cmp-lg/9511007 29 Nov 1995

Using Information Content to Evaluate Semantic Similarity in a Taxonomy

Philip Resnik*
Sun Microsystems Laboratories
Two Elizabeth Drive
Chelmsford, MA 01824-4195 USA
philip.resnik@east.sun.com

Abstract

This paper presents a new measure of semantic similarity in an IS-A taxonomy, based on the notion of information content. Experimental evaluation suggests that the measure performs encouragingly well (a correlation of $r = 0.79$ with a benchmark set of human similarity judgments, with an upper bound of $r = 0.90$ for human subjects performing the same task), and significantly better than the traditional edge counting approach ($r = 0.66$).

1 Introduction

Evaluating semantic relatedness using network representations is a problem with a long history in artificial intelligence and psychology, dating back to the spreading activation approach of Quillian [1968] and Collins and Loftus [1975]. Semantic similarity represents a special case of semantic relatedness: for example, cars and gasoline would seem to be more closely related than, say, cars and bicycles, but the latter pair are certainly more similar. Rada et al. [1989] suggest that the assessment of similarity in semantic networks can in fact be thought of as involving just taxonomic (IS-A) links, to the exclusion of other link types; that view will also be taken here, although admittedly it excludes some potentially useful information.

A natural way to evaluate semantic similarity in a taxonomy is to evaluate the distance between the nodes corresponding to the items being compared — the shorter the path from one node to another, the more similar they are. Given multiple paths, one takes the length of the shortest one [Lee *et al.*, 1993; Rada and Bicknell, 1989; Rada *et al.*, 1989].

A widely acknowledged problem with this approach, however, is that it relies on the notion that links in the taxonomy represent uniform distances. Unfortunately, this is difficult to define, much less to control. In real taxonomies, there is wide variability in the "distance" covered by a single taxonomic link, particularly when certain sub-taxonomies (e.g. biological categories) are much denser than others. For example, in WordNet [Miller, 1990], a broad-coverage semantic network for English constructed by George Miller and colleagues at Princeton, it is not at all difficult to find links that cover an intuitively narrow distance (RABBIT EARS IS-A TELEVISION ANTENNA) or an intuitively wide one (PHYTOPLANKTON IS-A LIVING THING). The same kinds of examples can be found in the Collins COBUILD Dictionary [Sinclair (ed.), 1987], which identifies superordinate terms for many words (e.g. SAFETY VALVE IS-A VALVE seems a lot narrower than KNITTING MACHINE IS-A MACHINE).

In this paper, I describe an alternative way to evaluate semantic similarity in a taxonomy, based on the notion of information content. Like the edge counting method, it is conceptually quite simple. However, it is not sensitive to the problem of varying link distances. In addition, by combining a taxonomic structure with empirical probability estimates, it provides a way of adapting a static knowledge structure to multiple contexts. Section 2 sets up the probabilistic framework and defines the measure of semantic similarity in information-theoretic terms; Section 3 presents an evaluation of the similarity measure against human similarity judgments, using the simple edge-counting method as a baseline; and Section 4 discusses related work.

2 Similarity and Information Content

Let $\mathcal{C}$ be the set of concepts in an IS-A taxonomy, permitting multiple inheritance. Intuitively, one key to the similarity of two concepts is the extent to which they share information in common, indicated in an IS-A taxonomy by a highly specific concept that subsumes them both. The edge counting method captures this indirectly, since if the minimal path of IS-A links between two nodes is long, that means it is necessary to go high in the taxonomy, to more abstract concepts, in order to find a least upper bound. For example, in WordNet, NICKEL and DIME are both subsumed by COIN, whereas the most specific superclass that NICKEL and CREDIT CARD share is MEDIUM OF EXCHANGE.[1] (See Figure 1.)

*Appears in Proceedings of IJCAI-95. Portions of this research were supported by an IBM Graduate Fellowship and grants ARO DAAL 03-89-C-0031, DARPA N00014-90-J-1863, NSF IRI 90-16592, and Ben Franklin 91S.3078C-1.

[1] In a feature-based setting (e.g. [Tversky, 1977]), this would be reflected by explicit shared features: nickels and

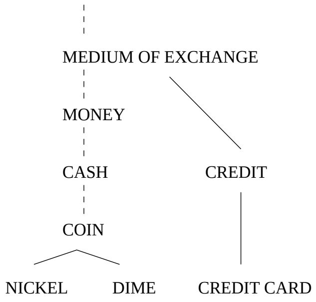

Figure 1: Fragment of the WordNet taxonomy. Solid lines represent IS-A links; dashed lines indicate that some intervening nodes were omitted to save space.

