Words with Attitude

Jaap Kamps and *Maarten Marx*

Abstract

The traditional notion of word meaning used in natural language processing is literal or lexical meaning as used in dictionaries and lexicons. This relatively objective notion of lexical meaning is different from more subjective notions of emotive or affective meaning. Our aim is to come to grips with subjective aspects of meaning expressed in written texts, such as the attitude or value expressed in them. This paper explores how the structure of the WordNet lexical database might be used to assess affective or emotive meaning. In particular, we construct measures based on Osgood's semantic differential technique.

1 Introduction

The traditional notion of word meaning used in natural language processing is literal or lexical meaning. This is the way the meaning of words is explained in dictionaries and lexicons. And, as may come as no surprise, the majority of research in natural language processing deemphasizes other aspects of meaning. Yet at the same time, we find a myriad of notions of meaning in the writings of philosophers, linguists, psychologists, and sociologists. This is not the place to have an extensive discussion on the meaning of meaning, but our aim will be to bring other notions of meaning into natural language processing. In particular, we will be interested in the differences between the relatively objective notion of lexical meaning, and more subjective notions of emotive or affective meaning.

Suppose we can evaluate the subjective meaning expressed in a text. This would allow us to classify documents on subjective criteria, rather than on their factual content. This can be as radical a change as categorizing the screws in a repair shop's inventory by their beauty, instead of their size and material. This may not be very practical for a repair shop, but document classification does not require a physical rearrangement of objects. As a result, it would simply provide an additional classification criterion. It is not difficult to envision cases in which precisely a subjective categorization is desirable and useful.

Our aim is to come to grips with aspects of the subjective meaning expressed in written texts, such as the attitude or value expressed in them. Of course, there are well-established approaches for this in the social and behavioral sciences. In particular, methods like surveys or test panels in which persons evaluate certain subjective criteria. However, the advent of the Internet gives us access to large numbers of documents and large corpora. Here, applying these traditional methods of evaluation is impractical: it is simply too time-consuming and very costly. For these reasons, we are interested in measures that can be evaluated automatically.

Our working hypothesis is that subjective aspects of meaning can be derived from the particular choice of words in a text. That is, there are indeed words with attitude or values. Prominent candidates for this are modifiers, such as descriptive adjectives like 'beautiful' or 'good' (and their antonyms 'ugly' and 'bad'). This paper explores how to assess more subjective aspects of meaning by using the structure of the WordNet lexical database (Miller, 1990; Fellbaum, 1998). At first glance, this may appear to be a bad choice because the words in WordNet are structured by their lexical meaning. In particular, the synonymy or SYNSET relation in WordNet represents the coincidence of lexical meaning. However, the organization of WordNet is not a conventional alphabetical list, but a large interconnected network of words (resembling the organization of human lexical memory). One of the design principles in

WordNet is a differential theory of meaning: the meaning of a concept is determined by its place relative to other concepts. It is precisely this larger WordNet structure that we want to exploit.

This paper is structured as follows. In §2, we will discuss a classical theory for measuring affective or emotive meaning. From this we take the major factors that differentiate between values or attitude. Then, in §3 we explore how we can translate the structure of WordNet into a measure for these factors. Next, in §4 we discuss how such measures can be implemented, and we end in §5 with conclusions and some discussion.

2 Affective Aspects of Meaning

Our aim is to measure the subjective meaning expressed in a text. For such an enterprise to be successful, there must be sufficient generality in the semantic dimensions used by individuals. This immediately prompts a number of questions: do such generic semantic dimensions exist at all? And if so, can we characterize these specific semantic dimensions?

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

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

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

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

3 Affective Meaning and WordNet

We will now investigate measures for the evaluative factor of meaning based on the WordNet lexical database (Fellbaum, 1998). The WordNet database has entries on the level of words (just as traditional dictionaries and lexicons). The unit of evaluation we are interested in is not individual words, but larger units of text, such as phrases, paragraphs, and larger units. We will proceed as follows: we will first investigate WordNet-based measures for individual words, and then consider ways of aggregating the scores of individual words to larger textual units. For example, an obvious way is to view a textual unit as a bag of words, and evaluate the text by combining the scores for the individual words in the text.

