Refining WordNet adjective dumbbells using intensity relations.

Vera Sheinman and **Takenobu Tokunaga**
vera@sheinman.org, take@cl.cs.titech.ac.jp
Tokyo Institute of Technology, Tokyo, 152-8552 Japan

Isaac Julien and **Peter Schulam** and **Christiane Fellbaum**
ijulien,pschulam,fellbaum@princeton.edu
Computer Science Department, Princeton University, New Jersey 08540, USA

Abstract

We propose a new semantic relation of intensity for gradable adjectives in WordNet and show its specific benefits for NLP. Intensity enriches the present, vague, *similar* relation with information on the degree with which different adjectives express a shared attribute. Using lexical-semantic patterns, we mine the Web for evidence of the relative strength of adjectives like "large", "huge" and "gigantic" with respect to their attribute ("size"). The pairwise orderings we derive allow us to construct scales on which the adjectives are located.

To represent the intensity relation among gradable adjectives in WordNet, we combine ordered scales with the current WordNet *dumbbells* based on the relation between a pair of central adjectives and a group of undifferentiated semantically *similar* adjectives. A new intensity relation links the adjectives in the dumbbells and their concurrent representation on the scale.

1 Introduction

A survey of publications on NLP work using WordNet shows that the more than $18,000$ adjective synsets are rarely part of a system, and many crosslingual wordnets do not include adjectives at all. This may be attributable to the role of adjectives as modifiers and carriers of arguably less essential information. But we conjecture that one principal reason for the current under-use is that the organization of adjectives in WordNet does not lend itself well to a clear determination of semantic similarity. The present work explores the semantics of scalar adjectives and outlines a novel way of representing such meanings in WordNet.

1.1 Adjectives in WordNet

WordNet originated as a model of human semantic memory. Specifically, it was designed to test then-current models of conceptual organization that supported a network structure (Collins and Quillian, 1969). Association data indicated that words expressing semantically similar concepts were stored in close proximity and strongly evoked one another. Thus, when presented with a stimulus word like "automobile", people overwhelmingly respond with "car"; the prevalent response to "celery" is "vegetable" and to "elephant", "trunk" (Moss and Older, 1996). Such data suggested the organization of words and concepts into a network structured around semantic relations like synonymy, meronymy (part–whole) and hyponymy (super/subordinates).

Most striking is the strong mutual association between members of antonymous adjective pairs like "wet–dry" and "dark–light", already discussed by (Deese, 1964) who noted that such pairs are acquired early by children. The *clang* association between antonymous adjectives might well be due to their high frequency and their shared contexts that indicate their common selectional restrictions. (Justeson and Katz, 1991) showed furthermore that members of an antonymous adjective pair co-occur in the same sentence far more often than chance would predict.

It seemed straightforward enough to represent the members of an antonym pair as opposite poles on an open-ended scale that encoded a particular attribute. But what about the many adjectives that are semantically similar to these adjectives yet are neither synonyms nor antonyms of a member of the pair?

(Gross et al., 1989) measured the time it took speakers to respond to questions like "Is small the opposite of large?", "Is miniature the opposite of large?" and "Is gigantic the opposite of miniature?" The first kind of question involved the members of an antonym pair and the latencies here were very short. The second kind of question involved one member of an antonym pair and an adjective that was similar to its antonym. People took measurably longer to affirm these questions. The third kind of question asked people's judgments about two adjectives that were each similar to one member of an antonym pair. In these cases, people either were hesitant to reply at all or they took a very long time to respond affirmatively.

These data inspired the representation of adjectives in WordNet by means of *dumbbells,* with antonyms as the centroids and semantically similar adjectives arranged in radial fashion around each antonym. Figure 1 depicts a schematic representation of a dumbbell.

1.2 Limitations of the Dumbbell Representation

While the dumbbells seemed well motivated psycholinguistically and distributionally, they do not lend themselves easily to Natural Language Processing and they stump systems designed to detect and quantify meaning similarity.

