# Finding Humour in the Blogosphere: The Role of WordNet Resources

**Antonio Reyes, Paolo Rosso** and **Davide Buscaldi**
Natural Language Engineering Lab - ELiRF
Departamento de Sistemas Informáticos y Computación
Universidad Politécnica de Valencia, Spain
{areyes,prosso,dbuscaldi}@dsic.upv.es

## Abstract

Humour is an amazing and challenging topic. Despite the several analyses and researches for understanding its complex mechanisms, it is not completely defined. The studies performed from the Natural Language Processing perspective have demonstrated that, taking into account linguistic resources, statistical methods, machine learning and corpus-based techniques, humour may be handled by means of computational systems in order to automatically generate and recognise it. In this paper we focused on studying humour in the blogosphere. The aim is to evaluate the importance of lexical resources linked to WordNet for recognising blogs with greater probability to contain humour. All the experiments were performed over a corpus integrated by 19,200 blogs. Moreover, a corpus of 16,000 humorous one-liners was used in order to evaluate our results. The findings are encouraging.

## 1 Introduction

The role of humour is very important in our lives. It impacts on physical, cognitive and social aspects causing, often, different positive effects: from a simple smile up to alleviating stress or improving our interpersonal relationships (Mihalcea, 2007). Also, besides producing well-being sensations, humour is a kind of catalyst which impacts on a broad spectrum of properties linked to cognitive and social information, realising our emotions and feelings, and providing knowledge about the human behaviour. For instance, Ruch (2001) has studied how humour appreciation is associated to personality and how, depending on this property, the kind of necessary stimuli to produce a response changes. However, despite its benefits on human health, humour is a very complex phenomenon whose multifactorial mechanisms can hardly be delimited, generalised and modeled.

The approaches performed in areas such as linguistics (Attardo, 1994), psychology (Ruch, 2001), sociology (Hertzler, 1970), etc., have provided valuable knowledge for explaining humour. Natural Language Processing has also contributed to supply a bit of information in this task through two perspectives: generation (Binsted and Ritchie, 2001) and recognition (Mihalcea and Strapparava, 2006). This paper regards to the second perspective: on the basis of applying different elements reported in the literature as humour features, we focus on studying the importance of lexical resources linked to WordNet for recognising humour in the blogosphere. The task was performed employing a corpus integrated by 19,200 documents divided in 8 sets (7 retrieved from LiveJournal and 1 retrieved from Wikipedia). The underlying aim is to assess the relevance of resources such as WordNet Domains (Bentivogli et al., 2004), WordNet-Affect (Strapparava and Valitutti, 2004) and SentiWordNet (Esuli and Sebastiani, 2006), as well as WordNet (Miller, 1995), for differentiating the sets with greater probability to contain humour. The results are evaluated taking as gold standard the corpus of one-liners (Mihalcea and Strapparava, 2006a) used in the main researches on automatic humour recognition.

The paper outline is organised as follows. Section 2 depicts the related work and establishes our objective. Section 3 describes the corpus and the experiments. Section 4 presents the evaluation. Finally, Section 5 concludes with some final remarks and addresses the future work.

## 2 Related Work

The interest for humour from a computational viewpoint comes from the last century. In the 90s, the researches in (Binsted, 1996; Binsted and Ritchie, 1997) showed the importance of linguistic patterns, especially phonetic and syntactic ones, for automatically generating humorous punning riddles. In this century, the findings reported in (Stock and Strapparava, 2005) demonstrated how incongruity and opposite concepts are important elements for producing funny senses. Recently, the researches described in (Mihalcea and Strapparava, 2006; Mihalcea and Pulman, 2007; Sjöbergh and Araki, 2007; Reyes et al., 2009a), have provided evidence for automatically recognising elements to characterise humour. Some of the elements reported as humour features are alliteration, antonymy or adult slang (Mihalcea and Strapparava, 2006), likewise similarity, style or idiomatic expressions (Sjöbergh and Araki, 2007). Also, in (Reyes et al., 2009a) it is argued that linguistic ambiguity is an important trigger of humour. All these studies have covered several layers of linguistic analysis: from lexical up to semantics. However, considering the advances achieved from both perspectives, it is clear that the results do not generate/recognise a natural and spontaneous joke. One of the reasons could be the kind of humour analysed: punning riddles and one-liners. The first ones are more related to children and, accordingly, humour tends to be not so funny for adults, whereas one-liners, given their own characteristics by one hand, and the similarity with proverbs by the other one, they often do not provoke a big laugh. Let us consider the following sentences for exemplifying these assertions.

