

Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach

Vassiliki Rentoumi*[†] *vrentoumi@iit.demokritos.gr*

George Giannakopoulos*[†] *ggianna@iit.demokritos.gr*

Vangelis Karkaletsis* *vangelis@iit.demokritos.gr* George A. Vouros[†] *georgev@aegean.gr*

Abstract

In this paper we propose a methodology for sentiment analysis of figurative language which applies Word Sense Disambiguation and, through an n-gram graph based method, assigns polarity to word senses. Polarity assigned to senses, combined with contextual valence shifters, is exploited for further assigning polarity to sentences, using Hidden Markov Models. Evaluation results using the corpus of the Affective Text task of SemEval'07, are presented together with a comparison with other state-of-the-art methods, showing that the proposed method provides promising results, and positive evidence supporting our conjecture: figurative language conveys sentiment.

1 Introduction

Metaphors and expansions are very common phenomena in everyday language. Exploring such cases, we aim at revealing new semantic aspects of figurative language. Detection of sentiments and implicit polarity, is within our focus. We thus propose a methodology for sentiment analysis that can be valuable in detecting new semantic aspects of language. Recent studies have shown that subjectivity is a language property which is directly related to word senses [14]. We believe that subjectivity and thus, senses related to meanings that convey subjectivity, are valuable indicators of sentiment. In order to prove this, we are led to the exploration of non-literal senses. Work presented in [11] provides evidence that non-literal senses, such as metaphors and expanded senses, tend to indicate subjectivity, triggering polarity. In order to capture the polarity that figurative language conveys [12], it is necessary to resolve ambiguities and detect the cases where words have non-literal meaning. Detecting this property for words, we can further assign polarity to the enclosing context, as it is shown in [12]. Towards this goal, other elements in discourse such as valence shifters [9] affect evaluative terms such as figurative expressions and modify the polarity of the whole context. In this paper we introduce a methodology for polarity detection that applies word sense disambiguation (WSD), exploits the assessed word senses and assigns

polarity to the enclosing sentences, exploiting other contextual features as well.

In Section 2 we briefly present related work while in Section 3 we present the theoretical background of this work, together with evidence for the conjectures made. Section 4 presents a detailed description of the overall methodology and of the specific methods used. Section 5 details the evaluation of the specific techniques used as well as the overall evaluation of the system. Section 6 concludes this article with a brief presentation of future research.

2 Related Work

Sentiment analysis aims to determine the exact polarity of a subjective expression. Specifically in [3] there is an effort using semi-supervised machine learning methods to determine orientation of subjective terms by exploiting information given in glosses provided by WordNet. In particular this approach is based on the assumption that terms with similar orientation tend to have similar glosses. Therefore, by means of glosses classification authors aim to classify the terms described by these glosses. Moreover in [16] the authors try to achieve a classification of phrases and sentences into positive/negative, by exploiting their context. In our approach we exploit the context where figurative expressions appear and perform disambiguation to reveal their senses which are considered as indicators of sentiment.

There are contextual elements in discourse that could modify the valence of words bearing opinion, thus affecting the overall polarity of a sentence. These contextual valence shifters are studied in [9].

Words, as shown in [14] can be assigned a subjective (with polarity) sense or an objective (neutral) sense. In this paper we support that we need to relate word senses with polarity, rather than the words themselves. It has also been shown through an empirical evaluation in [11], that especially metaphorical and expanded senses are strongly polar. Recent approaches follow this trend by developing sense tagged lists [4].

The methodology we propose aims to perform sentiment analysis on figurative language, detecting the writer's attitude. The suggested approach shows: (a) the necessity of WSD for assigning polarity to figurative expressions, (b) that figurative expressions combined with valence shifters drive the polarity of the sentence in which they appear, (c) that it seems

*National Centre for Scientific Research "Demokritos", Inst. of Informatics and Telecommunications, Greece

[†]University of the Aegean, Dpt. of Information and Communication System Engineering, Greece

promising even for the cases where the language is not figurative.