The evaluative dimension of Osgood is typically determined using the adjectives 'good' and 'bad' (other operationalizations are possible depending on the subject under investigation). Indeed, if we look up the meaning of these two evaluative adjectives in WordNet we find that they are each other's antonym. Our plan is to evaluate individual words by determining their relation to the words 'good' and 'bad' in the WordNet database. For this, we can use the synonymy relation (or a generalization of it) to establish the relatedness of two words. That is, WordNet's SYNSET relation may provide a handle to determine Osgood's evaluative factor.

We will define the notion of n-relatedness based on the SYNSET relation (this is similar to the graph-theoretic notion of connectedness).

Definition 1 *Two words w_0 and w_n are n-related if there exists an $(n+1)$-long sequence of words $\langle w_0, w_1, \ldots, w_n \rangle$ such that for each i from 0 to $n-1$ the two words w_i and w_{i+1} are in the same SYNSET (i.e., w_i and w_{i+1} are synonymous).*

For example, the adjectives 'good' and 'proper' are 2-related since there exists a 3-long sequence $\langle \mathsf{good}, \mathsf{right}, \mathsf{proper} \rangle$. Two words may of course be related by many different sequences, or by none at all. We will mainly be interested in the shortest sequences relating words. The minimal path-length (MPL) of two words w_i and w_j is n if there is an $(n+1)$-long sequence relating w_i and w_j and there is no sequence with length $\leq n$. If there is no sequence relating the two words, then the minimal path-length is undefined.

Definition 2 *Let MPL be a partial function such that $\mathsf{MPL}(w_i, w_j) = n$ if n is the smallest number such that w_i and w_j are n-related.*

The minimal path-length enjoys some of the geometrical properties we might expect from a distance measure.

Observation 1 *The minimal path-length is a* metric*, that is, it gives a non-negative number $\mathsf{MPL}(w_i, w_j)$ such that*

i) $\mathsf{MPL}(w_i, w_j) = 0$ *if and only if* $w_i = w_j$,

ii) $\mathsf{MPL}(w_i, w_j) = \mathsf{MPL}(w_j, w_i)$, *and*

iii) $\mathsf{MPL}(w_i, w_j) + \mathsf{MPL}(w_j, w_k) \geq \mathsf{MPL}(w_i, w_k)$.

The minimal path-length is a straightforward generalization of the synonymy relation. The synonymy relation connects words with similar meaning, so the minimal distance between words says something on the similarity of their meaning. For example, using WordNet we now find that

- $\mathsf{MPL}(\mathsf{good}, \mathsf{proper}) = 2$,
- $\mathsf{MPL}(\mathsf{good}, \mathsf{neat}) = 3$, and
- $\mathsf{MPL}(\mathsf{good}, \mathsf{noble}) = 4$.

This suggest that we can use the MPL distance measure for determining Osgood's evaluative dimension, for example by scoring words that are closely related to the words 'good' and 'bad' respectively. That is, we might consider using the distance to the word 'good' as a measure of 'goodness.' This makes sense considering the SYNSET relation in WordNet is representing similarity of meaning, and our MPL is a straightforward generalization of the SYNSET relation.

Figure 1 shows the minimal-path lengths of a selection of adjectives to the adjective 'good' based on the WordNet database.[1] Inspection of such a cloud of words gives us some confidence in the use of MPL as a measure for similarity of meaning. Note that we do not claim that the values obtained in this way are a precise scale for measuring degrees of goodness. Rather, we only expect a weak relation between the words used to express an positive opinion and their distance to words like 'good.'

However, further experimentation quickly reveal that this relation is very weak indeed. It turns out that the similarity of meaning waters down remarkably quick. A striking example of this is that we also find that 'good' and 'bad' themselves are closely related in WordNet.

