

# From Adjective Glosses to Attribute Concepts: Learning Different Aspects That an Adjective Can Describe

Omid Bakhshandeh<sup>1</sup> and James F. Allen<sup>1,2</sup>

<sup>1</sup>Computer Science Department, University of Rochester, Rochester, New York

<sup>2</sup>Institute for Human and Machine Cognition, Pensacola, Florida ,  
omidb@cs.rochester.edu, james@cs.rochester.edu

## Abstract

Adjectives are one of the major contributors to conveying subjective meaning to sentences. Understanding the semantics of adjectives is essential for understanding natural language sentences. In this paper we propose a novel approach for learning different properties that an adjective can describe, corresponding to the ‘attribute concepts’, which is not currently available in existing linguistic resources. We accomplish this by reading adjective glosses and bootstrapping attribute concepts from a seed of adjectives. We show that we can learn new attribute concepts using adjective glosses of WordNet adjectives with high accuracy as compared with human annotation on a test set.

## 1 Introduction

Adjectives have not been studied in lexical semantics as much as verbs and nouns. However, they have a very interesting polymorphic behavior in adding subtle meaning to a sentence. The main function of adjectives is the modification of words, such as nouns, by describing some properties of them. For instance, the adjective *fast* in the sentence ‘a fast car’ is describing the *speed* property of a car. Another example is the adjective *tall* in the sentence ‘he is tall’ which describes the ‘stature’ property of someone. Adjectives can describe a different value for the same property. For instance, *large* and *small* each describe the ‘size’ property of something. Such properties can be associated with an abstract scale, on which some adjectives are ordered by their semantic intensity. Generally, is it possible for an entity to have a quality which lies somewhere in the middle of a scale associated with a property. For instance, assuming a scale for size, it is possible to grade new adjectives so that they describe some new spot on the size scale. Such adjectives are called ‘gradable adjectives’. We generally determine whether an adjective is gradable by deciding whether or not graded usages with comparative modifiers (more, less, etc) is possible for it (Rusiecki, 1985). Gradable adjectives imply the existence of an underlying semantic scale (pertaining to a property). It has been argued that gradable adjectives denote functions from objects to intervals or points on a scale (Kennedy, 1999).

Not all adjectives are intended to be interpreted in a gradable sense. A famous example of the class of non-gradable adjectives is the set of color adjectives such as *red*. It is argued that when someone says ‘His tie is red’ it would scarcely make sense to follow up with the question ‘how red is the tie?’ (Young, 1984). However, we can all understand gradable usages of ‘red’ in sentences such as ‘the reddest I have ever seen’, which is because ‘red’ is being used as a quality adjective (Young, 1984). Many of the classic non-gradable adjectives can be seen to be used with degree-modifiers and comparatives or superlatives in various contexts, such as ‘This dog is the wildest’. By observing such adjectives deeper, one can see that there are some adjectives which specifically define a property, but that property does not really have other adjectives related to it. For instance, the adjective ‘intermural’ which means ‘between two or more institutions’ can be associated with ‘institution’ property, as one can say ‘some organization is more intermural than the other’. However, there are no other adjectives talking about ‘institution’ property.

Another example is the adjective *feminine*, which modifies the ‘femininity’ property, as someone can say ‘someone is more feminine than the other’. In language understanding knowing the underlying properties being modified by the adjective in a gradable sense is useful, so we try to come up with a property denoting attribute concept for all of the adjectives in general. We use ‘attribute identification’ to refer to the task of assigning a set of “attribute concepts” to an adjective. Hence, in the earlier examples, ‘size’ is the attribute concept of ‘large/small/tiny’, and ‘institution’ is the attribute concept of the ‘intermural’ adjective. Furthermore, there are adjectives that modify more than one property of an entity. For instance, the adjective *gangling* means “tall and thin and having long slender limbs”. It is evident from the meaning that *gangling* can modify the noun it is describing with respect to two different properties: ‘height’ and ‘thickness’. Another example is the adjective ‘squab’ which means ‘short and fat’ which again describes two distinct properties. Hence, the set of attribute concepts for both ‘gangling’ and ‘squab’ is {height, thickness}.