3 Theoretical Background and Corpus Study

We claim that metaphors and expansions drive the polarity of the whole sentence, as they constitute intense subjective expressions. The author, as shown in the following examples, transmits his opinion: (a) “Ice storm **cripples** central U.S”, (b) “Woman **fight**s to keep drunken driver in jail”. In (a) the author uses “cripple” to express implicitly his frustration about the physical disaster. The same is happening in (b) with the expanded sense of “fight”, where the writer expresses indirectly a positive attitude towards this woman.

We consider figurative language as the language which digresses from literal meanings. We conjecture that expanded senses and metaphors, being part of figurative language, can be used as expressive subjective elements since they display sentiment implicitly [12], [15]. To provide evidence for this, we used the corpus of the Affective Text task of Sem Eval '07¹ comprising 1000 newspaper headlines as it contains strongly personal and figurative language. We extracted headlines containing metaphorical and expanded expressions, using criteria inspired by Lakoff [6]. Lakoff's theory follows the principle of “from more concrete to more abstract meaning”. For this reason, we mostly considered as metaphorical those headlines whose word senses do not coincide with the default reading. The “default reading” of a word is the first sense that comes to mind in the absence of contextual clues, and it usually coincides with the literal one: the more concrete connotation of a word. In contrast, headlines containing figurative language invoke a deduction to a more abstract reading.

We manually extracted from the corpus 277 headlines in total²; 190 containing expanded senses (95 positive and 95 negative) and 87 containing metaphors (39 positive and 48 negative). These are annotated, as described in [13], according to a valence scale in which 0 represents a neutral opinion, -100 a negative opinion, and 100 a positive opinion. The average polarity assigned to headlines containing expansions and metaphors is above 40 which provides evidence that figurative language conveys significant polarity.

We consider that in the headlines, there can be contextual indicators, referred to as “valence shifters”, that can strengthen, weaken or even reverse the polarity evoked by metaphors and expansions. We first examine valence shifters that reverse the polarity of a sentence. Let us consider the following examples: (a) “Blazing Angels” for PS3 **fail** to **soar**, (b) “**Stop/halt/end**, violent nepal strikes”. In example (a) we observe, as is also claimed in [9], that “fail” has a negative connotation. On the other hand “to soar” in this context, has a positive connotation. The

evaluative term such as “to soar”, under the scope of “fail” will be neutralized, “fail” preserves its negative orientation and propagates it to the whole sentence. Moreover, in example (b) additional valence shifters are presented which are used as expanded senses, and they act as polarity reversers. These are the verbs “halt”, “end” and “stop”.

In the following examples we meet two more categories of valence modifiers: the diminishers and the intensifiers: (c) “Tsunami **fears ease** after quake”, (d) “New Orleans violence **sparks** tourism **fears**”. In example (c) “ease” functions as a valence diminisher for “fears”, as it means “to calm down”. The polarity of the whole sentence becomes less negative. In example (d), “spark” is used with its expanded sense denoting “to trigger a reaction”, thus strengthening the evaluative term it modifies.

There are words that always act as valence shifters, while certain polysemous words, act as valence shifters when they are used under a specific sense (i.e. as expanded senses or metaphors). We manually compiled a list of 40 valence shifters derived from our corpus. It contains common valence shifters (e.g. negations) and words used as such on a per context basis. In the near future we intend to exploit a WSD approach in order to detect the specific word senses of non literal expressions that act as valence shifters.

4 The Proposed Method For Sentiment Analysis

Our methodology involves three steps: (a) disambiguation of word senses (WSD). (b) assignment of polarity to word senses, based on the results derived from the WSD step. (c) polarity detection on a sentence level, by exploiting polarities of word senses and contextual cues such as valence shifters. The specific methods that implement these steps are presented in more detail in the subsequent sections.

4.1 First Step: Word Sense Disambiguation (WSD)

For WSD we chose an algorithm³, [8] that assigns to every word in a sentence the sense that is most closely related to the WordNet⁴ senses of its neighbouring words, revealing the meaning of that word. We used a context window of 8 words, as the mean length of each headline consists of 8-10 words. This WSD algorithm performs disambiguation for every word of each headline of our corpus, taking as input a headline and a relatedness measure [8]. Given such a measure, it computes similarity score for word sense pairs, created using every sense of a target word and every sense of its neighbouring words. The score of a sense of a target word is the sum of the maximum individual scores of that sense with the senses of the neighbouring words. The algorithm then assigns the sense with the highest score to the target word. The algorithm supports several WordNet based similarity measures, and among