1. What do you call a cold aunt? Aunty-freeze (Binsted, 1996).
2. Ah, nostalgia ain't what it used to be...[1]

Despite their simplicity, both kinds of structures have allowed researchers to make a computational humour treatment with success. On the basis of these achievements, we aim at assessing, given 8 different data sets, the role of WordNet resources for distinguishing the ones with greater probability to contain humour. This objective implies the following tasks:

i. to collect and to validate a corpus related to humour;
ii. to select, given the available resources, a set of features to represent humour;
iii. to measure the humour representativeness, if exists, in the corpus;
iv. to determine a gold standard in order to assess the results.

[1]One-liner taken from the examples given by Wikipedia.

The first task was accomplished by means of retrieving a corpus from one of the most important communities related to blogging: LiveJournal, besides a small set retrieved from Wikipedia. The corpus was evaluated using the measures proposed in (Pinto et al., 2009) for studying corpus features such as domain broadness, stylometry or structure.

The selection of features was performed considering the findings described in the literature: $stylistic$ features, (Mihalcea and Strapparava, 2006); $negative\ orientation$, (Mihalcea and Pulman, 2007); $affectiveness$, (Reyes et al., 2009b). According to each feature, we used WordNet Domains, SentiWordNet and WordNet-Affect, respectively.

The humour representativeness was measured estimating the amount of features per document, plus their semantic ambiguity.

Finally, the fourth task was achieved by means of using the most representative elements from the one-liners corpus, obtained through a clustering process, as gold standard.

## 3 Experiments on Features Extraction

### 3.1 Data Sets

The corpus was automatically collected from LiveJournal and Wikipedia. It contains 19,200 documents divided in 8 sets: $angry$, $happy$, $humour$, $sad$, $scared$, $others$, $general$ and $Wikipedia$. Every set contains 2,400 documents. The first 6 sets were retrieved taking advantage of the predefined mood tags[2] as well as the users tags[3]. The last two sets were considered as control sets, that is why we did not take into account any mood tag or any seed related to humour for collecting them. The $general$ one was also retrieved from LiveJournal considering just blogs related to topics such as news, politics, fashion, religion, technology, weather, computer and cars. For the $Wikipedia$ one, we just considered the articles related to technology. This corpus is available at: http://users.dsic.upv.es/grupos/nle/?file=kop4.php.

### 3.2 Corpus Evaluation

In order to promote the presence of humour in the corpus, except for the control sets, the documents

[2]LiveJournal provides 132 items organised in 15 categories. The sets used in the study correspond to the main categories: *angry, happy, sad, scared*. The set $others$ represents the rest of categories.

[3]We just considered the blogs labelled with tags such as humour and joke, which integrate the set $humour$.

| | Angry | Happy | Humour | Sad | Scared | Others | General | Wikipedia |
|---|---|---|---|---|---|---|---|---|
| Terms | 1,314.55 | 1,114.41 | 1,577.16 | 1,193.92 | 1,342.98 | 1,027.32 | 843.44 | 1,934.07 |
| CVS | 132.83 | 161.33 | 219.25 | 119.90 | 145.42 | 122.56 | 107.00 | 162.30 |
| DL | 604.39 | 542.56 | 720.50 | 567.73 | 625.81 | 483.44 | 410.44 | 937.96 |
| VL | 411.10 | 382.99 | 503.27 | 384.47 | 418.42 | 341.92 | 301.68 | 516.18 |
| VDR | 0.94 | 0.94 | 0.95 | 0.94 | 0.94 | 0.94 | 0.95 | 0.91 |
| UVB | 6.91 | 9.27 | 9.29 | 6.78 | 7.48 | 7.75 | 7.80 | 6.90 |
| SEM | 0.39 | 0.40 | 0.37 | 0.38 | 0.37 | 0.40 | 0.46 | 0.40 |

Table 1: Assessment corpus features per data set. Measures: corpus vocabulary size (CVS); document and vocabulary length (DL and VL, respectively); vocabulary and document length ratio (VDR); unsupervised vocabulary based measure (UVB); stylometric evaluation measure (SEM).

were retrieved if and only if, they contained keywords such as punch line, humour, funny, laughter, laugh line, gag, joke, and so on. The corpus was evaluated in terms of $shortness$, $broadness$ and $stylometry$ (Pinto et al., 2009), The results obtained are shown in Table 1.