Determining the attribute concepts of adjectives can improve our understanding of the natural language sentences. For instance, consider a sentence such as ‘The tap water here is lukewarm, but it is usually freezing uptown’. By having the knowledge of attributes and scales, one can have ordering of adjectives regarding their intensities, and then can understand how different the temperature property of tap water is in uptown. In general, learning the set of attribute concepts and associating adjectives to them is a pre-requisite for ordering adjectives and their polarity magnitude on scales. Moreover, having attribute concepts, one can disambiguate among various senses of an adjective such as *hot* in the sentence ‘our debate is so boring, but this topic is hot.’, by knowing that two adjectives usually pertain to the same attribute concept in order to be comparable.

Existing linguistic resources (dictionaries, WordNet (Miller, 1995), and thesauri) rarely contain information on adjectives being part of a scale, relating to an attribute concept, or being of a particular strength. In this paper, we present a novel approach for finding all the attribute concepts including scales that an adjective synset (here in WordNet 3.0) is graded on. Our approach is based on reading adjective glosses and bootstrapping attribute concepts from a seed of adjectives. Our approach builds on the fact that there are redundant syntactic and semantic patterns in definitions of words which enables bootstrapping (Yarowsky, 1995). To our knowledge, none of the earlier works have attempted to find more than one scale that an adjective can describe. In Section 4 we show that we can learn attribute concepts using adjective glosses with high accuracy as compared with human annotation on a sub-set of adjectives. Moreover, none of the earlier approaches have had as good coverage as our method, which is about 77% of WordNet adjectives. Last but not least, our bootstrapping algorithm can be generally employed in learning any kind of property from definitional texts for different parts of speech, not only adjectives. We focus on adjectives as property-denoting lexical entities, since they are valuable for acquiring concept representations for areas such as ontology learning. The result of this work can lead towards describing the semantics of adjectives contained in as part of a larger effort to develop new techniques for automatically acquiring deep lexical knowledge for capturing knowledge about concepts.

## 2 Adjectives in WordNet

The semantic organization of adjectives in the WordNet is not similar to the organization of nouns and verbs, as adjectives do not show a hierarchical organization (Mendes, 2006). In general, adjectives in WordNet are divided into *descriptive* and *relational* classes. Descriptive adjectives are the ones that ascribe a value of an attribute to a noun, i.e, they describe a property of a noun they modify. WordNet has links between some of the descriptive adjectives expressing a value of an attribute and the noun with which that attribute is lexicalized. For example, the adjective ‘tall’ is linked to the noun attribute ‘stature’.

Among all 18,156 adjective synsets in WordNet, only about 620 of them have the attribute link. Instead of the hypernymic relation that is used among nouns, the main relation used for descriptive adjectives is antonymy. Binary opposition of adjectives, which shows contrary values of an attribute, is represented by pointers with the meaning of ‘IS-ANTONIMOUS-TO’. For adjectives that do not seem

Class	Total in WordNet	Examples
Descriptive (head of dumbbell structures, with attribute links)	620	tall, warm, beautiful, heavy
Descriptive (similar to a head of a dumbbell, with indirect attribute links)	3,541	damp, scorching, lukewarm, massive
Descriptive (have antonyms, but no attribute links)	3,226	cheap, poor, dry, valuable
Relational (no direct or indirect antonym)	4,951	oxidized, racial, rat-like, beaten

Table 1: Statistics of different classes of adjectives in WordNet

to have a direct antonym, WordNet has a pointer with the meaning of ‘IS-SIMILAR-TO’ which points to an adjective which is similar in meaning, through which an indirect antonym is inherited. For example, wet and dry are direct antonyms and an indirect antonymic pointer exists between damp and dry since damp is similar to wet. WordNet encodes such structures in the form of clusters (adjective-sets), which are often called *dumbbells* or satellites. In dumbbells two antonymous synsets (head-words) are linked to a noun which they describe (attribute) and each head-word is linked to its similar adjectives. Dumbbells seemed well motivated psycho-linguistically and distributionally, but they are not sufficiently informative for detecting and quantifying meaning similarities among adjectives. For instance, there is no specific semantic distinction between all the similar-to links of a particular head-word, i.e., they are all treated in the same way. Moreover, only 3,541 adjectives are encoded in dumbbell structure, which is a limited coverage on all adjectives. There are 3,226 descriptive adjectives in WordNet which have antonyms and inherently ascribe a value to an attribute, but are not linked to any attributive nouns, e.g. ‘cheap’ and ‘expensive’, none of which has an attribute link in WordNet. Our main goal in this paper is to expand the set of attribute concepts to all WordNet adjectives, proposing a methodology for high coverage attribute learning. As will be explained in Section 4, our approach covers about 14,104 adjectives (77%) on average.

