Sentiment Analysis Exploring Metaphorical and Idiomatic Senses: A Word Sense Disambiguation Approach

Vassiliki Rentoumi[1] and **Vangelis Karkaletsis**[1] and **George A. Vouros**[2] and **Amalia Mozer**[3]

Abstract. Subjectivity and polarity are very important aspects of the language. We propose that these language properties are often expressed through non conventional utterances and in particular through metaphors, idioms and expanded senses. In this position paper, we bring empirical evidence that words that digress from their literal senses appear in subjective contexts and bear strong polarity. In order to capture these empirical results we propose a methodology for sentiment analysis which uses Word Sense Disambiguation techniques (WSD), and incorporates our linguistic principle in order to detect subjectivity of senses and then assign polarity scores to a phrase level. Moreover preliminary results concerning the adopted WSD methods for sentiment analysis are also presented.

1 Introduction

There is growing interest in the automatic detection of subjectivity in written speech, which is the linguistic expression of somebody's emotions, sentiments, evaluations, opinions, beliefs, speculations, and generally of private states. Moreover, there are efforts in the automatic extraction of sentiments, involving the expression of positive or negative opinion, emotion or evaluation towards a topic. The necessity of exploring word senses in order to solve ambiguity, constitutes a high priority need in NLP applications. Recent research efforts [16] have shown that subjectivity constitutes a property of the language that is correlated with word senses, as both are considered semantic properties of written speech. We also support that senses are a valuable indicator of sentiment. We believe that their role can reveal new semantic properties of language.

In order to investigate that, we need to examine different aspects of senses. We are intuitively guided to the investigation of the role of non literal senses, such as metaphors and idioms, as well as expanded senses. We aim to find out through the study of real data whether these types of senses tend to occur in subjective phrases, triggering subjectivity and polarity.

Moreover, in order to verify interaction among non literal senses, subjectivity and polarity, we propose a new methodology for automatic sentiment analysis which uses state-of-the-art methods for Word Sense Disambiguation (WSD).

[1] Inst. of Informatics and Telecommunications, NCSR "Demokritos", Greece, email:{vrentoumi| vangelis}@iit.demokritos.gr
[2] University of Aegean, Department of Information and Communication Systems Engineering, AI-Lab, Greece, email: georgev@aegean.gr
[3] School of Philosophy, Faculty of Philology, Department of Linguistics, National and Kapodistrian University of Athens, Greece, email: amozer@cc.uoa.gr

We aim for our proposed method for sentiment analysis to function as a complementary method for sentiment analysis to already existing ones, as it handles the subcategory of non literal senses which are thought to be more subtle and difficult to determine their polarity. Especially with respect to polysemous words seem ambiguous in terms that their polarity can vary depending on the sense they activate within a particular context.

In what follows, we start by presenting the State-of-the-art in section 2, while in section 3 we pose and verify our research hypothesis. In 3.1 we provide our current work that is the corpus evaluation and in 3.2 we present empirical results derived from the corpus. In section 4 we briefly present the steps of our proposed method for sentiment analysis. The steps of our methodology are analytically presented in sections 4.1, 4.2, and 4.3. Section 4.1 concerns already implemented work, and 4.2 as well as 4.3 concern our immediate next steps. Then in section 5 some preliminary results derived from the application of WSD methods of polysemous words are also presented and evaluated. Finally, section 6 concludes the paper sketching plans and future directions.

2 State Of The Art

Much work on sentiment analysis focuses on locating subjectivity by detecting individual subjective elements such as words (e.g nouns, adjectives) [15]. Moreover, other work focuses on classifying phrases, sentences into subjective or objective classes [14], by considering the context in which subjective expressions appear.

On the other hand analogous research is performed in locating sentiment, and particularly in detecting positive or negative orientation in a word or at phrase level. In some approaches they try to detect the orientation of terms, by exploiting information derived from glosses (e.g glosses provided by Wordnet) [3]. Moreover, in order to identify polarity at a phrase or in a sentence level, they try to capture the correlation of already found polar terms of the phrase or sentence by using statistical measures [7]. In these approaches the polarity classification is achieved in a phrase level regardless of the broader context of a text. Moreover, the authors in [18] try to achieve a classification into positive/negative, of phrases and sentences, within the context of a text.

