Affective 'this'

Christopher Potts and Florian Schwarz

Affective 'this'

CHRISTOPHER POTTS AND FLORIAN SCHWARZ, *Stanford University and University of Pennsylvania*

Abstract

Lakoff (1974) argues that affective demonstratives in English are markers of solidarity, with exclamative overtones deriving from their close association with evaluative predication. Focusing on *this*, we seek to inform these claims using quantitative corpus evidence. Our experiments suggest that affectivity is not limited to specific uses of *this*, but rather that it arises in a wide range of linguistic and discourse contexts. We also briefly extend our methodology to demonstrative *that* and to German *diese-* ('this').

1   Introduction

When we think of emotive language, we are apt to think of predicates of personal taste (*boring*, *tasty*), swears, (*bastard*, *damn*), exclamatives, honorifics, politeness markers, and other words and constructions that are more or less dedicated to the task of conveying information about our attitudes and emotions. For particular utterances in context, though, the sources of such expressive content are not always so apparent. Very commonly, they are hiding in plain sight, in the ways in which innocuous-seeming functional elements are used (Campbell and Pennebaker, 2003, Chung and Pennebaker, 2007).

Demonstratives provide a clear instance in which subtle forms of expressive content can arise from apparently colorless sources. Lakoff (1974) discusses three general uses: spatio-temporal deixis, anaphora, and emotional deixis. The first two are the best studied and arguably constitute the unmarked cases. The third, which we henceforth refer to as *affective* (Liberman, 2008), are Lakoff's focus (and ours). In English, both the proximal demonstratives (*this* and *these*) and the distal demonstratives (*that* and *those*) have affective uses, though with subtly different senses, as Lakoff observes.

Lakoff's central generalization is that affective demonstratives are markers of solidarity, indicating the speaker's desire to involve the listener emotionally and foster a sense of closeness and shared sentiment. For example, (1a) gestures at a common view of Kissinger (enthusiasm, exasperation) among the discourse participants, and the quasi-indefinite *this traveling salesman* in (1b) seems "to involve the addressee more fully" (p. 347) than an indefinite form would.

(1)   a. This Henry Kissinger is really something!

       b. There was this traveling salesman, and he . . .

The goal of the present paper is to inform and refine this generalization using quantitative corpus evidence, taking advantage of existing tools (Ramshaw and Marcus, 1995, Greenwood, 2005, Manning et al., 2008, Hatzivassiloglou and Wiebe, 2000b) and extending a set of sentiment corpora that we recently released (Constant et al., 2009b). In order to keep the scope of the paper manageable, we largely restrict attention to the proximal demonstrative *this*. We first discuss its affective uses in more detail (section 2), building not only on Lakoff (1974) but also on subsequent work by Prince (1981a), Bowdle and Ward (1995), Wolter (2006), Liberman (2008), and others. Section 3 reports on our corpus experiments, which support the generalization about solidarity and also further substantiate Lakoff's brief remarks about the connections between affective demonstratives and exclamatives. Section 4 tentatively extends our methods to distal demonstratives and to German, and our closing section 5 offers some general remarks about the usefulness

of corpora in pragmatic investigations. The appendix describes the data files associated with this paper.

The next section studies a few constructions that often or always give rise to affective readings, and we use corpora to gain a better understanding of them. For the large-scale corpus work of section 3, though, we do not seek to pin-point affective uses of *this*. We think it would be a mistake to try to restrict attention just to a few constructions. The affectivity in question is gradient, not categorical; individual uses can arrive with more or less affectivity, and with different shades of meaning. Thus, we aim to show that the affectivity (i) arises from the more basic spatio-temporal meaning of *this* and (ii) is strong enough to reveal itself across different linguistic and discourse contexts. Corpus experiments are ideally suited to these aims because they allow us to identify subtle patterns that hold across a vast range of real-world situations.

2   Affectivity effects

Lakoff summarizes her characterization of affective (emotional-deictic) *this* as follows (p. 347):

> Under this rubric I place a host of problematical uses, generally linked to the speaker's emotional involvement in the subject-matter of his utterance. Since emotional closeness often creates in the hearer a sense of participation, these forms are frequently described as used for 'vividness.' And since expressing emotion is — as I noted last year — a means of achieving camaraderie, very often these forms will be colloquial as well. This is used for several reasons, all linked to the achievement of 'closeness,' like spatio-temporal *this*, in a rather extended sense.

This maps out a fairly specific path from the basic meaning of *this* to its solidarity effects. The final sentence indicates, in a qualified way, that Lakoff sees both spatio-temporal and affective uses as potentially arising from a single core meaning. (Earlier in the paper, she suggests that this is just one example of a more general tendency for spatio-temporal expressions to have their "sphere of usability" (p. 346) extended into emotive realms.) The core meaning is something like 'proximity to the speaker', with the precise notion of proximity — physical, temporal, emotional, etc. — left open. Though one can imagine this closeness being exclusionary, it seems generally to pull the listener in, thereby fostering a feeling of shared sentiment and experience.

Affective uses arise in many morphosemantic and discourse contexts, so we cannot provide anything like an exhaustive look at such readings. Instead, we focus on a few environments in which such readings are prominent — proper names and generics (section 2.1) and existential constructions (section 2.2) — in an effort to get a better sense for the phenomenon.

2.1   Proper names and generics

Lakoff (1974) says that affective uses are common with proper names "that the speaker expects the hearer to be familiar with" (p. 347), as in (1a) above, and that they commonly involve referents that have not been mentioned previously in the discourse. She suggests that these demonstratives might be *uniformly* less felicitous if the referent has already been established, citing (2), but we have found exceptions to this (see also Wolter 2006: 83), as in (3), which is drawn from the corpus used in section 3 below. We suggest that (2) sounds odd at least in part because it is not clear why the speaker chose this particular sequence of referring devices (Vonk et al., 1992).

(2) $^?$The Secretary of State has made peace in the mideast. This Henry Kissinger really is something!

(3)   In "Darkly Dreamy Dexter" author Jeff Lindsay introduces us to the protagonist Dexter Morgan, a police criminologist working in Miami. More specifically, he's a blood splatter analyst who just happens to be revolted by blood because of the mess that it makes. Oh, and he's also a lifetime serial killer who lives by a code that only allows him to target other criminals. Some complex guy, this Dexter Morgan, eh?

Nonetheless, affective demonstratives often do lack overt antecedents in the discourse. This amounts to a presumption that the referent is, in the sense of Prince (1981b), hearer-old but not necessarily discourse-old. This status entails shared information, so felicitously making this demand (rather than avoiding it) is, in and of itself, a small gesture of solidarity.

Wolter (2006: 83) further supports the familiarity claim by showing that demonstrative-headed proper names sound marked if the addressee has established that he lacks familiarity with the referent of the name:[1]

(4)   a.   Who is John Smith (and what is he like)?

b.   ($^\sharp$This) John Smith is a really great guy who builds birdhouses!

Bowdle and Ward (1995) study a range of demonstratives with generic senses, as in (5), where the intended reading arises if we assume that the speaker is exclaiming about ThinkPads in general, as opposed to the specific one(s) present in the room.

(5)   a.   [in front of a computer] These IBM ThinkPads are amazing!

b.   [in front of a computer] This IBM ThinkPad is amazing!

Like Lakoff, Bowdle and Ward establish a link with speaker–hearer solidarity, concluding that these generics "mark the kind being referred to as

---

[1]Wolter's original example makes this point for *that*.

a relatively subordinate or homogeneous kind located among the speaker's and hearer's private shared knowledge" (p. 32), where the relevant notion of 'kind' is flexible enough to extend to certain kinds of events and predications. We can again apply Wolter's technique from (4) to test the strength of this familiarity effect:

(6) a. What is a ThinkPad?
     b. (♯These) ThinkPads are amazing new laptop computers!