Relational adjectives (pertainyms) are the ones that do not have an antonym and are related by derivation to specific nouns. If an adjective does not have a direct or indirect antonym, then it is relational and it has a pointer to the noun it relates to, i.e., is derived from. Relational adjectives are mostly known not to be gradable (Mendes, 2006), e.g., *atomic*, however, there are relational adjectives which pass the linguistic test for gradability, such as ‘nutritional’. We apply our approach to relational adjectives as well, and try to find a property-defining attribute concept for them. The results of this experiment can be found in section 4. Table 1 summarizes some statistics on adjectives in WordNet. The sum of total adjectives in this table is larger than 18,156, which is due to the intersection between adjectives with pertainym link and descriptive adjectives with no direct or indirect attribute link.

### 3 The Approach

Glosses, as a short meaning explanations for words, provide a rich pieces of information about the words they describe. In this paper, we present our novel approach on using WordNet glosses for understanding the semantics of adjectives, specifically the properties they can describe. As mentioned earlier, we aim to identify all the attribute concepts (including scales) that an adjective can be associated with. We propose a semi-supervised method for learning attribute concepts. The idea is to use the already encoded attribute links of 620 adjectives in WordNet as the initial seed and bootstrap attribute concept of the rest of adjectives by learning patterns. The high-level iterative bootstrapping algorithm is described in the Algorithm 1. In the upcoming subsections we will describe each phase of this algorithm in detail.

---

**Algorithm 1** Algorithm for learning attribute concepts

---

**Require:** set  $J$  = All adjectives

set  $\Theta^T$  = Features in iteration  $T$

set  $S$  = Seed adjectives with known attributes

set  $X_j$  = Candidate attributes for adjective  $j$

```
1: // Pre-processing and Candidate Attribute Extraction
2: for adjective  $j$  in  $J$  do
3:   Process  $j$ 's gloss, and get back its link-augmented-dependency tree (LAD)
4:   Process  $j$ 's gloss, and get back its set of candidate attributes,  $X_j$ 
5: end for
6: while not converged do
7:   //Feature Extraction and Model Training
8:    $\Theta^t$  = Extract features from  $S$ 
9:   Train a classifier  $F$ , using features  $\Theta^t$ 
10:  //Decoding and Updating the Seed
11:  for adjective  $j \in J$  and  $j \notin S$  do
12:    Using  $F$ , tag all candidates attributes in  $j$ 
13:    if Any of the candidates is tagged as 'attribute' then
14:      add  $j$  to  $S$ 
15:    end if
16:  end for
17: end while
```

---

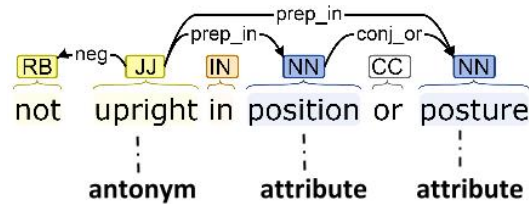


Figure 1: The link-augmented graph of the gloss of the adjective ‘unerec’

### 3.1 Pre-processing

For each adjective we should process its gloss first. The idea is to enrich the dependency structure of the gloss with the semantic links. First, we parse the gloss using Stanford dependency parser (de Marneffe et al., 2006). Here we use Collapsed CC-processed dependencies, which produces a graph for a sentence with a node for some of the words in the sentence and an edge for some predefined relations between the nodes. Second, we extract all the semantic links (such as ‘IS-ANTONIMOUS-TO’, ‘IS-SIMILAR-TO’ and ‘ATTRIBUTE’) that the adjective has and incorporate it as links into the dependency tree. We call the resulting data structure a ‘link-augmented-dependency tree’ (LAD). As an example, consider the adjective *unerec*, which means ‘not upright in position or posture’. Figure 1 shows the resultant LAD structure for this adjective. As you can see, the node *upright* is tagged as ‘IS-ANTONIMOUS-TO’ with directive *antonym*, and both nodes *position* and *posture* are tagged as ‘ATTRIBUTE’ with directive *attribute*.