So far we have referred to methods that perform either sentiment or subjectivity analysis. Recent research approaches [4], [19] confront the task of combining subjectivity as well as sentiment analysis. Subjectivity analysis is considered as the intermediate step for sentiment analysis. This assumption seems to be very realistic, as in order to explore opinion orientation, we first have to detect the opinionated region, that is a context which bears subjectivity.

It is further investigated whether words can have neutral as well as sentiment bearing senses. Therefore, we look for subjectivity and polarity in senses rather than in words. We support that in sentiment analysis research we have to focus on annotating senses with polarity. New approaches deal with this by developping methods for the construction of sense-tagged lists [5]. In [16], an automatic method is performed for sense tagging based on Lin's similarity measure [9].

Our methodology aims to perform both subjectivity and sentiment analysis. Contrary to the above methods and by improving on the work presented in [16], we propose a method for assigning subjectivity scores as well as polarity labels to word senses. We expect that this information will lead us to capture contextual polarity. Our final aim is to tackle the writer's attitude (pos/neg) towards a topic and since we are investigating journalistic written speech - as we are going to see below - we intend to reveal journalist's stance towards topics of every day life.

3 Position and Verification Of Research Hypothesis

Given the fact that a word obtains its sense within a context, it could be considered reasonable to say, that a sense depending on the context could be objective or subjective. In our approach we claim that subjectivity as well as polarity are often hidden, in metaphorical and expanded **senses** of polysemous words.

In order to achieve subjectivity and polarity analysis we first have to disambiguate metaphorical and expanded senses of polysemous words.

Let us take a look at the following examples:

- The government measures could provoke social **earthquake**. (earthquake = disruptive event)
- The politician **sat on the fence** and would not give his opinion about the tax issue. (to delay making a decision when you have to choose between two sides in an argument or a competition)

We observe, that opinions are expressed and that these opinions are strongly polar. The first constitutes a speech event expressing a private state, while the second one is an expressive subjective element. Both types are considered as main types of subjective expressions [17].

The metaphorical sense of "earthquake" and the idiomatic sense of the expression "sit on the fence" are expressive subjective elements, which demonstrate the authors' negative sentiment without explicitly stating so. Thus, metaphors as well as idiomatic expressions can be used as expressive subjective elements in order to display sentiment (opinion) implicitly. These metaphorical and idiomatic senses are also polar as they strongly express an opinion [14].

Are metaphors and idioms subjective expressions? Metaphors are a way to describe something. They have a figurative power, meaning that they set before the eyes the sense that they display. A metaphor, according to Aristotle, is the act of giving a thing a name that belongs to something else. The expression of metaphors carries as inherent characteristics a person's imagination and feeling. Thus, the choice of a metaphor is not a random lexical choice but an option which reflects the speaker's feelings, imagination, stance, as well as psychological situation. Therefore, it is truly subjective. An idiom is an expression, a figure of speech, the meaning of which cannot be deduced from the literal definitions and the arrangement of its parts, but rather refers to a figurative meaning that is known only through common use. Idioms could also be considered as colloquium metaphors in the sense that they demand as backround knowledge, a common cultural experience. The author using this kind of expressions aims to trigger this common experience to the reader, to lead him with a blink of the eye to the comprehension of his opinion. Moreover idioms express familiarity through which an author aims to approach his audience and to make them apprehend his intented figurative meaning which conveys idioms' subjective character.

