October 28, 2008 To appear in Sprache und Datenverarbeitung

The pragmatics of expressive content: Evidence from large corpora*

Noah Constant, Christopher Davis,
Christopher Potts, and Florian Schwarz
UMass Amherst

Abstract We use large collections of online product reviews, in Chinese, English, German, and Japanese, to study the use conditions of expressives (swears, antihonorifics, intensives). The distributional evidence provides quantitative support for a pragmatic theory of these items that is based in speaker and hearer expectations.

Keywords: expressives, intensives, antihonorifics, corpus pragmatics, logistic regression, Chinese, English, German, Japanese

1 Introduction

Under what circumstances do speakers resort to expressives — swears, honorifics, exclamations, polite terms — to help convey their intended messages? To answer this question, one can construct a wide variety of scenarios and see how people react to expressives in them. This is an important investigative strategy. It is, though, limited: one can only construct so many contexts, and it is hard to escape the influence of the experimental setting (Geurts 2007).

What we would like is a huge corpus of naturalistic data annotated with information about the context of use, so that we can probe what motivates speakers to use and avoid expressives. However, such a corpus would be prohibitively expensive to obtain, and it would end up tied closely to our preconceptions about what contextual information is important in this area. This might seem to drastically reduce the usefulness of corpus methods when studying expressives.

The Internet abounds with naturalistic texts, though, many of which have associated metadata that approximate features of the linguistic context (Pang et al. 2002; Pang and Lee 2005; Beineke et al. 2004; Potts and Schwarz 2008). Corpora of this sort are easy and inexpensive to acquire, and they can yield quantitative evidence for nuanced claims about linguistic use conditions. The present paper uses this methodology to help characterize the distinctively emotive qualities of a

* Our thanks to Chloe Gu, Sally McConnell-Ginet, Bitian Zhang, and audiences at ESSLLI 2008, Stanford, and UC Santa Cruz.

variety of expressive types, drawing on large collections of online product reviews in Chinese, English, German, and Japanese. We balance native-speaker intuitions about particular examples in context with frequency information relativized to rating categories, showing, in particular, that the two are well-aligned. This paves the way to using the frequencies to obtain an empirically grounded view of speaker and hearer expectations about words and constructions. Following Lewis (1969) and others, we put this information to work in developing a theoretical understanding of how expressive language shapes utterance interpretation.

Expressives pack a punch. We are attuned to their content and adept at leveraging it into a better understanding of the utterances we hear. This seems initially to be at odds with the fact that individual expressives admit of a wide range of uses, at wildly diverse points in the emotional spectrum. Our distributional evidence reveals deep regularities in how this language is used, though, and it provides the means for estimating just how reliable expressive signals are. Thus, we are able to help reconcile the extreme underspecification of expressives with the powerful messages they send in context.

The next section reviews our data and methodology, concentrating on how we obtain frequency information from our corpora and how we model such information. Sections 3–6 explore a wide range of expressives in our corpora: taboo intensifiers, emotionally charged common nouns and modifiers, politeness honorifics, and anti-honorifics. Our aim throughout is to provide insights into the particular contributions of these items, the general expressive classes that they represent, and the relationships among those classes. Finally, section 7 outlines an approach to incorporating information of this sort into a theory of formal pragmatics.

2 Data and methods

This paper is built around four collections of online product reviews. Three are from Amazon websites: Amazon.com (English), Amazon.de (German), and Amazon.co.jp (Japanese). The English collection is all book reviews; the German and Japanese collections are a mix of book, music, movie, and electronics reviews. The fourth collection is from MyPrice.com.cn (Chinese), where people review mainly electronics. The important unifying feature is that every text in these collections is tagged with a rating, one-star through five-stars.

In figure 1, we provide short examples from the English corpus, to convey a sense for what the texts from different rating categories are like. The appendix describes each of these data sets more fully. Taken together, they supply over 14 million words of review text drawn from over 110,000 reviews and 60,000 authors.

Our basic perspective is the log-odds distribution, as in (1), where count(x_n, R) is the number of tokens of x_n (a word-string of length n) in rating category R, and

Rating:	1 star
Review:	I bought this book and read it half way through. It was soo damn boring to read. Just opinions on what caused what. It was absolutely dreadful and was a good way to fall asleep. Sold it back on amazon for 8 bux.

(a) A short one-star review

Rating:	3 star
Review:	This is a good book to compare the various menu items and products for their calorie values. The one problem I have with it is that it only shows 3-4 good items and 3-4 bad items per restaurant/store/manufacturer. I would rather have had this broken into a series of books and gone into more detail.

