Similarity, Comparability and Analogy in WordNet:

Squaring the Analogical Circle with *Mondrian*

Tony Veale
School of Computer Science and Informatics
University College Dublin, Ireland.
Tony.Veale@UCD.ie

Mourad el Moueddeb
School of Computer Science and Informatics
University College Dublin, Ireland.
meddebmourad@gmail.com

Abstract

Similarity and comparability are complementary notions that are easy to confuse and difficult to tease apart. Semantic similarity makes comparison meaningful, while comparison is often the pragmatic means through which similarity is perceived and constructed. To date, WordNet has found wide application as a basis for modeling and measuring semantic similarity, but is lacking as a basis for well-formed comparability judgments. In this paper we describe a corpus-based approach to learning a model of sensible comparability, and show how this pragmatic model can be integrated with a WordNet-based notion of semantic similarity. We go on to show how this model of comparability, called *Mondrian*, provides a convenient and efficient means of supporting simple relational analogies between WordNet terms, and how analogies of this kind can be used for the automatic acquisition of enriching relational knowledge in WordNet.

1 Introduction

WordNet is much more than a large list of words, though it is often conveniently used as such in computational contexts. Indeed, it is also much more than an electronic dictionary, since it attempts to place an ontological order onto its inventory of word senses (Fellbaum, 1998). This organization, which is best realized in the taxonomic ordering of noun-senses as a relatively deep IS-A hierarchy, allows WordNet to be used as a lightweight knowledge representation in natural-language processing (NLP) and Artificial Intelligence (AI) systems (e.g., Veale, 2004). Despite any misgivings one might have about viewing WordNet as an ontological model of the world, this taxonomic organization allows WordNet to be used as a robust and efficient basis for semantic similarity judgments that are broadly consistent with human intuition (Miller and Charles, 1991).

The biggest advantage of this approach to similarity – and the reason it has been so widely embraced in NLP/AI applications – is that WordNet can be used to provide a numeric similarity judgment for any two terms that one cares to provide, no matter how dissimilar or oddly-paired they may seem (see Budanitsky and Hirst, 2006). Thus, whether one is comparing prawns and protons or galaxies and footballs, WordNet can be used to provide a reasonably sensible measure of their semantic similarity. These measures are relative, of course, so that one can know that protons are much more similar to electrons than they are to crustaceans. If used sensibly, with meaningful thresholds and cut-offs, WordNet-based similarity measures can play a vital role in many different NLP contexts.

The biggest *disadvantage* to this approach is also its biggest advantage: WordNet can be used to provide a numeric similarity judgment for any two terms at all, no matter how silly the comparison may seem. The space of meaningful comparisons is much smaller than the space of possible comparisons for which WordNet can provide a non-zero similarity judgment. This is fine in some contexts, in which we can trust the client application to only seek similarity measures for term pairs it has good reason to compare. In others, however, a strong semantic similarity score may be used to imply the comparability of terms that no human would ever sensibly compare.

Comparisons between ideas with a high semantic similarity are often fatuous, while comparisons between ideas with low semantic similarity can be highly insightful if used as the basis for a revealing metaphor or analogy (e.g., as when we compare the densely packed protons in an atom's nucleus to the plums in a fruit pudding). From the perspective of analogy *interpretation*, WordNet-based similarity offers limited insights into why two relational structures are comparable (see Veale, 2004, 2006; Turney, 2005), though a numeric measure of semantic similarity does go some way toward quantifying the creative tension exhibited by an analogy (e.g., the low score for plums and protons indicates a considerable analogical leap). In contrast, WordNet-based similarity offers nothing at all to the process of analogy *generation*, since a measure that assigns non-zero similarity to so many term-pairs can hardly be used to pick out a selective few for analogical consideration. A robust measure that is not selective in what it is willing to compare cannot then be selective in the suggestions it makes to a sensitive comparison mechanism like creative analogy generation.