3.1 Corpus Evaluation

In order to verify our research hypothesis we studied the SemEval 07' corpus[4] which consists of headlines, described in [1]. We chose this corpus, as headlines use most of the times non literal expressions, ambiguous phrases as well as metaphors and idioms, in order to attract the reader's attention. Moreover, headlines bear a dualism, that is the fact and the comment upon the fact. Thus we assume that in such a corpus we expect to find the types of expressions that we are looking for. Moreover in recent bibliography there is not so much work on detecting polarity in journalistic written speech. We aim to use, in the future, other types of corpora such as movie reviews which contain more complicated lexical structures in order to evaluate our methodology. Given the fact that there has been a lot of work concerning subjectivity and polarity detection of literal senses, our aim is to perform such a task for non literal senses such as metaphors, idioms and expanded senses. We assume that in these types of expressions the expressed polarity is not obvious but is rather hidden in the sense of the word in the given context. For example the word great is usually assigned with positive polarity but according to statistics in WordNet, used neutrally 75% of occurences. This fact leads us to the need of adopting word sense disambiguation techniques instead of blindly assigning subjectivity and polarity without taking into account the actual sense of the word in the given context. In particular we checked a sample of 200 headlines participating in the affective test valence gold standard corpus [1]. Our aim is to confirm that within the examined headlines which are annotated with polarity (pos/neg), according to the gold standard valence corpus, exist non literal senses, such as metaphors, idioms, and expanded senses, a fact which denotes that these types of expressions appear in subjective contexts expressing positive or negative attitude. In this way we verify that humans can assign to these peculiar types of expressions spontaneous polarity, using an inherent mechanism of understanding of these types of words. The gold standard annotation gives us a hint, that behind these polysemous words does indeed exist polarity, and more specifically by checking the valence rating of the headlines containing these words this polarity is significant. The corpus is annotated with a valence, using a scale from -100, to +100. We found out that 55 out of 200 headlines contain idioms, metaphors as well as expanded senses which also bear negative or positive polarity.

Let us examine the following example derived from the corpus:

- Visitors to Eiffel Tower **climb to record** in 2006.

The expression "climb to record" constitutes a metaphor. From a linguistic point of view, this phrase comprises two metaphors. In the first one, visitors denote the number of persons who "climb". This yields a kind of metaphor called "personification" [8], as the abstract entity "number" is represented as a person. The second one is another well known type of metaphor, the so-called "structural metaphor" [8]. In this kind of metaphor one concept is understood

[4] http://www.cse.unt.edu/~ rada/affectivetext/

and expressed in terms of another structured, sharply defined concept. Here, the process of "increase", the route to success, is given in terms of climbing which denotes the rising number. The author uses this metaphor in order to describe the fact of reaching the maximum number of visitors so far in a subjective way. This is further proved as according to the gold standard corpus this headline is annotated with positive polarity score (+61).

- Sleep more to **fight** childhood obesity
- Republicans plan to **block** Iraq debate.

In the above two examples we have again the use of non literal word senses. We observe the use of expanded senses of the verbs "fight" and "block" respectively. In the first example, the verb "fight" is used with the sense of "to prevent something or to get rid of something unpleasant that already exists", that is to prevent obesity. Notably, this sense digresses enough from the literal sense of the verb fight which literaly means "to take part in a war or battle". Similarly in the second example the verb "block" is used with the sense of "to stop something from hapenning, developping or succeeding". Again, a semantic expansion of the very first sense of the verb "block" is used. The literal sense of "block" implies the prevention of moving through the placement of physical obstacles. In this case the journalist makes this lexical choice in order to broadcast a news' item as well as to add his personal comment. The sentencial truth behind this phrase is that debate is prevented by republicans. This content denotes the fact itself, but the way this utterance is expressed using the semantic expansion "block", describes how the fact is perceived and expressed intentionally by the journalist, thus displaying the journalist's negative attitude. Both the above examples are annotated according to the gold standard corpus, with 31 (positive) and -57 (negative) valence respectively.

- Aquarium **puts ailing beluga whale to sleep**
- Firms **on alert** for letter bombs

Both the above phrases use idioms. The meaning of "put to sleep" denotes that the whale was killed gently, with euthanasia. Additionally the idiomatic expression "on alert", constitutes a warning. Specifically here, it describes that firms have to be prepared for danger, emergency. Both the above idiomatic phrases are annotated with significant negative valence -44 and -70 respectively. As it can easily be deduced, it is difficult to detect the sentiment expressed in all the above cases, as they are polysemous, idiomatic and metaphorical. It is obvious that in order to decide for their polarity we have to find their meaning, the senses that these expressions yield within the context. These senses can be obtained through disambiguation.