By associating probabilities with concepts in the taxonomy, it is possible to capture the same idea, but avoiding the unreliability of edge distances. Let the taxonomy be augmented with a function $\mathrm{p} : \mathcal{C} \rightarrow [0,1]$, such that for any $c \in \mathcal{C}$, $\mathrm{p}(c)$ is the probability of encountering an instance of concept c. This implies that p is monotonic as one moves up the taxonomy: if c_1 IS-A c_2, then $\mathrm{p}(c_1) \leq \mathrm{p}(c_2)$. Moreover, if the taxonomy has a unique top node then its probability is 1.

Following the standard argumentation of information theory [Ross, 1976], the *information content* of a concept c can be quantified as negative the log likelihood, $-\log \mathrm{p}(c)$. Notice that quantifying information content in this way makes intuitive sense in this setting: as probability increases, informativeness decreases, so the more abstract a concept, the lower its information content. Moreover, if there is a unique top concept, its information content is 0.

This quantitative characterization of information provides a new way to measure semantic similarity. The more information two concepts share in common, the more similar they are, and the information shared by two concepts is indicated by the information content of the concepts that subsume them in the taxonomy. Formally, define

$$\mathrm{sim}(c_1, c_2) \quad = \quad \max_{c \,\in\, S(c_1, c_2)} \left[-\log \mathrm{p}(c)\right], \qquad (1)$$

where $S(c_1, c_2)$ is the set of concepts that subsume both c_1 and c_2. Notice that although similarity is computed by considering all upper bounds for the two concepts, the information measure has the effect of identifying minimal upper bounds, since no class is less informative than its superordinates. For example, in Figure 1, COIN, CASH, etc. are all members of $S(\text{NICKEL}, \text{DIME})$, but the concept that is structurally the minimal upper bound, COIN, will also be the most informative. This can make a difference in cases of multiple inheritance; for example, in Figure 2, METAL and CHEMICAL ELEMENT are not structurally distinguishable as upper bounds of NICKEL' and GOLD', but their information content may in fact be quite different.

In practice, one often needs to measure word similarity, rather than concept similarity. Using $s(w)$ to represent the set of concepts in the taxonomy that are senses of word w, define

$$\mathrm{sim}(w_1, w_2) \quad = \quad \max_{c1,\, c2} \left[\mathrm{sim}(c_1, c_2)\right], \qquad (2)$$

where c_1 ranges over $s(w_1)$ and c_2 ranges over $s(w_2)$. This is consistent with Rada et al.'s [1989] treatment of "disjunctive concepts" using edge counting: they define the distance between two disjunctive sets of concepts as the minimum path length from any element of the first set to any element of the second. Here, the word similarity is judged by taking the maximal information content over all concepts of which both words could be an instance. For example, Figure 2 illustrates how the similarity of words *nickel* and *gold* would be computed: the information content would be computed for all classes subsuming any pair in the cross product of {NICKEL,NICKEL'} and {GOLD,GOLD'}, and the information content of the most informative class used to quantify the similarity of the two words.

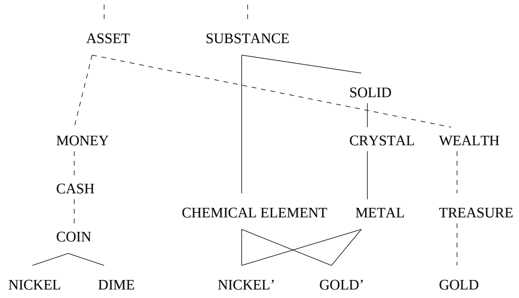

Figure 2: Another fragment of the WordNet taxonomy

3 Evaluation

3.1 Implementation

The work reported here used WordNet's (50,000-node) taxonomy of concepts represented by nouns (and compound nominals) in English.[2] Frequencies of concepts in the taxonomy were estimated using noun frequencies from the Brown Corpus of American English [Francis and Kučera, 1982], a large (1,000,000 word) collection of text across genres ranging from news articles to science fiction. Each noun that occurred in the corpus was counted as an occurrence of each taxonomic class containing it.[3] For example, in Figure 1, an occurrence of the noun *dime* would be counted toward the frequency of DIME, COIN, and so forth. Formally,

$$\mathrm{freq}(c) \quad = \quad \sum_{n \in \mathrm{words}(c)} \mathrm{count}(n), \qquad (3)$$

where $\mathrm{words}(c)$ is the set of words subsumed by concept c. Concept probabilities were computed simply as relative frequency:

$$\hat{\mathrm{p}}(c) \quad = \quad \frac{\mathrm{freq}(c)}{N}, \qquad (4)$$

dimes are both small, round, metallic, and so on. These features are captured implicitly by the taxonomy in categorizing NICKEL and DIME as subordinates of COIN.

