Using WordNet to Measure Semantic Orientations of Adjectives

Jaap Kamps and **Maarten Marx** and **Robert J. Mokken** and **Maarten de Rijke**
Language & Inference Technology Group
ILLC, University of Amsterdam
Nieuwe Achtergracht 166
NL-1018WV Amsterdam, the Netherlands
{kamps,marx,mokken,mdr}@science.uva.nl
http://lit.science.uva.nl/

Abstract

Current WordNet-based measures of distance or similarity focus almost exclusively on WordNet's taxonomic relations. This effectively restricts their applicability to the syntactic categories of noun and verb. We investigate a simple graph-theoretic model of WordNet's most important relation—synonymy—and propose measures that determine the semantic orientation of adjectives for three factors of subjective meaning. Evaluation against human judgments shows the effectiveness of the resulting measures.

1 Introduction

The ability to establish the relatedness, similarity, or distance between words and concepts is at the heart of computational linguistics. A case in point is the field of information retrieval, where experiments determine the ability of a retrieval systems to measure the similarity between queries and documents. Research in similarity measures dates back (at least) to the first conference in computational linguistics (Harper, 1965). There has been much interest in distances in semantic networks, originating with research on computational models of semantic memory (Quillian, 1967, 1968). This line of research has greatly profited from the advent of the WordNet lexical database (Miller, 1990; Fellbaum, 1998).

There is a broad range of distance or similarity measures based (completely or partially) on WordNet. Rada et al. (1989) use a simple edge-counting over taxonomy links (IS-A, Part-of, or WordNet's hyponymy relation). Hirst and St-Onge (1998) extend the path-length to all relations in WordNet (clustering them to horizontal, up, or down) and penalizing changes of direction. Leacock and Chodrow (1998) consider the path-length of hyponymy relations in WordNet, while reducing the distances by the depth in the hierarchy. Again, focusing on the hyponymy relation, Resnik (1995, 1999) extends the lexical hierarchy methods with a notion of information content, derived from word frequencies in the Brown Corpus, resulting in a hybrid measure combining WordNet's taxonomic hierarchy with corpus based methods. Lin (1998)'s information-theoretic notion of similarity is a theoretically motivated refinement of Resnik's measure. Budanitsky and Hirst (2001) give an overview of five measures, and evaluate their performance using a word association task (Miller and Charles, 1991).

A striking observation is that all these distance or similarity measures are only applicable to the hyponymy relations (the IS-A or HAS-PART relation in WordNet); a notable exception is (Hirst and St-Onge, 1998) whose method works for all syntactic categories in WordNet. The restriction to hyponymy makes distance or similarity measures only applicable to the syntactic categories of noun and verb. Thus, the measures proposed earlier do not apply to adjectives and adverbs. But precisely these syntactic categories can be crucial for some applications. In particular, they contain all the modifiers, adjectives and adverbs that modify or elaborate the meaning of other words. These words are of particular interest for determining the semantic orientation of subjective words (Hearst, 1992; Hatzivassiloglou and McKeown, 1997; Turney, 2002).

The aim of this paper is to develop WordNet-based measures for the semantic orientation of adjectives. The paper is structured as follows. In Section 2, we investigate a simple graph-theoretic model of WordNet, focusing on its most important relation—synonymy. In Section 3, we discuss the main factors of subjective meaning, define corresponding measures based on distances in the synonymy-graph, and evaluate the resulting measures against a human judged collection of words. Finally, in Section 4, we discuss our results and draw some conclusions.

2 Simple Distance Measures on WordNet

We will define a distance measure using elementary notions from graph theory (Harary, 1969). Here, we construct relations of the level of words.[1] The simplest approach here is just to collect all words in WordNet, and relate words that can be synonymous (i.e., they are in the same synset).