The speaker of (6a) indicates, via a general, unbiased question, that he doesn't know what a ThinkPad is. The reply is natural without *these* but seems infelicitous with it. Bowdle and Ward predict this effect: the unbiased question in (6a) clashes with the presumption of shared knowledge about ThinkPads that *these* carries in (6b).

Bowdle and Ward (1995) observe that "the predicate of a generic demonstrative is typically evaluative" (p. 33). Where it is not, the results tend to be odd, as in (7).

(7) ♯ These IBM ThinkPads have plastic cases!

If such uses sound reasonable in some contexts, it is likely because the predications can take on evaluative overtones.

The examples in (8) suggest that Bowdle and Ward's evaluative generalization extends to proper name versions as well. Where the predication is purely factual, the result sounds marked (8a), and the corpus examples in (8b-8d) all involve evaluative claims, either because of their lexical content (8b) or because their intended force is evaluative, as in the exclamative (8c) and the skeptical question (8d).

(8) a. ♯ This Henry Kissinger is Secretary of State!
     b. You don't deserve to be discouraged and lied to by a con artist. Which is what this Arthur Agatston is.
     c. Who is this William Young and where has he been? This wonderful work [...]
     d. And who is this John Perkins, who claims that he could confound the best economists of the World Bank and other aid institutions?

We sense also that the content of the evaluative predication involving the *this*-headed proper name is assumed by the speaker to be uncontroversial. In (3), for example, the speaker builds his case for the evaluative claim that Dexter is a "complex guy" before choosing to use *this Dexter Morgan* over the bare proper name. If this effect is real, then it too will reliably enhance solidarity, as it ensures not only agreement, but agreement about something subjective (for discussion, see Lasersohn 2005).

Lakoff's comments on the evaluativity effect are actually more specific than this, in that she ventures a connection with exclamatives and exclamativity (p. 354). In this light, it is noteworthy that many of the examples of affective *this* in the literature include exclamation points, as well as other markers of exclamativity (e.g., *really* as used in (1a)). Exclamatives are inherently evaluative, to such a degree that even non-evaluative predicates take on evaluative components inside of them, as in (9).

(9)   a. What a hotel/street/view!

      b. Boy, is it ever summertime!

      c. I would absolutely visit India!

This is not to say that affective demonstratives are exclamative constructions, but rather that the two morphosyntactic phenomena have something in common at a more abstract pragmatic level. Section 3.2 offers experimental evidence aimed at clarifying the connection.

2.2   Specific indefinites

Affective senses of *this* also arise when it is used in the specific-indefinite sense studied by Prince (1981a), as we saw in (1b). That example uses the existential-*there* construction to isolate the relevant reading of *this traveling salesman*. In (10)–(11), we illustrate with two more environments that require indefinite phrases and thus deliver indefinite readings for *this*: existential-*have* (Partee, 1999) and *warmth*-nominals (Lakoff, 1974).

(10)   a. Ed has {this/a/*the} grumpy old aunt who always calls him to complain.

       b. Ed has {these/two/*the} grumpy old aunts who always call him to complain.

(11)   He kissed her with {this/an/*the} unbelievable passion.

(Lakoff, 1974)

Prince (1981a) classifies these phrases as "unambiguously specific indefinite NPs". As one would expect from indefinites, they do not presuppose familiarity with the referent. In this sense, they contrast with proper names and kinds occurring with affective *this*, as discussed in the previous section. The solidarity effect does carry through to these uses, though. According to Lakoff (1974), *this* serves in these contexts "to give greater vividness to the narrative, to involve the addressee in it more fully" (p. 347).

We think the evaluative effect is in evidence as well, but more subtly than it was for the proper-name and generic cases. The majority of attested examples collected by Prince (1981a) contain evaluative language, including markers of exclamativity like *absolutely* and exclamation marks, but this is not an

absolute restriction; we can easily construct non-evaluative cases that sound natural. Thus, we performed a small corpus experiment to see if the evaluative effect is a reliable tendency.

Using the tagged data described in the next section, we heuristically identified all the existentials of the form *there BE {this/a(n)}*, allowing for optional adverbs immediately following *there* and the copula. There were 47 headed by *this* and 9,183 headed by *a(n)*.

To get an independent definition of evaluative language, we pooled a list of purely evaluative adjectives[2] with the adjective lists compiled by Hatzivassiloglou and Wiebe (2000a,b), which they show experimentally to be reliable predictors of subjectivity. There were 1,729 unique adjectives in the resulting list.

We measured the frequency of evaluative language in the *this*-headed examples vs. the *a(n)*-headed examples. In terms of absolute frequencies, the *this*-headed cases come out higher: 0.07 frequency for the *a(n)* cases versus 0.081 for the *this* cases.

Concerned that this might be in part due to the vast size differences between the two sets of sentences (Baayen, 2001), we also compared the frequencies for randomly selected 20-sentence subsets of each of these two sets over 100 trials. The results of this experiment are summarized in figure 1. The dark lines mark the median values, and the triangular indentations roughly delineate 95% confidence intervals for the medians. Here, the spans picked out by these indentations do not overlap, suggesting that the frequency difference is almost certainly real. Thus, though it is far from absolute, evaluative language is more common in *this*-headed cases, suggesting a deeper linguistic connection.

2.3 Summary

For a statement involving a *this*-headed proper name or generic *A*, we identify the following properties:

(12)   a. The speaker presumes that the hearer can identify the referent of *A*.

       b. The predication of the referent of *A* is evaluative.

       c. The evaluative predication is uncontroversial.

       d. The speaker intends to evoke solidarity with the hearer.

The effects for specific indefinites headed by *this* are somewhat different. Property (12a) does not hold, and indeed its negation seems to be true — we have a presumption on the part of the speaker that the hearer *cannot* identify the referent of *A*, which is in keeping with the indefiniteness of the phrase.

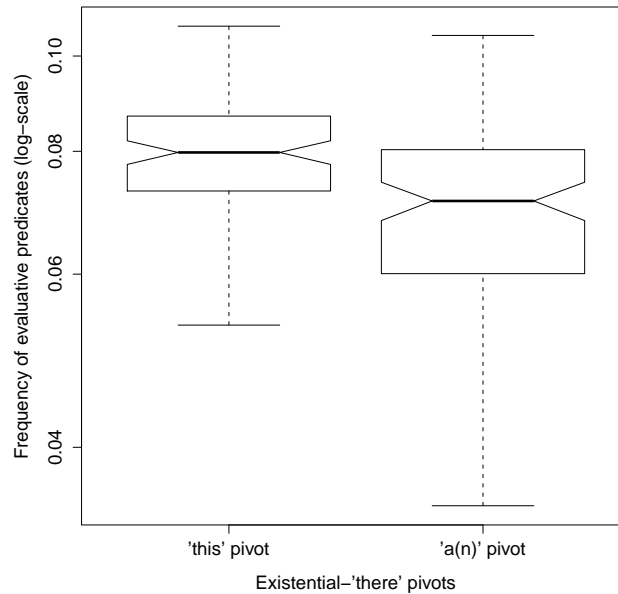[2]http://www.keepandshare.com/doc/view.php?u=12894

FIGURE 1　Box-and-whiskers plot of the frequency of evaluative language in *this-* and *a(n)*-headed existentials. For 100 trials, we picked random subsets of 20 sentences of each kind of corpus and calculated the relative frequency of evaluative language in those subsets. The result is strong evidence that evaluative language is more common in *this*-existential than *a(n)*-existentials.