First, relatively few adjectives are interconnected, which limits path-based Word Sense Disambiguation systems to the small number of adjectives that are classified as being either antonyms or semantically similar in a given cluster. Second, within a cluster, all semantically similar adjectives are arranged equidistantly from a centroid. As a result, the path length between the centroid and all similar adjectives is always one and that between two similar adjectives is invariably two, with each path connected via the centroid. This lack of encoding of independent meaning distinctions among the *similar* adjectives suggests that they are all equally similar to the centroid, which is intuitively not the case. For example, both "titanic" and "capacious" are represented as being equally similar to "large", as are "subatomic" and "gnomish" to "small". Moreover, the meaning differences among the similars themselves, such "titanic", "capacious", "monstrous" and "gigantic" on the one hand, and "subatomic", "gnomish", "dinky" and "pocket-size" on the other hand, are not represented. Finally, many similar adjectives are in fact misclassified as members of a same cluster, whereas based on their selectional restrictions, they should in many cases be assigned to different clusters. Thus, "hulking" describes entities with physical properties, while a related similar adjective like "epic" typically modifies abstract concepts like events ("epic battle", "epic voyage"). Likewise, adjectives that are currently classified as being similar to "small", for example "pocket-size" and "elfin", differ in their selectional restrictions: the former can be applied to objects like books, whereas the latter typically modifies people.

Semantically, the relation of the centroids to the similar adjectives as well as that among the similar adjectives themselves is underspecified and expressed only indirectly via antonymy. A second relation, labeled *see also* links different dumbbells via a shared centroid adjective that has a different but related sense in each dumbbell. It is often difficult to discern a motivated distinction between the similar and the see also relations and hence, among the adjectives they connect.

1.3 Scalar Adjectives

Our focus here is on adjectives that possess scalar properties. (Bierwisch, 1989) notes that dimensional adjectives like "long", "short", "wide", "narrow", "new" and "old" express a particular value on a scale or dimension. For example, while both "ancient" and "old" fall on the same scale ("age"), their relative placement on the scale represents the fact that "ancient" expresses a more intense degree of "age" and "old".

Some dimensional scales lexicalize many points ("large-small"), while others express few points besides paired polar antonyms ("tall–short"). Note that the scales are open-ended, and a stronger or weaker degree of the underlying shared attribute can always be conceived of, even if it is not independently lexicalized.

We propose a re-organization of the subset of adjectives that express different values of a gradable property (Bierwisch, 1989; Kennedy, 2001) using the AdjScales method (Sheinman and Tokunaga, 2009). For a given attribute, we construct scales of adjectives ordered according to the intensity with which they encode a shared attribute. The ordering will be based on corpus data.

2 AdjScales

The AdjScales method orders a set of related adjectives on a single scale using the intensity relation, as in the example $tiny \rightarrow small \rightarrow$

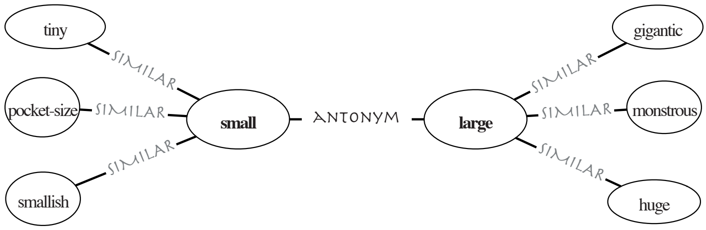

Figure 1: An illustration of WordNet's dumbbell structure.

$smallish \rightarrow large \rightarrow huge \rightarrow gigantic$.

The basic methodology of AdjScales is to extract patterns characterizing semantic relations from free text based on several word instances, and then use the extracted patterns for extraction of further instances of the relations of interest, or even for bootstrapping of additional patterns. Several techniques for extracting semantic similarity from corpora have been proposed.

Lexical-semantic patterns were first described by (Cruse, 1986), who notes that phrases like "xs such as ys" and "ys and other xs" identify x as a superordinate, or hypernym, of y. (Hearst, 1992) pioneered the identification and application of such phrases or patterns to the extraction of semantically related words from corpora as an efficient way to semi-automatically construct or enrich thesauri and ontologies. Her work was further extended by (Riloff and Jones, 1999; Chklovski and Pantel, 2004; Turney, 2008; Davidov and Rappoport, 2008; Snow et al., 2005).

Contextual or *distributional similarity* based approaches such as (Weeds and Weir, 2005; Lin, 1998) rely on the observation that words with similar meanings also share similar contexts. For instance, "rose" and "flower" constitute a hyponym-hypernym pair and thus one can expect some of the contexts of "rose" to appear in the same contexts of "flower"; differently put, semantically similar words are often mutually interchangeable.