### 3.2 Candidate Attribute Extraction

For each adjective, we need to have a set of candidate attributes,  $X_j$ . As all seed attribute concepts are nouns, we restrict our candidate set to only nouns. Experimenting on an initial development set, we found the following criteria for the candidate attributes of a given adjective  $j$ :

- All nouns appearing in the gloss of  $j$ .
- All nominalizations of the adjectives appearing in the gloss of  $j$ .
- All hypernyms of nouns appearing in gloss, up to two levels in WordNet hierarchy. For example, for the candidate ‘tallness’ we extract ‘height’.
- All hypernyms of nominalizations of the adjectives appearing in the gloss, up to two levels in WordNet hierarchy.
- All candidates of the adjectives appearing in the gloss of  $j$ . For example, for the adjective *gangling* which is defined as ‘tall and thin’ we recursively include all the candidates of ‘tall’ and ‘thin’ as the candidates for ‘gangling’.

Using all these criteria, we extract a set  $X_j$  for each  $j$ , which will be used in the training phase.

### 3.3 Feature Extraction

After pre-processing all glosses and making a LAD, in each iteration we extract features from our seed set. We mainly extract the syntactic and semantic features which are counted as evidence for pinpointing the attribute concepts. The rationale behind the forthcoming feature extraction is the observation that the glosses tend to look alike, so there are some hidden syntactic and semantic patterns among glosses. We use the paths of a given LAD for finding general lexical and semantic patterns in the glosses. The resulting patterns make our feature set. Following are the four types of features, each of which provides a different kind of information.

- **Lexical-semantic Features:** These are the features which use both lexical and semantic aspects of a path in LAD.
- **Semantic Features:** These are the features which only use semantic aspects of a path.
- **Lexical Features:** These features only use lexical aspects of a path.
- **POS Features:** These rules only use part of speech information of a path.

From Figure 1 consider the example path: not  $\xleftarrow{neg}$  upright  $\xrightarrow{prep\_in}$  position. Following is the four types of feature that we extract for this path:

- Lexical-semantic Feature: not(RB)  $\xleftarrow{neg}$  upright(ANT-JJ)  $\xrightarrow{prep\_in}$  ?(NN).
- Semantic Feature: not(RB)  $\xleftarrow{neg}$  \*(ANT-JJ)  $\xrightarrow{prep\_in}$  ?(NN).
- Lexical Feature: not(RB)  $\xleftarrow{neg}$  upright(JJ)  $\xrightarrow{prep\_in}$  ?(NN).
- POS Feature: \*(RB)  $\xleftarrow{neg}$  \*(JJ)  $\xrightarrow{prep\_in}$  ?(NN).

In all of the above features, the ‘?’ indicates the noun attribute concept for which we are extracting features. Also, the ‘\*’ shows any lexical item which satisfies the constraints provided inside the parentheses.

### 3.4 Model Training

For each adjective  $j$ , there is a set of candidates  $X_j$ , each of which should be tagged as either *attribute* or *non-attribute*. Therefore, the task of finding attribute concepts of an adjective becomes a tagging problem, where  $X$  is the set of candidates and  $Y$  is the set of possible tags. The two models we experiment with here are Logistic Regression and Conditional Random Fields:

**Logistic Regression:** The first model we experimented with is a binary classifier which attempts to tag each candidate attribute either as an *attribute* or *non-attribute*. The features used for training the model are the extracted features in the previous phase. Here we use a Logistic Regression model, which has one attribute candidate as its input. The probability of tagging attribute candidate  $x$  with tag  $y$  is as follows:

$$p_{\theta}(y|\mathbf{x}) = \frac{1}{Z_{\theta}(\mathbf{x})} \exp \left\{ \sum_{k=1}^K \theta_k f_k(y, x) \right\} \quad (1)$$

where  $f$  is the feature function and  $Z_{\theta}(\mathbf{x})$  is defined as follows:

$$Z_{\theta}(\mathbf{x}) = \sum_{y \in Y^T} \exp \left\{ \sum_{k=1}^K \theta_k f_k(y, x) \right\} \quad (2)$$

The results of training a classifier using a logistic regression model are presented in Section 4.