Definition 1 *Let* $\mathcal{G}(\mathcal{W}, \mathsf{Synonymy})$ *be a simple graph with* $\mathcal{W}$ *the set of nodes being all the words with associated part-of-speech in WordNet, and* Synonymy *the set of edges connecting each pair of synonymous words.*

We can immediately make some graph-theoretic observation on the simple graph $\mathcal{G}$: for example, the Synonymy relation is irreflexive and symmetric, and every set of synonymous words in WordNet (i.e., synset) is a clique of the simple graph $\mathcal{G}$. Next, we can look at *walks* (arbitrary sequences of nodes and lines), *trails* (walks with distinct edges), *paths* (trails with distinct nodes) in the WordNet graph $\mathcal{G}$.

Example 1 *Look at the graph in Figure 1. A* walk *in this graph is* $W_1 =$ n1, l1, n2, l3, n3, l3, n2*. A* trail *in this graph is* $W_2 =$ n1, l1, n2, l3, n3, l4, n4, l2, n2 *(and **not** W_1). A* path *in this graph is* $W_3 =$ n1, l1, n2, l3, n3, l4, n4 *(and **not** W_1 or W_2). Another* path *connecting* n1 *and* n4 *is* $W_4 =$ n1, l1, n2, l2, n4.

The length of a walk/trail/path is the number of lines occurring in it. We will be especially interested in

[1] Using similar techniques, one can investigate the dual graph of synsets (sets of synonymous words in WordNet parlance).

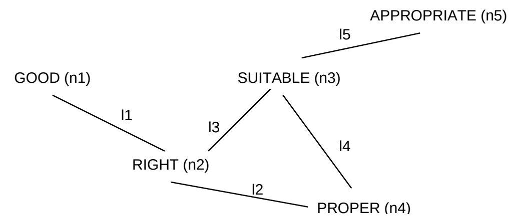

Figure 1: Walks, Trails, and Paths in the WordNet Graph.

the *geodesics*, i.e., in the shortest path between two nodes or words. The *geodesic distance*, or simply *distance*, between two nodes is the length of a shortest path. Recall the example above, here we have two paths connecting n1 and n4, W_3 of length 3 and W_4 of length 2, and only W_4 is a geodesic.

Definition 2 *The* distance $\mathsf{d}(w_i, w_j)$ *between two words* w_i *and* w_j *is the length of a shortest path between* w_i *and* w_j*, If there is no path between* w_i *and* w_j*, their distance is infinite.*

The minimal path-length enjoys some of the geometrical properties we might expect from a distance measure—it is a *metric*.[2]

We have determined a number of characteristic network results on the WordNet graph. The design strategy of WordNet was to have no relations across different syntactic categories (the separability hypothesis). Thus, the massive graph has disjoint subgraphs of nouns, verbs, adjectives, and adverbs. The degree sequence of the graph satisfies a power law distribution familiar from real networks like the Internet, cellular networks, or collaboration graphs (Aiello et al., 2001; Albert and Barabási, 2002).

For three syntactic categories, we find a giant component: In the noun-subgraph there is a connected component of size 10,922 (or 10% of all nouns); in the verb-subgraph there is a component of size 6,365 (or 57% of all verbs); and in the adjective-subgraph there is a component of size 5,427 (or 25% of all adjectives). In the adverbs-subgraph there are two large components of size 64 and 61. These are the fourth and fifth largest components in the en-

[2] The minimal path-length is a *metric*, that is, it gives a non-negative number $\mathsf{d}(w_i, w_j)$ such that
$\mathsf{d}(w_i, w_j) = 0$ if and only if $w_i = w_j$,
$\mathsf{d}(w_i, w_j) = \mathsf{d}(w_j, w_i)$, and
$\mathsf{d}(w_i, w_j) + \mathsf{d}(w_j, w_k) \geq \mathsf{d}(w_i, w_k)$.

tire WordNet—the second largest connected components of nouns, verbs, and adjectives contain 52, 14, and 30 words, respectively. This is in line with results from the theory of random graphs showing edge density (M/N) driving the emergence of a giant component (Erdős and Rényi, 1960; Janson et al., 2000).