Observation 2 *There exists a 5-long sequence $\langle \mathsf{good}, \mathsf{sound}, \mathsf{heavy}, \mathsf{big}, \mathsf{bad} \rangle$. So, we have that $\mathsf{MPL}(\mathsf{good}, \mathsf{bad}) = 4$.*

[1]To be more precise: these are all adjectives w with $\mathsf{MPL}(\mathsf{good}, w) \leq 2$ and word familiarity or polysemy count ≥ 2.

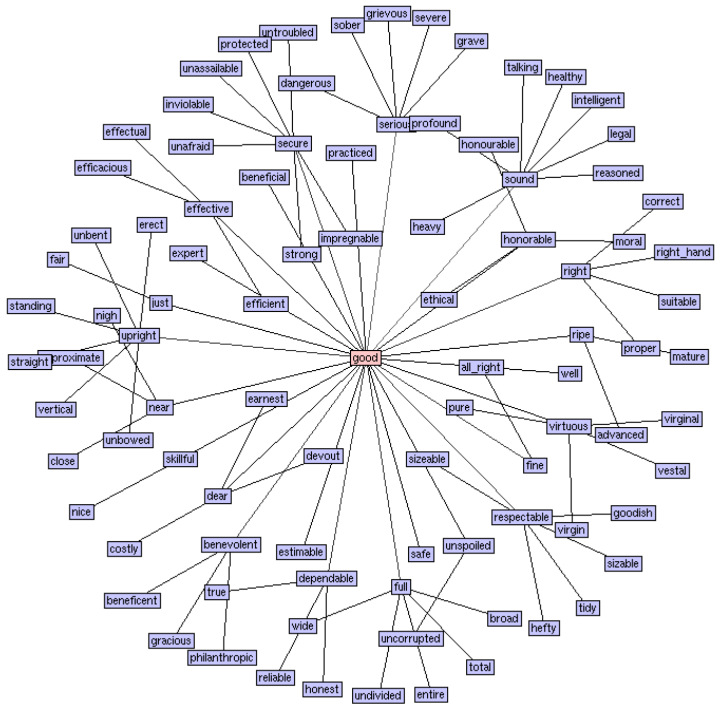

Figure 1: Part of the WordNet database from the vista point of adjective ‘good.’ The edges are SYNSET relations, nodes are only connected by a shortest path.

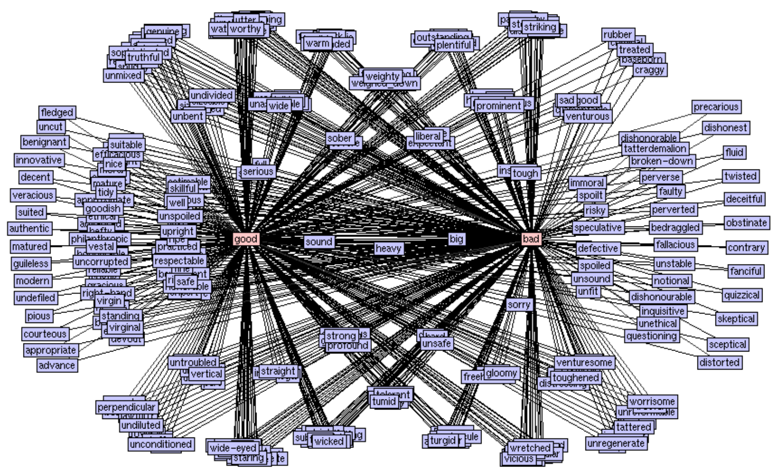

Figure 2: The MPL’s to adjectives ‘good’ and ‘bad’. Nodes are connected by edges of length corresponding to the MPL.

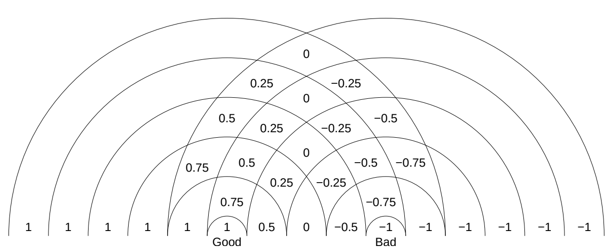

Figure 3: The values assigned by the EVA function.