However, we do find evidence for (12b), in the form of increased frequency for evaluative language. Property (12c) seems present, but passively: by the negation of (12a), the hearer is unfamiliar with the referent, so predications involving it are unlikely to be controversial. Property (12d) is in effect here as well, though.

The known lexical properties of *this*, proper names, generics, and specific indefinites bring us tantalizingly close to deriving these effects. Proper names always demand some degree of hearer familiarity (though the effect is stronger when *this* is present; (4)), and indefinites always resist hearer familiarity (though the effect is again stronger with *this*). And it is, at an informal level, no surprise that *this* should create solidarity — the effect seems like a natural extension of its basic uses as a marker of spatio-temporal proximity.

	Reviews	Words
1-star	8,434	1,537,917
2-star	7,545	1,517,654
3-star	10,083	1,899,550
4-star	28,186	4,609,556
5-star	64,147	9,094,340
total	118,395	18,659,017

TABLE 1 Basic numbers for the corpus we use throughout this paper.

3 Experiments

After describing our corpus (section 3.1), we turn to two experiments. The first informs the connections between *this* and evaluativity, by locating distributional parallels with exclamatives (section 3.2). The second probes for the solidarity effect, by studying how *this*-usage changes as the degree of basic sentiment uniformity changes (section 3.3). For these experiments, we study *this* in all its morphosyntactic environments, in an effort to show that its affectivity is in evidence at this general level, even though the data likely include many non-emotive uses.

3.1 Data

Our corpus is an extension of the English-language portion of the UMass Amherst Sentiment collection (Constant et al., 2009b), which consists of online product and hotel reviews drawn from Amazon and Tripadvisor websites (along with a smaller Chinese collection from MyPrice.com.cn). Each review has associated with it a short summary text (averaging about six words), a star rating, a date, and a username. In the experiments reported on here, we work with just the review texts, the dates, and the star ratings, which are given on a scale of 1 to 5 stars (1 the most negative, 5 the most positive).

We have extended the basic English portion primarily in an effort to gather more 1-star ratings. As the summary numbers in table 1 show, the bias for five-star reviews is still extreme (54% fall into this category), but we have enough low-rated reviews to support the inferences we aim to make.[3]

---

[3]The bias for five-star reviews is very general for this kind of corpus. The majority of online product reviews, across different websites and product areas, assign five-stars where that is the scale (Chevalier and Mayzlin, 2006, Hu et al., 2006, Pang and Lee, 2008). It was not easy for us to find products with majority 1-2 star reviews even when we went looking specifically for them (though movies, video games, and TV shows finally bore fruit). The best evidence for this bias known to us is the Chinese Amazon corpus in the UMass Amherst Sentiment collection. That corpus consists of essentially the whole of the Chinese Amazon website (as far as we can tell). 39% of its 527,794 reviews are 5-star reviews; see http://semanticsarchive.net/Archive/jQ0ZGZiM/readme.html.

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. [. . .]

(a) A short 1-star review with summary text "B double O ring".

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 3-star review with summary text "Good info, but no depth".

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 5-star review with summary text "Good stuff!".

TABLE 2 Sample reviews from the corpus we use throughout this paper.

The review texts themselves are relatively short (the average length is about 200 words), largely informal, and often richly emotive, particularly in the reviews that either pan the product under discussion (1-star) or praise it loudly (5-stars). Reviews in the middle of the scale use more temperate language that balances the good and the bad. Table 2 reproduces three short reviews exemplifying 1-star, 3-star, and 5-star reviews.

We take two perspectives on the use of demonstratives in these reviews: (i) determiner uses (*I like this book*) and (ii) pronominal uses (*I like this*). To find such phrases, we ran the Stanford Log-Linear Part-Of-Speech Tagger (Manning et al., 2008) on the sentences containing demonstratives, and then we ran Greenwood's (2005) NP chunker on that tagged output. The tagger was trained on Wall Street Journal data, and the chunker is a Java implementation of Ramshaw and Marcus's (1995) NP chunker, which is transformation-based in the sense described by Jurafsky and Martin (2009: §5.6). These tools were not designed for data of the sort found in our corpus, so it is worth briefly assessing how well they performed.

		Correct status	
		Det.	Pro.
Predicted	Det.	108	3
status	Pro.	2	37

(a) The results of assessing the NP chunker for a randomly-selected set of 150 tokens of *this* in our corpus. The tagger accurately tagged true determiner uses 98% of the time and true prominal uses 93% of the time.

		Correct POS	
		Comp.	Det.
Predicted	Comp.	113	19
POS	Det.	3	15

(b) The results of assessing the tags for a randomly-selected set of 150 tokens of *that* in our corpus marked with either the DT (determiner) tag or the IN (complementizer) tag, which subsumes both finite-clause and relative-clause *that*. The tagger accurately tagged true complementizers 97% of the time and determiners 44% of the time.

TABLE 3 Assessment of the POS-tagger and NP chunker on our corpus.

For *this*, we needed only to distinguish pronominal uses from determiner uses, since the string 'this' receives only one tag (DT) in our data. Though the chunker does not handle complex NPs, it captures a significant subset of simple and modified NP structures, with relatively few spurious bracketings for demonstrative-headed phrases (demonstrative-initial strings). Table 3(a) gives the results of our manual assessment of 150 randomly selected tokens of *this*; accuracy for both determiner and pronominal uses is very high.

For *that*, the situation is more challenging. Before chunking, the part-of-speech tagger must distinguish demonstratives from complementizers. The results for the part-of-speech tagging are not good, as table 3(b) shows. True complementizers are identified 97% of the time, but true demonstratives are identified just 44% of the time. This is potentially very problematic, since the effects we study below might interact in nontrivial ways with the sort of information that is commonly conveyed in English via clausal embedding — speaker attitudes, evidentiality, and factivity — and clausal embedding correlates highly with complementizer *that*. In section 4.1, we return briefly to demonstrative *that*, arguing that its affective uses are more variable than those of *this* and briefly discussing some suggestive results from our corpus methods.

3.2 Exclamativity

Our first experiment employs the methods of Potts and Schwarz (2008) and Constant et al. (2009a): we study the relationship between usage and the star-ratings. What emerges is a distributional affinity, across these ratings, between clear markers of exclamativity and *this*, on both its pronominal and its deter-

miner uses. We thus obtain quantitative evidence for Lakoff's (1974) proposed link between *this* and exclamativity, a connection that is unsurprising given the more general correlation between affective uses and evaluativity.

Methods

To get a grip on usage patterns, we rely on log-odds distributions of words and phrases:

Definition 1 (Rating-relative token counts). $\text{count}_T(w_n, r)$ is the number of tokens of w_n (a word-string of length n) in reviews assigning rating r, drawn from a corpus of reviews T.

Definition 2 (Rating-wide counts). $\text{count}_{T,n}(r)$ is the number of tokens of word-strings of length n in rating category r relative to a collection of reviews T.

Definition 3 (Rating-relative log-odds).

$$\text{log-odds}_T(w_n, r) \stackrel{def}{=} \ln\left(\frac{\text{count}_T(w_n, r)}{\text{count}_{T,n}(r) - \text{count}_T(w_n, r)}\right)$$

The log-odds distribution resembles a regular frequency distribution (obtained with $\text{count}_T(w_n, r)/\text{count}_{T,n}(r)$). Figure 2 sets the two side-by-side, here illustrating with the frequency of the positive modifier *superb* in our corpus. The overall shape is the same, except that the distance between the smallest values is lengthened in the log-odds distribution. This change is justified as follows: the approximate frequencies for the one-star and two-star categories here are 0.0000072 and 0.000026, respectively. To make these numbers more manageable, let's rescale them by multiplying each by 1 million, to yield 7.2 and 26. In absolute terms, these numbers are close together (18.8 apart). However, the second is 3.6 times larger than the first. In contrast, the rescaled 4-star and 5-star frequencies are 77 and 120, respectively. In absolute terms, they are farther apart than our first part. However, the second is just 1.56 times bigger than the first.