The giant component in the adjectives is of particular interest: it contains all modifiers used to express affective or emotive meaning. Linguistically, modifiers are words that provide a means to modify or elaborate the meaning of words, in particular, the sole function of adjectives is to modify nouns (like *good* and *exquisite* in *a good idea*, *an exquisite taste*). We can analyze the words in this connected component using the distance metric defined above.

3 Semantic Orientations of Adjectives

3.1 Which Subjective Orientations?

The classic work on measuring emotive or affective meaning in texts is Charles Osgood's Theory of Semantic Differentiation (Osgood et al., 1957). Osgood and his collaborators identify the aspect of meaning in which they are interested as

> a strictly psychological one: those cognitive states of human language users which are necessary antecedent conditions for selective encoding of lexical signs and necessary subsequent conditions in selective decoding of signs in messages. (Osgood et al., 1957, p.318)

Their semantic differential technique uses several pairs of bipolar adjectives to scale the responses of subjects to words, short phrases, or texts. That is, subjects are asked to rate their meaning on scales like active–passive; good–bad; optimistic–pessimistic; positive–negative; strong–weak; serious–humorous; and ugly–beautiful.

Each pair of bipolar adjectives is a factor in the semantic differential technique. As a result, the differential technique can cope with quite a large number of aspects of affective meaning. A natural question to ask is whether each of these factors is equally important. Osgood et al. (1957) use factorial analysis of extensive empirical tests to investigate this question. The surprising answer is that most of the variance in judgment could be explained by only three major factors. These three factors of the affective or emotive meaning are the *evaluative* factor (e.g., good–bad); the *potency* factor (e.g., strong–weak); and the *activity* factor (e.g., active–passive). Among these three factors, the evaluative factor has the strongest relative weight. All the three pairs of bipolar adjectives are in the giant adjective component described in Section 2.

3.2 Measures for Semantic Orientations

We will investigate measures based on the WordNet lexical database. The evaluative dimension of Osgood is typically determined using the adjectives 'good' and 'bad' (other operationalizations are possible depending on the subject under investigation). The geodesic distance (i.e., the minimal path connecting two words) is a straightforward generalization of the synonymy relation. The synonymy relation connects words with similar meaning, so the minimal distance between words says something on the similarity of their meaning. For example, using the part of WordNet graph shown in Figure 1 we now find that $\mathsf{d}(\mathsf{good}, \mathsf{right}) = 2$, $\mathsf{d}(\mathsf{good}, \mathsf{proper}) = 2$, $\mathsf{d}(\mathsf{good}, \mathsf{suitable}) = 2$, and $\mathsf{d}(\mathsf{good}, \mathsf{appropriate}) = 4$. This suggest that we can use the distance to the word 'good' as a measure of 'goodness.' Figure 2 shows the minimal-path lengths of a selection of adjectives to the adjective 'good' based on the WordNet database. Inspection of such a cloud of words gives us some confidence in the use of distance d as a measure for similarity of meaning. Note that we do not claim that the values obtained in this way are a precise scale for measuring degrees of goodness. Rather, we only expect a weak relation between the words used to express an positive opinion and their distance to words like 'good.'

However, further experimentation quickly reveals that this relation is very weak indeed. A striking example of this is that we also find that 'good' and 'bad' themselves are closely related in WordNet. There exists a 5-long sequence $\langle \mathsf{good}, \mathsf{sound}, \mathsf{heavy}, \mathsf{big}, \mathsf{bad} \rangle$. So, we have that $\mathsf{d}(\mathsf{good}, \mathsf{bad}) = 4$! Even though the adjectives 'good' and 'bad' have opposite meaning, they are still closely related by the synonymy relation. Although this is perhaps remarkable, it is not due to some error in the WordNet database (there exist sev-

