DISCUSSION NOTE

A note on quantifying "good" and "bad" prosodies*

PHILIP DILTS and JOHN NEWMAN

Semantic prosody, or *discourse prosody* as it is also known, has come to be a familiar dimension of corpus-based lexicology (see Louw 1993, Stubbs 1995, Sinclair 1991, Partington 2004 for examples), though it is not without its critics (cf. Whitsitt 2005). As Whitsitt demonstrates through his survey of the relevant literature, there are different emphases in the ways in which the term *semantic prosody* has come to be understood. The most common understanding that we seem to encounter, however, is that some words, or word groups, occur in contexts which are understood by the researcher to have "positive" or "negative" nuances, or *prosodies* (with KWIC displays of concordance lines facilitating the discovery of these prosodies). The prosodies are not simply to be equated with the nuances found at any one collocational position or with any one part of speech, but rather they emerge from miscellaneous lexical and phraseological phenomena in the context of usage of the word in question. In some interesting cases (e. g., the negative prosody associated with CAUSE, as discussed in Stubbs 1995), the prosody is not particularly obvious, or even evident at all, to the researcher or native speaker prior to the corpus-based analysis.

The methodology underlying such studies is not quite as sound as one might like, however. In particular, the researcher is required to make evaluative judgments in the absence of a set of principled criteria to guide the evaluation. Terms such as "good" or "bad", "positive" or "negative" etc. are introduced at will and without much care taken to explain the basis for the judgment. This aspect of semantic prosody would appear to warrant more attention than it has so far received. In fact, the issue is rarely even addressed in the typical prosody study. The following passage in Partington (1996: 73), occurring as part of a discussion of the prosody of *dealings*, is typical of the analytical discourse associated with this line of research: "Its concordance [i. e., the concordance of the word *dealings*] contained a striking paucity of modifiers

Corpus Linguistics and Linguistic Theory 2–2 (2006), 233–242 1613-7027/06/0002–0233
DOI 10.1515/CLLT.2006.011

expressing value judgments … However, if the wider co-text is taken into consideration, it becomes clearer that *dealings* generally indicates some unattractive or dishonest activity". How are we to understand "attractiveness" or "unattractiveness" of an activity and how would another researcher go about investigating these properties in the co-text of other words? In fairness to Partington (and other researchers working in this vein), it should be said that researchers generally make reasonable and convincing observations, and semantic prosody studies of this type have undeniably led to interesting results. The lack of agreed-upon criteria for making the positive versus negative evaluations remains a methodological problem, however, even if other aspects of the methodology can be made explicit, e. g., the number of concordance lines to be inspected, the sampling methods for obtaining concordance lines, the size of the window of co-text.[1]

It may be the case that the study of prosody is always likely to involve a certain degree of subjectivity on the part of the researcher. Nevertheless, if there are ways to explore prosody with less dependence on a researcher's subjective, evaluative judgments, then these deserve some discussion.[2] In this spirit, we propose and illustrate a method designed to eliminate the need for the researcher to be making their own evaluative judgments in assessing the positive or negative prosodies of words. Instead, we start with a set of experimentally obtained judgments relating to "goodness" or "badness" of concepts, following Osgood, Suci and Tannenbaum (1957: 47−64). Briefly, Osgood et al. identified 76 antonymous pairs of adjectives deemed by subjects and researchers to be representative of the semantic space of adjectives, such as *good-bad*, *wise-foolish*, *kind-cruel*, *clean-dirty*, *sober-drunk*, etc. Subjects were then asked to rate 20 concepts on scales defined by these 76 adjective pairs. The concepts used in this task were selected from 5 domains: person concepts (*Adlai Stevenson*), physical objects (*snow*), abstract concepts (*sin*), event concepts (*dawn*), and institutions (*America*). Through a factorial analysis ("centroid factor analysis"), the researchers found that subjects' judgments could be accounted for in terms of eight independent factors, of which an "evaluative" factor, exemplified by the *good-bad* pair, was the strongest predictor of responses. The evaluative factor of each of the 76 adjective pairs was measured as a correlation index, ranging from 0.1 (for *fast-slow*) to 1.0 (for *good-bad* itself). The 20 pairs with the strongest correlations are shown in Table 1.[3]

The numerical correlations with *good* and *bad* arrived at by Osgood et al. suggest, we contend, a natural, quantitative basis for studying the "goodness" or "badness" of co-occurring nouns and the concepts they represent. One could imagine a variety of ways to incorporate the correlation values from Osgood et al. into a formal measure of "good" or