Figure 3 plots the log-odds distribution of a variety of exclamative markers across the rating categories. In each case, the U-like shape of the distribution is evident from the empirical points (black dots). The U-shapes indicate that these phrases are uniformly more frequent in the extreme rating categories (1-star, 5-star) than they are in the more temperate middle. In the case of *absolutely*, *wow*, and the exclamation mark, the 1-star and 5-star values are about the same, reflecting the intuition that one uses these phrases when things are really good or really bad. For *damn*, the picture is roughly the same, but with a noticeable negative bias.[4]

---

[4]One might worry that the distribution is distorted by the differing sizes of the underlying rating categories seen in table 1. To address this (justified; Baayen 2001) concern, we created a
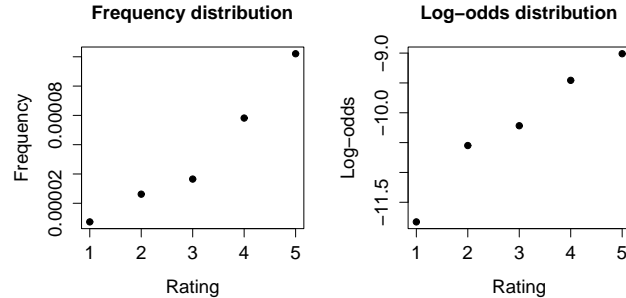
FIGURE 2 Comparing frequency and log-odds distributions, using *superb* as an example.

The curved lines represent logistic regression models, which we use to characterize the nature of these U shapes. For a phrase y, the model uses (13) to predict the probability $\Pr(y)$ of y being used given a rating x.

(13) $\Pr(y) = \text{logit}^{-1}(\text{intercept} + \beta_1 x + \beta_2 x^2)$

For example, the model for *absolutely* in our corpus is summarized in table 4. The coefficients determine the model in (14).

(14) $\Pr(absolutely) = -7.366 - 1.126x + 0.194x^2$

	coefficient estimate	standard error	p value
Intercept	-7.366	0.103	< 0.001
β_1	-1.126	0.075	< 0.001
β_2	0.194	0.012	< 0.001

TABLE 4 The fitted model for *absolutely*.

For the 5-star category, this predicts a frequency of $\text{logit}^{-1}(-7.366 + (-1.126 \cdot 5) + (0.194 \cdot 5^2)) = \text{logit}^{-1}(-8.141) = 0.00026$, which is close to the observed value of 0.00029. The predicted 3-star value is much lower: 0.00012 (cf. the observed value of 0.00011). That is, if you're writing a positive review, you are much more likely to use *absolutely* than you are if you are writing a lukewarm 3-star review.

Our primary concern is with the quadratic coefficient β_2, which determines the basic shape of the distribution. For each phrase in figure 3, the quadratic

---

balanced version with exactly 7,545 reviews in each rating category. All the basic results reported here hold for the balanced version as well.
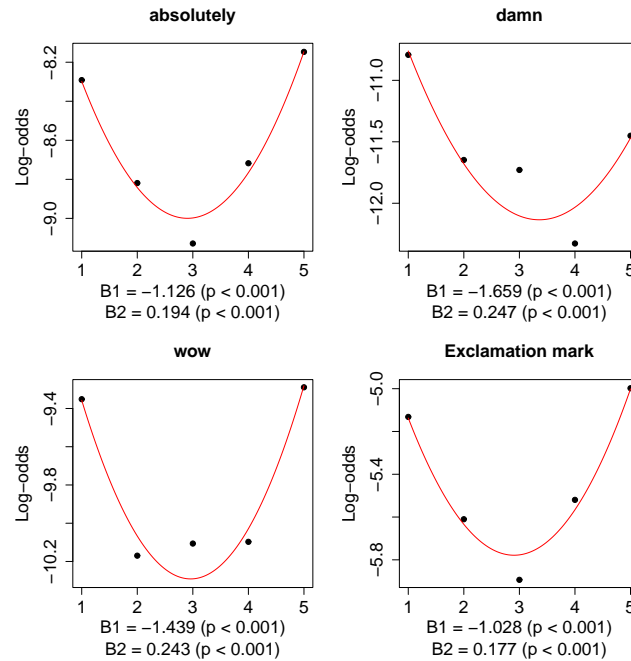
FIGURE 3 Exclamative log-odds relative to ratings. The U-shapes indicate that these expressions are common in the highly emotive 1-star and 5-star reviews and much less common in the middle-of-scale reviews.

coefficient is significant ($p < 0.001$). A positive coefficient corresponds to a U shape, a negative coefficient to a Turned-U shape. The absolute size of the coefficient determines its width, with larger coefficients corresponding to narrower curves. Thus, the deepest Us correspond to large positive quadratic coefficients. Throughout this picture, we see deep, statistically significant U shapes.

It is worth looking briefly at intuitively non-exclamative language, to see how it stacks up in comparison to the Us of figure 3. After all, if a lot of non-emotional language also displays a U-shape, then we might question whether these shapes are meaningful. Figure 4 should be reassuring in this regard. It depicts intuitively non-emotional language, with a uniform outcome: this language is most prominent in the more balanced middle-of-scale reviews, where the reviewer is weighing the good against the bad.[5] (For a rigorous

[5]Figure 2 highlights a different shape: a linear positive correlation with the rating scale, common for strongly positive modifiers. Strongly negative modifiers show an inverse correlation with the rating scale.
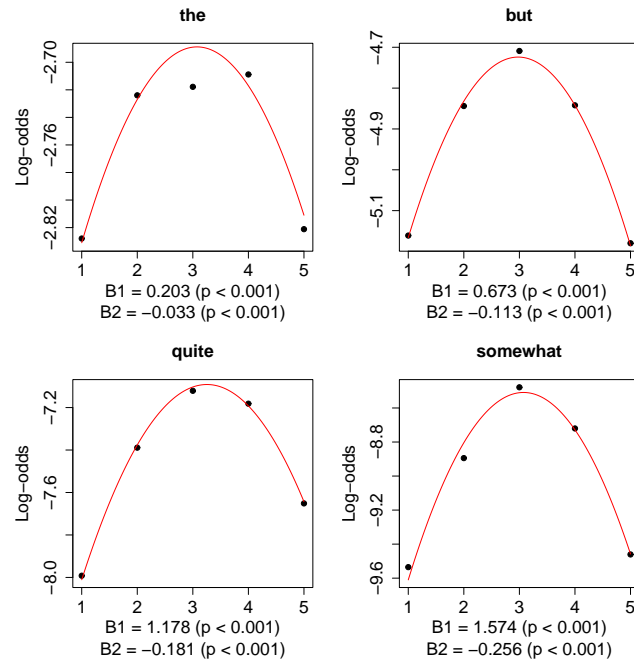
FIGURE 4 These nonemotional words are used more often in the 2-4 star reviews than at the extremes. This 'anti-exclamativity' puts them in direct contrast with expressions like those in figure 3.

look at the predictions this modeling makes for the whole of the vocabulary, see Potts and Schwarz 2008.)

