

Automatically identifying changes in the semantic orientation of words

Paul Cook, Suzanne Stevenson

Department of Computer Science
University of Toronto
Toronto, Canada
pcook@cs.toronto.edu, suzanne@cs.toronto.edu

Abstract

The meanings of words are not fixed but in fact undergo change, with new word senses arising and established senses taking on new aspects of meaning or falling out of usage. Two types of semantic change are amelioration and pejoration; in these processes a word sense changes to become more positive or negative, respectively. In this first computational study of amelioration and pejoration we adapt a web-based method for determining semantic orientation to the task of identifying ameliorations and pejorations in corpora from differing time periods. We evaluate our proposed method on a small dataset of known historical ameliorations and pejorations, and find it to perform better than a random baseline. Since this test dataset is small, we conduct a further evaluation on artificial examples of amelioration and pejoration, and again find evidence that our proposed method is able to identify changes in semantic orientation. Finally, we conduct a preliminary evaluation in which we apply our methods to the task of finding words which have recently undergone amelioration or pejoration.

1. Detecting changes in semantic orientation

Word senses are continually evolving, with both new words and new senses of words arising almost daily. Systems for natural language processing tasks, such as question answering and automatic machine translation, often depend on lexicons for a variety of information, such as a word's parts-of-speech or meaning representation. When a sense of a word that is not recorded in a system's lexicon is encountered in a text being processed, the system will typically fail to recognize the novel word sense as such, and then incorrectly draw on information from the lexical entry corresponding to some other sense of that word. The performance of the entire system will then likely suffer due to this incorrect lexical information. Ideally, a system could automatically identify novel word senses, and subsequently infer the necessary lexical information for the computational task at hand (e.g., the correct meaning representation for a novel word sense). Indeed, novel word senses present one of the most challenging phenomena in lexical acquisition (Zernik, 1991).

New word senses also present challenges in lexicography where determining how established words and senses have changed is recognized as an important, and very difficult, problem (Simpson, 2007). Dictionaries covering current language must be updated to reflect new senses of words (and indeed new words themselves) that have come into usage, and also changes in the usage of established words and senses. Nowadays, vast quantities of text are produced each day in a variety of media including traditional publications such as newspapers and magazines, as well as newer types of communication such as blogs and micro-blogs (e.g., Twitter, <http://twitter.com>). Lexicographers must search this text for new word senses; however, given the amount of text that must be analyzed, it is simply not feasible to manually process it all (Barnhart, 1985). Therefore, automatic (or semi-automatic) methods for identifying changes in a word's senses (such as new word senses) could be very helpful.

Early approaches to detecting novel word senses that rely

on rich lexical representations (e.g., Wilks, 1978) are not feasible in today's context since such resources are not available for large-scale vocabularies. However, words often change meaning in regular ways (Campbell, 2004), and this insight can be leveraged in computational systems. For example, Sagi et al. (2009) exploit knowledge of semantic widening and narrowing (extension and restriction of meaning) to automatically identify words which have undergone these changes. Although preliminary, their results suggest that focusing on specific types of semantic change is a promising direction for detecting new word senses.

One aspect of word-level semantics of great interest today is semantic orientation. Much recent computational work has looked at determining the sentiment or opinion expressed in some text (see Pang and Lee (2008) for an overview). A key aspect of many sentiment analysis systems is a lexicon in which words or senses are annotated with semantic orientation. Such lexicons are often manually-crafted (e.g., the General Inquirer, Stone et al., 1966). However, it is clearly important to have automatic methods to detect semantic changes that affect a word's orientation in order to keep such lexicons up-to-date. Indeed, there have been recent efforts to automatically infer polarity lexicons from corpora (e.g., Hatzivassiloglou and McKeown, 1997; Turney and Littman, 2003) and from other lexicons (e.g., Esuli and Sebastiani, 2006; Mohammad et al., 2009), and to adapt existing polarity lexicons to specific domains (e.g., Choi and Cardie, 2009). Similarly, since appropriate usage of words depends on knowledge of their semantic orientation, tools for detecting such changes would be helpful for lexicographers in updating dictionaries. Furthermore, diachronic studies in corpus linguistics have—to the best of our knowledge—not considered changes in the polarity of words, and have instead focused on topics such as changes in word frequency (e.g. Hilpert and Gries, 2009).