Table 1. *Top 20 adjective pairs which have been assigned a value by Osgood et al. (1957:53–61)*

Positive	Negative	Correlation with *good-bad*	Positive	Negative	Correlation with *good-bad*
good	*bad*	1.00	*positive*	*negative*	0.48
reputable	*disreputable*	0.68	*high*	*low*	0.45
wise	*foolish*	0.57	*clean*	*dirty*	0.45
beautiful	*ugly*	0.52	*progressive*	*regressive*	0.43
kind	*cruel*	0.52	*sociable*	*unsociable*	0.42
successful	*unsuccessful*	0.51	*meaningful*	*meaningless*	0.41
true	*false*	0.50	*sober*	*drunk*	0.40
harmonious	*dissonant*	0.49	*interesting*	*boring*	0.40
grateful	*ungrateful*	0.49	*important*	*unimportant*	0.38
sane	*insane*	0.48	*believing*	*skeptical*	0.38

"bad" prosodies of co-occurring nouns. We apply one such measure to nouns in the British National Corpus (BNC), without any claim that it is necessarily superior to others that could be proposed.[4] The basic idea underlying our measure is that the "good" or "bad" prosody of a noun can be understood as proportional to the number of "good" or "bad" adjectives (drawn from the full set of 152 adjectives in Osgood et al's study) that modify the noun in a corpus.[5] Our formalization of this idea can be expressed succinctly with the formula in (1).

(1) $$\text{Prosody}(n) = \sum_i \left[weight\,(adj_i) \times \left(\frac{p\,(adj_i, n)}{p\,(adj_i, n) + p\,(\neg adj_i, n)} \right) \right]$$ where:

- n denotes the noun under study
- $p\,(adj, n) = \dfrac{F\,(adj, n)}{F\,(adj)}$ where F (x) is the absolute frequency of x in the corpus
- adj_i represents whichever of the i-th pair of adjectives, the "good" or the "bad" adjective, has the higher proportional frequency in the adj + n construction (i. e., leads to the highest value of $p\,(adj_i, n)$)
- $\neg\, adj_i$ represents the antonym of adj_i in Osgood et al.'s pairing (and thus the adjective in the i-th pair with lower proportional frequency)
- $weight\,(adj_i)$ is the correlation strength of adj_i as given in Osgood et al. multiplied by -1 if and only if adj_i is a "bad" adjective.

To illustrate how (1) is applied, consider the word *analogy.* From the 152 adjectives measured by Osgood et al., *analogy* is modified by *true, false, good, bad,* and *cruel* in the BNC. In terms of proportional fre-

quency, the adjectives *bad* (p = 2/14389 = 0.00014), *false* (p = 2/3067 = 0.00067), and *cruel* (p = 1/1321 = 0.000757) outweigh their counterparts *good* (p = 8/75493 = 0.00011), *true* (p = 1/17301 = 0.00006), and *kind* (p = 0) in this construction, so adj_i, adj_2, and adj_3 are *bad*, *false*, and *cruel*, while $\neg adj_1$, $\neg adj_2$, and $\neg adj_3$ are *good*, *true*, and *kind*.[6] The more frequent adjectives are all on the "bad" side of Osgood et al.'s scales, so the value of weight (*adj*) for each adj_i will be a negative number. The prosody of *analogy* according to (1), then, is calculated as follows:

Prosody(analogy)
= Weight(false) * [(false analogy)/(false)] / [(true analogy)/(true) + (false analogy)/(false)]
+ Weight(bad) * [(bad analogy)/(bad)] / [(good analogy)/(good) + (bad analogy)/(bad)]
+ Weight(cruel) * [(cruel analogy)/(cruel)] / [(kind analogy)/(kind) + (cruel analogy)/(cruel)]
= (−0.5) * [2 / 3067] / [(1 / 17301) + (2 / 3067)]
+ (−1) * [2 / 14389] / [(8 / 75493) + (2 / 14389)]
+ (−0.52) * [1 / 1321] / [(0 / 1041) + (1 / 1321)]
= (−0.5) * (0.92)
+ (−1) * (0.67)
+ (−0.52) * (1)
= −1.55

To apply our method, a Perl program was written to extract nouns from the BNC collocating in the R1 position of each of the target adjectives and to calculate the "goodness" or "badness" of these nouns, based on our measure.[7] Table 2 lists the 25 nouns with the highest "good" and "bad" prosodies among the 27,528 adjective + noun pairs discovered. It is interesting to consider the nouns attracting the highest "good" and "bad" prosody scores in Table 2. Several of the adjectives with a positive estimated prosody in Table 2 do seem intuitively positive. *Support, home, model, performance*, *integration, family, design, development,* and *contact* all seem like plausible representatives of a set of positive nouns. Other nouns in Table 2, such as *government, study, information, alternative, groups, changes,* and *work*, might be considered arguably good, though a positive evaluation is far from compelling. At the very least, none of the nouns with positive prosody estimates denote clearly negative concepts.