Even though the adjectives 'good' and 'bad' have opposite meaning—they are antonyms—they are still closely related by the synonymy relation.[2] As a result of this, we must seriously question whether the relatedness to the word 'good' is a measure of 'goodness,' since any word related to 'good' is at most slightly less close-related to 'bad.'

Observation 3 *For any w, if* $\mathsf{MPL}(\mathsf{good}, w) = n$ *then* $n - 4 \leq \mathsf{MPL}(\mathsf{bad}, w) \leq n + 4$.

We seem to be at a dead-end: the WordNet database gives us similarity of meaning by its SYNSET relation, but its straightforward generalization MPL fails to provide a general measure of coincidence of meaning.

At this point several strategies present themselves. We might argue that, despite of observation 3, we may still expect some correlation between the opinion expressed in a text, and (a refined version of) a distance measure like MPL. Here, we will pursue an alternative strategy based on the fact that any word that is related to the adjective 'good' is also related to the adjective 'bad' (and *vice versa*). That is, we will use observation 3 to our advantage.

For each word, we can consider not only the shortest distance to 'good' but also the shortest distance to the antonym 'bad.' Figure 2 shows the minimal-path lengths of words to both the adjectives 'good' and 'bad.'[3] Inspection reveals that words neatly cluster in groups depending on the minimal path-lengths to 'good' and 'bad'. In short, this sort of graphs seems to resonate closely with an underlying evaluative factor (at least, much better than graphs based on a single distance measure such as figure 1).

We try to materialize this impression by defining a three argument function TRI that measures the relative distance of a word to two reference words.

Definition 3 *We define a partial function* TRI *of w_i, w_j, and w_k (with $w_j \neq w_k$) as*

$$\mathsf{TRI}(w_i; w_j, w_k) = \frac{\mathsf{MPL}(w_i, w_k) - \mathsf{MPL}(w_i, w_j)}{\mathsf{MPL}(w_k, w_j)}$$

If any of $\mathsf{MPL}(w_i, w_j)$, $\mathsf{MPL}(w_i, w_k)$, *or* $\mathsf{MPL}(w_k, w_j)$ *is undefined, then* $\mathsf{TRI}(w_i; w_j, w_k)$ *is undefined.*

We calculate the function TRI based on two reference words (w_j and w_k in definition 3). The maximal difference in minimal-path length to the two reference words depends on the MPL of the two reference words (by observation 3). Therefore, we divide the difference by the MPL of the two reference words, yielding a value in the interval $[-1, 1]$. In particular, we will be interested in the partial function TRI instantiated for the reference words 'good' and 'bad'. Recall that these two words correspond to Osgood's evaluative factor.

Definition 4 *We define a partial function* EVA *of w as* $\mathsf{EVA}(w) = \mathsf{TRI}(w; \mathsf{good}, \mathsf{bad})$.

We now have that every word, provided it is related to the adjectives 'good' and 'bad,' will be assigned a value ranging from -1 (for words on the 'bad' side of the lexicon) to 1 (for words on the 'good' side of the lexicon). Figure 3 shows how the EVA function assigns values based on the minimal-path lengths from adjectives 'good' and 'bad.'

For example, using WordNet we now find the following measures:

- $\mathsf{EVA}(\mathsf{proper}) = \mathsf{TRI}(\mathsf{proper}; \mathsf{good}, \mathsf{bad}) = \frac{\mathsf{MPL}(\mathsf{proper}, \mathsf{bad}) - \mathsf{MPL}(\mathsf{proper}, \mathsf{good})}{\mathsf{MPL}(\mathsf{good}, \mathsf{bad})} = \frac{6-2}{4} = 1$,
- $\mathsf{EVA}(\mathsf{neat}) = \frac{3-3}{4} = 0$,
- $\mathsf{EVA}(\mathsf{noble}) = \frac{5-4}{4} = 0.25$,

[2] Although this is perhaps remarkable, it is not due to some error in the WordNet database (there exist several paths of length 5). Part of the explanation seem to be the wide applicability of these two adjectives (WordNet has 14 senses of bad and 25 senses of good). Think of the small world problem predicting mean distance of 6 between arbitrary people (Milgram, 1967).