We focus here on *amelioration* and *pejoration*, common linguistic processes through which the meaning of a word changes to have a more positive or negative orientation.

Historical examples of amelioration and pejoration, respectively, include *nice*, which in Middle English meant ‘foolish’, and *vulgar*, originally meaning ‘common’. More recent examples are *sick* (now having the sense ‘excellent’, an amelioration), and *gay* (now having the sense ‘of inferior quality’, a pejoration, and often considered offensive). Our hypothesis is that methods for automatically inferring polarity lexicons from corpora can be used for detecting such changes in semantic orientation. If the corpus-based polarity of a word is found to vary significantly across two corpora which differ with respect to either timespan or geographic region, then that word is likely to have undergone amelioration or pejoration in one of these speech communities. In the case of corpora from different time periods, this approach could be used to find new word senses. Specifically, we adapt an existing web-based method for calculating corpus-based polarity (Turney and Littman, 2003) to work on smaller corpora (since our corpora will be restricted by timespan), and apply the method to words in the two corpora of interest.

2. Methods and materials

2.1. Determining semantic orientation

Turney and Littman (2003) present web and corpus-based methods for determining the semantic orientation of a target word. Their methods use either pointwise mutual information or latent semantic analysis to compare a target word to known words of positive and negative polarity. Here we focus on a variant of their PMI-based method.¹

Turney and Littman manually build small sets of known positive and negative seed words, and then determine the semantic orientation (SO) of a target word t by comparing its association with the positive and negative seed sets, *POS* and *NEG*, respectively.

$$SO-PMI(t) = PMI(t, POS) - PMI(t, NEG) \quad (1)$$

The association between the target and a seed set is then determined as below, where t is the target, $S = s_1, s_2 \dots s_n$ is a seed set of n words, N is the number of words in the corpus under consideration, and *hits* is the number of hits returned by a search engine for the given query.²

$$PMI(t, S) = \log \left(\frac{P(t, S)}{P(t)P(S)} \right) \quad (2)$$

$$\approx \log \left(\frac{N \cdot \text{hits}(t \text{ NEAR } (s_1 \text{ OR } s_2 \text{ OR } \dots \text{ OR } s_n))}{\text{hits}(t)\text{hits}(s_1 \text{ OR } s_2 \text{ OR } \dots \text{ OR } s_n)} \right) \quad (3)$$

¹In preliminary experiments we find a PMI-based method to out-perform a method using latent semantic analysis, and therefore choose to focus on PMI-based methods.

²This is Turney and Littman’s (2003) “disjunction” method for SO-PMI. Turney and Littman also present a variant referred to as “product” in which the association between a target t and seed set S is calculated as follows: $SO-PMI(t, S) = \sum_{s \in S} PMI(t, s)$. In preliminary experiments we find an adapted version of the disjunction variant to perform better, and we therefore focus on this method here.

In this study we do not use web data, and therefore do not need to estimate frequencies using the number of hits returned by a search engine. We therefore estimate $PMI(t, S)$ using frequencies obtained directly from a corpus, as below, where $freq(t, s)$ is the frequency of t and s co-occurring within a five-word window, and $freq(t)$ and $freq(s)$ are the frequency of t and s , respectively.³

$$PMI(t, S) \approx \log \left(\frac{N \sum_{s \in S} freq(t, s)}{freq(t) \sum_{s \in S} freq(s)} \right) \quad (4)$$

Turney and Littman focus on experiments using web data, the size of which allows them to use very small, but reliable, seed sets of just seven words each.⁴ However, their small seed sets can cause data sparseness problems when using the corpora of interest to us, which can be rather small since they are restricted in time period. Therefore, we use the positive and negative words from the General Inquirer (GI, Stone et al., 1966) as our seeds. Some words in GI are listed with multiple senses, and the polarity of these senses may differ. To avoid using seed words with ambiguous polarity, we select as seed words only those words which have either positive or negative senses, but not both. This gives positive and negative seed sets of 1621 and 1989 words, respectively, although at the cost of these seed words potentially being less reliable indicators of polarity than those used by Turney and Littman.