3.2 Empirical Results

According to the valence scale, with which the headlines are annotated, 0 represents a neutral headline, -100 represents a higly negative headline, and 100 represents a higly positive headline. This data set was independently annotated by six annotators. The annotators were instructed to follow their first intuition in order to annotate and to use the full range of annotation scale bar. In order to conduct interannotator agreement an agreement evaluation was followed using the Pearson correlation measure [1]. In order to measure the agreement among the annotators, the agreement between the first annotator and the average of the remaining five is first measured, followed by an average over the six resulting agreement figures.

In Table 1 we see that 55 out of 200 examined headlines bear metaphors, idioms as well as expanded senses.This constitutes a rather significant sample, even though it was extracted from a small corpus (200 headlines), showing the importance of capturing the polarity that such expressions exhibit. We are interested in capturing the polarity of such expressions since current techniques for sentiment analysis that do not take under consideration the words' senses within a context (e.g using pre annotated seed words with pos/neg), could easily fail or assign wrong polarity to these types of expressions. The rest of the headlines bear literal senses and thus are out of scope of our research hypothesis, since we assume that literal senses express more obvious polarity which can be handled efficiently using state of the art methods for polarity detection, without using so fine grained techniques as WSD.

Moreover in Table 1, average polarity (PosAvg/NegAvg) results are presented for each of the three categories (metaphors, idioms and expanded senses). We deduce from the results presented in Table 1 that the types of expressions detected display significant polarity. In the given phase of our research we do not examine the strength of polarity that these types display, but rather if they bear pos/neg polarity. Moreover if these phrases are assigned a significant polarity score according the valence scale adopted, this fact enforces our very first hypothesis that these types of expressions are used in subjective contexts and bear opinion. In particular, it is empirically observed that headlines containing metaphors bear in average the highest polarity score, and appear less often than expanded senses, which seem to be used more often and express at the same time milder polarity than metaphors. On the other hand idioms appear to be used less often and are accompanied by lower polarity.

	Total	Pos	PosAvg.	Neg	NegAvg.
Metaphors	19	12	40,33	7	-43,43
Idioms	9	4	28,25	5	-27,4
Expanded Senses	27	13	39	14	-42,8
TOTAL	55				

Table 1. Average Polarity on Examined Types Of Headlines

4 The Proposed Method For Sentiment Analysis

Below our proposed method for sentiment analysis is presented which could be developped within the following steps: the first step concerns the application of word sense disambiguation techniques - we tested two different approaches for WSD and we present some preliminary results. The second step concerns the assignment of subjectivity to word senses which is based on the results derived from WSD. The third step of our methodology concerns the description of polarity detection methodology in a phrase level, that is the author's attitude towards a topic as this is expressed through newspapers' headlines. The second and third step both constitute the immediate next steps of our research.

4.1 First step: Choose a state-of-the art method for WSD

In order to capture all the above cases of expanded senses, metaphors and idioms, we are going to employ state-of the-art WSD methods. First we used Lin's method [9], which is based on the distributional hypothesis that words that have similar senses tend to appear in similar contexts (distributionally similar words). We applied this method in order to find synonyms for expanded senses, metaphors, and idioms, that appear in the corpus. As the linguistic information derived from the headlines was not enough in order to obtain these "contextual synonyms", for the words that we want to disambiguate, we implemented Lin's methodology to the articles containing these headlines. We then applied this method in an expanded corpus by adding the MPQA corpus[5] consisting of news articles (535 texts, 11114 sentences) This method - being the first step of the disambiguation procedure - we hoped that would function as a filter, since the output consists of a list of words accompanied with similarity scores, from which we intented to select the top ranked synonyms in order to apply a sense similarity measure.

Then we used a second method for WSD, since the first method did not give adequate results (Section 5). We chose a method that assigns every word in a sentence the sense that is most related to the sense of its neighboring words. Disambiguation is performed in a phrase level according to a selected context window. WordNet is used as its sense inventory[6]. In particular, the selected method implements an algorithm for Word Sense Disambiguation that uses measures of semantic relatedness[7],[8]. The algorithm is an extension of an algorithm described by Pedersen, Banerjee, and Patwardhan in [12]. The difference is that this version disambiguates every word in the given context rather than just a single word. This algorithm disambiguates all the words in the specified context and returns them as a list. If a word cannot be disambiguated, then it is returned "as is". A word cannot be disambiguated if it is not in WordNet.