[3] To be more precise: these are all adjectives w with $\mathsf{MPL}(\mathsf{good}, w) \leq 3$ or $\mathsf{MPL}(\mathsf{bad}, w) \leq 3$, and with word familiarity or polysemy count ≥ 2.

- $\mathsf{EVA}(\mathsf{good}) = \frac{4-0}{4} = 1$, and
- $\mathsf{EVA}(\mathsf{bad}) = \frac{0-4}{4} = -1$.

Note that we do not claim that the EVA function assigns a precise measure of the 'goodness' or 'badness' of individual words (if such a thing is possible at all). Rather, we can only expect that it allows us to differentiate between words that are predominantly used for expressing positive opinions (values close to 1), or for expressing negative opinions (values close to -1), or for neutral words (values around 0).

Recall that EVA is a partial function that is undefined for words that are unrelated to the adjectives 'good' and 'bad.' The unrelatedness of such words is a sign that they are indifferent for assessing the evaluative factor. That is, unrelated words can be regarded as neutral for the EVA function. We will complete the partial function EVA in precisely this way, and define a complete function $\mathsf{EVA}^\star$ that returns a value for any arbitrary word.[4]

Definition 5 *The function* $\mathsf{EVA}^\star$ *is defined as follows:*

$$\mathsf{EVA}^\star(w) = \begin{cases} \mathsf{EVA}(w) & \textit{if defined} \\ 0 & \textit{if undefined} \end{cases}$$

Recall that we are mainly interested in evaluating larger textual units. A straightforward aggregation procedure is to view a text as a bag of words, evaluate each a these individual words, and simply add up their scores. Slightly abusing our notation, we will generalize the $\mathsf{EVA}^\star$ function to apply to arbitrary sequences of words.

Definition 6 *Let* $\langle w_1, \ldots, w_n \rangle$ *be a bag of words. We define the function* $\mathsf{EVA}^\star$ *as follows:*

$$\mathsf{EVA}^\star(\langle w_1, \ldots, w_n \rangle) = \sum_{i=1}^{n} \mathsf{EVA}^\star(w_i)$$

We now have a function $\mathsf{EVA}^\star$ that gives us a value for any arbitrary text. The precise interpretation of this value is not immediately clear, because it depends on how well our operationalization captures the concept of meaning we set out to measure (which was not very precisely defined to start with). Although $\mathsf{EVA}^\star$ function yields a specific value, we will be happy to use it as a coarse-grained ordinal scale. For example, by classifying text as positive, neutral, or negative, depending on the sign of the $\mathsf{EVA}^\star$ function.

4 Implementation and Evaluation

In the previous section, we have defined a function $\mathsf{EVA}^\star$ that gives a measure for the evaluative factor of meaning expressed in a text. To apply this measure in practice would require us to calculate a large number of minimal path-lengths between words (recall the definition of $\mathsf{EVA}^\star$ in terms of TRI and MPL). Calculating a large number of minimal path-lengths is far from trivial in a large network like the WordNet database. Especially since many words will not be related to the adjectives 'good' and 'bad,' which is the hardest case to establish. To make this problem feasible, we compile lists of words related to 'good' and 'bad,' either up to a particular MPL, or all related words. Words not occurring on this list have $\mathsf{EVA}^\star$ value zero, and can be safely ignored.

For this purpose, we have implemented a set of scripts that can efficiently generate related words by their MPL. The script starts with a particular word (such as 'good') and recursively generates all synonyms while filtering away words it has encountered earlier. That is, we start with a particular word w (i.e., having minimal path-length zero to itself), then generate all words w_i with $\mathsf{MPL}(w, w_i) = 1$,

[4]To be more precise, following WordNet we use the SYNSET relation only for words with the same part-of-speech (nouns, adjectives, verbs, adverbs), and only consider adjectives for EVA and $\mathsf{EVA}^\star$. That is, the $\mathsf{EVA}^\star$ of a verb or noun is zero.

then with $\mathsf{MPL}(w, w_i) = 2$, etcetera, until the search exhausts, or until we reach a given maximal value of MPL. The script has an additional argument that allows us to ignore words with a low polysemy count. By running this script on two related words (such as 'good' and 'bad'), we will have determined the minimal path-lengths needed for calculating the weight of all related words. The resulting list of rated words can be stored in a file for future use.