2.2. Corpora

In investigating this method for amelioration and pejoration detection, we make use of three British English corpora from differing time periods: the Lampeter Corpus of Early Modern English Tracts (Lampeter, Siemund and Claridge, 1997), approximately one million words of text from 1640–1740 taken from a variety of domains including religion, politics, and law; the Corpus of Late Modern English Texts Extended Version (CLMETEV, De Smet, 2005) consisting of fifteen million words of text from 1710–1920 concentrating on formal prose; and the British National Corpus (BNC, Burnard, 2007), one hundred million words from a variety of primarily written sources from the late 20th century. The size and time period of these three corpora is summarized in Table 1.

We first verify that our adapted version of Turney and Littman’s (2003) SO-PMI can reliably predict human polarity judgements on these corpora. Using a leave-one-out methodology, we calculate the polarity of each item in GI with frequency greater than five in the corpus under consideration, using all other GI words as (positive or negative)

³We do not smooth these estimates. In this study, we only calculate the polarity of a word t if its frequency is greater than five in the corpus being used, so the denominator is never zero. If t doesn’t co-occur with any seed word $s \in S$ the numerator is zero, in which case we simply set $PMI(t, S)$ to a very small number ($-\text{inf}$).

⁴Positive seeds: *good, nice, excellent, positive, fortunate, correct, superior*; negative seeds: *bad, nasty, poor, negative, unfortunate, wrong, inferior*.

Corpus	Percentage most polar items classified											
	Top 25%			Top 50%			Top 75%			100%		
	% acc.	BL	N	% acc.	BL	N	% acc.	BL	N	% acc.	BL	N
Lampeter	88	54	344	84	53	688	79	52	1032	74	50	1377
CLMETEV	92	61	792	90	59	1584	85	56	2376	80	55	3169
BNC	94	72	883	93	64	1767	89	59	2650	82	55	3534

Table 2: % accuracy (% acc.) for inferring the polarity of expressions in GI using each corpus. The accuracy for classifying the 25%, 50%, 75%, and 100% most polar items with frequency greater than five in the corresponding corpus is shown. In each case, the baseline (BL) of always choosing negative polarity and the number of items classified (N) is also shown.

Corpus	Time period	Approximate size (millions of words)
Lampeter	1640–1740	1
CLMETEV	1710–1920	15
BNC	Late 20th c.	100

Table 1: Time period and approximate size of each corpus.

seed words. The results are shown in Table 2. For each corpus, all accuracies are substantially higher than the baseline of always choosing the most frequent class, negative polarity. Moreover, when we focus on only those items with strong polarity—the top-25% most polar items—the accuracies are quite high, close to or over 90% in each case. Note that even with these restrictions on frequency and polarity, many items are still being classified—344 in the case of Lampeter, the smallest corpus. We conclude that using a very large set of potentially noisy seed words is useful for polarity measurement on even relatively small corpora.

3. Results

3.1. Identifying historical ameliorations and pejorations

We are compiling a dataset of words known to have undergone amelioration or pejoration to create an evaluation set for our methods. Although this research is in progress, here we apply our method to expressions collected thus far. Some examples are taken from etymological dictionaries (Room, 1986) and from textbooks discussing semantic change (Traugott and Dasher, 2002) and the history of the English language (Brinton and Arnovick, 2005). These expressions are restricted to those that are indicated as having undergone amelioration or pejoration in the eighteenth century or later (Room, 1986).⁵ We also search for additional test expressions in editions of Shakespearean plays that contain annotation as to words and phrases that are used differently in the play than they commonly are now (Shakespeare and Pearce, 2008a,b). Here we—perhaps naively—assume that the sense used by Shakespeare was the predominant sense at the time the play was written, and

⁵Note that historical dictionaries, such as the Oxford English Dictionary, do not appear to be appropriate for establishing the approximate date at which a word sense has become common because they give the date of the earliest known usage of a word sense, which could be much earlier than the widespread use of that sense.

consider these as expressions whose predominant sense has undergone semantic change. The expressions taken from Shakespeare are restricted to words whose senses as used in the plays are recorded in the Oxford English Dictionary (OED),⁶ and which both authors of this paper agree are ameliorations or pejorations.