(b) A short three-star review

Rating:	5 star
Review:	I never had an opinion either way about Tori. I thought she was fun in Scary Movie 2, and then got totally addicted to So NoTORIous. I couldn't believe how down to earth and self-deprecating she was, and it seemed so sincere. Her book confirmed that too! And I learned that in the end, little girls (rich or poor)are the same, and all we really want is that damn Barbie dream house :-)

(c) A short five-star review

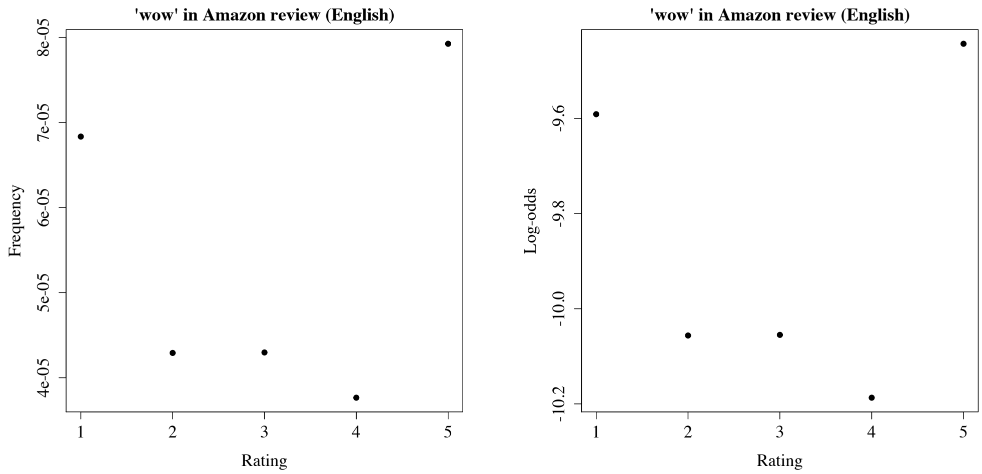
Figure 1: Example reviews from Amazon.com

$\text{count}_n(R)$ is the number of tokens of word-strings of length n in rating category R.

(1)
$$\text{log-odds}(x_n, R) \stackrel{def}{=} \ln\left(\frac{\text{count}(x_n, R)}{\text{count}_n(R) - \text{count}(x_n, R)}\right)$$

The log-odds distribution is closely related to a frequency distribution obtained by taking $\text{count}(x_n, R)/\,\text{count}_n(R)$ for each rating category R (given a phrase of interest x_n). Figure 2 compares the two kinds of distribution using *wow* in the English Amazon corpus. The overall shape is approximately the same. However, the log-odds perspective permits a more powerful statistical comparison between frequencies. For additional discussion of these distributions in general, we refer to Jaeger 2008. Our rationale for using log-odds is spelled out more fully in Potts and Schwarz 2008, a study of exclamatives in a similar data collection.

These frequency distributions already begin to illuminate the contribution of *wow*: it looks as though it is more common in the one-star and five-star reviews than in the two-, three, or four-star reviews. This makes intuitive sense: reviews at the extreme ends of this scale are more emotional — these are authors who either loved or hated what they are reviewing, and their linguistic choices reflect this. It also looks as though *wow* is not a reliable signal of which extreme the author is at: the numbers for the one- and five-star categories seem to be about the same.



(a) Frequencies

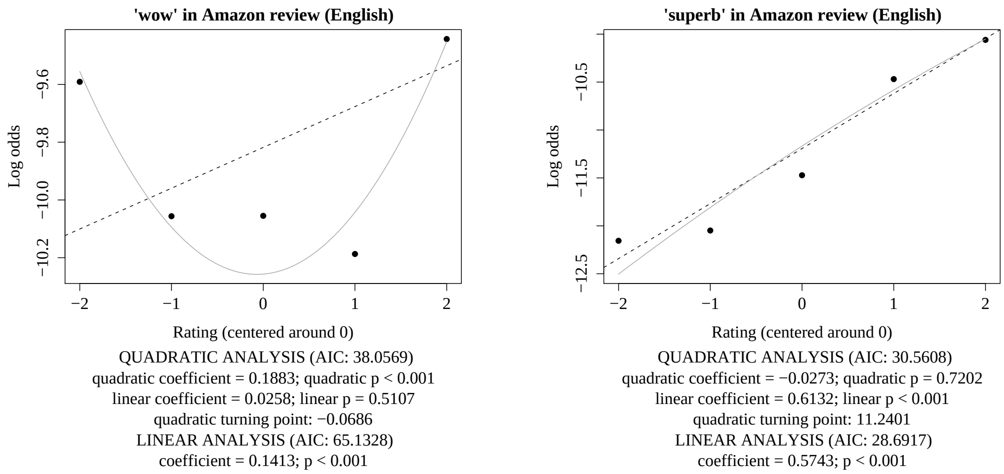
(b) Log-odds

Figure 2: Frequency and log-odds distributions for *wow*.

We can compare the log-odds values to quantify these observations. The empirical log-odds for *wow* in the one-star category (i.e., log-odds(*wow*, one-star)) is -9.591 and in the three-star category it is -10.055. To compare these values, we calculate $e^{-9.591+10.055} = 1.59$.[1] This means that *wow* is about 59% more frequent in the one-star reviews than in the five-star reviews. The contrast between one-star and five-star is comparatively much smaller: log-odds(*wow*, five-star) = -9.443, and thus the difference is $e^{-9.443+9.591} = 1.16$, i.e., about 16%.