This particular algorithm finds its root in the original Lesk's algorithm which disambiguates a polysemous word by selecting that sense of the target word whose definition has the most words in common with the definitions of other words in a given window or context. The intuition behind this measure is that related word senses are defined using similar words and overlaps in their definitions will indicate their relatedness. This algorithm is a generalized version which could perform disambiguation accepting as input any relatedness measure which returns a similarity score for a pair of senses. We made this choice as in recent bibliography [12], it is shown that some of the semantic-relatedness measures used by this algorithm give promising results concerning WSD.

The algorithm, takes as input a context window through which disambiguation will be performed. We used a context window of 8 words, - which means that we take under consideration four words before and three after a target word - as the mean length of each headline which contains the words to be disambiguated consists of 8-10 words. Given the fact that our words in which we were interested were located in the middle to the end of each headline, we wanted for the disambiguation to exploit the semantic relations (e.g. is-a relations, or relations between glosses) of our target word with all words in context. Then we had to specify as input a single sentence per run, e.g. "Sleep more to fight obesity". Disambiguation is performed only for content words of the sentence since WordNet do not provide information for function words. Thus in the above sentences, the words sleep, more, fight and obesity are to be disambiguated. Then the algorithm accepts as input a relatedness measure m.

Given a measure the algorithm computes a score for each sense of a target word (e.g fight1, fight2) as follows:

[5] http://www.cs.pitt.edu/mpqa/databaserelease/
[6] http://wordnet.princeton.edu/perl/webwn
[7] http://www.d.umn.edu/ tpederse/senserelate.html
[8] http://search.cpan.org.dist/WordNet-Similarity

For each neighboring word, relatedness measure m is used to get similarity score between fight1 and each sense of the words sleep, more, and obesity, m(fight1, more1), m(fight1, obesity1), m(fight1, sleep1), m(fight1, sleep2). For simplicity's sake "more" and "obesity" have one sense each and "sleep" has two senses. Then the highest of these scores are picked and added at the overall score of target's word sense. That is $score(fight1) = m(fight1, more1) + (fight1, obesity1) + max\{m(fight1sleep1), m(fight1sleep2)\}$ and $score(fight2) = m(fight2, more1) + (fight2, obesity1) + max\{m(fight2sleep1), m(fight2, sleep2)\}$. Then the algorithm compares the $score(fight1)$ and $score(fight2)$ and picks as proper sense for fight the one that provides the highest similarity score. Since this algorithm performs disambiguation for all words of an input sentence, this procedure is repeated in order to disambiguate every word in context.

This method exploits several Wordnet-based similarity measures, that are based on path, information content derived from a large corpus, and word sense glosses. Path based measures, given a subsumption hierarchy, measure the similarity of two senses, by using the shortest path between two words in WordNet graph. Moreover Information Content (IC) based measures, such as the ones we tested i.e. lin[10], res[13] and jcn[6] measure the specificity assigned to each concept of a hierarchy. A concept with a high information content is very specific to a particular topic, while concepts with lower information content are associated with more general, less specific concepts. The information content of a concept is estimated by counting the frequency of that concept in a large corpus and through that to determine its probability via a maximum likelihood estimation. The information content of a concept is defined as the negative log probability of the concept: $IC(concept) = -log(P(concept))$. Additionally the frequency of a concept includes the frequency of all its sub-concepts since the count we add to the concept is also added to its subsuming concepts. Thus, the more information content a concept has, the less probable is to appear in a corpus, and thus the most specific it is. This fact also means that it is a leaf concept in the hierarchy. According to these measures the similarity between two senses depends on the degree of their mutual specificity as well as on their position in the hierarchy. The third category of relatedness measures which was applied (vector[11], Lesk extended gloss overlaps [2]) measures the similarity of senses based on word sense glosses, and in particular, what determines similarity is gloss overlaps, i.e. how many consecutive words are identical between word sense glosses. Both Vectors and Extended Gloss Overlaps measures extend the above mentioned principle. The Gloss Vectors measure is analysed below. Extended Gloss Overlaps is based on the assumption that there are relations between concepts that are implicit, but can be found via gloss overlaps. Through Extended Gloss Overlaps it is possible to find overlaps not only between the definitions of two concepts being measured, but also among these concepts to which they are related.