For all the identified test expressions, we assume that their original meaning will be the predominant sense in Lampeter, while their ameliorated or pejorated sense will be dominant in CLMETEV. After removing expressions with frequency five or less in either Lampeter or CLMETEV, eight test items remain—six ameliorations and two pejorations. The results are shown in Table 3. Note that for seven out of eight expressions, the calculated change in polarity is as expected from the lexical resources; the one exception is *succeed*. The polarity of expressions calculated for the corpus expected to have higher polarity (CLMETEV in the case of ameliorations, Lampeter for pejorations) is significantly higher than that for the other corpus using a one-tailed paired t-test ($p = 0.024$).

3.2. Artificial ameliorations and pejorations

We would like to determine whether our method is able to identify known ameliorations and pejorations; however, as discussed in the previous section, the number of expressions in our historical dataset thus far is small. We can nevertheless evaluate our method on artificially created examples of amelioration and pejoration. For example, assume that *excellent* in Lampeter and *poor* in CLMETEV are in fact the same word. This would then be (an artificial example of) a word which has undergone pejoration. If our method assigns lower polarity to *poor* in CLMETEV than to *excellent* in Lampeter, then it has successfully identified this pejoration. This approach is similar to artificial data for word sense disambiguation evaluations, in which two distinct words are used to represent two senses of the same word (e.g., Schütze, 1992), except that in this case the distinct words are selected such that they have different polarity.

Since selecting appropriate pairs of words to compare is difficult, given the number of items, we average the polarity of all the positive/negative expressions in a given corpus with frequency greater than five and which co-occur at least once with both positive and negative seed words. These results are shown in Table 4. For each corpus, the positive

⁶OED Online. Oxford University Press. <http://dictionary.oed.com>

Expression	Change identified from resources	Change in polarity	Polarity in corpora	
			Lampeter	CLMETEV
ambition	amelioration	0.52	-0.76	-0.24
eager	amelioration	0.97	-1.09	-0.12
fond	amelioration	0.07	0.14	0.21
luxury	amelioration	1.49	-0.93	0.55
nice	amelioration	2.84	-2.48	0.36
*succeed	amelioration	-0.75	0.81	0.06
artful	pejoration	-1.71	1.33	-0.38
plainness	pejoration	-0.61	1.65	1.04

Table 3: The change in polarity, as well as polarity in each corpus, for each historical example of amelioration and pejoration. Note that *succeed* does not exhibit the expected change in polarity.

	Lampeter	CLMETEV	BNC
Ave. pos. pol.	0.58	0.50	0.40
Ave. neg. pol.	-0.74	-0.67	-0.76

Table 4: Average polarity of positive and negative words from GI in each corpus with frequency greater than five in the indicated corpus.

GI words have higher average polarity than the negative GI words in all other corpora. (All differences are strongly significant in unpaired t-tests: $p \ll 10^{-5}$.) Therefore, if we construct an artificial example of amelioration or pejoration, and estimate the polarity of this artificial word using any two of our three corpora, the expected polarity of the positive senses of that artificial word is higher than the expected polarity of the negative senses. This shows that our method can detect strong differences in calculated polarity across corpora, as expected for some ameliorations and pejorations.

3.3. Hunting for ameliorations and pejorations

Since we suggest our method as a way to discover potential new word senses that are ameliorations and pejorations, we test this directly by comparing the calculated polarity of words in a recent corpus, BNC, to those in an immediately preceding time period, CLMETEV. We consider the words with the largest increase and decrease in polarity between the two corpora as candidate ameliorations and pejorations, respectively, and then have human judges consider usages of these words to determine whether they are in fact ameliorations and pejorations.