1.1 Goals and Structure of this Paper

In this paper we seek to temper the broadness of WordNet-based semantic similarity to produce a more selective and pragmatically-guided measure of semantic *comparability*. In section 2 we review related work and ideas in re-making similarity as a pragmatic, corpus-based measure, before describing our own approach in this vein in section 3. Section 4 presents the *Mondrian* system, which uses comparability to support analogical reasoning, and in turn uses analogical reasoning (as supported by comparability) to support further knowledge-acquisition. In section 5 we then show how this measure can be used in a process of targeted knowledge-acquisition for enriching WordNet and other lexical ontologies with relational content. We conclude with a summary and some closing remarks in section 6.

2 Related Work and Ideas

Since WordNet organizes its noun-senses according to a hierarchical system of categories, this permits measures of semantic similarity to be operationalized in terms of the categories that are shared by two given terms. Measures differ in how these shared categories are exploited (see Budanitsky and Hirst for a menu of different possibilities): one might consider the minimum link-distance that must be traversed to find the most specific common category of two senses, or the size and generality of the shared categorization (where specific categories imply greater similarity than more general ones), or the *information content* of the categories concerned (e.g., it is more informative to say that two terms denote *mammals* than it is to say that they both denote *animals*). Resnick (1995) looks outside WordNet, to representative text corpora, to determine the extrinsic information content of a term like *animal* or *mammal*, while Seco, Veale and Hayes (2004) use WordNet itself as an intrinsic basis for determining information content, with comparable (but more convenient) results.

But one does not need a curated knowledge-source like WordNet to make sensible similarity judgments. The distributional hypothesis suggests that two words are similar to the extent that they are used in similar contexts (and co-texts) and with the same, or similar, lexical associations (see Weeds and Weir, 2005). Distributional similarity can cut across pre-defined taxonomic structures, to better reflect the dynamics of how words and concepts are actually used in context. For instance, the words "knife", "stove", "tent" and "backpack" are all used in camping contexts, and collectively represent the ad-hoc category *things one takes on a camping trip* (see Barsalou, 1983). An approach based on WordNet alone would be quite inadequate to the task of providing "more words like these" (such as "fishing-pole", "poncho" and "boots") because the context shared by these words, and which makes these words contextually similar, is not shared by WordNet.

The distributional approach can also be ratcheted up to the higher level of relational similarity required for analogical reasoning, in which one compares not terms but term pairs, each pair (such as *jury:verdict* or *courier:package*) representing an implicit relation (such as *delivers*) that must be matched. WordNet-based semantic similarity can only get one so far with these SAT-style analogical problems; Veale (2004) reports results in the 38-44% range using WordNet on a collection of 374 test analogies provided by Peter Turney and Michael Littman, while Turney himself (2005) uses a distributional approach to achieve higher scores approaching human levels of competence on the same test data. Turney uses an approach dubbed *Latent Relational Analysis* (LRA) in which a vector space of distributional features is derived from corpora or web text, and then smoothed using singular value decomposition (SVD).

Though convincing on SAT analogies and other interpretative/evaluative tasks, LRA is not a generative mechanism that can be used to suggest comparable relationships for given term pairs. That is, while it can provide strong results when evaluating comparisons posed by others, it is not (yet) a mechanism that can pose its own comparisons or suggest its own metaphors and analogies. In the sections to follow, we present a simpler corpus-based model of comparison that acquires knowledge of what terms can meaningfully be compared from textual evidence.

3 Learning to Compare from Corpora

WordNet-based similarity measures are decidedly semantic and objective, uninfluenced as they are by more subjective, pragmatic considerations. Any given measurement will implicate just a small number of static category structures, such as *mammal* when comparing *cats* and *dogs*, or *vehicle* when comparing *cars* and *buses*. In contrast, distributed corpus-based approaches implicitly capture the myriad contexts in which we experience two terms/ideas in similar ways. For instance, *pirates, astronauts* and *cowboys* are all semantically similar by virtue of being *human beings*, but are pragmatically similar for a variety of tacit cultural reasons, not least because they represent dashing heroic types that make for "cool" central characters in movies and books, while also making for "cool" costumes on Halloween. The distributed approach is successful because we cannot hope to articulate all the reasons why two terms are pragmatically comparable, much less express these reasons as static category structures in a system like WordNet.