We tested similarity measures of each category (Path, IC, Gloss Based) on our headline corpus (55 headlines). Path based and Information Content measures did not produce succesful results especially for verbs, since verb hierarchies in WordNet are shallow and numerous. As a result very few verbs occupy the same hierarchy and there are rarely paths between verb concepts. We will not present results from all measures tested, as most of Path based measures as well as Information Content based measures generally struggled. From the two gloss based measures tested, only Gloss Vectors results are presented, since Extended Gloss Overlaps did not perform well. We suppose that this is happenning as glosses in some cases may not contain enough overlaps to make a safe decision [12].

We will focus on measures that performed quite well and these were gloss based measures, namely Gloss Vectors measure which performs best for verbs, and Resnik's measure which is an Information Content based measure that performs quite well for nouns. Gloss Vectors measures the similarity of concepts by finding the cosine between their respective gloss vectors. More analytically, Gloss Vector measure treats the WordNet glosses as a corpus of plain text. The first step is to build a co-occurence matrix of the words that appear in the corpus. Every cell in the matrix represents the frequency of the co-occurence of any two words in a WordNet gloss. In order to measure relatedness of two concepts, a vector is created for each of the glosses as follows: first, every word in the gloss is represented as a vector, using the entire row for each word in the co-occurence matrix. After the creation of all these word vectors their average vector is found which in fact represents the meaning of the concept. The average vector, in fact, shows with which words in average, the gloss of the concept we are interested in co-occurs. Therefore, the similarity of two concepts is measured by the cosine similarity of their corresponding gloss vectors.

Moreover, Information Content and in particular Resnik's measure performed quite well for nouns. According to Resnik's measures two concepts are semantically related proportional to the amount of information they share in common. The common information is determined by the maximum information content of their lowest common subsumer. Information Content measure applies to subsumption relationships of WordNet which are meant to be the richest relations in WordNet. Moreover the relative success of Resnik's measure for nouns could be attributed to the fact that for nouns there are rather deep hierarchies in Wordnet which are joined in a single artificial root. In this way the hierarchy becomes richer permitting safer decisions upon similarity.

4.2 Second step: Subjectivity Assignment

In the second step of our methodology we intend to use statistical means, in order to assign subjectivity to word senses. Our approach, on the other hand, exploits the word senses, using linguistic as well as semantic empirical assumptions. It is based on the hypothesis that words, when they are used with their metaphorical or expanded meanings tend to participate in subjective contexts.

We can further model the above principle by assigning subjectivity to those senses that digress - according to a vocabulary - from their literal sense. These senses are going to be assigned a higher subjectivity score. For this reason, vocabularies which are structured according to the principle that senses are classified from their literal to more expanded senses will be used. In this way, objective senses will be filtered from subjective ones, and considering that subjective senses tend to participate in subjective contexts, we may filter subjective from objective phrases. The subjective states (e.g. headlines) will be further analysed into positive and negative expressions of sentiments.

4.3 Third Step: Polarity Detection

In the third step of our methodology we will use the features that have been extracted from the above two steps, in order to detect polarity. As we have already pointed out our final aim is polarity detection, in terms of detecting an author's negative or positive attitude towards a topic rather than measuring its strength. In particular, the definitions of word senses will be exploited considering that they are very good indicators of polarity orientation as they bear enough semantic and lexical information. We are going to exploit these definitions (and glosses) in order to extract the most useful linguistic features that would be good indicators for sentiment analysis.

We are going to adopt a combined method - linguistic approach with machine learning - that will exploit sense definitions, as semantic indicators of sentiment. More specifically, the proposed method will annotate a few patterns (adverbs-verbs, adjectives-nouns, etc.) from definitions and glosses and use bootstrapping algorithms to annotate all definitions with positive or negative sentiment.

Moreover, in order to capture the phrase level contextual polarity of a given word, we will learn patterns from context that denote polarity, and then combine them with sentiment-tagged senses.