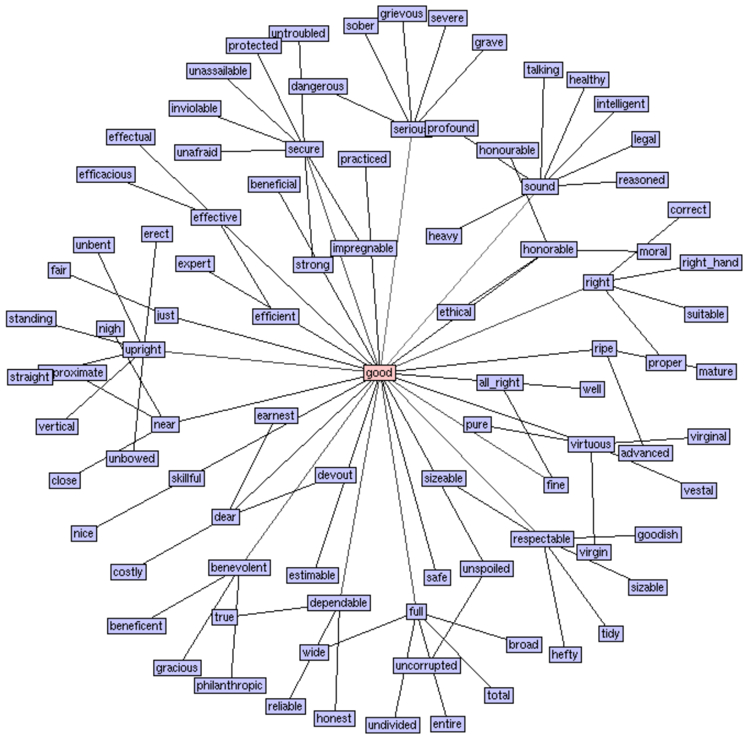

Figure 2: Part of the WordNet database from the vista point of adjective 'good.' The edges are synonymy relations, nodes are only connected by a shortest path.

eral paths of length 5). Part of the explanation seem to be the wide applicability of these two adjectives (WordNet has 14 senses of bad and 25 senses of good).[3]

Fortunately, we can use this fact to our advantage: For each word, we can now consider not only the shortest distance to 'good' but also the shortest distance to the antonym 'bad.' Figure 3 shows the minimal-path lengths of words to both the adjectives 'good' and 'bad.' Inspection reveals that words neatly cluster in groups depending on the minimal path-lengths to 'good' and 'bad'. In short, this sort of graphs seems to resonate closely with an underlying evaluative factor (at least, much better than graphs based on a single distance measure such as figure 2).

For Osgood's evaluative factor we operationalize this idea by defining a function EVA that measures the relative distance of a word to the two reference words 'good' and 'bad.' In symbols,

$$\mathsf{EVA}(w) = \frac{\mathsf{d}(w, \mathsf{bad}) - \mathsf{d}(w, \mathsf{good})}{\mathsf{d}(\mathsf{good}, \mathsf{bad})}.$$

The maximal difference in minimal-path length to the two reference words depends on the distance d between the two reference words. Therefore, we divide the difference by the distance between the two reference words, yielding a value in the interval $[-1, 1]$. We now have that for every word the EVA function assigns a value ranging from -1 (for words on the 'bad' side of the lexicon) to 1 (for words on the 'good' side of the lexicon).[4] For example, using WordNet the word 'honest' gets assigned the value 1:

$$\begin{aligned}\mathsf{EVA}(\mathsf{honest}) &= \frac{\mathsf{d}(\mathsf{honest}, \mathsf{bad}) - \mathsf{d}(\mathsf{honest}, \mathsf{good})}{\mathsf{d}(\mathsf{good}, \mathsf{bad})} \\ &= \frac{6-2}{4} = 1.\end{aligned}$$

[3]Think of the small world problem predicting mean distance of 6 between arbitrary people (Milgram, 1967).

[4]Recall that the geodesic distance function assigns infinity to unconnected words. If a word w has $\mathsf{d}(w, \mathsf{good}) = \infty$ then also $\mathsf{d}(w, \mathsf{bad}) = \infty$. This implies that $\mathsf{EVA}(w) = \frac{\infty - \infty}{4} = 0$, so unconnected words do not affect the evaluative factor.