Pattern-based extraction method identifies words that are *paradigmatically* related; approaches based on contextual similarity rely on the *syntagmatic* similarity of related words.

Both approaches to identifying semantically similar words should converge; automatically derived thesauri such as (Lin, 1998) show significant overlap with manual resources like WordNet. The AdjScales method exemplifies the phrase-based extraction approach.

AdjScales comprises two stages, preprocessing and scaling that are described in detail in (Sheinman and Tokunaga, 2009). The following section summarizes the method with an eye towards enriching adjectives in WordNet with intensity information.

2.1 Preprocessing: pattern extraction

The preprocessing step of the AdjScales handles extraction of patterns that later serve AdjScales for scaling of adjectives. Pattern extraction queries of the form "a $*$ b" are used, where a and b are seed words and "$*$" denotes a wildcard (zero to several words that may appear in its place). AdjScales extracts binary patterns of the form

$$p = [\mathrm{prefix_p} \quad x \quad \mathrm{infix_p} \quad y \quad \mathrm{postfix_p}]$$

from the snippets of the query results using a search engine, where x and y are slots for words or multiword expressions. A pattern p can be instantiated by a pair of words w_1, w_2 to result in a phrase "prefix$_p$ w_1 infix$_p$ w_2 postfix$_p$".

Let us consider an example pattern p_1 where $\mathrm{prefix_{p_1}} = \phi$, $\mathrm{infix_{p_1}} =$ "if not", and $\mathrm{postfix_{p_1}} = \phi$, if we instantiate it with the pair of words (large, gigantic) we will get a phrase $p_1(\mathrm{large,\ gigantic}) =$ "large if not gigantic".

If $p(w_1, w_2)$ appears in snippets that are returned by a search engine when querying it with a pattern-extraction-query, we refer to it as p is *supported-by* $(w_1,\ w_2)$. For the extraction purposes snippets are split into sentences and are cleaned from all kinds of punctuation. Up to this point, the notation and the method largely follow the work by (Davidov and Rappoport, 2008).

Differently from (Davidov and Rappoport, 2008) the seed word pairs for AdjScales are chosen in a supervised manner, so that $seed_2$ is more intense than $seed_1$. Consider, for instance the pair (“cold”, “frigid”), where “frigid” is more intense than “cold”. The relation *more-intense-than* is asymmetric. Therefore, AdjScales selects only the *asymmetric patterns* that are extracted consistently so that the less intense word in each supporting pair is only on the left side of the pattern (before the infix words) or so that the less intense word is only on the right side of the pattern (after the infix words). Unless all the supporting pairs of words share the same direction, the pattern is discarded. The former selected patterns are defined as *intense*, and the latter as *mild*.

AdjScales selects only the patterns supported by at least 3 seed pairs and requires a pattern instance by each supporting pair to repeat at least twice in the sentences extracted from the snippets to increase reliability. It also requires the patterns to be supported by adjectives describing different attributes (seed pairs should be selected accordingly). This constraint is important, because patterns that are supported by seeds that share the same attribute tend to appear in very specific contexts and are not useful for other attributes. For instance, [x even y amount] might be extracted while supported only by seeds sharing the “size” attribute, such as (“huge”, “astronomical”), (“large”, “huge”), (“tiny”, “infinitesimal”).

(Sheinman and Tokunaga, 2009) report on 16 English patterns that were extracted using this stage of the method. For the analysis of the English examples presented in this work, we did not reproduce the preprocessing stage, but used the 16 patterns reported in their work and augmented them with a set of 17 human constructed patterns. Table 1 lists all the patterns used in this work.

2.2 Scaling

For this step, we use AdjScales to process the dumbbell structure from WordNet to enrich it with intensity information. We process each one of the antonymous groups in the dumbell separately. For each pair (head-word, similar-adjective), we instantiate each pattern p in patterns that were extracted in the preprocessing stage to obtain phrases $s_1 = p$(head-word, similar-word) and $s_2 = p$(similar-word, head-word). We send $s1$ and $s2$ to a search engine as two separate queries and check whether $df^1(s_1) > weight \times df(s_2)$ and whether $df(s_1) > threshold$. The higher the values for the $threshold$[2] and $weight$[3] parameters, the more reliable are the results. If p is of the type *intense*, then a positive value is added to the similar-word’s score, otherwise if p is of the type *mild* a negative value is added. When all the patterns are tested, similar-words with positive values are classified as intense, while the similar-words with negative values are classified as mild. Words that score 0 are classified as *unconfirmed*. For each pair of words in each one of the subsets (mild and intense), the same procedure is repeated, creating further subsets of *mildest* words that have the most negative values within the mild subset, and *most intense* words for the words with the highest positive values within the intense subset.