[2] *Concept* as used here refers to what Miller et al. [1990] call a *synset*, essentially a node in the taxonomy.

[3] Plural nouns counted as instances of their singular forms.

where N was the total number of nouns observed (excluding those not subsumed by any WordNet class, of course).

3.2 Task

Although there is no standard way to evaluate computational measures of semantic similarity, one reasonable way to judge would seem to be agreement with human similarity ratings. This can be assessed by using a computational similarity measure to rate the similarity of a set of word pairs, and looking at how well its ratings correlate with human ratings of the same pairs.

An experiment by Miller and Charles [1991] provided appropriate human subject data for the task. In their study, 38 undergraduate subjects were given 30 pairs of nouns that were chosen to cover high, intermediate, and low levels of similarity (as determined using a previous study [Rubenstein and Goodenough, 1965]), and asked to rate "similarity of meaning" for each pair on a scale from 0 (no similarity) to 4 (perfect synonymy). The average rating for each pair thus represents a good estimate of how similar the two words are, according to human judgments.

In order to get a baseline for comparison, I replicated Miller and Charles's experiment, giving ten subjects the same 30 noun pairs. The subjects were all computer science graduate students or postdocs at the University of Pennsylvania, and the instructions were exactly the same as used by Miller and Charles, the main difference being that in this replication the subjects completed the questionnaire by electronic mail (though they were instructed to complete the whole thing in a single uninterrupted sitting). Five subjects received the list of word pairs in a random order, and the other five received the list in the reverse order. The correlation between the Miller and Charles mean ratings and the mean ratings in my replication was .96, quite close to the .97 correlation that Miller and Charles obtained between their results and the ratings determined by the earlier study.

For each subject in my replication, I computed how well his or her ratings correlated with the Miller and Charles ratings. The average correlation over the 10 subjects was $r = 0.8848$, with a standard deviation of 0.08.[4] This value represents an upper bound on what one should expect from a computational attempt to perform the same task.

For purposes of evaluation, three computational similarity measures were used. The first is the similarity measurement using information content proposed in the previous section. The second is a variant on the edge counting method, converting it from distance to similarity by subtracting the path length from the maximum possible path length:

$$\mathrm{sim}_{\mathrm{edge}}(w_1, w_2) = (2 \times \textsc{max}) - \left[\min_{c_1, c_2} \mathrm{len}(c_1, c_2)\right] \quad (5)$$

where c_1 ranges over $s(w_1)$, c_2 ranges over $s(w_2)$, MAX is the maximum depth of the taxonomy, and $\mathrm{len}(c_1, c_2)$ is the length of the shortest path from c_1 to c_2 . (Recall that $s(w)$ denotes the set of concepts in the taxonomy that represent senses of word w.) Note that the conversion from a distance to a similarity can be viewed as an expository convenience, and does not affect the evaluation: although the sign of the correlation coefficient changes from positive to negative, its magnitude turns out to be just the same regardless of whether or not the minimum path length is subtracted from $(2 \times \textsc{max})$.

[4] Inter-subject correlation in the replication, estimated using leaving-one-out resampling [Weiss and Kulikowski, 1991], was $r = .9026$, stdev $= 0.07$.

Similarity method	Correlation
Human judgments (replication)	$r = .9015$
Information content	$r = .7911$
Probability	$r = .6671$
Edge counting	$r = .6645$

Table 1: Summary of experimental results.

The third point of comparison is a measure that simply uses the probability of a concept, rather than the information content:

$$\mathrm{sim}_{\mathrm{p}(c)}(c_1, c_2) = \max_{c \in S(c_1, c_2)} [1 - \mathrm{p}(c)] \quad (6)$$

$$\mathrm{sim}_{\mathrm{p}(c)}(w_1, w_2) = \max_{c_1, c_2} \left[\mathrm{sim}_{\mathrm{p}(c)}(c_1, c_2)\right], \quad (7)$$

where c_1 ranges over $s(w_1)$ and c_2 ranges over $s(w_2)$ in (7). Again, the difference between maximizing $1 - \mathrm{p}(c)$ and minimizing $\mathrm{p}(c)$ turns out not to affect the magnitude of the correlation. It simply ensures that the value can be interpreted as a similarity value, with high values indicating similar words.