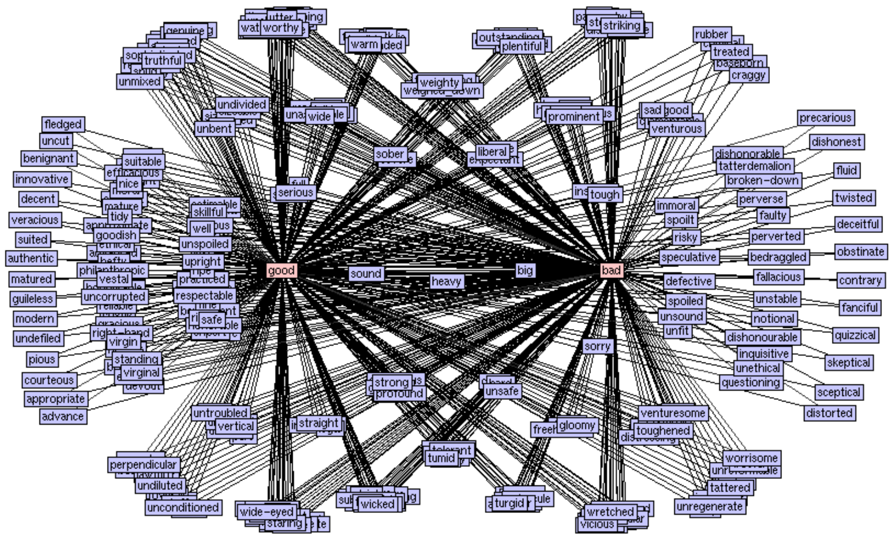

Figure 3: The geodesic distances to adjectives ‘good’ and ‘bad’. Nodes are connected by edges of length corresponding to the distance d.

In a similar vein, we can define measures for Osgood’s other dimensions. For the potency factor we define a function POT of w as

$$\mathsf{POT}(w) = \frac{\mathsf{d}(w, \mathsf{weak}) - \mathsf{d}(w, \mathsf{strong})}{\mathsf{d}(\mathsf{strong}, \mathsf{weak})}$$

and for the activity factor we define a function ACT of w as

$$\mathsf{ACT}(w) = \frac{\mathsf{d}(w, \mathsf{passive}) - \mathsf{d}(w, \mathsf{active})}{\mathsf{d}(\mathsf{active}, \mathsf{passive})}.$$

In fact, this allows us to define measure for any two connected words in WordNet.

3.3 Evaluation

We can evaluate our WordNet measures against the manually constructed lists of the General Inquirer (Stone et al., 1966; Stone, 1997). The General Inquirer is the classic system for content analysis—research techniques for making inferences by systematically and objectively identifying specified characteristics within text (Stone et al., 1966, p.5). The General Inquirer contains sets of words for all the three Osgood factors. These lists of words are derived from the Stanford Political Dictionary where, starting from a list of 3,000 most frequently used words in the English languages, three or more judges were asked to indicate which dimension were relevant to each word (Stone et al., 1966, p.189). After removing repeated occurrences due to multiple lexemes, there are 765 positive and 873 negative words for the evaluative factor; 1,474 strong and 647 weak words for the potency factor; and 1,568 active and 732 passive words for the activity factor. Additionally, there is a newer, extended set of words for the evaluative factor containing 1,634 positive and 2,004 negative words. The General Inquirer sets contain various syntactic categories, and does not indicate neutral words on the three dimensions. We evaluate on the intersection of words in the General Inquirer and our list of adjectives found in WordNet. Table 1 shows the number of words in the intersection of both lists, and the percentage of agreement between the two lists. In Table 1 we only treat words scoring 0 as neutral. If we consider a larger interval as neutral, the precision of our measure increases at the cost of a lower number of words in the intersec-

Factor	*Measure*	*# Words*	*Correct*
Evaluative	EVA	349	68.19%
Potency	POT	419	71.36%
Activity	ACT	173	61.85%
Evaluative II	EVA	667	67.32%

Table 1: Evaluation against the General Inquirer.

tion. For example, when treating $[-0.25, 0.25]$ as neutral, the score for the evaluative factor is 76.72% and 76.38% for the extended set, for the potency factor is 76.61%, and for the activity factor is 78.73%.