After the two parts of the dumbbell are processed, they are unified into a single scale. The unification attempts to order the adjectives from the half of the dumbbell with the less frequent centroid (starting from the most intense to the mildest) to the more frequent side (starting from the mildest to the most intense). Adjectives of similar intensity are grouped together.

The adjectives in a final scale are then linked

Table 1: Intense and mild patterns. x and y represent adjectives so that x is more intense than y.

Intense Patterns
(is / are) x but not y
(is / are) very x y
extremely x y
not x (hardly / barely / let alone) y
x (but / yet / though) never y
x (but / yet / though) hardly y
x (even / perhaps) y
x (perhaps / and) even y
x (almost / no / if not / sometimes) y
Mild Patterns
if not y at least x
not y but x enough
not y (just / merely / only) x
not y not even x
not y but still very x
though not y (at least) x
y (very / unbelievably) x

[1]df represents *document frequency*.

[2]*threshold* regulates the number of pages returned by the search engine that is considered sufficient to trust the result, and it was set to 20 in this work.

[3]*weight* regulates the gap between s_1 over s_2 that is required to prefer one over the other, and it was set to 15 in this work.

from the original adjective synsets in a dumbbell as illustrated in Figure 2. The unconfirmed adjectives on both sides of the dumbbell remain unlinked to the final scale.

Examples of scales extracted by applying AdjScales to the dumbbells in WordNet include:

- destitute → poor → broke → rich → loaded
- ice-cold → cold → chilly → tepid → warm → hot → (torrid, scorching)
- filthy → dirty → dingy → clean → spotless

2.3 Using the Web as a corpus

AdjScales is designed to extract fine-grained distinctions, and the relative sparseness of the lexical-semantic patterns with many of the less frequent adjectives mandates the use of a very large corpus. Second, the method requires a large, domain-independent corpus that reflects current language use and accommodates ever-shifting changes in meaning across diverse speaker communities. In particular, words with a strong flavoring tend to acquire a weaker connotation and reduced intensity with frequent use. While the Web has sometimes been criticized for being unreliable and unstable (Kilgarriff, 2007), it is a logical choice for our work, as corpora constructed for research purposes tend to be small (MASC), unbalanced (PropBank), and not representative of current language use (Brown Corpus, BNC). Finally, the method relies on the availability of a search engine that supports proximity search, provides an estimated number of page hits and snippets of the relevant Web pages.

3 Related Work

VerbOcean VerbOcean (Chklovski and Pantel, 2004) is a pattern-based approach to extracting fine-grained semantic relations among verbs from the Web. In contrast to other approaches, the patterns in VerbOcean are manually grammatically enhanced to be selective for verbs (see also (Fellbaum, 2002)). VerbOcean accounts for the frequency of the verbs as well as the frequency of the patterns themselves. Furthermore, VerbOcean distinguishes between symmetric and asymmetric semantic relations and utilizes this distinction. VerbOcean identifies six semantic relations among verbs, including *strength*, a subtype of *similarity*.

Strength, which is similar to *intensity* among adjectives, relates verb pairs in which one member denotes a more intense, thorough, comprehensive or absolute action than the other member, as in the case of "startle" and "shock".

A total of eight patterns were selected for extraction of the *strength* relation, including the patterns [x even y] and [not just $x$ed but $y$ed]. In the evaluation, the authors report that out of 14 sample pairs classified by VerbOcean as related by strength 75% were correctly classified.

Near Synonyms Differentiating between adjectives by their position on an intensity scale may fall into the research area of differentiation among *near-synonyms*. According to (Edmonds, 1999) near-synonyms are words that are alike in essential, language-neutral meaning (denotation), but possibly different in terms of only peripheral traits, whatever these may be. It is an open question whether true synonyms exist at all; WordNet defines membership in a synset as the property of being exchangeable in many, but not all contexts.