Yet, we desire a representation of similarity that explicitly links terms that are considered comparable within the representation. Recall that the space of pragmatically comparable terms is not the same as the space of semantically similar terms. Since humans only meaningfully compare a tiny subset of the terms that are semantically similar (i.e., that have a non-zero similarity score), we can represent this comparability space as a sparse matrix. This will allow the evaluation of inter-term similarity to be modelled as an efficient look-up of a given cell in the matrix, while the generation of similar terms can likewise be modeled as a look up of the entire (albeit sparse) row corresponding to a given comparison target.

Figure 1 presents a snapshot of such a sparse comparability matrix. Every row corresponds to a different term, as does every column, while the numeric value in each cell (0 … 1) corresponds to the semantic similarity of the corresponding row/column terms. The similarity of a term to itself is 1.0, while the semantic similarity of all other terms is pre-computed using any one of the WordNet measures (at the developer's discretion) described in Budanitsky and Hirst (2006). As such, the numeric values in Figure 1 are all *semantic* similarity scores. But while semantics, not pragmatics, determines these numeric values, the choice of cells to fill is entirely determined by pragmatic factors. Of the 64 cells shown in Figure 1, 42 (or 65%) contain zeros, not because the corresponding term pairs have no semantic similarity, but because they are not deemed to be comparable. The matrix is large, but very sparse.

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**…**
A	1.0	0	0	0	0	0	0	0	…
B	0	1.0	0.7	0	0	0.2	0.3	0	…
C	0	0.7	1.0	0	0.4	0	0	0.15	…
D	0	0	0	1.0	0	0	0	0.2	…
E	0	0	0.4	0	1.0	0	0	0	…
F	0	0.2	0	0	0	1.0	0	0.65	…
G	0	0.3	0	0	0	0	1.0	0	…
H	0	0	0.15	0.2	0	0.65	0	1.0	…
…	…	…	…	…	…	…	…	…	…

Figure 1. A sparse comparability matrix. Cells contain WordNet-based similarity scores, and are filled on the basis of explicit linguistic evidence.

A given cell holds a non-zero value if the corresponding terms have a non-zero similarity score *and* there is linguistic evidence that the terms are comparable. A myriad subtle factors influence whether one term is comparable to another: are they defined at the same level of specification? Are they inter-changeable in some respect? Do they denote cultural counterparts of one another? We short-circuit this complexity by noting that if any of these (or similar) criteria hold, then we should observe that the terms will tend to be used in the same linguistic contexts by speakers. More specifically, we should observe that the terms will be clustered into the same ad-hoc sets. For instance, we can observe coordinations such as "*scientists and artists*", "*robots and clones*", "*imams and priests*", "*mosques and synagogues*" and "*pirates and cowboys*". Sets like these indicate that a speaker believes the given elements to belong to the same semantic/pragmatic category, even if, in a resource like WordNet, the elements do not share a direct hypernym.

Set-building linguistic constructs such as coordination provide evidence of fine-grained pragmatically-motivated categorizations that a resource like WordNet cannot. Such constructions provide the basis for *Google Sets*, an online tool that allows Google to perform on-demand set completion (see Tong and Dean, 2008). Given a sampling of terms, such as “hamburger” and “pizza”, *Google Sets* can infer the implicit category and flesh out the set with additional members such as “taco” and “hotdog”. Google also uses this set completion functionality in its online spreadsheet (part of *Google Docs*), allowing a user to specify some values in a column before asking Google to automatically fill it with other values from the same implicit category.