In particular, we can run these scripts exhaustively on the adjectives 'good' and 'bad.' As it turns out, this generates a list of 5410 adjectives (or a component in graph-theoretical terms).

Observation 4 *The set of words which are n-related to the adjectives 'good' and 'bad' (for some n) consists of 5410 adjectives.*

The adjective cluster in which 'good' and 'bad' reside, contains 25% of the adjectives in the WordNet database.[5] For each of these words, we can assign a weight corresponding to the evaluative factor of meaning: the EVA function assigns a value in the interval $[-1, 1]$, with positive values for words on the 'good' side and negative values for words on the 'bad' side. Note that this exhaustive list will completely determine the $\mathsf{EVA}^\star$ function: all words not on this list will have $\mathsf{EVA}^\star$ value zero. This allows us to efficiently calculate the $\mathsf{EVA}^\star$ function of a text.

The exhaustive list of adjectives related to 'good' and 'bad' is also useful in its own right. We can use such lists for further evaluation of the constructed measures. In particular, one may suspect there to be a bias towards one of the bipolar adjectives, simply by the number of words in the WordNet database. This is not unlikely considering that the WordNet database gives 35 synonyms of the adjective 'good,' and only 15 synonyms of 'bad.' Using the exhaustive list of all 5410 adjectives related to 'good' and 'bad,' we can simply add up each word's assigned value. Recall that these values range from -1 to 1, so if the amounts of positive and negative words are completely balanced, the grand total will be zero, making the mean value assigned to a word zero as well. It turns out that the total score over 5410 words is -48.25, yielding a mean value of $\frac{-48.25}{5410} = -0.0089$.[6] This is only a marginal deviation, so we may conclude that the list of words is well-balanced between the two opposite words. In light of the resemblance of the WordNet database structure to human lexical memory, this finding increases our confidence that the $\mathsf{EVA}^\star$ measure is corresponding to an evaluative aspect of meaning. This relates to one of the problems left unsolved in Osgood et al. (1957, p.327).

> One of the most difficult methodological problems we have faced—unsuccessful so far—is to demonstrate that the polar terms we now use are true psychological opposites, i.e., fall at equal distances from the origin of the semantic space and in opposite directions along a straight line passing through the origin.

Almost half a century later, our measure based on the WordNet database provides some indirect evidence for this.[7] In this sense, our work can also be viewed as a partial evaluation of Osgood's original semantic differential technique.

The same set of scripts also allows us to compile lists for the other factors of meaning. For Osgood's potency factor, the prototypical operationalization is using the adjectives 'strong' and 'weak.' As it turns out, 'strong' and 'weak' are antonyms in WordNet, but also related by the synonymy relation.

Observation 5 $\mathsf{MPL}(\mathsf{strong}, \mathsf{weak}) = 6$

We can define a function $\mathsf{POT}^\star$ as follows:

---

[5] Our version of WordNet, 1.7, has 21365 adjectives (i.e., when counting unique strings), so the cluster surrounding 'good' and 'bad' is 25.32% of the total collection of adjectives. Generating the exhaustive list for adjectives 'good' and 'bad' takes 6 minutes and 19 seconds on a Pentium-III 800Mhz with 512 MB memory running Red Hat Linux 7.0.

[6] Perhaps we can make this more clear by estimating the number of words with the 'wrong' sign. Since negative words range from -1 to 0, the average weight of a negative word is -0.5. So we may estimate the excess of negative words to be $\frac{-48.25}{-0.5} = 96.5$ words, which is 1.78% of the total number of words in the list. This amounts to flipping the sign of 48 words in the list.

[7] At least, this seems to be the case for the English language, it is unclear whether there are significant differences in other languages or cultures. This could be investigated using the multi-lingual versions of EuroWordNet (Vossen, 1998).

