers that are less frequent and longer (and therefore are more costly to utter) also tend to be interpreted as having stronger meanings, at least when used to modify *expensive*. Furthermore, the relationship appears to be linear in surprisal and length (though with an interaction), as predicted. This is consistent with the M-implicature proposal introduced above.

Experiment 2

The M-implicature account described above implies that there is no semantic interaction between the intensifier and the adjective it is applied to. Instead an intensifier should contribute similar cost, and therefore meaning, to the different adjectival phrases in which it occurs[9]. To explore this issue, we would like to extend our results to additional adjectival scales. However, most scales are not so easily quantifiable as price; we require a different dependent measure in order to probe them. For Experiment 2 we used a forced-ranking dependent measure, which allows us to consider additional adjectival scales. This dependent measure has the added benefit of providing a more sensitive measure of the differences in degrees between similar adjectival phrases.

Method[10] 30 participants with US IP addresses were recruited through Amazon's Mechanical Turk and paid $0.40 for participation. 2 participants wer excluded from the analysis for admitting that they did not think they followed the instructions in a post-experiment survey.

We asked participants to order (by clicking and dragging) various adjective phrases with the same adjective but different intensifiers according to strength of meaning. Because arranging these phrases required participants to be aware of the full set of adjective phrases and access all of them on the same computer screen (which might vary in size for different participants), not all of our 40 intensifiers could effectively be presented at once. We divided the 40 intensifiers from Experiment 1 into four lists of 10 intensifiers. 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 with one adjective. Participants 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 completed four such trials, 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 constant, so that the same 10 intensifiers were always shown together to simplify data analysis.

Results Our results for Experiment 2 are shown in Figure 4. We ran an ordinal regression with centered surprisal and syllable lengths and their interaction as fixed effects. As in Experiment 1, we found strong main effects of surprisal ($\beta = 0.301 SE = 0.0259, t = 11.6, p < 5e-16$) and syllable length ($\beta = 0.456, SE = 0.0695, t = 6.65, p < 5e-10$), with

[9] If the bigram frequency of the modified adjective ("very expensive") deviated from that expected based on independent word frequencies our frequency-based cost account would predict an interactive effect on meaning. This would be a relatively small effect, and the relevant bigrams were too sparse in our corpora to pursue.

[10] The full experiment can be found at `http://cocolab.stanford.edu/cogsci2015/intensifiers/Experiment2`

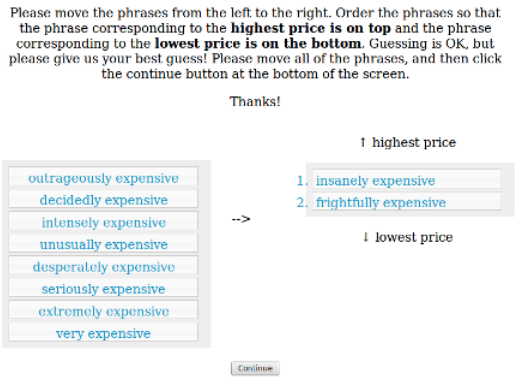

Figure 3: Screenshot from Experiment 2 target question.

only a trending interaction ($p = 0.0799$). In other words, we again found that participants assign stronger interpretations to intensifiers with higher surprisals and/or higher syllable lengths, extending now across four different adjectival scales.

Discussion

In these first two experiments we provided evidence that intensifier meanings depend systematically on the length and frequency of distribution of those word forms. While it is unlikely that this accounts for all intensifier meaning, it does suggest that a major portion of meaning comes not from arbitrary, conventional association of signal to sign (de Saussure, 1916), but from features of the word's form and distribution.

Since this is a correlational study, such a relationship does not confirm that an intensifier's cost *causes* it to have a given meaning.Rarity in particular might be correlated with strength of meaning merely because more extreme meanings refer to less probable things in the world, are therefore talked about less, and therefore the words with those meanings will necessarily be rarer. Although it seems reasonable to suspect that word frequencies reflect the probabilities of the real-world concepts they describe, it might also be the case that improbable things are more likely to be commented on, and so to a certain extent the frequencies of words that describe rare concepts will be inflated. Syllable length in turn might depend on the frequency or simplicity of a word[11], either because words that are frequently used get shortened over time cite someone or because words that refer to simpler or more common concepts enter the lexicon sooner (when more shorter word forms remain unassigned to meanings). It is therefore possible that these measures of cost have no causal influence on the meanings of intensifiers within a particular communicative act.

In order to address the question of whether utterance cost might cause people to interpret an intensifier as stronger, we ran Experiments 3 and 4, where we directly manipulated one of our measures of cost – length – in novel intensifiers which have no conventional meaning associated to them.