These comparisons lend credence to the notion that the distributions in figure 2 are U-shaped. To test whether this visual impression is statistically significant, we build logistic regression models of the data (Baayen 2008; Bresnan and Nikitina 2008; Jaeger 2008). Figure 3(a) illustrates with *wow*; the black dots are the log-odds distribution of figure 2(b). We have included both a quadratic regression line (gray) and a linear regression line (dashed), along with important summary numbers for these fits. These statistical models substantiate our intuitive idea that the distribution is U-shaped, as follows. First, the quadratic fit is excellent. The plot makes this clear: with the possible exception of the middle rating, the empirical values hug close to the predicted values. The p value is, in turn, extremely small. Second, the turning point of the quadratic curve, the point at which it reaches its lowest value, is at the middle of our rating scale. Again, though this is visually obvious, we would like to back it up statistically. To do this, we have centered the rating categories around 0, by subtracting 3 from each. After this shift, the turning point is $x = -0.069$, i.e., just

1 This is equivalent to dividing the two frequencies obtained with $\text{count}(x_n, R)/\,\text{count}_n(R)$.



(a) The empirical distribution (black dots) is U-shaped: the quadratic regression (gray) is a good fit with a near-0 turning point, and the linear regression (dashed) is a poor fit.

(b) The empirical distribution rises steadily; the linear model is an excellent fit.

Figure 3: Two expressive shapes

to the left of the middle of our rating scale. What's more, the linear coefficient has a high p value (0.511), indicating that this turning point is not significantly different from 0.

We have shown the linear regression line as well (dashed). The statistical fit is good here as well, with a p value near 0. However, it is clear from the plot itself that the quadratic model is a much better fit. It captures the curvature that we see in the distribution across rating categories, whereas the linear model does not. The superiority of the quadratic model for these data is further supported by the Akaike Information Criterion (AIC; Cohen et al. 2003: 509; Johnson 2008: 90–91), which balances model complexity with goodness of fit. Though the quadratic model is more complex, it fares much better by this criterion than the linear model (quadratic AIC: 38.057; linear AIC: 65.133; lower is better), in virtue of the fact that it is closer to the empirical observations.

We see lots of other shapes in the data sets. For identifying and studying expressives, the U-like shapes are important, both when they are balanced, as with *wow* above, and also when they reveal a bias for one end of the scale or the other (often with J or Reverse-J appearances). Equally important are the linear distributions with steep slopes. Those that run up from one-star to five-star show a positive bias, whereas those that run down from one-star to five-star show a negative bias. Figure 3(b) continues the positive vibe of this introductory section (there is plenty of negativity to come) by illustrating a linear-increase shape with the scalar-endpoint modifier *superb*. Again, we give the quadratic and linear regression lines. Here, there is no doubt that the linear model is a good fit.[2] Moreover, the slope is steep, a reflection of the fact that, empirically, *superb* is vastly more frequent in the five-star reviews than in the one-star reviews ($e^{\text{log-odds}(\textit{superb},\text{five-star})-\text{log-odds}(\textit{superb},\text{one-star})} = e^{-10.06+12.156} = 8.136$.)

Over the next few sections, we highlight particular expressive types, using their distributional shapes to better understand their emotive qualities. As the above discussion suggests, we work under the assumption that the rating categories provide approximate information about the speaker's emotional or attitudinal state. Our specific assumption is (2) (see also Potts and Schwarz 2008).

(2) Speakers writing one-star reviews are (or seek to create the impression that they are) in negative emotional states, and speakers writing five-star reviews are (or seek to create the impression that they are) in positive emotional states.

This hypothesis connects the rating categories with emotional states. Linguistic and cognitive investigations have already made great progress in establishing that expressives correlate with emotional states (Kaplan 1999; Jay 2000; Jay and Janschewitz 2007; Jay et al. 2008). Thus, (2) straightforwardly predicts that expressivess will have distributions with one or two frequency peaks relative to our rating scale, always at the extreme ends of that scale. This is uniformly what we find in our data sets, and it paves the way to studying the particular contribution of expressives in terms of their distribution in these data sets.

3 Taboo intensives

Many languages have *intensive* particles that serve to highlight or amplify specific pieces of information (Beaver and Clark 2008). In English, many uses of *totally*, *really*, stressed *SO*, and *absolutely* fall into this category:

(3)
a. I could absolutely/totally/SO jump over that fence.
b. MTV like totally gave us TWO episodes back to back. It was like so random. The more the merrier, but it's like waay too much for one recap. (Web example from Beaver and Clark 2008: 74)

2 It is noteworthy that the more powerful quadratic model has a high p value, though the fit looks good. With such a small quadratic coefficient, even if it were significant, the difference between the quadratic and similar fits in the range we're looking at is so small that it doesn't make sense to read anything into the curvature.

From our perspective, the most interesting intensives are the taboo ones: *damn*, *fucking*, *bloody*, and variants thereof. Though many are ambiguous between literal and expressive meanings, it is generally possible to isolate the expressive uses:[3] they are blocked in predicative position, they can intensify other adjectives (*damn*/**surprising good*), and they are semantically unrestrictive. (For additional discussion of their morphosyntax, see Potts 2005; Potts et al. 2007.) While all the items in this class are taboo to some degree, *damn* is relatively unencumbered in this way. Speakers use it quite freely in a variety of texts and contexts. Thus, we can gather enough data to address the question, Under what circumstances do speakers choose *damn* to convey their intended messages?