4 Discussion and Conclusions

In this paper, we developed a simple distance measure on WordNet, and showed how it can be used to determine the semantic orientation of adjectives. Current WordNet-based measures of distance or similarity focus almost exclusively on taxonomic relations. This effectively restricts their application to the noun and verb categories in WordNet. An important exception is (Hirst and St-Onge, 1998), extending (Morris and Hirst, 1991) dealing with *Roget's Thesaurus*, which uses all relations coded in WordNet. Although this distance measure can be applied to the adjective category in WordNet, it is unsuitable for determining the semantic orientation of adjectives. Hirst and St-Onge (1998, p.308) include the antonymy relation as one of the three *strong* relations between words. However, all the pairs of adjectives used to measure subjective meaning are directly related by the antonymy relation. As a result, this destroys the bipolarity of the concepts we are interested in.

It seems clear that the choice of similarity or distance measure greatly depends on the type of task at hand. First, there are differences in applicability. Similarity measures using the taxonomic hyponymy relation can only be applied to the noun and verb categories. Our distance measure using the synonymy relation can only be applied to words in connected components. Second, there are differences in the level of relations. Most of the WordNet relations are between synsets or concepts. The synonymy relation is the only WordNet relation on words. Third, there are differences in granularity. The taxonomic relations WordNet can be coarse-grained when compared to the fine-grained synonymy relation (Edmonds and Hirst, 2002). Our choice is motivated by the aim to determine the semantic orientation of adjectives. Already Quillian (1968, p.228) has it that

> One issue facing the investigator of semantic memory is: exactly what is it about word meanings that is to be considered? First, the memory model here is designed to deal with exactly complementary kinds of meaning to that involved in Osgood's "semantic differential" (Osgood et al., 1957). While the semantic differential is concerned with people's feelings in regard to words, or the words possible emotive impact on others, this model is explicitly designed to represent the nonemotive, relatively "objective" part of meaning.

We have shown in this paper how a measure for the affective meaning studied by Osgood et al. can be derived from a representation of the relatively "objective" meaning as represented in the WordNet database. The effectiveness of the resulting measures may come as no surprize given that the initial set of words in WordNet were from the Brown corpus plus "all the adjective pairs that Charles Osgood had used to develop the semantic differential" (Fellbaum, 1998, p.xix).

The measure for the evaluative factor of adjectives is related to work on text understanding; research in this area, such as (Hearst, 1992), has been looking at the *directionality* (e.g., is the agent in favor of, neutral, or opposed to the event) as a contrasting criterion to topicality. Automatically assigning positive or negative semantic orientation based on a large corpus, the Wall Street Journal corpus, is investigated in (Hatzivassiloglou and McKeown, 1997; Hatzivassiloglou and Wiebe, 2000). The authors use the ingenious method of analyzing conjoined adjectives, e.g., 'and' indicates agreement of alignment (good *and* beautiful) and 'but' indicates disagreement of alignment (friendly *but* dangerous). Given a list of candidate words, such as lists of modifiers, one may also use collocation statistics, including maximum likelihood estimators (Dunning, 1993) and point-wise mutual information (Manning and Schütze, 1999; Turney, 2001). The statistical estimations can be obtained from a large corpus, or from Internet search engine's hit counts.[5] Turney (2002) calculates the orientation of a text by the similarity between a word or phrase and two spe-

[5]The use of this technique for determining subjective meaning was pioneered in so-called sucks-rules-o-meters on the Internet, e.g., `http://srom.zgp.org/`, using search engine hit counts on names of operating systems (i.e., Windows, Linux, MacOS) combined with either the word 'rules' or 'sucks'.

cific words, 'excellent' and 'poor.' We believe that all of these methods can be extended fruitfully to the other factors of subjective meaning as identified by Osgood. Our current research focuses on refinements of the measures for semantic orientation, both by incorporating corpus statistics, as well as by using multidimensional scaling techniques on the high-dimensional space of words in WordNet.