We use the coordination construction to find evidence that two terms can reside in the same ad-hoc set, and thus the same pragmatic category. Google’s database of web n-grams (see Brants and Franz, 2006) provides a very large corpus from which to harvest these coordinations. In effect, we harvest all plural coordinations of the form “*Xs and Ys*” (where X and Y are common nouns, as in “*cats and dogs*” or “*zoos and circuses*”) and all singular proper coordinations of the form “*X and Y*” (where X and Y are capitalized proper nouns, as in “*Paris and London*” or “*Zeus and Jupiter*”). For each pairing X and Y, we calculate a WordNet-based similarity score and populate the comparability matrix accordingly. In all, the n-grams yield coordinations involving 35,019 different terms, producing a matrix with 35,019 rows and 35,019 columns.

In practice, the matrix is sparse and only a small fraction of these cells are populated. In fact, just 1,363,184 cells have non-zero values, giving the matrix a density of just 0.1%. This matrix is thus compact enough to store in memory, yet contains all of the most plausible comparisons a system is ever likely to consider.

4 Analogical Comparisons in Mondrian

Broadly speaking, coordination patterns provide two different kinds of associations that can be useful for making and understanding comparisons: *substitutive* associations and *contiguous* associations. Substitutive associations are those that suggest that one term might be used as a substitution for another in a simile, metaphor or analogy. For instance, the coordinations “*priests and scientists*”, “*scientists and artists*”, “*artists and anarchists*”, “*churches and mosques*”, “*angels and demons*”, “*spires and minarets*” and “*rituals and experiments*” all seem to suggest cross-domain equivalences while suggesting alternate ways of looking at a given term / concept. In some contexts then. it might be meaningful (if only figuratively) to view scientists as priests, or experiments as rituals, or artists as anarchists. While substitutive associations often cross domain boundaries, *contiguous* associations relate a term to another term in the same domain. For instance, the coordinations “*mosques and minarets*”, “*imams and mosques*”, “*artists and studios*”, “*priests and sacrifices*” and “*scientists and laboratories*” each express a kind of semantic relatedness rather than strict semantic similarity. Mosques are not similar to imams (at least not in the way that imams are similar to priests), but they are highly related to imams. As such, contiguous associations are better suited to the generation of metonymies than they are to similes, metaphors or analogies.

The comparability matrix contains both kinds of association in abundance; since both arise from the same coordination construction, the matrix does not (and cannot) distinguish the substitutive from the contiguous variety. However, as a rough heuristic, substitutive associations will exhibit high semantic similarity scores (e.g., > .6), while contiguous associations will exhibit much lower similarity scores (e.g., < .25). Figure 2 highlights elements of the Figure 1 matrix that exhibit substitutability because of high similarity (shown with bold lines) and contiguity because of low similarity (shown with dashed lines).

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**…**
A	1.0	0	0	0	0	0	0	0	…
B	0	1.0	0.7	0	0	0.2	0.3	0	…
C	0	0.7	1.0	0	0.4	0	0	0.15	…
D	0	0	0	1.0	0	0	0	0.2	…
E	0	0	0.4	0	1.0	0	0	0	…
F	0	0.2	0	0	0	1.0	0	0.65	…
G	0	0.3	0	0	0	0	1.0	0	…
H	0	0	0.15	0.2	0	0.65	0	1.0	…
…	…	…	…	…	…	…	…	…	…

Figure 2. Analogical connections in a comparability matrix. B is highly similar to C, just as H is to F, while B is related to F and C to H.

In Figure 2 we see that B and C have strong similarity (= 0.7), and are seen as comparable because of the linguistic evidence “*Bs and Cs*”. F and H are likewise strongly similar (= 0.65). In turn, B and F are weakly similar (= 0.2), as are C and H (= 0.15), while the patterns "*Bs and Fs*" and "*Cs and Hs*" suggests that B and F, as well as C and H, are contiguous in the same domains. In other words, a squaring relationship holds between B, C, F and H: B is contiguous to F, which is strongly similar to H, which is contiguous to C, which is strongly similar to B. This associative square pattern is visible in Figure 2.