**Conditional Random Fields:** In order to better estimate the tag for each attribute candidate it is essential to know the other candidates. The Logistic Regression model does not take into account the set of specific candidates derived from the gloss of one specific adjective. A common approach for taking into account such a sequential tagging is to model the problem with Conditional Random Fields (CRF) (Lafferty et al., 2001). We decided to use a linear chain CRF, which produces promising results. The results of training a classifier using a CRF model are presented in Section 4.

### 3.5 Decoding and Updating the Seed

The bootstrapping algorithm described earlier iterates until convergence, which is until the seed set gets stabilized. At each iteration, we try to find all attribute concepts for some new adjectives which are not covered in the seed, decoding using the trained model parameters. At each iteration, the seed gets updated with a new set of adjectives, the ones which have at least one of their attribute candidates tagged as *attribute* by the decoder. Changes in seed set result in extracting new features for the next iterations, hence making a more general classifier. Table 2 shows some of the adjectives classified under the attribute concepts ‘size’, ‘magnet’, ‘price’ and ‘moisture’ using our algorithm. We will discuss the evaluation results of our approach in the next section.

Attribute	Adjective	Gloss
<b>Size</b>	minor	limited in size or scope
	great	relatively large in size or number or extent
	mountainous	like a mountain in size and impressiveness
<b>Magnet</b>	attractable	capable of being magnetized or attracted by a magnet
	magnetic	having the properties of a magnet
<b>Price</b>	cheap	relatively low in price or charging low prices
	expensive	high in price or charging high prices
<b>Moisture</b>	dry	lacking moisture or volatile components
	wet	containing moisture or volatile components

Table 2: Example of the learned attribute concepts for adjectives using our algorithm

## 4 Evaluation and Experimental Results

We evaluate the quality of our extracted attribute concepts in two stages: evaluation of the tagging and the full iterative approach.

## 4.1 Evaluation of the Tagging

The first stage attempts to evaluate the classifiers together with feature sets independent from the bootstrapping method. In order to compare the classifiers’ performance standalone, we sampled a test set from the set of gold-standard attribute links of WordNet, which have not been used as in the initial seed set. We compared our learned attribute concepts with this gold-standard test set. As discussed earlier, we mainly use two classifiers for finding attribute concepts. The first method is Logistic Regression (*LR*) and the second method is (*CRF*). We measure precision, recall and F1-score for each of these classifiers, for both *attribute* and *non-attribute* tags. The results of this experiment are depicted in tables 3 and 4.

Method	Precision	Recall	F1-Score
<i>CRF</i>	87%	62%	72%
<i>LR</i>	79%	63%	70%

Table 3: Attribute Tagging Results

Method	Precision	Recall	F1-Score
<i>CRF</i>	71%	91%	80%
<i>LR</i>	70%	83%	76%

Table 4: Non-Attribute Tagging Results

As the results show, the classification method using our feature set has high accuracy for both *attribute* and *non-attribute* tagging. The CRF classifier outperforms the LR classifier mainly in tagging as *attribute*. This was expected, as the CRF model takes into account the set of candidates for a given adjective, which results in better predictions. Overall, the classifiers have higher accuracy on assigning *attribute* tag, i.e., true positive cases. Also, the lower recall in attribute-tagging is affected by the fact that most of the candidates are non-attributes and only one or few candidates for each adjective are tagged as attribute.