The expressions with the ten largest increases and decreases in polarity from CLMETEV to BNC (restricted to expressions with frequency greater than five in each corpus) are presented in Table 5. An expression with an increase in polarity from CLMETEV to BNC is a candidate amelioration (left panel of Table 5—A), while an expression with a decrease from CLMETEV to BNC is a candidate pejoration (right panel—B). We extract ten random usages of each expression—or all usages if the word has frequency lower than ten—from each corpus, and then pair each usage from CLMETEV with a usage from BNC. This gives ten pairs of usages (or as many as are available) for each expression,

resulting in 190 total pairs.

We use Amazon Mechanical Turk (AMT, <https://www.mturk.com/>) to obtain judgements for each pair of usages. For each pair, a human judge is asked to decide whether the usage from CLMETEV or BNC is more positive/less negative, or whether the two usages are equally positive/negative. A sample of this AMT polarity judgement task is presented in Table 6. We solicit responses from ten judges for each pair, and pay \$0.05 per judgement.

The judgements obtained from AMT are shown in Table 5. For each candidate amelioration or pejoration the proportion of responses that the usage from CLMETEV, BNC, or neither is more positive/less negative is shown. For each expression, the majority response is indicated in boldface. In the case of both ameliorations and pejorations, four out of ten items are correct; these expressions are also shown in boldface. We also consider the average proportion of responses for each category (CLMETEV usage is more positive/less negative, BNC usage is more positive/less negative, neither usage is more positive or negative) for the candidate ameliorations and pejorations (shown in the last row of Table 5). Here we note that for candidate ameliorations the average proportion of responses that the BNC usage is more positive is higher than the average proportion of responses that the CLMETEV usage is more positive, and vice versa for candidate pejorations. This is an encouraging result, but in one-tailed paired t-tests it is not found to be significant for candidate ameliorations ($p = 0.12$), although it is marginally significant for candidate pejorations ($p = 0.05$).

We also consider an evaluation methodology in which we ignore the judgements for usage pairs for which the judgements are roughly uniformly distributed across the three categories. For each usage pair, if the proportion of judgements of the most frequent judgement is greater than 0.5 then this pair is assigned the category of the most frequent judgement, otherwise we ignore the judgements for this pair. We then count these resulting judgements for each candidate amelioration and pejoration. In this alternative evaluation, the overall results are quite similar to those presented in Table 5.

These are only preliminary evaluations which we intend to improve on in the future. In particular, in our current evaluation, each usage participates in only one pairing; the exact

(A) Candidate ameliorations				(B) Candidate pejorations			
Expression	Proportion of judgements for corpus of more positive usage			Expression	Proportion of judgements for corpus of more positive usage		
	CLMETEV	BNC	Neither		CLMETEV	BNC	Neither
bequeath	0.25	0.28	0.47	adornment	0.43	0.27	0.33
coerce	0.38	0.20	0.42	disavow	0.37	0.22	0.41
costliness	0.41	0.24	0.35	dynamic	0.43	0.27	0.30
disputable	0.30	0.43	0.27	elaboration	0.26	0.38	0.36
empower	0.30	0.29	0.40	fluent	0.25	0.34	0.41
foreboding	0.19	0.39	0.42	gladden	0.39	0.12	0.49
hysteria	0.26	0.39	0.35	outrun	0.30	0.38	0.31
slothful	0.24	0.44	0.31	skillful	0.43	0.27	0.29
thoughtfulness	0.21	0.50	0.29	synthesis	0.41	0.19	0.40
verification	0.27	0.27	0.46	wane	0.33	0.34	0.33
Average	0.28	0.34	0.37	Average	0.36	0.27	0.36

Table 5: (A) Expressions with top 10 increase in polarity from CLMETEV to BNC—candidate ameliorations; (B) Expressions with top 10 decrease in polarity from CLMETEV to BNC—candidate pejorations. For each expression, the proportion of human judgements for each category is shown: CLMETEV usage is more positive/less negative (CLMETEV), BNC usage is more positive/less negative (BNC), neither usage is more positive or negative (Neither). Majority judgements are shown in boldface, as are correct candidate ameliorations and pejorations according to the majority responses of the judges.

pairings chosen therefore heavily influence the outcome. In the future we will include each usage in multiple pairings in order to reduce this effect. We also have not measured the inter-annotator agreement for our AMT task. Our preliminary observation is that on average the most frequent judgement for a pairing corresponds to 55% of the judgements for that pairing. In the future we intend to further explore this to have a more clear assessment of the judges’ agreement on each pair.