3.3 Results

Table 1 summarizes the experimental results, giving the correlation between the similarity ratings and the mean ratings reported by Miller and Charles. Note that, owing to a noun missing from the WordNet taxonomy, it was only possible to obtain computational similarity ratings for 28 of the 30 noun pairs; hence the proper point of comparison for human judgments is not the correlation over all 30 items ($r = .8848$), but rather the correlation over the 28 included pairs ($r = .9015$). The similarity ratings by item are given in Table 3.

3.4 Discussion

The experimental results in the previous section suggest that measuring semantic similarity using information content provides quite reasonable results, significantly better than the traditional method of simply counting the number of intervening IS-A links.

The measure is not without its problems, however. One problem is that, like simple edge counting, the measure sometimes produces spuriously high similarity measures for words on the basis of inappropriate word senses. For example, Table 2 shows the word similarity for several words with *tobacco*. *Tobacco* and *alcohol* are similar, both being drugs, and *tobacco* and *sugar* are less similar, though not entirely dissimilar, since both can be classified as substances. The problem arises, however, in the similarity rating for *tobacco* with *horse*: the word

n1	n2	sim(n1,n2)	class
tobacco	alcohol	7.63	DRUG
tobacco	sugar	3.56	SUBSTANCE
tobacco	horse	8.26	NARCOTIC

Table 2: Similarity with *tobacco* computed by maximizing information content

horse can be used as a slang term for *heroin*, and as a result information-based similarity is maximized, and path length minimized, when the two words are both categorized as narcotics. This is contrary to intuition.

Cases like this are probably relatively rare. However, the example illustrates a more general concern: in measuring similarity between words, it is really the relationship among word *senses* that matters, and a similarity measure should be able to take this into account.

In the absence of a reliable algorithm for choosing the appropriate word senses, the most straightforward way to do so in the information-based setting is to consider *all* concepts to which both nouns belong rather than taking just the single maximally informative class. This suggests redefining similarity as follows:

$$\mathrm{sim}(c_1, c_2) = \sum_i \alpha(c_i)[-\log \mathrm{p}(c_i)], \qquad (8)$$

where $\{c_i\}$ is the set of concepts dominating both c_1 and c_2, as before, and $\sum_i \alpha(c_i) = 1$. This measure of similarity takes more information into account than the previous one: rather than relying on the single concept with *maximum* information content, it allows *each* class to contribute information content according to the value of $\alpha(c_i)$. Intuitively, these α values measure relevance — for example, $\alpha(\text{NARCOTIC})$ might be low in general usage but high in the context of a newspaper article about drug dealers. In work on resolving syntactic ambiguity using semantic information [Resnik, 1993b], I have found that local syntactic information can be used successfully to set values for the α.

4 Related Work

Although the counting of edges in IS-A taxonomies seems to be something many people have tried, there seem to be few published descriptions of attempts to directly evaluate the effectiveness of this method. A number of researchers have attempted to make use of conceptual distance in information retrieval. For example, Rada et al. [1989; 1989] and Lee et al. [1993] report experiments using conceptual distance, implemented using the edge counting metric, as the basis for ranking documents by their similarity to a query. Sussna [1993] uses semantic relatedness measured with WordNet in word sense disambiguation, defining a measure of distance that weights different types of links and also explicitly takes depth in the taxonomy into account.

The most relevant related work appears in an unpublished manuscript by Leacock and Chodorow [1994]. They have defined a measure resembling information content, but using the normalized path length between the two concepts being compared rather than the probability of a subsuming concept. Specifically, they define

$$\mathrm{sim}_{\mathrm{ndist}}(w_1, w_2) = -\log \left[\frac{\min_{c_1, c_2} \mathrm{len}(c_1, c_2)}{(2 \times \mathrm{MAX})} \right]. \qquad (9)$$

(The notation above is the same as for equation (5).) In addition to this definition, they also include several special cases, most notably to avoid infinite similarity when c_1 and c_2 are exact synonyms and thus have a path length of 0. Leacock and Chodorow have experimented with this measure and the information content measure described here in the context of word sense disambiguation, and found that they yield roughly similar results. More significantly, I recently implemented their method and tested it on the task reported in the previous section, and found that it actually outperforms the information-based measure. This led me to do a followup experiment using a different and larger set of noun pairs, and in the followup study the information-based measure performed better.[5] The relationship between the two algorithms will thus require further study. For now, however, what seems most significant is that both approaches take the form of a log-based (and hence information-like) measure, as originally proposed in [Resnik, 1993a].