5 Preliminary Experimental Results

Table 2 presents preliminary results derived from the evaluation of Lin's method which is based on distributional similarity. We evaluate this method using first the corpus consinsting only of 60 articles containing the examined headlines. Second, we augment the corpus with MPQA articles and repeat the experiment (Table 3). The method is tested on 46 polysemous words indicating metaphors and expanded senses. Senses for idioms would be extracted immediately through a dictionary of idioms. Tables 2 and 3 present the synonyms derived for some of the polysemous words appearing in headlines. Indicatively we present 5 synonyms for each of the randomly selected five polysemous words. Table 4 presents the results derived from the evaluation of semantic relatedness measures presented in section 4.1. We particularly tested Gloss Vectors and Resnik's semantic relatedness measures on the headlines corpus, trying to disambiguate 46 polysemous words, 21 verbs, 21 nouns and 4 adjectives. The results in Table 4 show the accuracy achieved for each part of speech (Verbs, Nouns and Adjectives) and the types of expressions (Metaphors, M, Idioms, I, and Expanded Senses E) that are identified each time, are shown in the parentheses.

bid	**clash**	**locked**	**fight**	**hit**
quarantine	faith	outmanoeuvred	urinate	overcome
ferocity	sunshade	recorded	pitch in	hold over
endeavour	ineptitude	projected	bluff	abuse
regret	oxidation	commisionned	predict	halt
crosscourt	solace	located	negate	identify

Table 2. Preliminary Experimental Results using the corpus consisting of articles. Polysemous words are presented with bold and in columns their synonyms are given

bid	**clash**	**locked**	**fight**	**hit**
groundwork	overcapacity	manufactured	pitch in	hold over
bouncer	precipition	crouched	whizz	surpass
crosscourt	immigration	represented	sorrow	abuse
prescribed	pirate	prescribed	teeter	identify
semidarkness	unilateralist	assembled	audit	pack

Table 3. Preliminary Experimental Results using the expanded corpus consisting of the articles and the MPQA coprus. Polysemous words are presented with bold and in columns their synonyms are given

As it is deduced from the results presented in Tables 2, 3 Lin's approach didn't give adequate results, as the most of the times it assigned irrelevant words, a high similarity score for a given target

POS	Gloss Vectors	Res	Sum
Verbs	**12**(8M&4E)	9 (7M&2E)	21
Nouns	9 (3M&6E)	**10**(2M&6E)	21
Adj.	3 (1M&2E)	3 (1M&2E)	4

Table 4. Preliminary Experimental Results using measures of semantic relatedness

word. We suppose that this is happening since Lin's similarity measure is based on words' common triplets, - words' common syntactic structures- it finds that a polysemous word shares most of the triplets with words whose syntax could vary. For this reason these kind of words could have in common the most syntactic structures with the word to be disambiguated. These words are often common enough in terms that they could appear in several contexts. On the other hand we suppose that the polysemous words whose synonyms we want to derive could also appear in multiple syntactic structures, meaning that they do not figure rare syntactic structures.

In Table 4 we observe that Gloss Vectors performs best for verbs as it identifies correctly 12 out of 21 verbs, a fact which shows that Gloss Vectors seem promising in disambiguating verbs. This assumption needs to be verified by evaluating a larger set of verbs. The first intuition behind Gloss Vectors' good performance concerning verbs is that similarity is deduced through a larger data set since this method takes under consideration the words that co-occur with the words in the glosses. Thus more contextual information in our task leads to safer and more accurate results. Moreover concerning nouns it is observed through the table's results that Resnik's measure performs a little better than Gloss Vectors in nouns, as 10 out of 21 nouns are assigned the proper sense. As it is said before Resnik's measure presents some limitations, in that quite a few concepts in WordNet hierarchy may share the same least common subsumer and thus have identical values of similarity assigned to them. This fact can lead to wrong decisions concerning similarity, as given a different corpus these concepts could be very different in meaning, depending on the context.