4. Amelioration or pejoration of the seeds

Our method for identifying ameliorations and pejorations relies on knowing the polarity of a large number of seed words. However, a seed word itself may also undergo amelioration or pejoration, and therefore its polarity may in fact differ from what we assume it to be in the seed sets. Here we explore the extent to which noisy seed words—i.e., seed words labelled with incorrect polarity—affect the performance of our method. We begin by randomly selecting $n\%$ of the positive seed words, and $n\%$ of the negative seed words, and swapping these items in the seed sets. We then conduct a leave-one-out experiment, using the same methodology as in Section 2.2., in which we use the noisy seed words to calculate the polarity of all items in the GI lexicon with frequency greater than five in the corpus under consideration. We consider each n in $\{5, 10, 15, 20\}$, and repeat each experiment five times, randomly selecting the seed words to change the polarity of in each trial. The average accuracy over the five trials is shown in Figure 1. We observe a similar trend for all three corpora: the average accuracy decreases as the percentage of noisy seed words increases. However, with a small amount of noise in the seed sets, 5%, the reduction in absolute average accuracy is small, only 1–2%, for each corpus. Furthermore, when the percentage of noisy seed words is increased to 20%, the absolute average accuracy is lowered by only 5–7%. We conclude that by aggregating information from many seed

Instructions:

- Read the two usages of the word *disavow* below.
- Based on your interpretation of those usages, select the best answer.

A: in a still more obscure passage he now desires to DISAVOW the circular or aristocratic tendencies with which some critics have naturally credited him .

B: the article went on to DISAVOW the use of violent methods :

- *disavow* is used in a more positive, or less negative, sense in A than B.
- *disavow* is used in a more negative, or less positive, sense in A than B.
- *disavow* is used in an equally positive or negative sense in A and B.

Enter any feedback you have about this HIT. We greatly appreciate you taking the time to do so.

Table 6: A sample of the Amazon Mechanical Turk polarity judgement task.

words, our method for determining semantic orientation is robust against a small amount of noise in the seed sets.

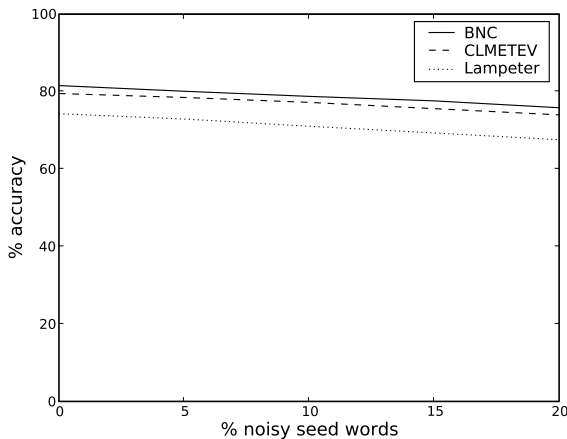


Figure 1: Average % accuracy for inferring the polarity of the items in GI for each corpus as the percentage of noisy seed words is varied.

5. Conclusions

The results of these preliminary experiments on identifying ameliorations and pejorations encourage further research in this direction. We plan to improve and extend this work in a number of ways.

We intend to improve our methods for measuring semantic orientation by incorporating syntactic information, such as the target expression’s part-of-speech, as well as linguistic knowledge about common patterns that indicate polarity; for example, adjectives co-ordinated by *but* often have opposite semantic orientation.

The corpora used in this study, although all consisting of British English, are not comparable, i.e., they were not constructed using the same or similar sampling strategies. It is possible that any differences in polarity found between these corpora can be attributed to differences in the composition of the corpora. In future work, we intend to evaluate our methods on more comparable corpora; for example, the Brown Corpus (Kucera and Francis, 1967) and Frown Corpus (Hundt et al., 1999)—comparable corpora of American English from the 1960s and 1990s, respectively—could be used to study changes in polarity between these time periods in American English. We are also excited about applying our methods to very recent corpora to identify new word senses.