## 4.2 Evaluating the Full Iterative Approach

The second scenario aims at evaluating the full approach, including the iterative bootstrapping. For this purpose, we made a dataset of randomly selected 250 WordNet adjectives (either descriptive or relational) which did not have an attribute link and were associated with an attribute concept by our approach. We attempted to manually annotate this dataset with a gold-standard attribute concept. We asked 20 human judges to perform the evaluation task. The judges were provided with a guideline defining the notions of attribute-concept and gradability for adjectives. Then for each adjective, we provided the human judges with a few lines of information regarding the adjective in question, together with our extracted attribute concept for that adjective. As an example, the evaluation task for the adjective ‘*dry*’ is as follows:

```
* [Synset: dry (lacking moisture or volatile components)
  example usage: "dry paint"]
>> Is this a gradable adjective? (y,n)
* ATTRIBUTE CONCEPT: moisture
>> Does the above attribute concept sound correct to you? (y,n)
>> If not, what is your suggestion? ...
>> Was it vague to assign an attribute to this adjective? (y,n)
```

The annotation agreement using Fleiss’ kappa measure (Fleiss, 1981) was  $\kappa = 74\%$ , which shows a substantial agreement. Only the adjectives with substantial agreements on all annotation questions were included in the gold-standard set, resulting in about 210 adjectives. Given the fact that our algorithm

has found an attribute concept for all the gold-standard data, we only report on Precision score. Also, it is important to note that the bootstrapping algorithm is sensitive to the adjectives added to the seed on every iteration, so we tune our classifiers to have better precision than recall. That is, we would rather have a robust feature extraction phase which accounts for true positives and fewer false positives. Table 5 summarizes the results of evaluation of the two methods *LR* and *CRF* against the created gold-standard dataset under *Precision<sub>all</sub>* column.

As the results show, 57% of our approach’s decisions on determining the attribute concepts were correct. Among the adjectives for which the system has determined wrong answers, 82% were annotated as vague or hard to annotate by the human judges, most of which were relational adjectives. An example of an incorrect answer of the algorithm is for the adjective *abdominal*, in a noun phrase such as ‘abdominal muscles’. In order to correctly determine the scalability and gradability of such adjectives, we need to have world knowledge about various concepts, knowing that abdominal refers to a part of body. The algorithm’s selected attribute for this adjective was *abdomen*, which does not seem natural to a human annotator, but from the algorithm’s viewpoint it is the main property that this adjective is modifying. Most such adjectives are classified as being vague and mostly non-gradable by the human annotators.

We compute the descriptive precision (*Precision<sub>descriptive</sub>*) score by removing the adjectives which have been tagged as vague. As the results in Table 5 show, the accuracy of finding the attribute concepts on mostly descriptive adjectives is high. This shows that our approach performs very well on pinpointing the gradable adjectives, learning new attribute concepts for about 77% of adjectives in WordNet. As mentioned earlier in section 2, many of the relational adjectives can be considered gradable, e.g., for the adjective ‘nutritional’ it is plausible to say ‘Milk is more nutritional than soda’. Hence, including relational adjectives in the attribute concept dataset can be very useful.

<b>Method</b>	<i>Precision<sub>all</sub></i>	<i>Precision<sub>descriptive</sub></i>
<i>CRF</i>	57%	<b>85%</b>
<i>LR</i>	48%	72%

Table 5: Evaluation results on hand annotated random test set.

## 5 Related Work

Typically adjectives play a significant role in conveying opinion and subjectivity of language. There have been many works concerning the semantics of adjectives in the field of opinion mining which can relate to our work. Hatzivassiloglou and McKeown (1993) performed the first attempt towards automatic identification of adjective scales. They presented an approach for clustering adjectives in the same scale based on their positive or negative orientation. Another work (Hatzivassiloglou and McKeown, 1997) proposes to classify the semantic polarity of adjectives based on their behavior in conjunction with other adjectives in a news corpus. They employ the existing clustering algorithms for this task. Turney and Littman (2003) decide on semantic orientation of a word (positive, negative combined with mild or strong) using statistical association with a set of positive and negative paradigm words. OPINE (Popescu and Etzioni, 2005), a system for product review mining, ranks opinion words by their strength. Our work differs fundamentally from these works in that it does not attempt to assign positive or negative polarities to adjectives. All such works focus on detection of semantic orientation of adjectives, and do not report on extracting attribute concepts or scales for adjectives. The information on orientation of adjectives is very helpful for understanding their semantics, but it is not sufficient for deep understanding which can enable further inference.