Finally, in the context of current research in computational linguistics, the approach to semantic similarity taken here can be viewed as a hybrid, combining corpus-based statistical methods with knowledge-based taxonomic information. The use of corpus statistics alone in evaluating word similarity — without prior taxonomic knowledge — is currently an active area of research in the natural language community. This is largely a reaction to sparse data problems in training statistical language models: it is difficult to come up with an accurate statistical characterization of the behavior of words that have been encountered few times or not at all. Word similarity appears to be one promising way to solve the problem: the behavior of a word is approximated by smoothing its observed behavior together with the behavior of words to which it is similar. For example, a speech recognizer that has never seen the phrase *ate a peach* can still conclude that *John ate a peach* is a reasonable sequence of words in English if it has seen other sentences like *Mary ate a pear* and knows that *peach* and *pear* have similar behavior.

The literature on corpus-based determination of word similarity has recently been growing by leaps and bounds, and is too extensive to discuss in detail here (for a review, see [Resnik, 1993a]), but most approaches to the problem share a common assumption: semantically similar words have similar distributional behavior in a corpus. Using this assumption, it is common to treat the words that co-occur near a word as constituting features, and to compute word similarity in terms of how similar their feature sets are. As in information retrieval, the "feature" representation of a word often

[5] In the followup study, I used netnews archives to gather highly frequent nouns within related topic areas, and then selected noun pairings at random, in order to avoid biasing the followup study in favor of either algorithm.

takes the form of a vector, with the similarity computation amounting to a computation of distance in a highly multidimensional space. Given a distance measure, it is not uncommon to derive word classes by hierarchical clustering. A difficulty with most distributional methods, however, is how the measure of similarity (or distance) is to be interpreted. Although word classes resulting from distributional clustering are often described as "semantic," they often capture syntactic, pragmatic, or stylistic factors as well.

5 Conclusions

This paper has presented a new measure of semantic similarity in an IS-A taxonomy, based on the notion of information content. Experimental evaluation was performed using a large, independently constructed corpus, an independently constructed taxonomy, and previously existing human subject data. The results suggest that the measure performs encouragingly well (a correlation of $r = 0.79$ with a benchmark set of human similarity judgments, against an upper bound of $r = 0.90$ for human subjects performing the same task), and significantly better than the traditional edge counting approach ($r = 0.66$).

In ongoing work, I am currently exploring the application of taxonomically-based semantic similarity in the disambiguation of word senses [Resnik, 1995]. The idea behind the approach is that when polysemous words appear together, the appropriate word senses to assign are often those that share elements of meaning. Thus *doctor* can refer to either a Ph.D. or an M.D., and *nurse* can signify either a health professional or someone who takes care of small children; but when *doctor* and *nurse* are seen together, the Ph.D. sense and the childcare sense go by the wayside. In a widely known paper, Lesk [1986] exploits dictionary definitions to identify shared elements of meaning — for example, in the Collins COBUILD Dictionary [Sinclair (ed.), 1987], the word *ill* can be found in the definitions of the correct senses. More recently, Sussna [1993] has explored using similarity of word senses based on WordNet for the same purpose. The work I am pursuing is similar in spirit to Sussna's approach, although the disambiguation algorithm and the similarity measure differ substantially.

References

[Collins and Loftus, 1975] A. Collins and E. Loftus. A spreading activation theory of semantic processing. *Psychological Review*, 82:407–428, 1975.

[Francis and Kučera, 1982] W. N. Francis and H. Kučera. *Frequency Analysis of English Usage: Lexicon and Grammar.* Houghton Mifflin, 1982.

[Leacock and Chodorow, 1994] Claudia Leacock and Martin Chodorow. Filling in a sparse training space for word sense identification. ms., March 1994.

[Lee *et al.*, 1993] Joon Ho Lee, Myoung Ho Kim, and Yoon Joon Lee. Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of Documentation*, 49(2):188–207, June 1993.