Definition 7 *The function* $\mathsf{POT}^\star$ *of a word* w *is defined as follows:*

$$\mathsf{POT}^\star(w) = \begin{cases} \mathsf{TRI}(w; \mathsf{strong}, \mathsf{weak}) & \textit{if defined} \\ 0 & \textit{otherwise} \end{cases}$$

Let $\langle w_1, \ldots, w_n \rangle$ *be a bag of words. We define the function* $\mathsf{POT}^\star$ *of a tuple* $\langle w_1, \ldots, w_n \rangle$ *as:*

$$\mathsf{POT}^\star(\langle w_1, \ldots, w_n \rangle) = \sum_{i=1}^{n} \mathsf{POT}^\star(w_i)$$

The third major factor of meaning, Osgood's activity factor, is usually operationalized using the two adjectives 'active' and 'passive.' Again, adjectives 'active' and 'passive' are antonyms in WordNet, but also related by the synonymy relation

Observation 6 $\mathsf{MPL}(\mathsf{active}, \mathsf{passive}) = 12$

We define a function $\mathsf{ACT}^\star$ just like $\mathsf{EVA}^\star$ and $\mathsf{POT}^\star$ but now with the reference words 'active' and 'passive.' Specifically, we will use the $\mathsf{TRI}(w; \mathsf{active}, \mathsf{passive})$ function yielding a value 1 for $\mathsf{ACT}^\star(\mathsf{active})$, and -1 for $\mathsf{ACT}^\star(\mathsf{passive})$.

Similar to the evaluative factor, our set of scripts generates lists of all related adjectives for the potency and activity factors. Investigating these three lists, we immediately discover the following, remarkable finding.

Observation 7 *All three lists corresponding to* $\mathsf{EVA}^\star$, $\mathsf{POT}^\star$, *and* $\mathsf{ACT}^\star$ *functions single-out the same cluster of* 5410 *related adjectives in WordNet.*

This cluster of words appears to have a special status: it contains all the important modifiers used to express emotive or affective meaning—to use our slogan, these are "words with attitude." Although the three measures use the same set of words, the distribution of weights is radically different. These weights for each of the three measures is calculated from different words, giving rise to different minimal path-lengths, and thus different values. For example, we find that

- $\mathsf{EVA}^\star(\mathsf{proper}) = 1.00$;
- $\mathsf{POT}^\star(\mathsf{proper}) = 0.50$; and
- $\mathsf{ACT}^\star(\mathsf{proper}) = 0.09$.

Our future research is to evaluate the measures of this paper, and refinements decorated with polynomial constants. Ideally, this should be done on a test corpus that has been rated on the affective or emotive meaning expressed in the texts. So far we have been unable to locate such a corpus, and are investigating ways to construct one ourselves. We have also done initial tests on texts found on Internet discussion sites (without taking into account negation, nor parsing the texts to determine syntactic categories of words). The first initial observation on this small test set is that there is correspondence between the measures and the meaning expressed. On the one hand, the measures are not flawless when considering individual texts. This is hardly surprizing since sometimes none or very few of the special adjectives occur in these short texts. On the other hand, however, over larger sets of texts the measure gives a much better impression (i.e., the false positives and false negatives seem to cancel out each-other). We need extensive empirical tests in order to qualify what the particular value means (i.e., can we distinguish degrees of goodness instead of more coarse-grained distinctions). Since scores increase with the length of a text, it is clear that some normalization for the length of a text is needed for considering the value to indicate the degree of goodness. Another initial observation is that, although the set of words is well-balanced between the opposing sides, there appears to be a bias towards the good-side of the evaluative factor. That is, there seems to be a tendency to expound negative judgments more concisely than positive judgments. The existence of an asymmetry between positive and negative deviations is also known from judgments under uncertainty, think of prospect theory (Tversky and Kahneman, 1974). If a similar bias in positive word-choice exists, we can easily compensate for it by the relative weight we assign to words.