Nouns in the negative prosody column conform much more poorly to intuition than those in the positive column, however. Some prominent, intuitively negative nouns are present (*beast, mistakes, cough, bruise, shock, killer,* and *drag*) but they do not form the majority of the results.

Table 2. *Top 25 positive-prosody and top 25 negative-prosody words in the BNC*

Rank	Positive prosody			Negative Prosody		
	Noun	Prosody (Max 19.97)	Num. Collocates	Noun	Prosody (Min −19.97)	Num. Collocates
1	*support*	7.64	31	*joke*	−3.26	16
2	*government*	6.78	33	*grin*	−3.19	13
3	*home*	6.31	30	*jokes*	−3.00	12
4	*atmosphere*	5.99	31	*beast*	−2.83	10
5	*model*	5.91	26	*thoughts*	−2.72	23
6	*programme*	5.90	26	*mistakes*	−2.72	9
7	*study*	5.85	26	*laughter*	−2.69	14
8	*information*	5.81	37	*head*	−2.68	32
9	*function*	5.80	21	*shoulder*	−2.44	13
10	*performance*	5.66	31	*dreams*	−2.43	16
11	*alternative*	5.53	24	*cough*	−2.40	9
12	*setting*	5.48	24	*mouth*	−2.40	20
13	*groups*	5.44	25	*eyes*	−2.39	31
14	*integration*	5.36	15	*chest*	−2.31	13
15	*range*	5.32	25	*bruise*	−2.26	6
16	*family*	5.32	41	*shock*	−2.24	14
17	*changes*	5.31	32	*killer*	−2.22	6
18	*manner*	5.21	49	*hole*	−2.16	11
19	*use*	5.11	32	*lump*	−2.15	13
20	*design*	5.11	32	*calf*	−2.15	8
21	*development*	5.08	35	*bounce*	−2.12	6
22	*contact*	5.05	24	*drag*	−2.12	7
23	*arrangements*	5.03	17	*legacy*	−2.08	13
24	*assessment*	5.02	31	*leg*	−2.08	19
25	*work*	5.02	40	*gods*	−2.08	7

There is, for one thing, a preponderance of nouns denoting body parts, including *head, shoulder, mouth, eyes, chest,* and *leg* (and possibly *calf*). Still more curious, however, are nouns that seem to denote happy or positive concepts: *thoughts, dreams, legacy,* and *gods* describe intuitively positive abstract concepts, and *joke, grin, jokes,* and *laughter* all represent humorous activity.

One might ask whether the method adopted here leads to results which align in any way with results from other approaches to prosody. The most obvious kind of approach to compare ours with would be to consider adjective collocates of the nouns in Table 2, based on the BNC, without restricting the adjectives to those utilized in the study by Osgood et al. While a study of all adjective collocates of these nouns is still more restrictive than a typical prosody study, it is comparable to what we have carried out. We therefore undertook a collocate analysis of selected nouns in Table 2 to compare results. We chose *laughter*, *grin*, and *joke*

as potentially interesting nouns to study. At first glance, it might seem odd that they should emerge as having relatively strong negative prosody using our method. We were curious, therefore, to see whether a collocate analysis looking at all the adjectives modifying each of these words produces a similar negative skewing or not. We carried out a "collexeme analysis", a method of measuring association strengths between words and the constructions they occur in (Stefanowitsch and Gries 2003). More precisely, we considered the adjectives most attracted to the specific constructions Adj + *laughter*, Adj + *grin*, and Adj + *joke*.[8] The association or "collostructional" strength (measured according to the Fisher-p test) of all adjectives occurring more than three times in this construction was calculated using Coll.Analysis 3 (Gries 2004).[9] Following this method, we use the total number of Adj + *laughter*, Adj + *grin*, and Adj + *joke* combinations in the BNC as the relevant corpus size, rather than the total number of words. The words most strongly attracted to each construction, the "collexemes", are shown in Tables 3–5. The results can be summarized as follows:

laughter
While the strongest collexeme (*hysterical*) is not obviously negative, and the weakest collexeme (*delighted*) seems fairly positive, the majority of the adjectives in the set of strong collexemes (*raucous, mocking, suppressed, helpless, incredulous,* and *loud*) would most likely be judged to be negative.