According to these values, the $shortness$ of the data sets is low, both in terms of documents and vocabulary. The vocabulary and document ratio (VDR) measure indicates that, in terms of frequency, all the sets imply high complexity. Regarding to the $broadness$, the unsupervised vocabulary based (UVB) measure indicates that, broadly, all the sets tend to restrict their topics to specific contents, especially, the happy and humour ones. With respect to the $stylometry$, the stylometric evaluation measure (SEM) indicates interesting information related to a common expression style among some sets. Considering these results, we can point out that all the data sets are distinguishable collections with sufficient information for assessing our hypothesis.

### 3.3 Humour Average

The features for measuring the humour average in the data sets were selected accordingly to the available WordNet resources and, especially, to the results reported in the literature. These ones were: $stylistic$ elements, $negative\ orientation$ and $affectiveness$.

The first one, according to (Mihalcea and Strapparava, 2006), represents one of the most relevant features for discriminating humour. The information which better describes the feature is the sexual one. We used WordNet Domains for obtaining the elements related to sexual information. All the words labelled with the tag “sexuality” were retrieved from WordNet Domains.

In (Mihalcea and Pulman, 2007), the authors reported how $negative\ orientation$ may help for discriminating humorous from non humorous data. We used SentiWordNet, for labelling the data sets. We focused on identifying the negative elements, centering on the morphosyntactic categories: nouns, adjectives and verbs, if and only if, they passed an empirically founded threshold $\geq$ 375 in their negative scores.

$Affective$ information is also considered a focus for identifying humour (Reyes et al., 2009b). We computed, according to WordNet-Affect, the amount of affective elements in every set. It is important to note that we did not considered all the WordNet-Affect categories but only the most informative ones. They were obtained by means of applying an information gain filter (Witten and Frank, 2005). The most informative categories, according to this measure, were: attitude (att), behaviour (beh), cognitive state (cog), emotion (emo) and trait (tra)[4].

A numerical value, plus an ID depending on the feature they belong, were assigned to every element. Then, each document within the set was represented through a feature vector. We just considered the documents, whose sum of features was $\geq$ 15, i.e., at least 5 elements per feature. The amount of documents which matched this criterion is the following: angry (2,091), happy (1,966), humour (2,074), sad (1,997), scared (2,052), others (2,019), general (1,955), and Wikipedia (1,955). The humour average was obtained summing all the elements per document and dividing the result by the number of documents per set. In order to avoid any tendency in the further experiments, we just considered the 1,000 documents with greater humour average. In Figure 1 we depicted the humour average per set according to these 1,000 documents.

[4] All the information about the concepts symbolised by these categories appears in (Strapparava and Valitutti, 2004).

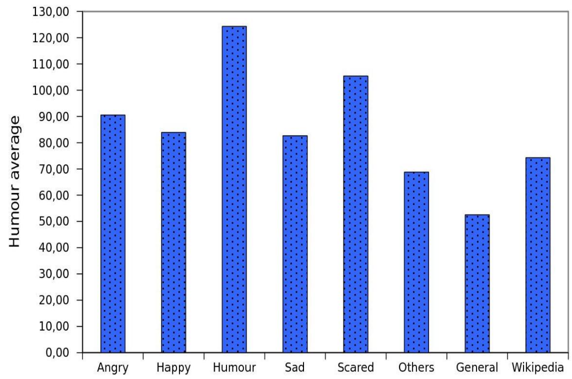


Figure 1: Humour average per set using WordNet Resources.