How does *this* compare to the exclamatives of figure 3 and the non-emotional words of figure 4? Figure 5 provides the same perspective on *this*, showing both its pronominal and determiner uses. The same U-shape is evident for both, and the quadratic coefficient is again significant ($p < 0.001$). Figure 6 plots the coefficients for the items we have seen so far, to give a better sense of the relative size differences. The coefficients for determiner and pronominal *this* are not nearly as large as for the exclamative markers, suggesting that the exclamative effects are less pronounced for them (as one would expect), but the quadratic coefficient is still significant ($p < 0.001$).

Discussion

We connect the distributional evidence with pragmatics via two claims. First, speakers writing one- and five-star reviews are in heightened emotional states (or intend to create that impression); they loved or loathed what they were
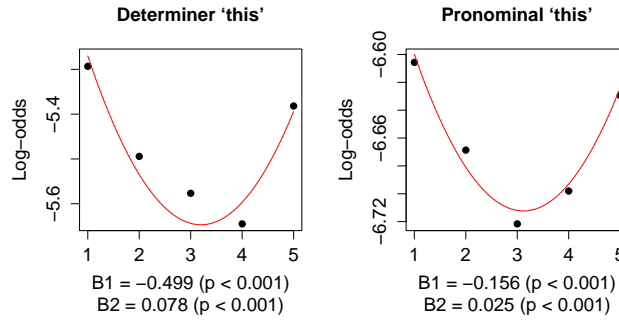
Determiner 'this'   **Pronominal 'this'**

B1 = −0.499 (p < 0.001)   B1 = −0.156 (p < 0.001)
B2 = 0.078 (p < 0.001)   B2 = 0.025 (p < 0.001)

FIGURE 5 Determiner and pronominal *this* pattern with exclamatives (figure 3) and contrast with *the* and other nonemotional words (figure 4).
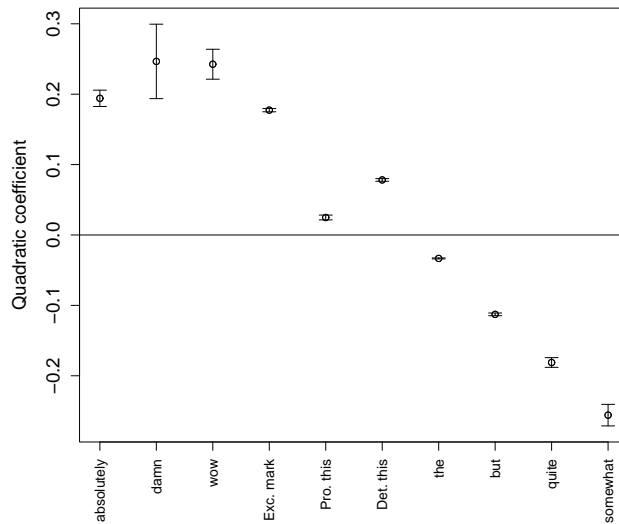
FIGURE 6 Quadratic coefficients, with ±1 standard error bars. Above 0, higher values correspond to increased exclamativity. Below 0, smaller values correspond to increased 'anti-exclamativity', as measured by frequency in different rating categories.

writing about, and they aimed to convey this. This is amply supported by the texts: scalar-endpoint reviews are full of emotional language, whereas those in the 2-star to 4-star range offer more nuanced opinions, balancing the good and the bad (Potts and Schwarz, 2008). Second, theoretical work on the

context-dependent semantics of exclamatives tells us that they signal heightened emotions (of many kinds). These two claims predict that exclamatives will be more frequent in extreme reviews, which is exactly what we find.

Pronominal and determiner *this* might be said to have 'milder' versions of the same exclamative profile. Like exclamatives, they are more common at the extremes, but the contrast with middle-of-the-scale reviews is less dramatic, as evidenced by the coefficients in figure 5, which are relatively small compared to those of other exclamatives, as shown in figure 6. We can think of two factors that might contribute to this difference: (i) *this* travels often with exclamative markers, but it does not typically determine much exclamativity on its own (though certain intonational tunes might suffice to make it punchier); and (ii) relatively few uses of *this* involve perceptible exclamativity, so the overwhelming percentage of uses are not sensitive to the rating-context. In general, though, we see emotivity and evaluativity shining through, not just when we restrict attention to the specialized environments of section 2, but also when we step back to look at all uses.

3.3 Solidarity

Our second experiment aims to better understand the solidarity effects of *this*. To do this, we take a different perspective on our corpus data. The products in our collection have between 1 and 1,068 reviews associated with them. For some products, the reviews are basically univocal in their praise or condemnation of the product. For others, the sentiments are more mixed.

Review authors can gauge the room quickly: on both Tripadvisor and Amazon, the review pages provide a barplot summarizing the distribution of reviews to date. Thus, the tenor of the room is immediately apparent, and review authors have the chance, for rooms with largely uniform sentiment, to align with that majority opinion or to dissent from it. We expect this to influence *this*-usage, as effects related to solidarity are more likely to arise when a reviewer joins in with the majority opinion.

Methods

For this experiment, we looked just at the reviews of products with at least 15 reviews.[6] This yielded 836 products and 116,312 individual reviews, totaling 17,609,549 words. For each product *A*, we gathered the following basic data:

1. The distribution of ratings across the reviews for *A*.

2. The number of occurrences of determiner and pronominal *this*, *that*, *these*, and *those* in reviews for *A* relative to individual rating categories, along with the same counts for *the* (which has only determiner uses).

---

[6]Fifteen seemed sufficient, given our goals, for avoiding unreliable frequencies resulting from overly sparse data. We obtained very similar results with this threshold set at 10, 50, and 100.

1	2	3	4	5	6	7
Product	Rating	Reviews	Pro. *this*	Det. *this*	Words	Pr(Product, Rating)
A	1	5	1	6	571	0.021
A	2	3	1	4	401	0.013
A	3	2	1	2	325	0.009
A	4	9	0	13	1828	0.039
A	5	214	85	294	42835	0.918
B	1	4	2	3	516	0.045
B	2	4	1	5	824	0.045
B	3	7	1	10	2528	0.08
B	4	11	7	10	3075	0.125
B	5	62	20	43	10466	0.705

TABLE 5 Basic data on *this* for two products, here labeled A and B. For each product, there are 5 rows, one for each rating category. For each row, Reviews is the number of reviews for Product that assign Rating. Columns 4 and 5 give the number of tokens for *this* in reviews for Product with Rating. Column 6 gives the total number of word tokens in reviews for Product assigning Rating. Column 7 contains the percentage of reviews in each rating category for each product, i.e., the row's column 3 value divided by all the column 3 values for the product in question.

3. The number of words in reviews for *A* relative to individual rating categories.

Table 5 illustrates with counts for pronominal *this* for two products.

In the exclamativity experiment of section 3.2, we took a corpus-level view, calculating overall distributional information for each of the rating categories. For this experiment, we take a product-level view instead, which means that, for each product, there are five rows, one for each rating category. Our primary interest is the relationship between the frequencies of determiner and pronominal *this* and the overall level of agreement among reviewers about the product in question.

We again calculate log-odds values for individual words and phrases relative to rating categories, as in definition 3. The only change is that we now relativize these values, not to the whole corpus, but rather to the review-sets for individual products. More specifically: let T_A be the set of reviews of product A, and let w_n be a word-string of length n. Then the log-odds value for w_n relative to T_A and rating category r is log-odds$_{T_A}(w_n, r)$ as in definition 3. For example, the log-odds of determiner *this* in 1-star reviews for the first product in table 5 is calculated using the values in row 1, column 4 (tokens of pronominal *this* in 1-star reviews for this product), and row 1, column 6 (total number of words in 1-star reviews for this product): log-odds$_{T_A}(this, 1\text{-star}) = \log(1/(571 - 1)) = -6.346$, which corresponds to a

frequency of $\text{logit}^{-1}(-6.346) = 0.002$.