(Edmonds, 1999) introduces an extensive model to account for the differences among near-synonyms, classifying the distinctions into *denotational*, *expressive*, *stylistic*, and *collocational*. Thus, stylistic distinctions include differences in *formality* [4]. For example, "motion picture" is a more formal expression than "movie" which in turn is more formal the "flick".

The AdjScales method indirectly takes into consideration some of the criteria for synonymy in (Edmonds, 1999). The nature of the lexical-semantic patterns is such that they retrieve snippets in which an adjective pair necessarily modifies the same noun; the narrow context moreover assures stylistic homogeneity of the scalemates.

4 Limitations of the AdjScales method

The AdjScales method promises to grant insight into a relatively underexplored corner of the lexicon by providing empirical evidence for subtle intuitions about the intensity of gradable adjectives. Scales constructed on corpus data may reflect the lexical organization of a broad community of language users. At the same time, the distinctions among the adjectives on a given scale can be very fine-grained, and speakers' explicit judgments do

[4]WordNet's *domain* labels encode some register and usage distinctions, but the categories are notoriously fuzzy. (Maks and Vossen, 2010) talk in detail about the differences between synset members in WordNet and propose remodeling solutions to overcome this problem.

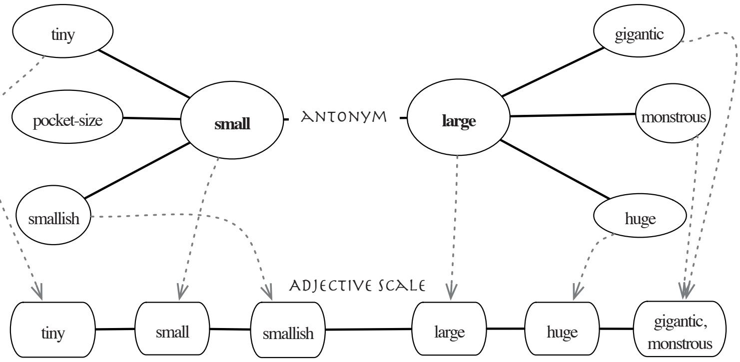

Figure 2: Illustration of the proposed structure of an adjective scale linked from some adjectives in a dumbbell. Note that "pocket-size" has more specific selectional restrictions that the other, more generically applicable adjectives in the dumbbell. It remains unconfirmed and not linked to the scale. "Smallish" is determined to be less intense than the centroid "small". "Gigantic" and "monstrous" are recognized to be of similar intensity relatively to "huge" and "large".

not always conform to the scales constructed on the basis of the corpus data. In the evaluation reported by (Sheinman and Tokunaga, 2009) annotators agreed with each other for only 63.5% of the adjective pairs when judging whether an adjective is milder, similar in intensity, or more intense than another adjective.

AdjScales method in particular, and pattern based methods in general, may suffer from low coverage. (Sheinman and Tokunaga, 2009) report that out of total of 5,378 distinct descriptive adjectives, only 763 were selected as suitable for further scaling, because the remainder could not be extracted in sufficient numbers in the patterns produced by the AdjScales' preprocessing stage, which requires at least 3 seed pairs. This limitation calls for further refinement of the method, such as the extraction of a wider selection of patterns.

Another weakness of the method is its poor ability to determine the place of adjectives in the neutral area of an adjective scale. For example, "tepid", "smallish", and "acceptable" are difficult to properly locate on their corresponding scales, and the weakness of method here is reflected in lower human agreement. Extending our work to a larger number of attributes will show whether this problem is specific to the limited number of scales tested or more general.

Currently we apply the AdjScales method on each half of a dumbbell and unify the results into a single scale. This approach relies on the assumption that each dumbbell can produce a single scale, which is not necessarily the case. The reason is that in many cases, WordNet currently subsumes semantically heterogeneous adjectives in a single dumbbell. Consider the adjectives "chilly, frosty, cutting, unheated" and "raw", which are all part of a dumbbell centered around (one sense of) "cold". But due to their different selectional restrictions, the Web does not return snippets like "∗ he ate his food unheated but not arctic" and "∗ a cutting, even refrigerated wind". We plan to examine the members of dumbbells for their semantic similarity and refine the clusters such that they lend themselves better to placement on scales. The AdjScales method will help in the identification of semantically homogeneous adjectives, leading to a cleaner representation in WordNet.

5 Applications of AdjScales in WordNet

We discuss a representative sample of applying AdjScales to gradable adjectives below.