[11] Acknolwedge that this totally happens. cite someone

Manipulation experiments: novel intensifiers of varying lengths

Although the meanings of our existing English intensifiers could have influenced their lengths and frequencies over time, novel intensifiers have no meaning already associated with them. Therefore, if we found a relationship between the length of a novel intensifier and its interpreted meaning, we would have evidence that length can causally influence meaning. In the following two experiments, we directly manipulate the lengths of novel intensifiers and show that longer novel intensifiers are interpreted as having stronger meanings.

Experiment 3[12]

motivate

Method 30 participants with US IP addresses were recruited through Amazon's Mechanical Turk and paid $0.80 for their participation. 2 participants were excluded from the analysis for admitting that they did not think they followed the instructions in a post-expeirment survey.

Experiment 3 was identical to Experiment 1, except that we included only a subset of the intensifiers from Experiment 1[13] and each participant also saw one novel intensifier, randomly mixed in with the rest.

We varied the novel intensifier between participants from a set of 6 novel intensifiers, three of which were relatively short (*lopusly*, *ratumly*, and *bugornly*) and three of which shared the same "root" but were two CVCV syllables longer (*fepolopusly*, *gaburatumly*, and *tupabugornly*).

Participants again estimated prices for objects of three different categories paired with all of the intensifiers. The order of the questions was randomized between and within participants.

Results In Experiment 3, we included as filler a subset of the intensifiers we tested in Experiment 1, and so we first confirmed our findings from Experiment 1. As in Experiment 1, we ran a linear mixed effects regression with centered fixed effects of syllables, surprisal, and their interaction, and random intercepts and slopes for syllables and surprisal for both participant and object, and we used the logarithm of participants' price estimates as the dependent variable. Replicating our findings from Experiment 1, we found significant main effects of surprisal ($\beta = 0.09481$, $SE = 0.02503$, $t = 3.787$, $p < 0.05$) and syllable length ($\beta = 0.1622$, $SE = 0.02716$, $t = 5.972$, $p < 0.0005$), and a significant interaction ($\beta = 0.03071$, $SE = 0.009156$, $t = 3.353$, $p < 0.005$).

We then ran a linear mixed effects model on only the novel intensifiers, with length ("long" or "short") as a fixed effect, random intercepts and slopes for objects, and random intercepts for the three different "roots". We found a significant ef-

[12] The full experiment can be found at `http://cocolab.stanford.edu/cogsci2015/intensifiers/Experiment3`

[13] We chose this subset of 9 intensifiers to get a wide range of surprisals and syllable lengths (*colossally*, *phenomenally*, *mightily*, *extraordinarily*, *amazingly*, *terribly*, *notably*, *significantly*, *quite*)

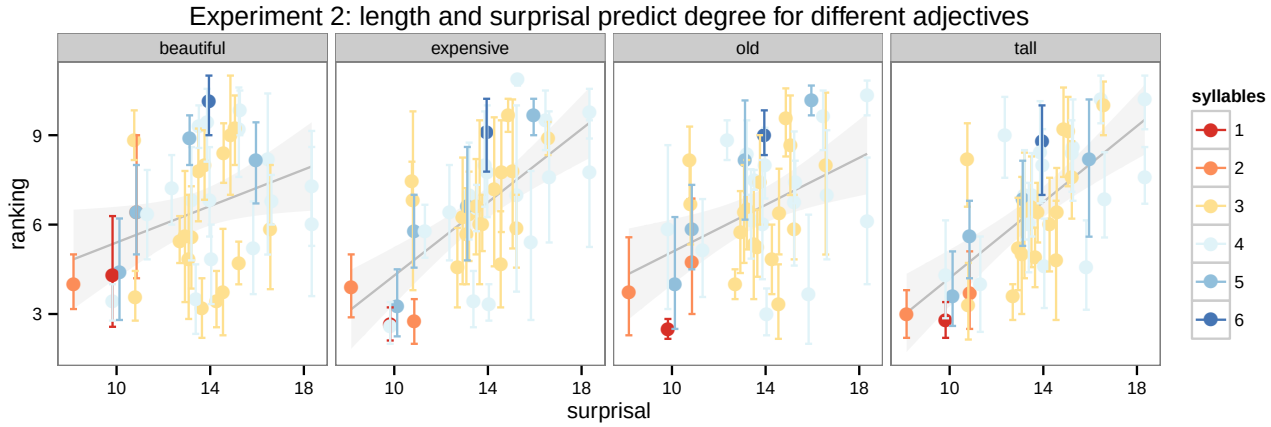

Figure 4: Results of Experiment 2. As surprisal and length in syllables increase, participants' rankings increased.