¹ <http://www.cse.unt.edu/~rada/affectivetext/>

² The SemEval 07 corpus subset, annotated with metaphors and expansions can be downloaded from: <http://www.iit.demokritos.gr/~vrentoumi/corpus.zip>

³ <http://www.d.umn.edu/~tpederse/senserelete.html>

⁴ <http://wordnet.princeton.edu>

these, Gloss Vector (GV) performs best for non literal verbs and nouns [11]. GV creates a co-occurrence matrix of words. Each cell in this matrix indicates the number of times the words represented by the row and the column occur together in a WordNet gloss. Every word existing within a WordNet gloss is represented in a multi-dimensional space by treating its corresponding row as a vector. A context vector is created for each word in the gloss, using its corresponding row in the co-occurrence matrix. Then the gloss of each word sense is represented by a GV that is the average of all these context vectors. In order to measure similarity between two word senses, the cosine similarity of their corresponding gloss vectors is calculated. The input to the algorithm is the corpus enriched with Part-of-Speech (POS) tags performed by the Stanford POS tagger ⁵.

4.2 Second Step: Sense Level Polarity Assignment

This step detects polarity of the senses computed during the first step. To do this, WordNet senses associated with words in the corpus are mapped to models of positive or negative polarity. These models are learned by exploiting corresponding examples from the General Inquirer (GI). GI ⁶ is a lexical resource containing 1915 words labeled “positive” and 2291 labeled “negative”. Results from preliminary experiments (Section 5) show that GI provides correct information concerning polarity of non literal senses. On the other hand, SentiWordNet [4] is a resource for opinion mining assigning every synset in WordNet three scores, which represent the probability for a sense of a word to be used in a positive, negative or objective manner. Our method assumed binary classification of polarity while SentiWordNet performs ternary polarity classification. In the training procedure for SentiWordNet’s construction a seed list consisting of positive and negative synsets was compiled and was expanded iteratively through WordNet lexical relations. Our method uses GI’s positive and negative terms together with their definitions, instead. Moreover, after experimental evaluation with various WordNet features in order to detect sense level polarity for non literal senses, we decided that the best representative combination in judging the polarity for non literal senses is the combination which consists of synsets and GlossExamples (GES) of non literal senses. On the other hand SentiWordNet uses synsets and the whole gloss of every WordNet sense, in order to judge sense level polarity.

To compute the models of positive and negative polarity and produce mappings of senses to these models, we adopt a graph based method based on character n-grams [5], which takes into account contextual (neighbourhood) and sub-word information.

A (character) n-gram S^n contained in a text T can be any substring of length n of the original text. The *n-gram graph* is a graph $G = \{V^G, E^G, L, W\}$, where V^G is the set of vertices, E^G is the set of edges, L is a function assigning a label to each vertex and edge, and W is a function assigning a weight to every edge.

⁵ <http://nlp.stanford.edu/software/tagger.shtml>

⁶ <http://www.wjh.harvard.edu/~inquirer/>

n-grams label the vertices $v^G \in V^G$ of the graph. The (directed) edges are labeled by the concatenation of the labels of the vertices they connect in the direction of the connection. The edges $e^G \in E^G$ connecting the n-grams indicate proximity of these n-grams in the text within a given window D_{win} of the original text [5]. The edges are weighted by measuring the number of co-occurrences of the vertices’ n-grams within the window D_{win} . Subsequent paragraphs explain how the models of polarity are generated from the General Inquirer and how mappings of WordNet senses to these models are calculated by exploiting n-gram graph representations.

4.2.1 Constructing models using n-gram Graphs

To compute models of polarity using n-gram graphs, we have used two sets of positive and negative examples of words and definitions provided by the General Inquirer (GI). To represent a text set using n-gram graphs, we have implemented an *update/merge* operator between n-gram graphs of the same rank. Specifically, given two graphs, G_1 and G_2 , each representing a subset of the set of texts, we create a single graph that represents the merging of the two text subsets: $update(G_1, G_2) \equiv G^u = (E^u, V^u, L, W^u)$, such that $E^u = E_1^G \cup E_2^G$, where E_1^G, E_2^G are the edge sets of G_1, G_2 correspondingly.