5.1 Language pedagogy

Adjective scales in WordNet will provide learners of English with a more subtle understanding of the meanings of adjectives. By contrast, WordNet's current dumbbell representation and standard thesauri do not give clear information about the meaning distinctions among similar adjectives. We plan to develop a new interface that lets users visualize the unidimensional scales and gain an intuitive access to the meanings with a single glance.

5.2 Crosslingual encoding

Constructing and encoding scales with gradable adjectives for languages that have this lexical category would allow one to compare crosslinguistic lexicalizations with respect to questions like: which languages populate a given scale more or less richly? How do the members of corresponding scales line up? Mapping scales across languages could well support fine-grained human and machine translation.

(Schulam and Fellbaum, 2010) extracted patterns from the large COSMAS-II[5] German corpus using the process described in Section 2.1.

5.3 Reading textual entailment

Modeling the understanding of implicit and entailed information is a a major focus of current research in NLP. The PASCAL Recognizing Textual Entailment task challenges automatic systems to evaluate the truth or falsety of a statement (the Hypothesis) given a prior statement (the Text). For example, a system must decide whether or not H is true or false given T:

- T: Frigid weather sweeps across New Jersey
- H: The Garden State experiences cold temperatures

(Clark et al., 2007; Clark et al., 2008; Fellbaum et al., 2008) show that the semantic knowledge encoded in WordNet can be harnessed to extract information that is not present on the surface. Thus, WordNet tells that "New Jersey" and "the Garden State" are synonymous, increasing the probability that H is true. But knowing that "frigid" unilaterally entails "cold" would allow a more confident evaluation of H. If T and H were switched, the symmetric synonymy relation between the nouns would not facilitate a correct evaluation of H, whereas the downward entailing intensity relation would evaluate a Hypothesis containing "frigid" to be false if the Text referred to "cold". An RTE system with access to a resource that encodes intensity relations among its adjectives is thus potentially more powerful.[6]

[5] http://www.ids-mannheim.de/cosmas2

[6] Currently, WordNet encodes entailment relations among some verbs, but it doesn't provide a distinction between finer-grained subtypes such as *backward presupposition* ("know" must happen before "forget") vs. *temporal inclusion* ("step" is part of the action of "walk") (Fellbaum et al., 1993). Extracting instances of specific fine-grained relations, including intensity (may → should → must) using computational methods such as those in VerbOcean may provide further enrichment of WordNet.

5.4 Identifying spam product reviews

(Julien, 2010) examines how AdjScales might be used as a tool for detecting spam product reviews. Spam reviews are online reviews of products written for either deceptive or unhelpful purposes. For instance, company owners or employees may write a positive review of their product to boost the chances that customers will buy it; conversely, negative review of a competitor's product to discourage sales. Such reviews are more likely than genuine ones to contain highly intense adjectives.

5.5 Comparing nouns with AdjScales

(Schulam, 2011) develops a prototype of a system called SCLE (Semantic Comparison of Linguistic Entities), which uses the AdjScales algorithm to build adjective scales to compare the values represented by nouns modified by scalar adjectives.

Consider the phrases "warm day" and "hot day." Without knowledge of the relative intensity of adjectives that ascribe different values of "temperature" to the nouns, a system may know only that both nouns are modified by semantically similar adjectives. SCLE accesses adjective scales to infer which of the two days is characterized by a higher "temperature".

6 Conclusion

We propose a new semantic relation for WordNet's currently under-used adjective component. The *intensity* relation holds among gradable adjectives that fall on different points along a scale or dimension. Identifying and encoding this relation relies crucially on AdjScales (Sheinman and Tokunaga, 2009), a method for extracting and applying lexical-semantic patterns to a corpus. The patterns differentiate semantically similar adjectives in terms of the intensity with which they express a shared attribute and make it possible to construct scales where the adjectives are ordered relative to one another based on their intensity.

While only gradable adjectives express varying degrees of intensity, they constitute a highly frequent and polysemous subset of adjectives that are richly encoded crosslinguistically. We propose a model for representing scales in WordNet such that they supplement and co-exist with the current dumbbells. The principal improvement will be an empirically supported refinement of the present vague *similar* relation among many adjectives arranged around a shared centroid. The encoding of fine-grained intensity relations among presently undifferented adjectives will greatly enhance WordNet's potential for a wide range of diverse applications.