As shown by Veale and Keane (1997), this squaring pattern is capable of efficiently leveraging local relationships into global analogical structures. Moreover, the squaring approach, captured here in a model we name *Mondrian*, explicitly views analogical mapping as a data-mining process, in which small regularities of structure are identified in large masses of knowledge that are largely irrelevant to the given analogy. In other words, Mondrian views analogical reasoning as a problem of data-mining in the comparability matrix. Given a starting point B, say, Mondrian can seek out terms that complete the square C, F and H; likewise, given a contiguous relation B:F, Mondrian can find the corresponding relation C:H by examining all comparable terms B and F that also happen to be contiguous. In the most restricted, and efficient, cases, Mondrian can complete the analogical square B:F::C:?, as in the following examples:

priest : church	:: imam :	?	(A: mosque)
mosque : minaret	:: church :	?	(A: spire)
chef : recipe	:: scientist :	?	(A: formula)
school : bus	:: hospital :	?	(A: ambulance)

Analogical completion is not a deterministic process, and even the most constrained examples above can yield several competing answers (e.g., perhaps a tower or a steeple is a church's answer to a minaret). So it is necessary to rank potential analogies according to their overall similarity. For instance, we might estimate the quality of an analogy based on the quality of the substitutions it involves, as in the following measure:

$$\mathit{subst\_sim}(\mathrm{B{:}F{::}C{:}H}) = \mathit{sim}(\mathrm{B, C}) \times \mathit{sim}(\mathrm{F, H})$$

However, this *subst_sim* measure does not take into account the actual nature of the contiguous relations between B and F or C and H. For example, the *mosque:minaret::church:spire* analogy implicitly hinges on the fact that minarets are tall, slender parts of a mosque, while spires are likewise tall slender parts of a church. The *part_of* relation is not explicitly coded here, but the pragmatic comparability of minarets and spires means that they are largely interchangeable in many contexts. Nonetheless, without knowledge of the specific relationships between B and F or C and H, we cannot be sure that the analogy is sound. But whatever this relation happens to be, we can expect that if it imposes specific selectional preferences, then the relative similarity of B to F will be comparable to that of C to H. So while mosques relate to minarets in different ways than they relate to imams or mullahs, we can exploit the fact that they are also more similar to minarets than to imams or mullahs. An analogy can thus hinge on a higher-level equivalence of lower-level similarities, suggesting a measure of *balance* such as the following:

$$\mathit{balance}(\mathrm{B{:}F{::}C{:}H}) = \frac{\min(\mathrm{sim}(B,\ F),\ \mathrm{sim}(C,\ H))}{\max(\mathrm{sim}(B,\ F),\ \mathrm{sim}(C,\ H))}$$

In other words, if B relates to F in the same way that C relates to H (and we don't know the actual relation), and if this relation imposes specific semantic restrictions on its arguments, then we should expect F to be as similar to B as H is to C, and the balance factor above will be close to 1. An unbalanced analogy, in which B and F have a different relationship than C and H, will have a score closer to 0. Combining both measures, we can now judge the quality of analogy as follows:

$$\mathit{quality}(\mathrm{B{:}F{::}C{:}H}) = \mathit{subst\_sim}(\mathrm{B{:}F{::}C{:}H}) \times \mathit{balance}(\mathrm{B{:}F{::}C{:}H})$$

This *quality* measure uses only the contents of the comparability matrix as the basis for its judgments. This is a heuristic that often works, since in many cases, semantic relations can be differentiated by their similarity profiles. Nonetheless, to be sure that a proposed analogy is indeed sound, one needs to know the specific semantic relations involved, rather than just the similarity distributions of the terms they are used to relate.