The weights of the resulting graph’s edges are calculated as follows: $W^i(e) = W^1(e) + (W^2(e) - W^1(e)) \times l$. The factor $l \in [0, 1]$ is called the learning factor: the higher the value of learning factor, the higher the impact of the second graph to the first graph. The model construction process for each class (e.g. of the positive/negative polarity class) comprises the initialization of a graph with the first document of a class, and the subsequent update of this initial graph with the graphs of the other documents in the class using the union operator. As we need the model of a class to hold the average weights of all the individual graphs contributing to this model, functioning as a representative graph for the class documents, the i -th graph that updates the class graph (model) uses a learning factor of $l = \frac{i-1}{i}, i > 1$.

When the model for each class is created, we can determine the class of a test document by computing the similarity of the test document n-gram graph to the models of the classes: the class whose model is the most similar to the test document graph, is the class of the document. More specifically, for every sense x of the test set, the set of its synonyms (synsets) and Gloss Example Sentences (GES) extracted from WordNet, are being used for the construction of the corresponding n-gram graph X for this sense.

4.2.2 Graph Similarity

To represent a character sequence or text we use a set of n-gram graphs, for various n-gram ranks (i.e. lengths), instead of a single n-gram graph.

To compare two graph sets $\mathbb{G}_1, \mathbb{G}_2$ (one representing a sense and the other the model of a polarity class) we first use the *Value Similarity (VS)* for every n-gram rank [5], indicating how many of the edges contained in

graph G^i of rank n are also contained in graph G^j also of rank n , considering also the weights of the matching edges. In this measure each matching edge e having weight w_e^i in graph G^i contributes $\frac{\text{VR}(e)}{\max(|G^i|, |G^j|)}$ to the sum, while not matching edges do not contribute i.e. if an edge $e \notin G^i$ then $w_e^i = 0$. The *ValueRatio* (VR) scaling factor is defined as $\text{VR}(e) = \frac{\min(w_e^i, w_e^j)}{\max(w_e^i, w_e^j)}$. Thus, the full equation for VS is:

$$\text{VS}(G^i, G^j) = \frac{\sum_{e \in G^i} \frac{\min(w_e^i, w_e^j)}{\max(w_e^i, w_e^j)}}{\max(|G^i|, |G^j|)} \quad (1)$$

VS is a measure converging to 1 for graphs that share their edges and have identical edge weights.

The overall similarity VS^O of the sets $\mathbb{G}_1, \mathbb{G}_2$ is computed as the weighted sum of the VS over all ranks:

$$\text{VS}^O(\mathbb{G}_1, \mathbb{G}_2) = \frac{\sum_{r \in [L_{\min}, L_{\text{MAX}}]} r \times \text{VS}^r}{\sum_{r \in [L_{\min}, L_{\text{MAX}}]} r} \quad (2)$$

where VS^r is the VS measure for extracted graphs of rank r in \mathbb{G} , and L_{\min}, L_{MAX} are arbitrary chosen minimum and maximum n-gram ranks. For our task we used $L_{\min} = 3$ and $L_{\text{MAX}} = 5$.

4.3 Third Step: Sentence Level Polarity Detection

For the sentence level polarity detection we train two HMMs [10] - one for the positive, and one for the negative cases. The reason behind the choice of HMMs was that they take under consideration transitions among observations which constitute sequences. In our case the POS of a word combined with the word's polarity constitutes an observation. This information is provided by the POS tagging, and the graph based polarity assignment method upon metaphorical and expanded senses of the input sentences. The transitions among these observations yield the polarity of the sentential sequences. Structured models have been exploited for polarity detection showing promising results [2]. To exploit valence shifters, these are manually annotated in the corpus: they are assigned a predefined value depending on whether they revert, strengthen or weaken the polarity, in order to be integrated in the HMM.

This choice of features is based on the assumption that polarity of sentences is implied by patterns of parts of speech appearance, by the polarity assigned to specific senses in the specific context, and by the presence of valence shifters types in a sentence. We need to emphasize that the sequences are constructed using only non literal senses and valence shifters (if present), as we believe that these sequences enclose and determine the polarity of the sentences in which they participate. Having trained two HMM's, one for positive and one for negative cases, we determine the polarity of each headline sentence by means of the maximum likelihood of the judged observations given by each HMM. In order to evaluate this method of classifying headlines containing metaphors and expanded senses into positive and negative, we have used a 10-fold cross validation method for each of the two subsets.