6 Conclusions

In this position paper, we examined the research hypothesis that when words digress from their literal meanings they tend to appear in opinionated contexts and express significant polarity. We also proposed a methodology which exploits this hypothesis, in order to perform sentiment analysis using word sense disambiguation techniques. We claim that the major indicator for subjectivity is the sense of the word and not the word itself. From the manual corpus evaluation, our initial research hypothesis is verified, as the headline corpus which is already annotated with positive or negative polarity contains enough expanded senses, metaphors and idioms. This fact shows us that these phrases are used in subjective contexts yielding also significant polarity.

Preliminary results derived from the evaluation of the Word Sense Disambiguation method which is based on measures of semantic relatedness, show that WSD can be precious in identifying metaphors, idioms, and expanded senses, a fact that is very useful as it can lead to subjectivity and sentiment analysis. We also aim for this method to act complementarily to other existing methods for sentiment analysis in order to ameliorate current system's performances. We believe that the advantage of the methodology proposed is that, since it models linguistic principles, it is heavily language oriented and, thus, we expect to deduce valuable conclusions for the language function. We expect to see if our methodology for sentiment analysis can help us reveal the internal correlation of these kinds of expressions with the cognitive function, and more specifically, if sentiment constitutes an inherent property of figurative expressions.

REFERENCES

[1] 'Semeval-2007 task 14: Affective text'.

[2] S. Banerjee and T. Pedersen, 'Extended gloss overlaps as a measure of semantic relatedness', *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 805–810, (2003).

[3] A. Esuli and F. Sebastiani, 'Determining the semantic orientation of terms through gloss analysis', *Proc. CIKM-2005*, (2005).

[4] A. Esuli and F. Sebastiani, 'Determining term subjectivity and term orientation for opinion mining', *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, IT. Forthcoming*, (2006).

[5] A. Esuli and F. Sebastiani, 'Sentiwordnet: A publicly available lexical resource for opinion mining', *Proceedings of LREC*, 417–422, (2006).

[6] J.J. Jiang and D.W. Conrath, 'Semantic similarity based on corpus statistics and lexical taxonomy', *Proceedings of International Conference on Research in Computational Linguistics*, **33**, (1997).

[7] S.M. Kim and E. Hovy, 'Determining the sentiment of opinions', *Proceedings of COLING*, **4**, 1367–1373, (2004).

[8] G. LAKOFF and M. JOHNSON, 'Metaphors we live by', *Cultural Metaphors: Readings, Research Translations, and Commentary*, (2001).

[9] D. Lin, 'Automatic retrieval and clustering of similar words', *Proceedings of COLING-ACL*, **98**, 768–773, (1998).

[10] D. Lin, 'An information-theoretic definition of similarity', *Proceedings of the 15th International Conference on Machine Learning*, 296–304, (1998).

[11] S. Patwardhan, 'Incorporating dictionary and corpus information into a context vector measure of semantic relatedness', *Master of Science Thesis, Duluth, MN: Department of Computer Science. Duluth: University of Minnesota*, (2003).

[12] T. Pedersen, S. Banerjee, and S. Patwardhan, 'Maximizing semantic relatedness to perform word sense disambiguation', *Supercomputing institute research report umsi*, **25**, (2005).

[13] P. Resnik, 'Using information content to evaluate semantic similarity in a taxonomy', *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, **1**, 448–453, (1995).

[14] E. Riloff and J. Wiebe, 'Learning extraction patterns for subjective expressions', *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, 105–112, (2003).

[15] E. Riloff, J. Wiebe, and T. Wilson, 'Learning subjective nouns using extraction pattern bootstrapping', *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, 25–32, (2003).

[16] J. Wiebe and R. Mihalcea, 'Word sense and subjectivity', *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, 1065–1072, (2006).

[17] J. Wiebe, T. Wilson, and C. Cardie, 'Annotating expressions of opinions and emotions in language', *Language Resources and Evaluation*, **39**(2), 165–210, (2005).

[18] T. Wilson, J. Wiebe, and P. Hoffmann, 'Recognizing contextual polarity in phrase-level sentiment analysis', *Proceedings of HLT/EMNLP 2005*, (2005).

[19] H. Yu and V. Hatzivassiloglou, 'Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences', *Proceedings of EMNLP*, **3**, 129–136, (2003).