The stereotypical uses of *damn* seem to involve emotions like anger, frustration, and aggression, as in the following examples:[4]

(4) a. Try answering the damn question.
    b. Sounds like another damn politician to me.
    c. Just a damn minute! What history books did you read?

However, it would be a mistake to assume that these uses characterize the contribution of *damn*. It is easy to find examples in which it is intended to convey solidarity, exuberance, trust, excitement, sadness, and others, as illustrated briefly in (5).

(5) a. funnest damn movie i've seen all year[5]
    b. Linnea Faris, a woman from Michigan who was wearing a "Remember Alex" T-shirt, shook her head in disbelief. [. . .] "I've spent hours crying over that damn bird."[6]

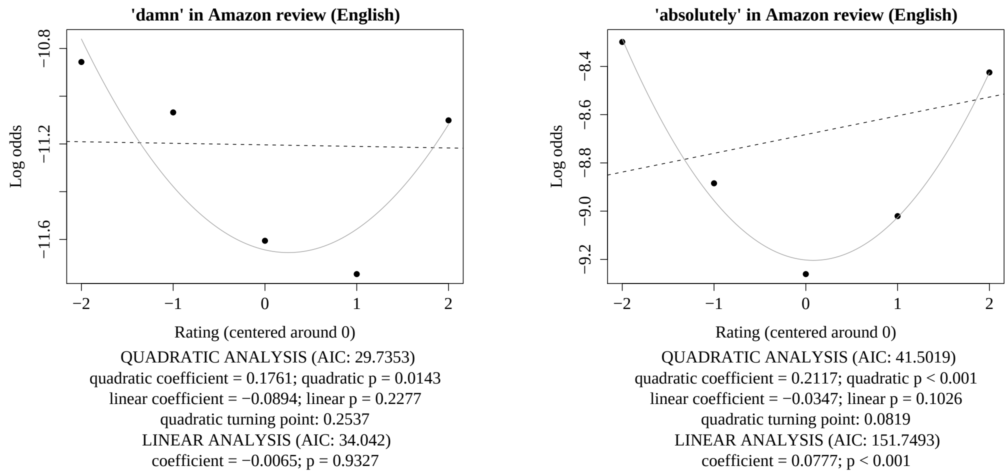
This small sample suggests that *damn* is compatible with a variety of specific emotions, contra the claims of Potts (2005, 2007). This conclusion is reenforced by quantitative assessment. Figure 4(a) summarizes an analysis of *damn* in the Amazon collection, using the statistical methods described in section 2. What we see strongly resembles the plot for *wow* that we looked at earlier. The quadratic regression is a good fit, and the turning point is not significantly different from 0. Here, the linear model is plainly inadequate, as is evident both from the high p value and the high AIC value. The difference between the two extreme categories is small (*damn* is 27% more likely to appear in a one-star than a five-star review), especially when

3 This is not so easily done with raw text, though, and we make only minimal efforts to do it in this paper. Where this is an issue, we regard it as a tolerable introduction of noise into the data.

4 These are drawn from the 20_newsgroups corpus, which is widely available on the Internet.

5 Pang et al. 2002

6 The New Yorker Magazine, May 12, 2008, p. 64



(a) The taboo intensive *damn*.

(b) The intensive *absolutely*.

Figure 4: Like *wow*, these are U-shaped distributions: the quadratic fit is good, with a positive coefficient and a turning point near 0, and, the linear model is disfavored, because the U is too deep to be modeled as a line.

compared with differences between the extremes and the middle (it is about 66% more likely to be in a five-star review than a three-star one). To further support the connection between these taboo items and non-taboo intensives, we offer an analysis of *absolutely* in figure 4(b), from the same data set and using the same methods.[7]

In sum, *damn* is an exclamative, indicating heightened emotion without biasing in one direction or another. This is not to say, though, that individual uses of *damn* don't clearly convey positivity or negativity. On the contrary, one striking thing about (4) and (5) is how easily we identify the intended emotional polarity. This suggests that the syntactic and semantic environment conditions these effects. Our corpora provide preliminary evidence that this is so. Using an independently gathered classification of 151 subjective adjectives into 'positive' and 'negative',[8] we counted the occurrences of *damn POS-ADJ* and *damn NEG-ADJ* in the English Amazon corpus. Though there are not quite enough tokens for a reliable statistical analysis, the numbers we did get for *damn POS-ADJ* are suggestive: 26 tokens in the five-star category, 2 in the four-star category, and just 2 more tokens in the remaining three low-end categories. Thus, it looks like the lexical content of the

7 Individual intensives present different shapes, though the shapes are largely consistent across the Amazon and Tripadvisor corpora used in Potts and Schwarz 2008. We conjecture that this is largely a function of lexical ambiguity and the difficulty inherent in isolating intensive senses.

8 `http://www.keepandshare.com/doc/view.php?u=12894`

adjacent phrase is a good indicator of emotional polarity.