5 Experimental Results

We evaluated the performance of the whole system (Table 4), but also the distinct steps comprising our method in order to verify our initial hypotheses, (a) WSD helps in polarity detection of non literal sentences and (b) the polarity of figurative language expressions drives the overall polarity of the sentences where these are present.

To evaluate the WSD method selected, we manually annotated the metaphorical and expanded cases with their corresponding senses in WordNet, indicated by their synsets and glosses. Two annotators were instructed to assign the most appropriate senses derived from WordNet according to the semantics of each headline's context and a third one refereed any disagreement. In order to evaluate the polarity assignment to senses, we manually aligned the senses of metaphors and expansions indicated by the WSD step, with the corresponding senses existing in GI, in order to assign to them the polarity provided by the latter.

In assigning polarities to senses, three annotators participated. Two of them mapped senses to GI, and the third refereed any disagreement. The annotators aligned metaphorical and expanded senses from WordNet, considering synsets and GES, with the corresponding senses from GI and took into account the polarity orientation (pos/neg) assigned to these senses. As synsets and GES denote the contextual use of the given sense, they can also reveal its polarity orientation in a given context. For each corpus subset (metaphors and expansions) there were two sets of senses, one extracted manually and one using automatic WSD. The annotators performed the alignment of all four of these sense sets with GI, which was exploited in the experimental evaluation. The results for the four polarity alignment tasks concern disagreement upon polarity alignment between annotators. In particular for metaphors, annotators disagreed in 10 senses for manual and 13 senses for automatic disambiguation, out of a total of 128 senses. Moreover for expansions annotators disagreed in 20 senses for manual and 24 senses for automatic disambiguation, out of a total of 243 senses. In preliminary research we performed an extra alignment with GI in order to detect if the figurative senses investigated are polar. Results show us that according to GI, the majority of metaphors (positive: 38.28%, negative: 35.15%) and expansions (positive: 31.27% negative: 37.8%) are polar. This verifies our initial hypothesis concerning the polarity of metaphors and expansions.

5.1 Evaluation of WSD in Polarity Detection

We first defined a baseline WSD method. In this method all senses were assigned the first sense given by WordNet. Since WordNet ranks the senses depending on their frequency, the first sense indicates the most common sense. We then compared the performance of the baseline method against a method without WSD for the polarity detection process and we observed that the former performs better than the latter. In Table 1 results are presented, in terms of recall and precision (prec), concerning polarity classification of

headlines containing metaphors (Met) and expansions (Exp), with WSD (GVbasedWSD/baselineWSD) and without the use of WSD (nonWSD). The results presented in Table 1 are based on automatic WSD and Sense level polarity assignment. It is deduced that even crude WSD, like the baseline method, could help the polarity classification of non literal sentences. Table 1 shows that polarity detection using the GV based WSD method performs much better than the one without WSD (nonWSD), for both categories of headlines. We further performed t-tests [7], to verify if the performance boost when GV based WSD is exploited is statistically significant over that when WSD is absent. Indeed, the method exploiting GV based WSD is better - within the 1% statistical error threshold (p-value 0.01).

Table 1 shows that polarity classification using GV based WSD outperforms the one using baseline WSD for both subsets of headlines. For metaphors, the GV based WSD method is better than the one using baseline WSD, within the 1% statistical error threshold (p-value 0.01). On the other hand, for expanded senses we cannot support within 5% statistical error that GV based WSD performs better than baseline (p-value 0.20). The above results verify our initial hypothesis that WSD is valuable in the polarity classification of headlines containing non literal senses.

In order to evaluate the GV based WSD method, we first compared the output of the GV based WSD step with that of the manually performed WSD. This comparison is performed for both metaphorical and expanded cases. We detected that in the WSD process for metaphorical senses, GV based WSD had a 49.3% success and for expanded senses 45%. The mediocre performance of GV based WSD is attributed to the fact that disambiguation of metaphors and expansions is itself a difficult task. Additionally, the restricted context of a headline makes the process even more difficult. In order to find out to which extent the errors introduced in the disambiguation step can affect the performance of the whole system, we compare two versions of our system. These versions are differentiated by the automation of different components of the system.