References

William Aiello, Fan Chung, and Linyuan Lu. 2001. A random graph model for power law graphs. *Experimental Mathematics*, 10:53–66.

Réka Albert and Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97.

Alexander Budanitsky and Graeme Hirst. 2001. Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In *Workshop on WordNet and Other Lexical Resources*. Second meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh.

Ted Dunning. 1993. Accurate methods for the statistics of surprize and coincidence. *Computational Linguistics*, 19:61–74.

Philip Edmonds and Graeme Hirst. 2002. Near-synonymy and lexical choise. *Computational Linguistics*, 28:105–144.

Paul Erdős and Alfred Rényi. 1960. On the evolution of random graphs. *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 5:17–61.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*, Language, Speech, and Communication Series. The MIT Press, Cambridge MA.

Frank Harary. 1969. *Graph Theory*. Addison-Wesley Series in Mathematics. Addison-Wesley Publishing Company, Reading MA.

Kenneth E. Harper. 1965. Measurement of similarity between nouns. In *International Conference on Computational Linguistics*.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the ACL 1997*, pages 174–181. Morgan Kaufmann Publishers.

Vasileios Hatzivassiloglou and Janyce M. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *International Conference on Computational Linguistics (COLING-2000)*. Morgan Kaufmann Publishers.

Marti A. Hearst. 1992. Direction-based text interpretation as an information access refinement. In Paul S. Jacobs, editor, *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval*, pages 257–274. Lawrence Erlbaum Associates, Hillsdale.

Graeme Hirst and David St-Onge. 1998. Lexical chains as representation of context for the detection and correction of malapropisms. In Fellbaum (1998), chapter 13, pages 305–332.

Svante Janson, Tomasz Luczak, and Andrzej Rucinski. 2000. *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics. Wiley Interscience, New York NY.

Claudia Leacock and Martin Chodrow. 1998. Combining local context and WordNet similarity for word sense identification. In Fellbaum (1998), chapter 11, pages 265–284.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning*. Madison WI.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge MA.

Stanley Milgram. 1967. The small world problem. *Psychology Today*, 2:61–67.

George E. Miller and Walter A. Charles. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6:1–28.

George A. Miller. 1990. WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312. Special Issue.

Jane Morris and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Lin-*

guistics, 17:21–48.

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

M. Ross Quillian. 1967. Word concepts: a theory and simulation of some basic semantic capabilities. *Behavioral Science*, 12:410–430.

M. Ross Quillian. 1968. Semantic memory. In Marvin Minsky, editor, *Semantic Information Processing*, chapter 4, pages 227–270. The MIT Press, Cambridge MA.

Roy Rada, Hafedh Mili, Ellen Bicknell, and Maria Blettner. 1989. Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 19:17–30.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453. Morgan Kaufmann.

Philip Resnik. 1999. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130.

Philip J. Stone, Dexter C. Dunphy, Marschall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: a computer approach to content analysis*. M.I.T. studies in comparative politics. MIT Press, Cambridge MA.

Philip J. Stone. 1997. Thematic text analysis: new agendas for analyzing text content. In Carl Roberts, editor, *Text Analysis for the Social Sciences*, pages 35–54. Lawrence Erlbaum Associates, Mahwah NJ.

Peter D. Turney. 2001. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In Luc De Raedt and Peter Flach, editors, *Proceedings of the European Conference on Machine Learning (ECML2001)*, volume 2167 of *Lecture Notes in Artificial Intelligence*, pages 491–502. Springer, Berlin.

Peter D. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, pages 417–424.