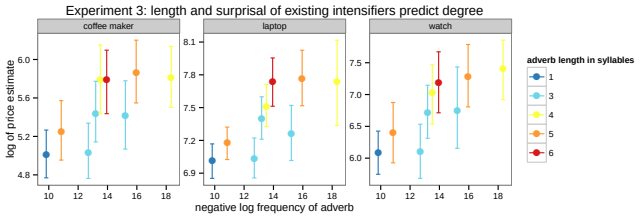

Figure 5: In Experiment 3, we replicated our findings from Experiment 1.

fect of length condition (β("short") $= -1.3955$, $SE = 0.3794$, $t = -3.678$, $p < 0.0005$), indicating that people use the length of an intensifier in the moment in order to interpret its meaning, even for novel intensifiers with no conventional meaning.

Experiment 4

motivate

Method[14] 56 participants with US IP addresses were recruited through Amazon's Mechanical Turk and paid \$0.X0 for their participation. 5 participants were excluded from the analysis for admitting that they did not think they followed the instructions in a post-experiment survey.

Experiment 3 was identical to Experiment 2, except that we included only two of the four adjectives (*expensive* and *tall*) – which we varied *between* participants – and the intensifiers from Experiment 3. Each participant saw exactly one novel intensifier, and we varied the novel intensifier between participants.

As in Experiment 2, participants were asked to drag and drop adjective phrases into order from highest to lowest on the adjectival scale. However, in Experiment 4, each participant saw exactly one of two adjectives (*expensive* or *tall*, varied between participants) and only the set of intensifiers from Experiment 3. This set included one novel intensifier, which we varied between participants. As in Experiment 2, adjective phrases for each intensifier-adjective pairing were initialized in random order.

Results With our filler intensifiers for Experiment 4, we replicating our findings from Experiment 2 of significant effects of both surprisal ($\beta = 0.781$, $SE = 0.0499$, $t = 15.7$, $p < 5e-16$) and syllable length ($\beta = 1.111$, $SE = 0.0974$, $t = 11.4$, $p < 5e-16$) on the order in the list that participants chose for the intensifiers. In this replication, we also found a significant interaction ($\beta = 0.189$, $SE = 0.0351$, $t = 5.39$, $p < 5e-7$). This is consistent with the findings from all of our previous experiments.

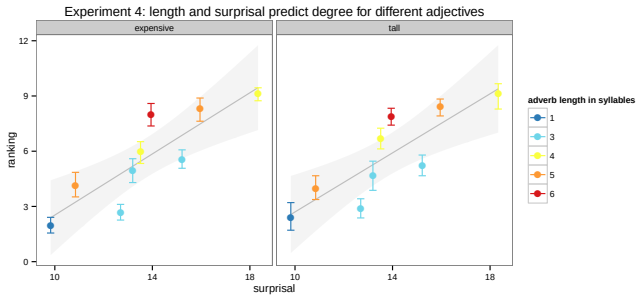

Figure 7: In Experiment 4, we replicated our finding from Experiment 2: longer and less frequent intensifiers are ranked higher than shorter and more frequent ones.

For the novel intensifiers, we ran an ordinal regression on the rankings (relative to the filler intensifiers) and found a significant effect of length condition (β("short") $= -1.17$, $SE = 0.525$, $t = -2.24$, $p < 0.05$).

[14]The full experiment can be found at `http://cocolab.stanford.edu/cogsci2015/intensifiers/Experiment4`

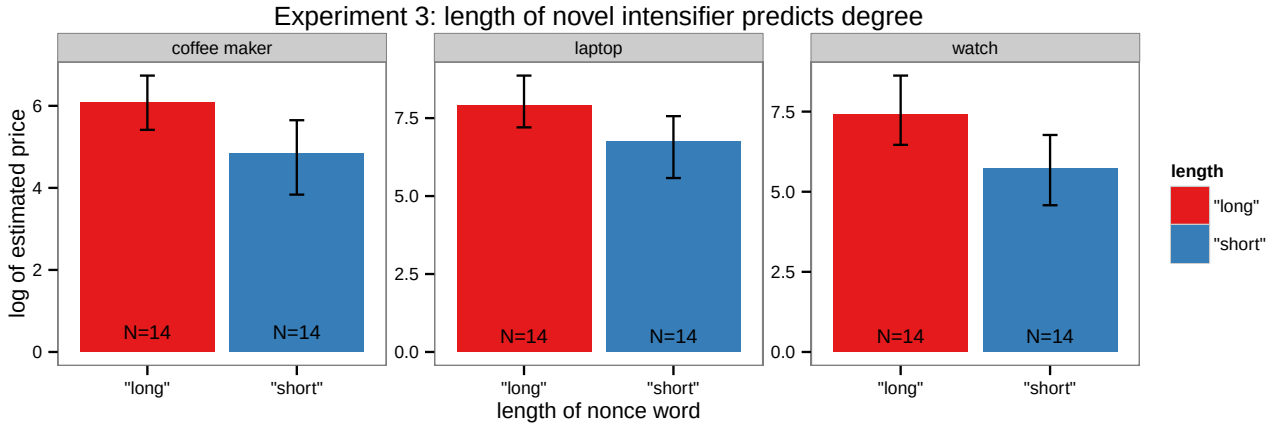

Figure 6: In Experiment 3, we found a significant effect of length for all novel intensifiers.