Not all intensives have the balanced emotional polarity of *damn*. In this sense, we find an interesting contrast with the Chinese swear *tāmā* (literally ‘his mom’), which usually appears as *tāmāde*. In many ways, *tāmā* resembles *damn*. Both have been bleached of any literal content, both can be used as interjections and as morphosyntactically integrated modifiers, and, in terms of the intuitive (albeit vague) measure of emotive strength, both are mild. The prominent early-20th century writer Lu Xun called *tāmā* China’s ‘national swear’ in his affectionately disapproving essay ‘On (the swear) “Your mother. . . ” ’[9]. As we might expect of an item whose “frequency of use may not be less than that of the polite *nin hao ya* [‘hello’]”, forms of *tāmā* are common in our MyPrice corpus. Some representative examples:

(6) a. shēngyīn dī, shēngyīn hěn dī, shēngyīn fēicháng dī, shēngyīn
sound low sound very low sound extremely low sound
zhēn tāmā de dī
really **tāmā** DE low
‘The sound is soft. . . very soft. . . extremely soft. . . really damn soft.’

b. mǎi lājī dōu bùyào mǎi tiāngéxīn de chǎnpǐn, mǎi le jiù
buy trash all don’t buy Topsec DE product buy ASP immediately
méiyou rén guǎn, zhēn **tāmā** piànzi
not-have people care really **tāmā** swindler
‘Even if you buy trash, don’t buy Topsec’s products. After you buy, they just don’t care, real damn swindlers.’

In light of these data, we might be tempted to offer *damn* and *tāmā* as a translation pair. However, the two have very different distributions in our corpora. Figure 5 summarizes our distributional analysis of *tāmā*. It mirrors *superb* in figure 3(b). Here, we have a linear inverse correlation with rating categories, reflecting the fact that *tāmā* closely associates with negativity. The quadratic model predicts a sharp drop-off at the highest points in the rating scale, but that model is not a good fit overall, suggesting that it would be hasty to draw conclusions from the nature of this curve. The linear model is superior, and it predicts frequencies that we might characterize as ‘mild’. It is possible to use *tāmā* throughout the rating scale — it is not so negative, for example, that it is excluded from positive reviews — but the expressive nonetheless does associate with negativity: *tāmā* is vastly more frequent in one-star reviews than in five-star reviews ($e^{-7.447+10.948} = 33.142$), and the frequencies drop off steadily at every interval in between.

9 http://singaporeangle.blogspot.com/2005/07/lu-xun-on-chinese-national-swear.html

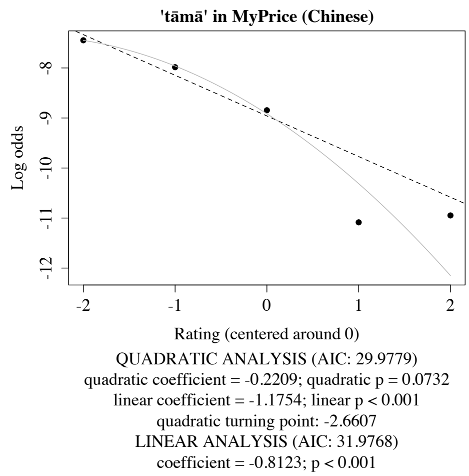


Figure 5: The Chinese swear *tāmā*. The linear model, which is more reliable here than the quadratic one, captures the distribution well. We see a steady drop off in frequencies as we move up the rating scale.

Thus, we are not saying that *tāmā* must always be used negatively, just that it creates expectations of negativity. As the distribution above indicates, our corpus does contain positive uses. We illustrate in (7).

(7) tài **tāmā** de zhèngdiǎn le
too **tāmā** DE right-on LE
'Absolutely friggin' right on!"

A speaker who uses *tāmā* positively takes a much greater risk than a speaker who uses *damn* that way. Our experiences prime us to expect negativity from this item, and those expectations have to be overcome somehow.

4 An emotionally-layered Chinese noun

Taboo intensives like *damn* and *tāmā* are easy targets for an expressive analysis. Dictionaries flag them as taboo, and they do little else but imbue the utterances containing them with expressivity. The goal of this section is to show that our data are sensitive to subtler forms of expressive content as well. Our central example is the Chinese common noun *dōngxi*. It is the most general and informal word for things, usually smallish things, and has unemotional uses, as in the following from our MyPrice corpus:

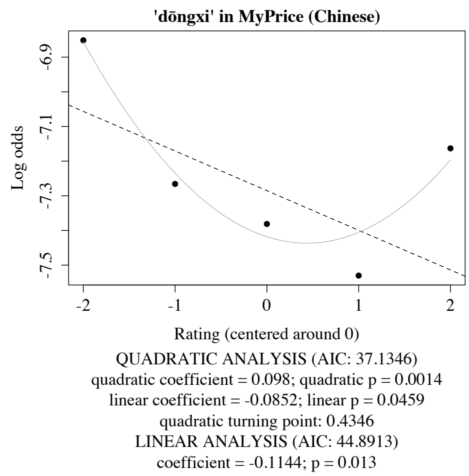


Figure 6: The Chinese common noun *dōngxi*. The linear fit, while significant, fails to capture the Reverse-J shape. The quadratic fit brings it out well.