In this regard, we have two options. In the first, we use the *quality* measure above to identify strong candidates for analogical squaring in the comparability matrix, and then use LRA or a similar technique to more rigorously evaluate these candidates from a relational standpoint. In other words, we can use *Mondrian* as a generative precursor to an evaluative technology like LRA, so that the combined system can both generate and evaluate its own relational analogies. In the second option, we attempt to acquire specific semantic relations for the unspecified contiguous pairings we find in the comparability matrix. As we shall see in the next section, we acquire these relations to support analogical reasoning, but analogical reasoning can itself be used to hasten and direct the acquisition of these relations.

5 Layered Knowledge Acquisition

The notion of a contiguous association is highly underspecified. Though the Google n-grams allow us to determine that *scientist* is contiguous with *laboratory, theory, experiment, research* and *grant*, and WordNet allows us to associate a specific similarity score with each pairing, a different semantic relationship holds in each case. It is necessary to do more than heuristically separate comparable terms into substitutive and contiguous groups, and to know the actual relationships (there may be many) that hold in each case.

Interestingly, though contiguous associations lack a specific semantics, they do at least tell us which associations are worthy of semantic description. That is, of all the term pairings one can imagine, contiguous associations indicate those that are most deserving of further elaboration. Contiguous associations can thus drive the process of knowledge acquisition, either in directing a human user's attention to associations that are likely to be important, or in directing the efforts of an automated approach to knowledge acquisition. A middle-ground approach is also tenable here, in which the system hypothesizes a set of candidate relationships for each contiguous association, before asking a human user to choose amongst this set of potential relations. This approach can also work well in a web-based setting, where anonymous contributors volunteer their time and knowledge in updating the system. In such a setting, users are not asked to suggest their own relationships (a request that can elicit an anarchic response) but to vet relationships that the system already considers plausible.

In this semi-automatic approach, knowledge can be acquired in successive layers. Given the contiguous association X:Y (gleaned from the 3-gram "*Xs and Ys*"), the system can look for 3-grams of the form "Xs preposition Ys". For instance, "*imams and mosques*" prompts one to find "*imams in mosques*", which yields the relation "in" as a linkage for *imam:mosque*. Likewise we find, "*priests in churches*", "*artists in studios*", "*scientists in laboratories*" and "*chefs in kitchens*", all comparable (and analogous) uses of the relation "in".

But prepositions like "in" are pseudo-relations at best: they are vague and highly polysemous, and we really desire an additional verb to lock down their relational meaning. Looking to the 3-grams again, this time focusing on patterns of the form "*verb preposition Xs*", we can identify more specific relations that apply to a given object, such as "*work in laboratories*", "*preach in mosques*" and "*cook in kitchens*". These *verb+preposition* combinations can then be presented to a human user as a menu of candidate relations to elaborate a pseudo-relation like "*imams in mosques*".

To summarize, unspecified contiguous associations can be automatically elaborated into preposition-based pseudo-relations by again mining the Google 3-grams. With the computer-guided input of human volunteers, these pseudo-relations can then be further elaborated into specific relations like *work_in* or *preach_in*. At this point, the system can use analogical reasoning to drive the acquisition of further relationships. So, once a system acquires *imam:work_in:mosque* and *imam:preach_in:mosque*, it can reason via the analogy *imam:mosque::priest:church* that the comparable relationships *priest:work_in:church* and *priest:preach_in:church* may also hold true. Conversely, if the system knows the relationship *priest:minister_in:church*, it can use the same analogy to suggest *imam:minister_in:mosque.* If a human user validates the analogical hypothesis *priest:work_in:church*, a system can use another analogy to suggest *teacher:work_in:school* (and conversely, *priest:teach_in:church*), and so on.

The key point here is that a lightweight approach to analogical reasoning, based only on pragmatic comparability and semantic similarity (outlined in section 4 as the *Mondrian* system), can support meaningful analogies even before a specific relational semantics is acquired. Analogy is both the chicken and the egg in this circular situation: lightweight analogies (the egg) drive the acquisition of specific knowledge (the chicken) that in turn supports acquisition of further knowledge in a virtuous cycle of analogy-driven hypothesis generation and validation.