grin
Several negative-sounding adjectives are among the collocates of *grin*: *mischievous, rueful, wolfish, sardonic, mirthless, manic, evil, crooked,* and *wicked* are the most clearly negative adjectives. Of the remaining adjectives, some (e. g., *wry, sly, sheepish*) may be considered positive

Table 3. *Collostruction strength of adjectives in [Adj* + laughter*] construction. Adjectives with coll.strength < 3.0 were removed*

Words	Word.freq	Obs.freq	Exp.freq	Relation	Coll.strength
hysterical	401	20	0.27	attraction	30.2
raucous	140	12	0.10	attraction	21.2
mocking	325	14	0.22	attraction	20.4
suppressed	159	9	0.11	attraction	14.4
helpless	791	9	0.54	attraction	8.2
incredulous	171	4	0.12	attraction	5.2
loud	1521	8	1.04	attraction	4.9
delighted	2545	8	1.73	attraction	3.4

Table 4. *Collostruction strengths of the [Adj + grin] construction. Adjectives with coll.strength < 3.0 were removed*

Words	Word.freq	Obs.freq	Exp.freq	Relation	Coll.strength
sheepish	82	16	0.09	attraction	31.1
wry	420	21	0.44	attraction	27.7
toothy	31	12	0.03	attraction	27.6
mischievous	260	15	0.27	attraction	20.9
lop-sided	40	10	0.04	attraction	20.9
lopsided	47	9	0.05	attraction	17.7
toothless	83	9	0.09	attraction	15.3
rueful	138	9	0.15	attraction	13.3
gap-toothed	24	6	0.03	attraction	12.8
cheeky	326	10	0.34	attraction	11.4
wolfish	32	5	0.03	attraction	9.6
sly	310	8	0.33	attraction	8.7
sardonic	234	7	0.25	attraction	8.1
mirthless	39	4	0.04	attraction	7.0
broad	4777	19	5.02	attraction	5.9
wide	11018	29	11.59	attraction	5.0
boyish	159	4	0.17	attraction	4.6
manic	216	4	0.23	attraction	4.0
evil	1473	8	1.55	attraction	3.7
crooked	314	4	0.33	attraction	3.4
wicked	1036	6	1.09	attraction	3.0

Table 5. *Collostruction strengths of the [Adj + joke] construction. Adjectives with coll.strength < 3.0 were removed*

Words	Word.freq	Obs.freq	Exp.freq	Relation	Coll.strength
cruel	1321	14	1.16	attraction	10.6
practical	7612	27	6.67	attraction	8.8
sick	4209	19	3.69	attraction	7.9
macabre	144	5	0.13	attraction	6.7
good	77335	28	67.79	repulsion	8.3
great	43121	11	37.80	repulsion	6.9
old	52275	20	45.82	repulsion	5.2

or negative, depending on whether one takes the point of view of the grinner or the one grinned at. Many of the remaining adjectives (*toothy, toothless, gap-toothed,* and possibly *cheeky*) might inherit their high collostructional strengths from their literal interpretations: these adjectives describe mouths, so they might be expected to have a greater-than-average chance of occurring with nouns related to the mouth.

joke
Here, three of the four adjectives attracted to the construction with collostructional strength greater than 3 are negative: *cruel, sick,* and *macabre*. Moreover, of the three adjectives repulsed from the construction with collostruction strength greater than three, the two strongest might be called very positive: *good* and *great*.

Interestingly, these three humor-related nouns arguably have negative semantic prosodies using this kind of collocate analysis. This result accords with the results calculated by the method developed in this paper. While our comparison of results for these three nouns does not allow us to draw general conclusions about all the nouns in Table 2, the comparison encourages us to think that the results shown in Table 2 may match quite nicely what might be found using other prosody techniques.

In this paper, we offer a method of quantifying "good" or "bad" prosodies without relying on the subjective judgments of the analyst. Instead, we base our method on experimentally measured judgments of *good*ness and *bad*ness obtained prior to, and independently of, corpus-based studies of prosody. An additional benefit of our method is the automatic identification of interesting keywords. In typical prosody studies, there is no algorithmic procedure for identifying the key words or phrases which occur with positive or negative prosodies. Instead, it is left to the researcher to intuit or otherwise arrive at the words or phrases of interest. By contrast, our method has the advantage of identifying, from the bottom up, nouns distinguished by high "good" and "bad" prosodies. In this way, we were able to establish the negative prosodies associated with words such as *laughter*, *grin*, and *joke*, words which do not, out of context, have any obvious negative connotation and thus would not be natural starting points for prosody research.