Another ongoing research project is on adjective intensity ordering, where researchers aim at ordering/ranking similar adjectives based on their intensity, such as lukewarm < warm < hot. Sheinman and Tokunaga (Sheinman and Tokunaga, 2009a) attempt to automatically learn adjective scaling patterns using seed adjectives and Web hits. They collect lexico-semantic patterns via bootstrapping from seed



adjective pairs to obtain pairwise intensities. AdjScale (Sheinman and Tokunaga, 2009a,b) proposes a new semantic relation for gradable adjectives in WordNet, which encodes the information on the intensity of different adjectives which share the attribute link. They use lexical-semantic patterns for mining the Web for evidence of the relative strength of adjectives such as ‘large’ and ‘gigantic’ with respect to their attribute ‘size’. Then they can derive pairwise partial ordering on adjectives which share an attribute. They apply the extracted patterns on WordNet dumbbells, for which they get new intensity relation links among the similar adjectives in the dumbbells. Finally, Melo and Bansal (de Melo and Bansal, 2013) present an unsupervised approach that uses semantics from Web-scale data (patterns such as ‘good but not excellent’) to rank words by assigning them positions on a continuous scale. They use Mixed Integer Linear Programming to determine the ranks on scales, as opposed to pairwise ranking in earlier works. All the mentioned works have only taken into account the descriptive adjectives in WordNet, i.e., the ones having an attribute link in the dumbbell structures – which provides a limited coverage, only about 22% of all adjectives. These works come short on extracting non-existing attribute concepts for naming new scales. Our work can find attribute concepts for all WordNet adjectives and associate relevant adjectives to the same attribute concept, which is a big step towards high quality ordering of adjective intensities. Moreover, none of these works investigate finding more than one scale that an adjective is graded on. Our work determines not only one, but all different aspects that an adjective could describe.

Another close body of work is the research on assigning attributes to adjective-noun phrases. The compositional semantics of adjective-noun phrases can be modeled in terms of selective binding (Pustejovsky, 1995), i.e., the adjective selects one of possibly several roles or attributes from the semantics of the noun. For example, given the phrase (1) *warm weather*, the semantic representation is ‘ $TEMPERATURE(weather) = warm$ ’. Hartung and Frank (Hartung and Frank, 2010a) attempt to extract the attribute for adjective-noun phrases, by selecting the semantics of the noun that is selected by the adjective. Mainly, this kind of knowledge has been extracted from corpora by searching for patterns (Almuhareb, 2006; Cimiano, 2006). An instance of pattern is [the  $x$  of the  $y$  is  $z$ ], where  $x$  is an attribute,  $y$  is a noun, and  $z$  is an adjective that paraphrases (1), e.g. *the color of the car is blue*. However, linguistic patterns that explicitly relate triplet co-occurrences of nouns, adjectives, and attributes are very rare – and in many cases, may not even provide sufficient evidence to determine an attribute for a given noun-adjective pair. Hartung and Frank (Hartung and Frank, 2010b) propose an alternative method which is doublet co-occurrences. They first search for noun-attribute co-occurrences, then adjective-attribute co-occurrences. However, doublet co-occurrences do not result in significant boost in web hits for patterns and their approach still lacks breadth in identifying adjective attributes.

## 6 Conclusion

In this paper we presented a new approach for comprehensive identification of the attribute concepts of adjectives in WordNet. The main idea is to learn the attribute concepts by bootstrapping using the adjective glosses. Our results show that our approach can identify the attribute concepts of about 77% of WordNet adjectives with high accuracy. Our algorithm can be generalized in order to be applied to various applications which require finding certain properties from definitional texts. Our work on determining the adjective attribute concepts could also benefit the research on sentiment orientation (positive/negative) of adjectives.