In Table 2, we see the performance of both versions in terms of recall and precision for polarity classification of headlines containing metaphors and expansions. The first version is based on manual WSD and manual sense level polarity assignment (manual/manual(a)), and the second is based on automatic WSD and manual polarity assignment upon the senses that the automatic WSD method indicated (auto/manual(b)). Both versions use HMM models for headlines classification. We can also deduce from these results, that errors introduced during automatic WSD do not affect the system significantly. This is attributed to the fact that the prototypical core sense of a word remains, even if the word can be semantically expanded, acquiring elements from relative semantic fields. This core sense can bear a very subtle polarity orientation, which becomes stronger and eventually gets activated as the word sense digresses (as in the cases of expansions and metaphors) from the core one. There also exists the rare case when the polarity of a word is reversed because of a semantic change. The

	GVbasedWSD		nonWSD		baselineWSD	
	recall	prec	recall	prec	recall	prec
Met	72.6	76.5	46.75	48.5	54.20	57.00
Exp	67.9	68.0	48.1	48.0	63.12	63.00

Table 1: Polarity classification results of headlines containing metaphors (Met) and expansions (Exp) with or without the use of WSD

WSD	manual		auto		manual	
Sense Pol	manual(a)		manual(b)		n-gram graphs	
	recall	prec	recall	prec	recall	prec
Met	78.5	81.5	71.00	74.5	62.57	66.00
Exp	79.1	79.5	75.36	75.5	62.53	62.95

Table 2: Evaluation of GV based WSD (auto) and n-gram graphs steps, Sense Pol: sense level polarity, manual(a): manual polarity assignment on the manually disambiguated senses, manual(b): manual polarity assignment on the automatically disambiguated senses

above lead to the deduction that WSD helps indeed to improve sentiment analysis, even though the “exact” sense is not always correct.

5.2 Evaluation of n-gram graphs for sense level polarity assignment

In order to evaluate the n-gram graph method used for sense level polarity assignment, we first compare the output of n-gram graphs, with that of the manual procedure. The n-gram graphs scored 60.15% success for metaphorical senses, and 67.07% for expanded senses.

We present in Table 2 the performance of the system, with n-gram graphs (manual/n-gram graphs) and with manual sense level polarity assignment (manual/manual(a)). The significant drop of the performance when n-gram graphs are used, led to the assumption that sense level polarity assignment errors affect our system’s performance because they change the input of the decisive HMM component.

5.3 Metaphors and Expansions: Enough for sentence level polarity detection

We also performed experiments to verify that figurative language expressions represent the polarity of the sentences in which they appear. For that we performed two more experiments - all steps of which are performed automatically - one for metaphors and one for expansions, where we trained HMMs with input

headlines with met				headlines with exp			
met		all words		exp		all words	
rec	prec	rec	prec	rec	prec	rec	prec
72.6	76.5	55.9	57.45	67.89	68.0	59.4	59.8

Table 3: Evaluation of the system for polarity classification of headlines containing metaphors and expansions (headlines with met/headlines with exp) using only non literal expressions (metaphors(met) and expansions(exp)) vs using all words

Our System						CLaC		CLaC - NB	
head. met.		head. exp.		1000 headlines		1000 headlines			
rec	prec	rec	prec	rec	prec	rec	prec	rec	prec
72.6	76.5	67.9	68.0	55.60	55.65	9.20	61.42	66.38	31.18

Table 4: *Evaluation of the performance of the system*

sequences containing all the words of each headline instead of only the non literal expressions. In Table 3 the system’s performance for polarity classification, in terms of recall and precision, is presented, for each subset. Results are shown for the cases when all words of the sentence are used and for the cases where only the non literal expressions are used. The experiments with only the non literal expressions have much better results. This verifies our initial hypothesis that the polarity of figurative language expressions can represent the polarity of the sentence in which they appear.