Acknowledgments

Fellbaum, Schulam and Julien's work is supported by grant CNS 0855157 from the U.S. National Science Foundation; Fellbaum is additionally supported by the Tim Gill Foundation.

References

Manfred Bierwisch. 1989. The semantics of gradation. In *Dimensional adjectives*, pages 71–261, Berlin: Springer.

Timothy Chklovski and Patrick Pantel. 2004. Verbocean: Mining the web for fine-grained semantic verb relations. In *EMNLP-04*, pages 33–40, Barcelona.

Peter Clark, William R. Murray, John Thompson, Phil Harrison, Jerry Hobbs, and Christiane Fellbaum. 2007. On the role of lexical and world knowledge in rte3. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, RTE '07, pages 54–59, Stroudsburg, PA.

Peter Clark, Christiane Fellbaum, Jerry Hobbs, Phil Harrison, William Murray, and John Thompson. 2008. Augmenting wordnet for deep understanding of text. In *Proceedings of the 2008 Conference on Semantics in Text Processing*, Stroudsburg, PA.

A. M. Collins and M. R. Quillian. 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:240 – 247.

D. A. Cruse. 1986. *Lexical Semantics*. Cambridge University Press, New York.

Dmitry Davidov and Ari Rappoport. 2008. Unsupervised discovery of generic relationships using pattern clusters and its evaluation by automatically generated SAT analogy questions. In *Proceedings of ACL-08, HLT*, pages 692–700, Ohio.

James Deese. 1964. The associative structure of some common english adjectives. *Journal of Verbal Learning and Verbal Behavior*, 3(5):347–357.

Philip Edmonds. 1999. *Semantic Representation of Near-synonyms for automatic lexical choice*. Ph.D. thesis, University of Toronto.

Christiane Fellbaum, Derek Gross, and Katherine Miller. 1993. Adjectives in wordnet. In *Five papers on WordNet*. MIT Press, Princeton.

Christiane Fellbaum, Peter Clark, and Jerry Hobbs. 2008. Towards improved text understanding with wordnet. In *Text Resources and Lexical Knowledge*, Berlin.

Christiane Fellbaum. 2002. Parallel hierarchies in the verb lexicon. In *Proceedings of the Ontolex02 Workshop*, ELRA, Paris.

Derek Gross, Ute Fischer, and George A. Miller. 1989. Antonyms and the representation of adjectival meanings. *Journal of Memory and Language*, 28:1:92 – 106.

M. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *14th Conference on Computational Linguistics*, pages 539–545.

Isaac Julien. 2010. Linguistic analysis with adjscales as a tool for predicting spam product reviews. Technical report, Department of Computer Science, Princeton University.

John S. Justeson and Slava M. Katz. 1991. Co-occurrences of antonymous adjectives and their contexts. *Computational Linguistics*, 17:1–19.

Chris Kennedy. 2001. Polar opposition and the ontology of degrees. *Linguistics and Philosophy*, 24:33–70.

Adam Kilgarriff. 2007. Googleology is bad science. *Computational Linguistics*, 33(1):147–151.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *CoLing-98*.

Isa Maks and Piek Vossen. 2010. Modeling attitude, polarity and subjectivity in wordnet. In *Proceedings of the 5th Global WordNet Conference*.

Helen Moss and Lianne Older. 1996. *Word Association Norms*. Psychology Press, Hove, U. K.

Ellen Riloff and Rosie Jones. 1999. Learning dictionaries for information extraction by multi-level bootstrapping. In *AAAI-99*.

Peter F. Schulam and Christiane Fellbaum. 2010. Automatically determining the semantic gradation of german adjectives. In *Semantic Approaches to Natural Language*, page 163, Saarbruecken.

P.F. Schulam. 2011. Scle: A system for automatically comparing gradable adjectives. Senior Thesis.

Vera Sheinman and Takenobu Tokunaga. 2009. Adjscales: Visualizing differences between adjectives for language learners. *IEICE Transactions on Information and Systems*, E92-D(8):1542–1550.

R. Snow, D. Jurafsky, and A.Y. Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. *Advances in neural information processing systems*, 17:1297–1304.

Peter D. Turney. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. In *CoLing-08*, Manchester, UK.

Julie Weeds and David Weir. 2005. Co-occurrence retrieval: A flexible framework for lexical distributional similarity. *Computational Linguistics*, 31(4):439–475.