To measure levels of agreement among reviewers for a given product A, we use the distribution of reviews for A across the rating categories, as in definition 4.

Definition 4 (Product-relative review distributions)**.** Let rcount(A, r) be the number of reviews of product A that assign rating r, and let rcount(A) be the set of all reviews for product A. The review distribution for A is the probability distribution $\Pr(A)$ such that

$$\Pr(A, r) = \frac{\text{rcount}(A, r)}{\text{rcount}(A)}$$

These distributions can be calculated using table 5. For example, product A has review counts $\langle 5, 3, 2, 9, 214 \rangle$. We turn this into a probability distribution by dividing each count by the sum of all these counts. The values are given in column 7 of table 5.

To see the relevance of these probabilities for solidarity effects, consider an individual reviewer who signs on intending to write a review that assigns rating r. The barplot near the top of the webpage gives her a sense for what the room is like. She might feel solidarity with the people who assigned r. Where those people are numerous relative to the full set of reviewers, solidarity should be enhanced. Conversely, if our reviewer is one of just a few advocates for a given rating category, solidarity should give way to opposition and its associated language. The solidarity hypothesis for *this* says that its usage should increase as solidarity increases, which straightforwardly predicts a positive correlation between *this* usage and agreement.

More specifically, our expectation, in light of the solidarity effects observed in section 2, is that $\text{log-odds}_{T_A}(\textit{this}, r)$ and $\Pr(A, r)$ will be correlated with for all products A and rating categories r. To test this, we again fit generalized linear models for both determiner and pronominal *this* with the $\Pr(\text{Product}, \text{Rating})$ values as the sole predictor. Figure 7 plots these relationships for both determiner and pronominal *this*, and table 6 summarizes the statistical model for the determiner version (the pronominal model is qualitatively the same). The effect is significant in both cases ($p < 0.001$), though the figure makes it clear that we explain only a small amount of the variance. We address possible causes of this in the discussion section just below.

Section 3.2 showed that *this* has a U-shaped profile relative to the rating categories: it is most frequent at the extreme ends of the scale. Rating also seems to affect the strength of the effect of solidarity. There is a significant positive interaction between Rating and $\Pr(\text{Product}, \text{Rating})$, which indicates that the effect is stronger in higher rating categories. The relationship may not just be linear, however: we created a variable grouping extreme ratings (1-star, 5-star) against the others. There is again a significant interaction between this
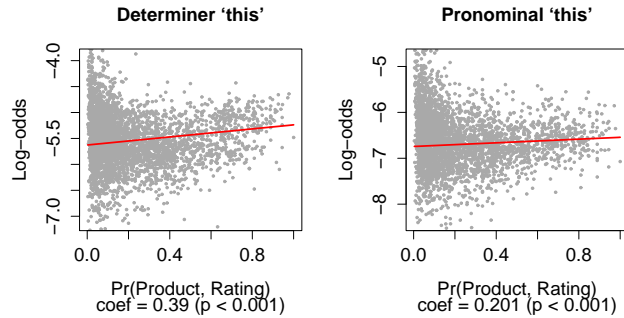
Determiner 'this'   **Pronominal 'this'**

Pr(Product, Rating)
coef = 0.39 (p < 0.001)

Pr(Product, Rating)
coef = 0.201 (p < 0.001)

FIGURE 7 Testing for solidarity effects across all ratings and all products. The *x*-axis tracks Pr(Product, Rating), and the *y*-axis measures log-odds values for *this*. These two values are positively correlated, as predicted by the solidarity hypothesis: the more agreement there is for a rating category, the more frequent *this* becomes.

	coefficient estimate	standard error	*p* value
Intercept	-5.627	0.006	< 0.001
Pr(Product, Rating)	0.39	0.011	< 0.001

TABLE 6 Summary of the fitted model for determiner *this* testing the solidarity hypothesis using the review distributions obtained from definition 4 to measure levels of solidarity.

variable and Pr(Product, Rating), which suggests that the effect of sentiment uniformity on *this*-frequency is stronger in the extreme rating categories and weaker in the mid-range ones.

Discussion

The above experiment provides evidence for the solidarity effect we argued for in section 2 based on a small number of carefully chosen examples. The relative weakness of the effect may be due to the presence of many non-emotive uses, but it could also, at least in part, be due to weaknesses in our approach rather than the pragmatic generalization itself. Our measures of solidarity (the distribution of reviews) no doubt obscure much of the complexity and richness of solidarity, and the "discourses" on which it is based — sets of reviews for a single product — are atypically fragmented, with little real interaction. (Reviewers sometimes comment on other reviews, creating something like dialogue, but these seem rarely to blossom into conversations.) Thus, our results might be taken simply as an indication that it would be promising to pursue the solidarity effect for *this* using other data and methods.

Our measure of solidarity is not the only plausible one, so one might worry that our result depends heavily on the specific choices we made. How-

ever, we tried a number of different measures with nearly the same results for all choices. One important alternative to consider is the Shannon entropy of the review distribution (Cover and Thomas, 1991). This too reveals the same basic pattern, though with weaker results. We conjecture that this is because each product has just one entropy value associated with it (as it is a global measure of the entire distribution), which obscures many potential anti-solidarity effects. Thus, for example, a distribution of reviews $\langle 5, 5, 5, 5, 20 \rangle$ shows a high amount of agreement, as indicated by its low entropy value, but the situation is very different for the reviewers giving stars one through four than it is for reviewers giving rating five, though the two groups are equally numerous. The Pr(Product, Rating) values appropriately distinguish reviewers in the one-to-four-star group (many of whom might stand in opposition to the five-star-givers) from the five-star group.[7]

We also experimented with combining any rating categories within one standard deviation of the mean rating, as a way of capturing the fact that, for example, a distribution like $\langle 1, 1, 1, 45, 50 \rangle$ shows high agreement at the positive end of the scale, with the 45 four-star reviews and the 50 five-star reviewers presumably feeling pretty much the same about the product. This too delivered the correlation.

One might expect the result to improve if we paid more attention to the temporal ordering of the reviews. If the entropy of the review distributions changed radically over time, then the solidarity effects would wax and wane as well, which could obscure the effect. We did experiments in which we calculated the Pr(Product, Rating) values for each review based on the distribution at the point it was written, but the result again didn't change. We suspect the intuition about temporal dependence is sound but ignorable here due to the fact that the review distributions for individual products tend to be very stable throughout our data. There is of course some flux at the start, but it quickly settles out.

One might also worry that the effect we observe is due to the nature of the star-rating categories and the associated extreme bias for five-star reviews. To address this concern, we ran a version of the solidarity experiment, as described above, on a corpus of comments from the website Opposing Views (`http://www.opposingviews.com`). Each comment is associated with a specific question (e.g., "Should we legalize marijuana?", "Does yoga harm your body", "When does life begin?") to which a variety of experts have provided commentary. The experts are divided up by the side they take on the issue. As with the review distributions at Amazon.com and Tripadvisor.com, the distribution of sides is visually apparent when a user logs

---

[7]We also attempted to create a combined measure, consisting of the product of Entropy and Pr(Product, Rating). This turned out to be almost perfectly correlated to Pr(Product, Rating) alone, and yielded all the same effects.

in to comment, so commenters get an intuitive visual sense for whether they align with a majority or minority view. The solidarity hypothesis predicts that this should influence demonstrative use. We fit the same kind of statistical model as above, with the ratings now replaced by sides, which are generally non-scalar and vary from question to question. We again found a significant positive correlation between *this* usage (combining determiner and pronominal uses, since we did not chunk the data) and our measures of solidarity. The sides typically are not ordered, nor is there a bias in the corpus for any particular side. These results therefore help to bolster the view that the effects we observe trace to genuine usage conditions. We hope to more fully discuss this and other experiments with the Opposing Views data in future work.