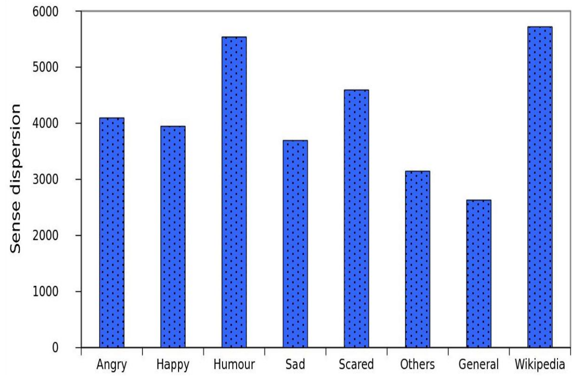


Figure 2: Semantic ambiguity per set according to the sense dispersion measure.

The results corroborate the *a priori* hypothesis, i.e., the sets which took advantage of the mood and users tags are the ones with greater humour average, whereas the control sets are just comparable with the most heterogeneous set: others.

### 3.4 Semantic Ambiguity

Ambiguity is one of the characteristics whose impact on humour is very important (Attardo, 1994; Mihalcea and Strapparava, 2006; Sjöbergh and Araki, 2007). From a phonetic level up to a pragmatic one. We decided to estimate this characteristic from a semantic viewpoint. The *sense dispersion* formula in (1), described in (Reyes et al., 2009a), was applied for measuring the semantic ambiguity in the data sets.

This formula is based on the hypernym distance among synsets. Its underlying aim is to quantify the semantic ambiguity through measuring the differences among the senses of a word. The hypothesis is that a word with senses that differ significantly is more likely to be used to create humour than a word with senses that differ slightly.

$$\delta(w_s) = \frac{1}{P(|S|, 2)} \sum_{s_i, s_j \in S} d(s_i, s_j) \qquad (1)$$

where S is the set of synsets $(s_1, \ldots, s_n)$ for the word $w$; P($n$, $k$) is the number of permutations of $n$ objects in $k$ slots; and $d(s_i, s_j)$ is the length of the hypernym path between synsets $(s_i, s_j)$.

We estimated the sense dispersion for the 1,000 documents with greater humour average from all the sets according to the formula in (2):

$$\delta_{TOT} = \sum_{w_s \in W} \delta(w_s) \qquad (2)$$

where $W$ is the set of nouns in the collection $N$. The results are depicted in Figure 2.

According to the results showed in this figure, the fact of considering the semantic ambiguity corroborates the results described in the previous section. It seems that semantic ambiguity would improve a humour recognition task over these data sets, although the improvement is not so significant. On the other hand, we noted that the behaviour of the Wikipedia set is quite different from what expected. Being a control set in which humour does not appear, we would have expected another result. We think that this behaviour is due to the kinds of discourses considered and, especially, to the lexicon employed in the blogs and in the articles. It supposes that the last ones contain less words out of vocabulary and consequently, they have more nouns to estimate their dispersion (since this measure takes only into account nouns), whereas the blogs often contain more misspellings, neologisms, etc., therefore, the amount of elements to be measured decreases.

## 4 Evaluation

In order to verify the relevance of the results obtained, we decided to evaluate them by means of estimating the similarity between every set (only the 1,000 documents with greater humour average) and a gold standard, in such a way to avoid a subjective and personal assessment[5]. The corpus of one-liners (Mihalcea and Strapparava, 2006a) was employed as gold standard. This corpus has been the main source for the most important research works related to humour recognition. It

[5] Let us remember the fuzzy nature of humour: what is humorous for some people could even be offensive for other persons.

| | Descriptive | Discriminating |
|---|---|---|
| Angry | 199.21 | 113.11 |
| Happy | 170.70 | 106.81 |
| Humour | 174.19 | 117.72 |
| Sad | 173.26 | 143.16 |
| Scared | 184.66 | 128.70 |
| Others | 171.28 | 130.03 |
| General | 164.42 | 101.14 |
| Wikipedia | 151.63 | 79.61 |

Table 2: Evaluation results.

contains 16,000 humorous one-liners automatically retrieved from the web. The processes performed for evaluating consisted in two phases:

i. we divided every set, including the gold standard, in 10 clusters in order to extract the 20 most descriptive and discriminating items from each one[6]. The 40 items per set were obtained using Cluto (Karypis, 2003);

ii. we applied the Resnik measure implemented in WordNet::Similarity (Pedersen et al., 2004) in order to compare every set of descriptive and discriminating items and to determine how much similarity existed among our sets and the gold standard.