In the present study we have only considered amelioration and pejoration across time. However, words may have senses which are specific to a particular speech community. In the future, we intend to apply our methods to comparable corpora of the same language, but different geographical regions, such as the International Corpus of English (<http://ice-corpora.net/ice/>) to identify words with differing semantic orientation in these speech communities.

Finally, we are working to enlarge our dataset of expressions known to have undergone amelioration or pejoration to enable more thorough evaluation of our methods. We also intend to conduct more extensive manual evaluation of

our methods, along the lines of the evaluation presented in Section 3.3.

Acknowledgements

This research was financially supported by the Natural Sciences and Engineering Research Council of Canada, the University of Toronto, and the Dictionary Society of North America.

References

- David K. Barnhart. 1985. Prizes and pitfalls of computerized searching for new words for dictionaries. *Dictionaries*, 7:253–260.
- Laurel Brinton and Leslie Arnovick, editors. 2005. *The English Language: A Linguistic History*. Oxford University Press.
- Lou Burnard. 2007. Reference guide for the British National Corpus (XML Edition). Oxford University Computing Services.
- Lyle Campbell. 2004. *Historical Linguistics: An Introduction*. MIT Press, Cambridge, MA.
- Yejin Choi and Claire Cardie. 2009. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, pages 590–598. Singapore.
- Henrik De Smet. 2005. A corpus of Late Modern English texts. *International Computer Archive of Modern and Medieval English*, 29:69–82.
- Andrea Esuli and Fabrizio Sebastiani. 2006. SENTIWORDNET: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation*, pages 417–422. Genoa, Italy.
- Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the Eighth Conference of the European Chapter of the Association for Computational Linguistics*, pages 174–181. Madrid, Spain.
- Martin Hilpert and Stefan Th. Gries. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing*, 24(4):385–401.
- Marianne Hundt, Andrea Sand, and Paul Skandera. 1999. Manual of information to accompany the Freiburg - Brown Corpus of American English (‘Frown’). <http://khnt.aksis.uib.no/icame/manuals/frown/INDEX.HTM>.
- Henry Kucera and W. Nelson Francis. 1967. *Computational Analysis of Present Day American English*. Brown University Press, Providence, Rhode Island.

- Saif Mohammad, Cody Dunne, and Bonnie Dorr. 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP-2009)*, pages 599–608. Singapore.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2):1–135.
- Adrian Room. 1986. *Dictionary of Changes in Meaning*. Routledge and Kegan Paul, London, New York.
- Eyal Sagi, Stefan Kaufmann, and Brady Clark. 2009. Semantic density analysis: Comparing word meaning across time and space. In *Proceedings of the EACL 2009 Workshop on GEMS: GEometrical Models of Natural Language Semantics*, pages 104–111. Athens, Greece.
- Hinrich Schütze. 1992. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.
- William Shakespeare and Joseph Pearce. 2008a. *Hamlet (Ignatius Critical Editions)*. Ignatius Press, San Francisco.
- William Shakespeare and Joseph Pearce. 2008b. *King Lear (Ignatius Critical Editions)*. Ignatius Press, San Francisco.
- Rainer Siemund and Claudia Claridge. 1997. The Lampeter Corpus of Early Modern English Tracts. *International Computer Archive of Modern and Medieval English*, 21:61–70.
- John Simpson. 2007. Neologism: The long view. *Dictionaries*, 28:146–148.
- Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie, editors. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.
- Elizabeth C. Traugott and Richard B. Dasher. 2002. *Regularity in Semantic Change*. Cambridge University Press.
- Peter D. Turney and Michael L. Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, 21(4):315–346.
- Yorick Wilks. 1978. Making preferences more active. *Artificial Intelligence*, 11(3):197–223.
- Uri Zernik, editor. 1991. *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*. Lawrence Erlbaum Associates, Hillsdale, NJ.