4 Extensions

We have so far restricted our quantitative work to English *this*. We feel that we have the best grip on its affectivity of all the demonstratives, it is both numerous in our data (unlike *these*, *those*), and it is easy to identify (unlike *that*; table 3). Nonetheless, it is worth briefly looking at the result of applying the above methods more broadly. Section 4.1 looks at English *that*, and section 4.2 extends our methods outside of English (see also Davis and Potts To appear).

4.1 Distal demonstratives

Lakoff (1974), Bowdle and Ward (1995), and Wolter (2006) analyze affective proximal and distal demonstratives, identifying common elements of affectivity across both. Lakoff remarks that, surprisingly, there are solidarity effects with *that*: "These are perhaps the most curious semantically, since the distance marker *that* seems to establish emotional closeness between speaker and addressee" (p. 351). Commenting on Liberman (2008), Barbara Partee points out that this closeness effect is often exploited by advertisers, who foster a "fake familiarity" with phrases like *that certain someone* and (in the context of public radio fundraising) *Dont put it off any longer — make that phone call right now*.

These characterizations might lead one to expect *that* to have roughly the same distributional profile as *this*. This is not what we find, however. Figure 8 applies the methodology of section 3.2 to *that*, giving its log-odds distribution relative to rating categories. Both models are good ($p < 0.001$), and the quadratic coefficients are negative, making these items comparable to *the* in their 'anti-exclamativity', though with a notable bias for the negative end of the rating scale.

In retrospect, this might not be so surprising. Though it is clear that *that* can work as a marker of solidarity and exclamativity, these uses might be in the minority. Citing the Oxford English Dictionary, Liberman (2008) offers
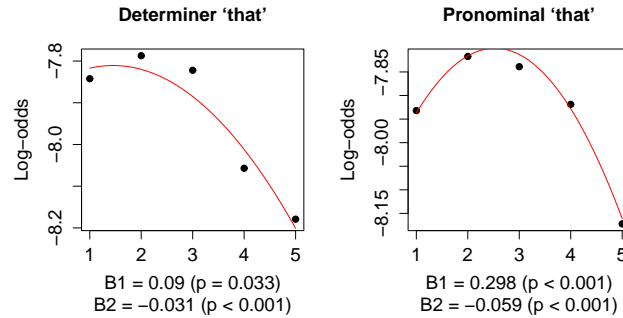
FIGURE 8 Demonstrative *that* more closely resembles the anti-exclamatives of figure 4 than it does *this* (cf. figure 5).

a mixed assessment of then-Vice Presidential candidate Sarah Palin's usage: "Often, there seems to me to be a tint of 'censure, dislike, or scorn' (or 'commendation or admiration') in Gov. Palin's use of demonstrative determiners". Some of its constructional uses strike us as almost uniformly negative (sarcastic, ironical) in current usage despite positive origins. For example, Lakoff cites examples like (15a) as sympathetic, but a Google search for the quoted string how's that in June, 2009, turns up lots of negative things like (15b-15d), which all relate to President Obama's economic stimulus program. Similarly, though the "fake familiarity" of the advertisements Partee cites do seem to be intended as friendly, there is arguably an element of scolding in them as well.

(15) a. How's that throat?
      b. How's that porkulus working out?
      c. How's that "hope" and "change" working out for you?
      d. How's that $787 billion stimulus plan working out?

The methods we used in section 3.3 give results that contrast with the ones we saw for *this* but that are in keeping with figure 8. Here again, we used $\Pr(\text{Product}, \text{Rating})$ as the sole predictor. Figure 9 plots the results of this modeling. The coefficient for $\Pr(\text{Product}, \text{Rating})$ is -0.475 with a standard error of 0.039 ($p < 0.001$). What's noteworthy here is really that the effect goes in the other direction from *this*. Notably, there again is evidence suggesting an interaction with Rating, with the effect becoming milder in higher rating categories.

4.2 German demonstratives

We have suggested that affective uses of *this* arise from its more basic meaning as a marker of spatio-temporal proximity. We do not offer a theoretical
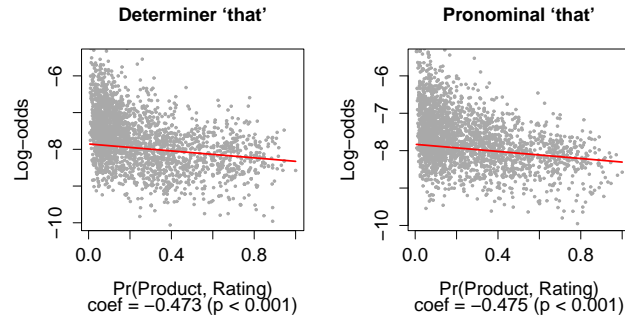
FIGURE 9 Solidarity effects with *that*. Here, there is an inverse correlation between *that* usage and Pr(Product, Rating) values, a marked contrast with *this* suggesting that the two give rise to different kinds of affectivity.

account of how the underlying semantics can be this flexible, but our experimental evidence, which looks at all uses of *this*, supports this kind of account.
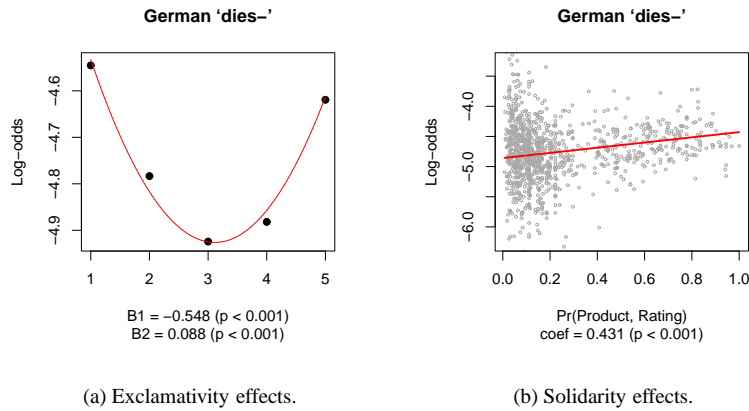
If this is correct, then we expect affectivity for proximal demonstratives quite generally across languages. Wolter (2006) explicitly takes a different view, saying "There is no guarantee that every language will have emotive demonstratives" (p. 85). We think it is fair to say that we are advocating the opposite view. To begin testing this, we now look briefly at the German proximal demonstrative *dies-* ('this') in the UMass Amherst Sentiment Corpora. We have not tagged or chunked that data, so we look at the occurrences of this word, without distinguishing its pronominal and determiner uses as we have for English.

Figure 10(a) gives the analysis of *dies-* relative to the rating categories, with results that are very much like those we have for *this*. Davis and Potts (To appear) find the same basic pattern for the Japanese proximal determiners *kono* (determiner 'this'), *kore* (prominal 'this'), and *konna* ('this kind of'), though with noteworthy contrasts within the paradigm.

Figure 10(b) identifies a positive correlation between the use of *dies-* and the percentage of agreeing reviews. The picture is the same as we saw for English (figure 7). As was the case for *this*, the effect appears to be stronger in higher rating categories, as there is a positive interaction between Rating and Pr(Product, Rating). This kind of cross-linguistic correspondence is expected if affective uses derive from a more basic proximal meaning.