5 Conclusions and Discussion

In this paper, we investigated measures for affective or emotive aspects of meaning derived from the structure of the WordNet lexical database. Such a project presupposes that subjective aspects of meaning can be derived from the choice of words in a text. That is, there are indeed words with attitude or values. This is not undisputed, some philosophers have been skeptical whether different people's words can mean the same (Quine, 1960).[8] Our focus on texts, rather than other modes of communication, gives some confidence that certain aspects of the expressed meaning can be derived from the particular word choice. One of the types of texts we consider interesting are texts on Internet discussion sites: here there is a strong incentive for the writer to make sure that a reader can grasp the intended meaning from the textual content. Even the mere existence of such discussion sites can be viewed as evidence that this is the case.

Mainstream research in natural language processing deemphasizes more subjective aspects of meaning (Manning and Schütze, 1999; Jurafsky and Martin, 2000). Our work can be viewed as an attempt to rectify this. A consequence of going beyond the established notion of lexical meaning, is that there is no consensus on notions of affective or emotive meaning. So it is not immediate clear what notions to use. We decided to go back to one of the seminal works on measuring affective meaning, Osgood et al. (1957)'s semantic differential technique. From this, we took some of the most important factors of affective meaning, the evaluative, potency, and activity factors. The second crucial ingredient is our use of the WordNet lexical database (Miller, 1990; Fellbaum, 1998). The basic notion of meaning used in WordNet is lexical meaning, and WordNet's main SYNSET relation is denoting coincidence of lexical meaning. However, it is important to stress that WordNet is partly inspired by psycholinguistic theories of human lexical memory. That is, the meaning of words is also determined by its place in the larger structure of the database. Also note that this larger structure shows some resemblance with our own lexical memory. In this paper, we have translated this structure back into concrete measures for the Osgood factors of meaning. All the three resulting measures single-out the same cluster of 5410 adjectives, which is 25% of the adjectives in WordNet. This cluster appears to have a special status: it contains all the important modifiers used to express affective or emotive meaning—these are words with attitude.

As it turned out, the measures we constructed are based on a distance metric. This relates our work to the ubiquity of measures of distance, similarity, or relatedness in natural language processing. To name a few, the use of path-length as a measure of similarity can be traced to (Quillian, 1968). The use of path-length as similarity metric also discussed in (Rada et al., 1989). A recent evaluation of five distance measure can be found in (Budanitsky and Hirst, 2001). Most measures of relatedness use more than just the synonymy relation. For our purposes, this is not useful because it destroys the bipolarity of the concepts we are interested in. For example, all our pairs of adjectives are directly related by the antonymy relation, and one may suspect a close common hypernym. Although there is similarity with the traditional distance measures use in NLP, it is important to stress that we use these measures for different purposes. Already Quillian (1968, p.228) has it that

[8] This reminds of (Carroll, 1871, Chapter 6):

> "When *I* use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean—neither more nor less."
> "The question is," said Alice, "whether you *can* make words mean so many different things."
> "The question is," said Humpty Dumpty, "which is to be master—that's all."

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

We have shown in this paper how a measure for the affective meaning studied by Osgood can be derived from a representation of the relatively "objective" meaning as represented in the WordNet database.

Acknowledgments This research was supported by the Netherlands Organization for Scientific Research (NWO, grants # 400-20-036 and # 612-62-001). Thanks to Patrick Blackburn, Michael Masuch and Ivar Vermeulen for their comments, and to Rob Mokken for suggesting Osgood's semantic differential technique. All data is derived from Princeton WordNet 1.7, using Perl and Dan Brian's excellent `Lingua::Wordnet` module.

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION
UNIVERSITY OF AMSTERDAM
NIEUWE ACHTERGRACHT 166
1018WV AMSTERDAM
THE NETHERLANDS
{kamps,marx}@illc.uva.nl

References

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

Lewis Carroll. 1871. *Through the looking-glass, and what Alice found there*. Macmillan, London.

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

Daniel Jurafsky and James H. Martin. 2000. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, Upper Saddle River NJ.

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

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

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

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

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

Willard Van Orman Quine. 1960. *Word and Object*. The MIT Press, Cambridge MA.

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

Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.

Piek Vossen. 1998. Special issue on EuroWordNet. *Computers and the Humanities*, 32(2/3):73–251.