(8) wǒ bù gǎn yòng lái bǎocún zhòngyāo **dōngxi**
1SG not dare use for store important **thing**
'I'm scared to use it to store anything important.' (said of a hard drive)

Since its meaning is broad and potentially unemotional, the expressivity of *dōngxi* might not be immediately evident. However, the item has a distinctly expresssive profile in our corpus, as figure 6 shows.

In many ways, this is the most complex distribution we have seen so far. Both the quadratic and linear models are good, with the AIC slightly favoring the quadratic model. It is easy to see, though, why the linear model is reasonable in this case: the turning point of the quadratic is shifted fairly far to the right — it is, in fact, marginally statistically different from 0. It looks as though the frequency data are almost linearly correlated with the rating categories, but the rise in the five-star category is really too significant to ignore. In the terms of Potts and Schwarz 2008, this is a Reverse-J distribution.

To understand what is happening linguistically, it helps to look at a sample of common phrases containing *dōngxi*:

(9)
a. hǎo dōngxi (good + thing, 'nice one')
b. pò dōngxi (worn out + thing, 'piece of crap')
c. huài dōngxi (bad + thing, 'scoundrel')
d. gǒu dōngxi (dog + thing, 'bastard')
e. lǎo dōngxi (old + thing, 'geezer')

The positivity of (9a) is predictable from the nature of the modifier. Uses of *dōngxi* in positive sentences may contribute some sense of intimacy with the (non-human) object in question. The examples in (9b–e) are negative. In a few cases, it is clear why: the modifier *dōngxi* combines with has negative qualities. In others, though, the result seems to be a quirk of expressivity. For example, (9e) has a derogatory flavor along the lines of 'geezer', though there is nothing inherently negative about *lǎo* ('old').

As the sample in (9) suggests, it is easier to find negative phrases containing *dōngxi* than it is to find positive phrases containing it. Furthermore, in phrases like the ones in (9), *dōngxi* seems to add a layer of heightened emotion. This bias is visible in the significant quadratic component of the distribution.

One general question that emerges from our discussion of *dōngxi* is the following. How can the expressive contribution a lexical item brings with it be distinguished from the overall expressive profile of the phrases it most commonly occurs in? That is, does figure 6 characterize an aspect of the meaning of *dōngxi* itself, or does it reflect facts about the relative frequencies of fixed phrases like those in (9)? These questions are relevant not only for the analyst, but also for language users who have to make decisions about producing and interpreting emotional content. In future work, we hope to study the use conditions of unmodified *dōngxi*, as well as the items with which it combines, with the goal of pin-pointing the locus (or loci) of expressivity.

5 A German positive expressive

Expressive uses of German *hammer*, as in the Amazon examples in (10), are overwhelmingly positive.

(10) a. Dieses Album ist **der Hammer**
This album is **the hammer**
'This album is cool.' (≈ 'the bomb', colloquially)

b. amys stimme mischt sich mit **hammer** beats
Amy's voice mixed self with **cool** beats
'Amy's voice mixed with cool beats.'

c. Was für ein **Hammer** Album
what for a **cool** album
'What a cool album!'

Unlike *damn* and its ilk, *hammer* has predicative uses, both with the article (as in (10a)) and without (*Das video is hammer!*). However, its obligatory lack of inflection

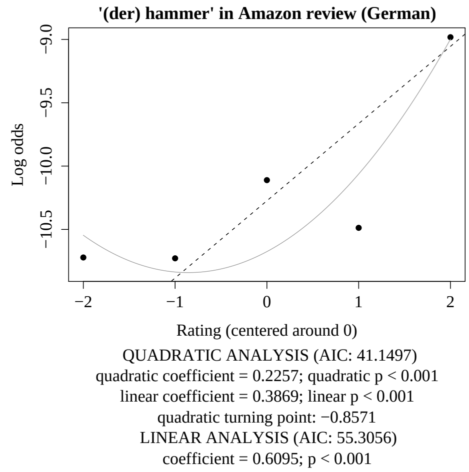


Figure 7: The German positive expressive *(der) hammer*

(*ein schönes*/**hammeres Album*, 'a beautiful/hammer album') is a hallmark of its membership in the expressive lexicon (Potts and Roeper 2006).

None of the examples in (10) contains a modifier that would contribute positivity of its own, so we can trace the emotional polarity to *hammer* itself. (Not surprisingly, though, *hammer* frequently co-occurs with other positive expressives like *geil* ('cool') and uninflected *absolut*, which bring their own positive biases.) The positivity carries through our entire data set, as figure 7 shows. This figure includes spurious hits (references to real tools, and uses of *hammer* in the idiom *wo der hammer hängt*, 'to display exemplary abilities'), but the majority of uses are expressive. In this case, the difference between the two ends of the scale is extreme: *hammer* is 470% more frequent in the five-star category than in the one-star category. Both the quadratic and linear models are good, and the turning point of the quadratic curve is quite far to the left. All these measures point to the conclusion that *hammer* is strongly associated with the most positive reviews. It is arguably more superlative even than *superb*, which remained frequent even in the relatively temperate parts of the scale.

6 A Japanese antihonorific

Our Japanese corpus contains a wealth of both honorifics and antihonorifics, many with noteworthy statistical profiles. In this section, we look briefly at the Japanese antihonorific form *te simaw*, consisting of the infinitival suffix *te* followed by the auxiliary verb *simaw*.[10] An example of this form from our Japanese Amazon corpus is given in (11).