5   Conclusion

With the experiments of section 3, we sought to use large corpora to test linguists' intuition-based hypotheses about affective uses of *this*. Our evidence

German 'dies–'

Log-odds

B1 = −0.548 (p < 0.001)
B2 = 0.088 (p < 0.001)

(a) Exclamativity effects.

German 'dies–'

Log-odds

Pr(Product, Rating)
coef = 0.431 (p < 0.001)

(b) Solidarity effects.

FIGURE 10 German *dies-*.

helps confirm the generalizations of Lakoff (1974), Bowdle and Ward (1995), and Wolter (2006), and others that affective *this* both evokes speaker–hearer solidarity and correlates strongly with evaluativity of the sort we find with uncontroversially exclamative words and constructions. We are able to build robust models for the exclamativity effect. The solidarity effects that we found are more subtle. This might, in part, be due to the likely presence of many non-emotive uses, but also might trace to the necessarily approximative nature of our methods rather than the true linguistic strength of the effect. It is furthermore clear that many more factors affect determiner choice, so that the emotive dimension accounts for only a small part for the overall variation in this regard.

We hope that these experiments help to build the case that corpus methods and corpus tools are valuable to pragmatic theory. The quantitative perspective that they provide allows us to identify important factors and correlations even where there are a great many other influences getting in the way of such generalizations. Corpus methods also provide the opportunity to look at thousands of different contexts, identifying the commonalities in them and using that information to gain a deeper understanding of the rich and complex messages that speakers send and hearers receive when using language.

Appendix

Data files

This appendix briefly describes the data files that were used for the experiments in this paper and that are distributed with the paper. The file `generate-all-plots.R` generates all the plots for the paper, placing them in the `plots/` subdirectory.

1 `unigrams.english.csv`

This is the basis for the section 3.2 experiment. It includes distributional information for all the unigrams in our corpus with at least 15 tokens. Here is the basic format, illustrating with *absolutely*:

Word	Rating	TokenCount	RatingWideCount
absolutely	1	378	1508620
absolutely	2	220	1487793
absolutely	3	202	1861874
absolutely	4	739	4515665
absolutely	5	2585	8928094

For each row, TokenCount is the number of tokens of Word in reviews assigning Rating, and RatingWideCount is the number of tokens of all words in reviews assigning Rating. The associated R script (R Development Core Team, 2008) is `exclamativity.R`, which builds generalized linear models and associated plots using this format.

2 `dems.csv`

This file contains the demonstrative counts for English and German. The format is the same as that of `unigrams.english.csv`

3 `byproduct.english.dems.csv`

This is the basis for the section 3.3 experiments. The basic format is described in table 5, except that the full version includes columns for *the* as well as *that*,

these, and *those*, on their determiner and pronominal uses. The associated R script is `solidarity.R`, which builds generalized linear models and associated plots using this format.

4 `byproduct.dies.csv`

Same format as `byproduct.english.dems.csv` but for German *dies*-using the German portion of the UMass Amherst Sentiment Corpora (Constant et al., 2009b). The script `solidarity-plots.R` works for this file as well.

5 Existentials files

The files `existential-sentences-(a|this).txt` contain existential sentences extracted from our corpora, in the manner described at the end of section 2. The file `sentiment-adjectives.txt` contains the list of subjective adjectives (derived from `http://www.keepandshare.com/doc/view.php?u=12894`) that we used to supplement the lists distributed by Hatzivassiloglou and Wiebe (2000b). The script `generate-all-plots.R` contains the particular values that we used to create the plot in the paper.

Acknowledgments

References

Baayen, R. Harald. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer Academic Publishers.

Bowdle, Brian F. and Gregory Ward. 1995. Generic demonstratives. In J. Ahlers, L. Bilmes, J. S. Guenter, B. A. Kaiser, and J. Namkung, eds., *Proceedings of the Twenty-First Annual Meeting of the Berkeley Linguistics Society*, pages 32–43. Berkeley Linguistics Society.

Campbell, Sherlock R. and James W. Pennebaker. 2003. The secret life of pronouns: Flexibility in writing style and physical health. *Psychological Science* 14(1):60–65.

Chevalier, Judith A. and Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3):345–354.

Chung, Cindy and James W. Pennebaker. 2007. The psychological function of function words. In K. Fiedler, ed., *Social Communication*, pages 343–359. New York: Psychology Press.

Constant, Noah, Christopher Davis, Christopher Potts, and Florian Schwarz. 2009a. The pragmatics of expressive content: Evidence from large corpora. *Sprache und Datenverarbeitung* 33(1–2):5–21.

Constant, Noah, Christopher Davis, Christopher Potts, and Florian Schwarz. 2009b. UMass Amherst sentiment corpora.

Cover, Thomas M. and Joy A. Thomas. 1991. *Elements of Information Theory*. New York: Wiley.

Davis, Christopher and Christopher Potts. To appear. Affective demonstratives and the division of pragmatic labor. In *Proceedings of the 17th Amsterdam Colloquium*. Amsterdam: Kurosio.

Greenwood, Mark. 2005. NP chunker v1.1.

Hatzivassiloglou, Vasileios and Janyce Wiebe. 2000a. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the 18th International Conference on Computational Linguistics*, pages 299–305. San Fransisco, CA: Morgan Kaufmann.

Hatzivassiloglou, Vasileios and Janyce Wiebe. 2000b. Lists of gradable, polar, and dynamic adjectives.

Hu, Nan, Paul A. Pavlou, and Jennifer Zhang. 2006. Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In *Proceedings of Electronic Commerce (EC)*, pages 324–330. New York, NY, USA: ACM.

Jurafsky, Daniel and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 2nd edn.

Lakoff, Robin. 1974. Remarks on 'this' and 'that'. In *Proceedings of the Chicago Linguistics Society 10*, pages 345–356.

Lasersohn, Peter. 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and Philosophy* 28(6):643–686.

Liberman, Mark. 2008. Affective demonstratives. Language Log, http://languagelog.ldc.upenn.edu/nll/?p=674.

Manning, Christopher D., Dan Klein, Huihsin Morgan, William Tseng, and Anna N. Rafferty. 2008. Stanford log-linear part-of-speech tagger, version 1.6.

Pang, Bo and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1):1–135.

Partee, Barbara H. 1999. Weak NP's in HAVE sentences. In J. Gerbrandy, M. de Rijke, and Y. Venema, eds., *JFAK: Essays dedicated to Johan van Bentham on the Occasion of his 50th Birthday*. Vossiuspers: Amsterdam University Press.

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

Prince, Ellen. 1981a. On the inferencing of indefinite 'this' NPs. In B. L. Webber, I. Sag, and A. Joshi, eds., *Elements of Discourse Understanding*, pages 231–250. Cambridge: Cambridge University Press.

Prince, Ellen F. 1981b. Toward a taxonomy of given–new information. In P. Cole, ed., *Radical Pragmatics*, pages 223–255. New York: Academic Press.

R Development Core Team. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramshaw, Lance and Mitchell P. Marcus. 1995. Text chunking using transformation-based learning. In D. Yarowsky and K. Church, eds., *Proceedings of the Third ACL Workshop on Very Large Corpora*, pages 82–94. The Association for Computational Linguistics.

Vonk, Wietske, Lettica G. M. M. Hustinx, and Wim H. G. Simons. 1992. The use of referential expressions in structuring discourse. *Language and Cognitive Processes* 7(3/4):301–333.

Wolter, Lynsey Kay. 2006. *That's That: The Semantics and Pragmatics of Demonstrative Noun Phrases*. Ph.D. thesis, UC Santa Cruz.