[Lesk, 1986] Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 1986 SIGDOC Conference*, pages 24–26, 1986.

[Miller and Charles, 1991] George A. Miller and Walter G. Charles. Contextual correlates of semantic similarity. *Language and Cognitive Processes*, 6(1):1–28, 1991.

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

[Quillian, 1968] M. Ross Quillian. Semantic memory. In M. Minsky, editor, *Semantic Information Processing.* MIT Press, Cambridge, MA, 1968.

[Rada and Bicknell, 1989] Roy Rada and Ellen Bicknell. Ranking documents with a thesaurus. *JASIS*, 40(5):304–310, September 1989.

[Rada *et al.*, 1989] Roy Rada, Hafedh Mili, Ellen Bicknell, and Maria Blettner. Development and application of a metric on semantic nets. *IEEE Transaction on Systems, Man, and Cybernetics*, 19(1):17–30, February 1989.

[Resnik, 1993a] Philip Resnik. *Selection and Information: A Class-Based Approach to Lexical Relationships.* PhD thesis, University of Pennsylvania, December 1993.

[Resnik, 1993b] Philip Resnik. Semantic classes and syntactic ambiguity. In *Proceedings of the 1993 ARPA Human Language Technology Workshop.* Morgan Kaufmann, March 1993.

[Resnik, 1995] Philip Resnik. Disambiguating noun groupings with respect to WordNet senses. In *Third Workshop on Very Large Corpora.* Association for Computational Linguistics, 1995.

[Ross, 1976] Sheldon Ross. *A First Course in Probability.* Macmillan, 1976.

[Rubenstein and Goodenough, 1965] Herbert Rubenstein and John Goodenough. Contextual correlates of synonymy. *CACM*, 8(10):627–633, October 1965.

[Sinclair (ed.), 1987] John Sinclair (ed.). *Collins COBUILD English Language Dictionary.* Collins: London, 1987.

[Sussna, 1993] Michael Sussna. Word sense disambiguation for free-text indexing using a massive semantic network. In *Proceedings of the Second International Conference on Information and Knowledge Management (CIKM-93)*, Arlington, Virginia, 1993.

[Tversky, 1977] A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.

[Weiss and Kulikowski, 1991] Sholom M. Weiss and Casimir A. Kulikowski. *Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems.* Morgan Kaufmann, San Mateo, CA, 1991.

Word Pair		Miller and Charles means	Replication means	sim	$\mathrm{sim}_{\mathrm{edge}}$	$\mathrm{sim}_{\mathrm{p}(c)}$
car	automobile	3.92	3.9	8.0411	30	0.9962
gem	jewel	3.84	3.5	14.9286	30	1.0000
journey	voyage	3.84	3.5	6.7537	29	0.9907
boy	lad	3.76	3.5	8.4240	29	0.9971
coast	shore	3.70	3.5	10.8076	29	0.9994
asylum	madhouse	3.61	3.6	15.6656	29	1.0000
magician	wizard	3.50	3.5	13.6656	30	0.9999
midday	noon	3.42	3.6	12.3925	30	0.9998
furnace	stove	3.11	2.6	1.7135	23	0.6951
food	fruit	3.08	2.1	5.0076	27	0.9689
bird	cock	3.05	2.2	9.3139	29	0.9984
bird	crane	2.97	2.1	9.3139	27	0.9984
tool	implement	2.95	3.4	6.0787	29	0.9852
brother	monk	2.82	2.4	2.9683	24	0.8722
crane	implement	1.68	0.3	2.9683	24	0.8722
lad	brother	1.66	1.2	2.9355	26	0.8693
journey	car	1.16	0.7	0.0000	0	0.0000
monk	oracle	1.10	0.8	2.9683	24	0.8722
food	rooster	0.89	1.1	1.0105	18	0.5036
coast	hill	0.87	0.7	6.2344	26	0.9867
forest	graveyard	0.84	0.6	0.0000	0	0.0000
monk	slave	0.55	0.7	2.9683	27	0.8722
coast	forest	0.42	0.6	0.0000	0	0.0000
lad	wizard	0.42	0.7	2.9683	26	0.8722
chord	smile	0.13	0.1	2.3544	20	0.8044
glass	magician	0.11	0.1	1.0105	22	0.5036
noon	string	0.08	0.0	0.0000	0	0.0000
rooster	voyage	0.08	0.0	0.0000	0	0.0000

Table 3: Semantic similarity by item.