(11)

harii	pottaa	to	wa	nagai	o-tukiai	de,	owat-**tesimat**-te	totemo
Harry	Potter	with	TOPIC	long	HON-acquaintance	be	end-**ANTIHON**-INF	very

kanasii	omoi-ga	si-masu
sad	thought-NOM	do-HON

'I had a long acquaintance with the Harry Potter series, and was sad to see it end.'

Potts and Kawahara (2004) study *te simaw* under the rubric of expressive content. In their examples, it appears as the phonologically reduced form *chimaw*:

(12) nesugoshi-**chimat**-ta.
overslept-**ANTIHON**-PAST
'I overslept, which sucks.'

This phonologically reduced version was not robust in our corpus, so the results reported here are for the unreduced form *te simaw*.

Figure 8 is our analysis of this morpheme, which has 4,525 tokens in our data.[11] Here, the choice of model is easy: the linear fit (dashed line) is extremely good, whereas the quadratic fit is not significant. We are not sure why the empirical frequency for the two-star category is larger than it is for the one-star category (it is about 40% bigger). The linear model predicts that the one-star log-odds value is about −6.19, whereas the predicted log-odds for the five-star category value is about −6.55. Thus, the overall prediction is that *te simaw* is 43% more frequent at the negative extreme than at the positive one, which fits well with Potts and Kawahara's (2004) claim that it signals that the speaker "has contempt for the proposition expressed by the clause in which it appears".

7 Reliable signals and lexical pragmatics

The goal of this section is to set up a general framework for incorporating the above distributional information into a theory of the lexical pragmatics of expressives. The central idea is that the frequencies tell us about the strength and nature of hearer expectations. Speakers, attuned to these expectations, have a strong incentive to work with them, rather than against them (Lewis 1969).

10 The *w* at the end of *te simaw* is the assumed consonantal ending of the root of the verbal suffix, which never shows up as such in overt morphology.

11 This count is based on the regular expression `^て（しまい|しまう|しまっ)$`.

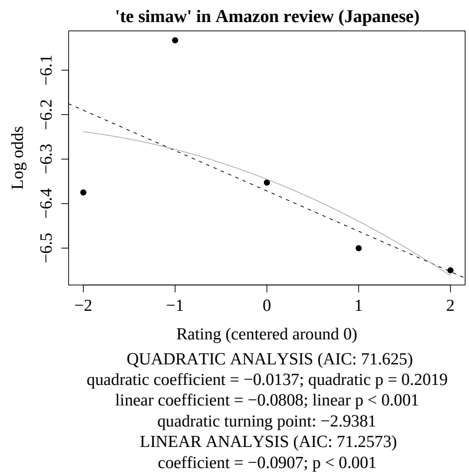


Figure 8: The Japanese antihonorific *te simaw*.

We assume that the corpus evidence is representative of our linguistic experiences. Thus, for example, it is part of English speakers' shared implicit knowledge of their language that *damn* signals heightened emotion. In turn, one uses this item only when one is in such a state (or wishes to create that impression). Usage that runs counter to this is suboptimal because it is so likely to be misunderstood. Thus, *damn* is a reliable signal of heightened emotion. By itself, though, it says little about the emotional polarity of that emotion. In that sense, it is an unreliable signal. Intuitively, this reflects the fact that my utterance of (13) might leave you genuinely puzzled as to my feelings.

(13) Sam bought that damn bike.

If you know I've been trying to discourage Sam from wasting $6,000 on a bike he doesn't need, then you will likely understand that my usage is negative. Similarly, if we've been losing races to Sam and he has just bought an even faster bike, then you'll perceive a resigned solidarity. However, if you know that I am delighted by the prospect of perhaps getting a ride on Sam's spiffy new bicycle, then *damn* will probably convey exuberance. Knowing nothing about the context, though, you will perceive only heightened emotion.

We identified many expressives that do strongly bias in one direction or another, though we do not claim that this is a categorical bias. For example, Chinese *tāmā* is overwhelmingly negative, but we do find positive uses. The frequency data tell a specific story here as well: the default hearer assumption is that *tāmā* is negative. Speakers wishing to use it positively must take care to provide the right contextual support, lest a quite unintended emotive message shine through.

In what sense is this information lexical? On the one hand, it is intimately connected with specific lexical items (and constructions; Potts and Schwarz 2008. On the other hand, there is no sense in which it has lost its inherently stochastic, defeasible qualities; we think there are positive and negative uses of all the items we discussed, and that those uses can vary considerably in their overall strength. A great advantage of the frequency-based approach we have taken is that we need not push these items into fixed categories. We can, instead, take advantage of their variability.

8 Conclusion

Using large corpora of product reviews in Chinese, English, German, and Japanese, we have investigated a wide variety of expressive content items. The statistical models we build from this evidence allow us to establish abstract connections between expressives, to juxtapose them, and to quantify both their emotive strength and the reliability with which they indicate specific emotions. And, though we have not really touched on this here, it facilitates the identification of new expressives (Potts and Schwarz 2008). Above all else, it yields a rich source of empirical evidence for claims about the precise contribution of these highly elusive items.