6 Conclusions

WordNet offers a variety of different semantic relations to weave its word-senses together, though the weft is decidedly patchy in parts. Consequently, WordNet's richest resource by far is its structured hierarchy of noun-senses. This hierarchy underpins numeric judgments of semantic similarity that are robust and efficient, and which accord well with human intuitions. In this paper we have teased apart the related notions of similarity and comparability, and provided a convenient, corpus-based approach to determining which term-pairs can be sensibly grouped together and compared. The result is a comparability matrix whose structure is determined by lexical distribution patterns in corpora, and whose numeric content is determined by WordNet-based semantic similarity scores.

Analogy is a knowledge-hungry process, but one that plays a vital educational role in knowledge transfer among humans. As such, analogy can be used to extend the knowledge already possessed by a system (e.g., see Speer, Havasi and Lieberman, 2008), and if properly harnessed, analogy can drive a virtuous cycle of knowledge-acquisition and hypothesis generation. We have shown here how a comparability matrix populated with WordNet-based similarity judgments can provide a lightweight foundation for analogical reasoning, in the absence of a rich relational semantics. Furthermore, we have shown how this lightweight approach can drive a knowledge-acquisition process in a highly targeted fashion, so that an agent acquires precisely the kind of cross-domain knowledge that results in sound, well-structured analogies. By incorporating analogy into the acquisition process at so early a stage, we can ensure that the resulting knowledge-base is not just analogy-rich, but consistent and well-balanced.

The system described here, named *Mondrian* after its penchant for squaring patterns, is available as an on-line demo at *http://Afflatus.UCD.ie* under *Current Projects*. The full comparability matrix, as derived from Google n-grams, is available for browsing here, as are the many analogies that have been mined from this matrix. Ongoing and future research concerns the development of an editor, named *EdMond*, in which *Mondrian*'s contiguous associations can be elaborated through interactions with web-users in the structured fashion outlined in section 5.

References

Lawrence W. Barsalou. 1983. Ad hoc categories. Memory and Cognition, 11:211–227.

Thorsten Brants and Alex Franz. 2006. *Web 1T 5-gram Version 1*. Linguistic Data Consortium, Philadelphia.

Alexander Budanitsky and Graeme Hirst. 2006. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics,* 32(1):13-47.

Christiane Fellbaum (ed.). 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA.

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

Resnick, Philip. 1995. Using Information Content to Evaluate Semantic Similarity in a Taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence.*

Nuno Seco, Tony Veale and Jer Hayes, 2004. An Intrinsic Information Content Metric for Semantic Similarity in WordNet. Proceedings of ECAI'04, *the European Conference on Artificial Intelligence*.

Robert Speer, Catherine Havasi, and Henry Lieberman. 2008. AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge. *Proceedings of AAAI 2008, the 23rd AAAI conference on Artificial Intelligence*.

Simon Tong and Jeff Dean. 2008. System and methods for automatically creating lists. *US Patent 7,350,187* (assigned to Google; granted March 25, 2008).

Peter Turney. 2005. Measuring semantic similarity by latent relational analysis. *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 1136-1141.

Tony Veale and Mark T. Keane. 1997. The Competence of Sub-Optimal Structure Mapping on 'Hard' Analogies. *Proceedings of the 15th International Joint Conference on Artificial Intelligence.*

Tony Veale. 2004. WordNet sits the SAT: A knowledge-based approach to lexical analogy. *Proceedings of ECAI'04, the European Conference on Artificial Intelligence*, 606-612.

Tony Veale. 2006. Re-Representation and Creative Analogy: A Lexico-Semantic Perspective. *New Generation Computing* 24:223-240.

Julie Weeds and David Weir. 2005. Co-occurrence retrieval: A flexible framework for lexical distributional similarity. *Computational Linguistics*, 31(4):433–475.