We should emphasize that our method is not a *general* procedure allowing the calculation of the prosodies of all words in a corpus. It is limited in three significant ways: (i) the method can only be used to measure prosodies of nouns; (ii) the method considers only adjectives immediately preceding nouns; (iii) only the 152 adjectives with correlation indices are considered. Our intention here has been to show how a procedure based on the findings of Osgood et al. could be constructed, rather than develop a general methodology for the study of prosody. The scope of the present study is also limited, neglecting possible register differences and considering only a single corpus of a single variety of English. Clearly, these considerations merit further study. While acknowledging the limited scope of what we set out to do here, we feel that our methodology is of interest insofar as it proposes an empirically grounded, replicable procedure where none has yet been offered.

Notes

* We would like to thank two anonymous reviewers for their valuable comments and insights on an earlier draft of this paper.

1. Even these aspects, which could be standardized for the study of prosody, have not been, adding to the variability in the procedures followed and the difficulty of comparing results from different studies.
2. Substantial research has been carried out into computational methods relating to evaluative language (e. g., Wiebe et al. 2001). An automated measure of semantic prosody such as the one developed here can be considered a contribution to this endeavor.
3. More recent studies (Morgan and Heise 1988; Wurm and Vakoch 1996) have confirmed the robustness of Osgood et al.'s results and shown that the scales developed play a role in online processing tasks.
4. Indeed, we would like to thank an anonymous reviewer for recommending an alternate measure: the average of the weight co-efficients after Fisher-z transformation. We are currently exploring this measure.
5. Since we narrow the scope of our co-occurrences to adjectives immediately preceding a noun, it could be considered a measure of semantic *preference* (in the sense of Stubbs 2001: 65) for negative or positive adjectives rather than semantic prosody. This distinction between semantic prosody and semantic preference is not clearly drawn by all corpus linguists.
6. Unless otherwise noted, adjectives were counted only when used as adjectives, i. e., when tagged AJ0 (positive), AJC(comparative), or AJS (superlative). Ambiguous uses (e. g., *believing* [AJ0-VVG]) were excluded.
7. As with the adjectives, only nouns tagged N** (neutral/singular/plural common noun or proper noun) were included in our calculations. Ambiguous uses (e. g., *support* [NN0-VVB] were excluded.
8. Results for this section were found using http://view.byu.edu. The tag [AJ*] was used to select adjectives. Note that this will catch some ambiguously tagged adjectives (e. g., [AJ0-VVD]) ignored by the script used to obtain Table 2.
9. Adjectives appearing in the target construction fewer than three times were excluded for a practical reason. It reduced the number of adjectives whose token frequency was required for collexeme analysis from 659 to 66. Unlike prosody calculation, pre-processing for collexeme analysis was conducted manually using VIEW, making such a simplification highly desirable.

References

Gries, Stefan Th.
2004 Coll.analysis 3. A program for R for Windows 2.x.

Louw, Bill
1993 Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In Baker, Mona, Gill Francis and Elena Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair.* Amsterdam/Philadelphia: John Benjamins, 157–176.

Morgan, Rick L. and David Heise
1988 Structure of emotions. *Social Psychology Quarterly* 51 (1), 19–31.

Osgood, Charles E., George J. Suci and Percy H. Tannenbaum
1957 *The Measurement of Meaning.* Urbana: University of Illinois Press.

Partington, Alan
1996 *Patterns and Meanings*. Amsterdam: John Benjamins.
2004 Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics* 9 (1), 131−156.
Sinclair, John
1991 *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
Stefanowitsch, Anatol and Stefan Th. Gries
2003 Collostructions: Investigations in the interaction of words and constructions. *International Journal of Corpus Linguistics* 8 (2), 209−243.
Stubbs, Michael
1995 Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language* 2 (1), 23−55.
2001 *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
Whitsitt, Sam
2005 A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics* 10 (3), 283−305.
Wiebe, Janyce, Rebecca Bruce, Matthew Bell, Melanie Martin and Theresa Wilson
2001 A corpus study of evaluative and speculative language. *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue*, 1−10.
Wurm, Lee H. and Douglas A. Vakoch
1996 Dimensions of speech perception: Semantic associations in the affective lexicon. *Cognition and Emotion* 10 (4), 409−423.