References

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge University Press.

Beaver, David and Brady Zack Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Oxford: Wiley-Blackwell.

Beineke, Philip; Trevor Hastie; Christopher Manning; and Shivakumar Vaithyanathan. 2004. Exploring sentiment summarization. In Qu Yan; James G. Shanahan; and Janyce Wiebe, eds., *Exploring Attitude and Affect in Text: Theories and Applications*, Papers from the AAAI Spring Symposium, 4–7.

Bresnan, Joan and Tatiana Nikitina. 2008. The gradience of the dative alternation. In Linda Uyechi and Lian Hee Wee, eds., *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. CSLI Publications.

Cohen, Jacob; Patricia Cohen; Stephen G. West; and Leona S. Aiken. 2003. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. London: Lawrence Erlbaum Associates, 3rd edition.

Geurts, Bart. 2007. Scalar implicatures and local pragmatics. Ms., University of Nijmegen.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* .

Jay, Timothy. 2000. *Why We Curse: A Neuro-Psycho-Social Theory of Speech*. Philadelphia/Amsterdam: John Benjamins.

Jay, Timothy; Catherine Caldwell-Harris; and Krista King. 2008. Recalling taboo and nontaboo words. *American Journal of Psychology* 121(1):83–103.

Jay, Timothy and Kristin Janschewitz. 2007. Filling the emotion gap in linguistic theory: Commentary on Potts' expressive dimension. *Theoretical Linguistics* 33(2).

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.

Kaplan, David. 1999. What is meaning? Explorations in the theory of *Meaning as Use*. Brief version — draft 1. Ms., UCLA.

Lewis, David. 1969. *Convention*. Cambridge, MA: Harvard University Press. Reprinted 2002 by Blackwell.

Pang, Bo and Lillian Lee. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, 115–124. Ann Arbor, Michigan: Association for Computational Linguistics.

Pang, Bo; Lillian Lee; and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 79–86. Philadelphia: Association for Computational Linguistics.

Potts, Christopher. 2005. *The Logic of Conventional Implicatures*. Oxford Studies in Theoretical Linguistics. Oxford: Oxford University Press.

Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33(2):165–197.

Potts, Christopher; Luis Alonso-Ovalle; Ash Asudeh; Rajesh Bhatt; Seth Cable; Christopher Davis; Yurie Hara; Angelika Kratzer; Eric McCready; Tom Roeper; and Martin Walkow. 2007. Expressives and identity conditions. To appear in *Linguistic Inquiry*.

Potts, Christopher and Shigeto Kawahara. 2004. The performative content of Japanese honorifics. In Kazuha Watanabe and Robert B. Young, eds., *Proceedings of Semantics and Linguistic Theory 14*, 235–254. Ithaca, NY: CLC Publications.

Potts, Christopher and Tom Roeper. 2006. The narrowing acquisition path: From declarative to expressive small clauses. In Ljiljana Progovac; Kate Paesani; Eugenia Casielles-Suárez; and Ellen Barton, eds., *The Syntax of Nonsententials: Multi-Disciplinary Perspectives*, 183–201. Amsterdam: John Benjamins.

Potts, Christopher and Florian Schwarz. 2008. Exclamatives and heightened emotion: Extracting pragmatic generalizations from large corpora. Ms., UMass Amherst.

Appendix: Data sets

The Japanese corpus was preprocessed using the morphological analyzer MeCab. The ratings for the Chinese corpus are an average of a number of a few different five-star rating categories (‘value for the price’, ‘service and support’, ‘quality and reliability’, ‘features’), not all of which appear consistently with all reviews, probably due to a site update at some point; to calculate the overall rating of a review, we average all the scores which were entered for that review, rounding to the nearest integer (reviews without any rankings were ignored).

Chinese (MyPrice.com.cn)						
	1 star	2 star	3 star	4 star	5 star	total
reviews	2115	3042	8007	2,055	2,294	17,513
characters	73798	111,659	236,184	65,264	56847	543,752
Unique authors (by user-supplied name): not known						
English (Amazon.com)						
	1 star	2 star	3 star	4 star	5 star	total
reviews	3,323	2,687	3,994	8601	34,952	53,557
words	570,687	512,643	767,958	1,513,776	4,769,921	8,134,985
vocab	27,352	26,239	32,818	46,306	80,569	112,323
Unique authors (by user-supplied name): 40,625						
German (Amazon.de)						
	1 star	2 star	3 star	4 star	5 star	total
reviews	2,987	1,881	2,647	4,431	15,784	27,730
words	407,888	319,341	467,556	788,915	2,205,666	4,189,366
vocab	35,177	30,835	37,644	53,539	95,453	144,418
Unique authors (by user-supplied name): 16,623						
Japanese (Amazon.co.jp)						
	1 star	2 star	3 star	4 star	5 star	total
reviews	971	759	1,609	3,504	11,031	17,874
words	127,049	123,312	277,857	636,067	1,805,764	2,970,049
vocab	9,574	9,909	16,247	24,902	39,948	49,054
Unique authors (by user-supplied name): 12,747						