Another semantic relation between adjectives that is not considered by WordNet is gradation, i.e., the intensity of adjectives as compared with one another, going from a weak strength to a strong one. Once we have the extensive attribute concepts for most of the WordNet adjectives, we can attempt to order the adjectives classified under the same concept based on their degree of intensity. Our work aims at producing a large, unrestricted number of individual intensity scales (attribute concepts) for different qualities and hence can help in fine-grained sentiment analysis with respect to very particular aspects. As the next step of this work, we are planning to order all adjectives associated with the attribute concepts that we identified in this work. Also, given the promising results using linear chain CRF, we are planning to experiment with other structured CRF models as a future work. Moreover, we are planning to release

the derived attribute concepts dataset, which could be helpful for various language understanding tasks.

## Acknowledgments

We would like to thank anonymous reviewers for their valuable feedback. This work is sponsored by Nuance Foundation and The Office of Naval Research under grant number N000141110417.

## References

- Almuhareb, A. (2006). *Attributes in Lexical Acquisition*. Ph. D. thesis.
- Cimiano, P. (2006). *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.
- de Marneffe, M.-C., B. MacCartney, and C. D. Manning (2006). Generating typed dependency parses from phrase structure parses. In *In Proc. INTL Conf. On Language Resource and Evaluation (LREC)*, pp. 449–454.
- de Melo, G. and M. Bansal (2013). Good, great, excellent: Global inference of semantic intensities. *TACL 1*, 279–290.
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* (2nd ed.). New York: Wiley-Interscience.
- Hartung, M. and A. Frank (2010a). A semi-supervised type-based classification of adjectives: Distinguishing properties and relations. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odiijk, S. Piperidis, M. Rosner, and D. Tapias (Eds.), *LREC*. European Language Resources Association.
- Hartung, M. and A. Frank (2010b). A structured vector space model for hidden attribute meaning in adjective-noun phrases. In *Proceedings of the 23rd International Conference on Computational Linguistics, COLING '10*, Stroudsburg, PA, USA, pp. 430–438. Association for Computational Linguistics.
- Hatzivassiloglou, V. and K. R. McKeown (1993). Towards the automatic identification of adjectival scales: Clustering adjectives according to meaning. In *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics, ACL '93*, Stroudsburg, PA, USA, pp. 172–182. Association for Computational Linguistics.
- Hatzivassiloglou, V. and K. R. McKeown (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics, ACL '98*, Stroudsburg, PA, USA, pp. 174–181. Association for Computational Linguistics.
- Kennedy, C. (1999). Projecting the adjective: The syntax and semantics of gradability and comparison. In *In Proc. INTL Conf. On Language Resource and Evaluation (LREC)*. New York: Garland Press.
- Lafferty, J. D., A. McCallum, and F. C. N. Pereira (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning, ICML '01*, San Francisco, CA, USA, pp. 282–289. Morgan Kaufmann Publishers Inc.
- Mendes, S. (2006). Adjectives in wordnet.pt. In *In: Proceedings of the GWA 2006, Global WordNet Association Conference, Jeju Island, Korea*.
- Miller, G. A. (1995, November). Wordnet: A lexical database for english. *Commun. ACM* 38(11), 39–41.

- Popescu, A.-M. and O. Etzioni (2005). Extracting product features and opinions from reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05*, Stroudsburg, PA, USA, pp. 339–346. Association for Computational Linguistics.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Rusiecki, J. (1985). *Adjectives and comparison in english*. Longman.
- Sheinman, V. and T. Tokunaga (2009a). Adjscales: Differentiating between similar adjectives for language learners. In J. A. M. Cordeiro, B. Shishkov, A. Verbraeck, and M. Helfert (Eds.), *CSEU (1)*, pp. 229–235. INSTICC Press.
- Sheinman, V. and T. Tokunaga (2009b). Adjscales: Visualizing differences between adjectives for language learners. *IEICE Transactions 92-D(8)*, 1542–1550.
- Turney, P. D. and M. L. Littman (2003, October). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Trans. Inf. Syst.* 21(4), 315–346.
- Yarowsky, D. (1995). Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics, ACL '95*, Stroudsburg, PA, USA, pp. 189–196. Association for Computational Linguistics.
- Young, D. J. (1984). *Introducing English Grammar*. Howard Jackson.