5.4 System Evaluation, Comparison with state-of-the-art systems

Table 4 presents results for our system as well as two state-of-the-art systems CLaC and CLaC-NB [1], in terms of recall and precision, compared to the Gold Standard polarity annotation. For our system we present in Table 4 three sets of results, for headlines containing metaphors (head. met), expanded senses (head. exp) and for the whole corpus (1000 headlines). The last set of results is presented in order to have a comparison with the two other systems under a common data set. As mentioned in the beginning of this paper, our system aims to fill the gap for polarity detection in sentences containing figurative language. The results for the individual cases (exp. and met.) show that the system performs well under these circumstances. This leads to the result that the specific combination of GV based WSD, n-gram graphs and HMMs works well for these two subsets. As results in Tables 4 and 2 show, the performance of the overall method, compared to configurations where some of the steps are performed manually, is very promising.

When our system is applied to the overall corpus (Table 4), although the peak values are lower than the two other systems, they are high enough to spark further research. We can see that although precision in the CLaC system is quite high, it suffers a low recall value. The authors attribute this to the fact that the system was based on an unsupervised knowledge-based approach in which they aimed to achieve results of higher quality, thus missing a lot of sentiment bearing headlines [1]. On the contrary CLaC-NB has a high recall and relatively low precision. This system is based on supervised machine learning and the authors attributed this behaviour to the lack of significant corpus to train the statistical classifier. The strength of our system is that we achieved relatively high values in both recall and precision.

6 Conclusions and Future Work

This paper presents a new methodology for polarity classification of non literal sentences. We showed

through experimental evaluation that WSD is valuable in polarity classification of sentences containing figurative expressions. Moreover, we showed that polarity orientation hidden in figurative expressions prevails in sentences where such expressions are present and combined with contextual valence shifters, can lead us to assessing the overall polarity for the sentence. So far evaluation results of our methodology, seem comparable with the state-of-the art methodologies tested upon the same data set. Testing our methodology in a more extended corpus is our next step.

Acknowledgments.

We would like to thank Giota Antonakaki, Ilias Zavitanos, Anastasios Skarlatidis, Kostas Stamatakis and Jim Tsarouhas for their invaluable help.

References

- [1] A. Andreevskaia and S. Bergler. Clac and clac-nb: Knowledge-based and corpus-based approaches to sentiment tagging. *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 119–120, 2007.
- [2] Y. Choi, E. Breck, and C. Cardie. Joint extraction of entities and relations for opinion recognition. In *Proc. EMNLP*, 2006.
- [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. Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of LREC*, pages 417–422, 2006.
- [5] G. Giannakopoulos, V. Karkaletsis, G. Vouros, and P. Stamatopoulos. Summarization system evaluation revisited: N-gram graphs. *ACM Transactions on Speech and Language Processing (TSLP)*, 5(3), 2008.
- [6] G. Lakoff and M. Johnson. *Metaphors we live by*. *Cultural Metaphors: Readings, Research Translations, and Commentary*, 2001.
- [7] M. O’Mahony. *Sensory evaluation of food: Statistical methods and procedures*. CRC Press, 1986.
- [8] T. Pedersen, S. Banerjee, and S. Patwardhan. Maximizing semantic relatedness to perform word sense disambiguation. *Supercomputing institute research report umsi*, 25, 2005.
- [9] L. Polanyi and A. Zaenen. Contextual valence shifters. *Computing Attitude and Affect in Text: Theory and Applications*, pages 1–9, 2004.
- [10] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.
- [11] V. Rentoumi, V. Karkaletsis, G. Vouros, and A. Mozer. Sentiment analysis exploring metaphorical and idiomatic senses: A word sense disambiguation approach. *Proceedings of International Workshop on Computational Aspects of Affectual and Emotional Interaction (CAFEEi 2008)*, 2008.
- [12] E. Riloff and J. Wiebe. Learning extraction patterns for subjective expressions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2003)*, pages 105–112, 2003.
- [13] C. Strapparava and R. Mihalcea. SemEval-2007 Task 14: Affective Text. In *Proceedings of the 3rd International Workshop on Semantic Evaluations (SemEval 2007)*, Prague, Czech Republic, pages 70–74, 2007.
- [14] 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*, pages 1065–1072, 2006.
- [15] J. Wiebe, T. Wilson, and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2):165–210, 2005.
- [16] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. *Proceedings of HLT/EMNLP 2005*, 2005.