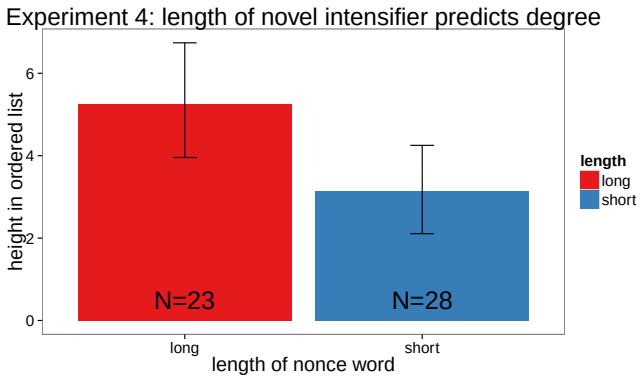

Figure 8: In Experiment 4, the lengths of novel intensifiers predicted participants' rankings of that intensifier.

Discussion

Overall, in Experiments 3 and 4 we found that length is a significant predictor of interpretation strength for novel intensifiers. These novel intensifiers have no established meaning, so the relationship between their length and strength cannot be a direct consequence of the lexicon becoming more efficient over time. It may be that participants are inferring the meanings of the novel intensifiers pragmatically, as in the M-implicature account sketched above. It may also be that participants have learned a relationship between lenth and meaning of intensifiers in English, and are generalizing this meta-linguistic knowledge to the new words they encounter cite someone. unpack meta-linguistic versus pragmatic ambiguity a bit more? Either way, at least some inference must happen online as people see these words that influences their interpreations.

General Discussion

Motivated by a recent probabilistic model of scalar adjectives (Lassiter & Goodman, 2013), we showed how adjectival intensifiers could potentially get their meaning through a pragmatic M-implicature, despite having vacuous literal meaning to add to an adjective. Our model predicted a linear relationship between the intensity of an intensifier and its cost, measured here in terms of length and log-frequency. In four experiments we provided evidence that intensifier meanings do depend systematically on the length and frequency of distribution of those word forms and that this relationship might be causal, with length actually causing the interpretation. While it is unlikely that this accounts for all intensifier meaning, it does suggest that a major portion of meaning comes not from arbitrary, conventional association of signal to sign (de Saussure, 1916), but from features of the word's form and distribution.

A number of issues remain to future work, including the mechanism by which people use length to interpret novel adverbs, whether other cost factors affect the interpretation of novel adverbs, and the other aspects of intensifier meaning (such as polarity or affective color). However we believe that the preliminary results presented in this paper already have interesting implications. For the semantics of adverbial modifiers, we have shown how pragmatic mechanisms could be central in establishing flexible contributions to sentence meaning. For the broader question of form-meaning mapping, we have suggested a source of non-arbitrary association based on both properties of the word form and of its distribution.

References

Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use..

Bolinger, D. (1972). *Degree words*. Paris: Mouton.

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

Davis, R. (1961). The fitness of names to drawings: a cross-cultural study in tanganyika. *British Journal of Psychology*.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147.

de Saussure, F. (1916). *Nature of the linguistic sign.*

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*.

Goodman, N. D., & Stuhlmüller, A. (2013). Knowledge and implicature: Modeling language understanding as social cognition. *Topics in cognitive science*.

Holland, M., & Wertheimer, M. (1964). Some physiognomic aspects of naming, or *maluma* and *takete* revisited. *Perceptual and Motor Skills*.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*.

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and philosophy*.

Köhler, W. (1947). *Gestalt psychology* (Second ed.). Liveright.

Koriat, A., & Levy, I. (1979). Figural symbolism in chinese ideographs. *Journal of Psycholinguistic Research*.

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and Linguistic Theory (SALT) 23.*

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*.

Lewis, M., Sugarman, E., & Frank, M. C. (2014). *The structure of the lexicon reflects principles of communication.*

McCusker, L. (1977). Some determinants of word recognition: Frequency. In *24th annual convention of the southwestern psychological association, fort worth, tx.*

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*.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). A comprehensive grammar of the english language.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies*.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*.

Wheeler, S. C. (1972). Attributives and their modifiers. *Noûs*.