In Table 2 is displayed, based on the most descriptive and discriminating items, the similarity between every single set and the gold standard.

According to the evaluation, the sets whose discourse is more similar to the discourse profiled in the gold standard is the angry (descriptive items) and the sad ones (discriminating items). This means that humour tends to appear often in the sets which denote negative moods (do we laugh for not suffering?). On the other hand, the results corroborate, in some manner, the characterisation performed with the WordNet resources, which indicates a greater probability to find humour in some sets. For instance, both with the descriptive and the discriminating items, the control sets (general and Wikipedia) obtained the worst similarity scores. With respect to the humour set, we would have expected a better similarity score, given the results achieved in the experiments reported in Section 3. However, the results are quite different. Perhaps we have to consider more descriptive and discriminating items in order to verify whether the results change or not.

[6]For instance, items such as bed, friend, fun (descriptive), or class, college, summer (discriminating) are examples about the clusters generated in the humour set.

## 5 Conclusions and Future Work

In this paper we have focused on analysing the role of WordNet resources (WordNet Domains, SentiWordNet, WordNet-Affect and WordNet) for identifying, given 8 different data sets, the ones with greater probability to contain humour. In order to obtain an indicator about the presence of humour in the sets, we characterised all the documents in terms of the features reported in the literature as fundamental in the manner of expressing verbal humour: $stylistic$ elements (WordNet Domains), $negative\ orientation$ (SentiWordNet), and $affectiveness$ (WordNet-Affect). Moreover, we tried to measure the semantic ambiguity using WordNet (sense dispersion) in order to provide more elements related to humour and to enhance the probability of finding signs of humour in our data sets.

An evaluation over the 1,000 documents with greater humour average was performed in order to verify the similarity between every single set and the gold standard. The results corroborate some of the findings achieved using the WordNet resources, besides indicating that, on the basis of the most descriptive and discriminating items, the angry and sad sets are the ones whose discourses are more similar to the one profiled in the gold standard.

Finally, we plan in the future to verify the results applying more features, assigning weights to every feature depending on their relevance, and assessing other similarity measures.

## Acknowledgments

The TEXT-ENTERPRISE 2.0 (TIN2009-13391-C04-03) research project has partially funded this work. The National Council for Science and Technology (CONACyT - Mexico) has funded the research work of Antonio Reyes.

## References

S. Attardo. 1994. *Linguistic Theories of Humor*. Mouton de Gruyter.

L. Bentivogli, P. Forner, B. Magnini, and E. Pianta. 2004. Revising the wordnet domains hierarchy: semantics, coverage and balancing. In Gilles Sérasset, editor, *COLING 2004 Multilingual Linguistic Resources*, pages 94–101. COLING.

K. Binsted and G. Ritchie. 1997. Computational rules for punning riddles. *Humour*, 10:25–75.

K. Binsted and G. Ritchie. 2001. Towards a model of story puns. *Humour*, 14:275–292.

K. Binsted. 1996. *Machine humour: An implemented model of puns*. Ph.D. thesis, University of Edinburgh, Edinburgh, Scotland.

A. Esuli and F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006)*, pages 417–422.

J. Hertzler. 1970. *Laughter: A social scientific analysis*. Exposition Press.

G. Karypis. 2003. Cluto. a clustering toolkit. technical report 02-017. Technical report, University of Minnesota, Department of Computer Science.

R. Mihalcea and S. Pulman. 2007. Characterizing humour: An exploration of features in humorous texts. In *8th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2007*, volume 4394, pages 337–347.

R. Mihalcea and C. Strapparava. 2006. Learning to Laugh (Automatically): Computational Models for Humor Recognition. *Journal of Computational Intelligence*, 22(2):126–142.

R. Mihalcea and C. Strapparava. 2006a. Technologies that make you smile: Adding humour to text-based applications. *IEEE Intelligent Systems*, 21(5):33–39.

R. Mihalcea. 2007. Multidisciplinary facets of research on humour. In *3rd Workshop on Cross Language Information Processing, CLIP-2007, Int. Conf. WILF-2007*, volume 4578, pages 412–421.

G. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

T. Pedersen, S. Patwardhan, and J. Michelizzi. 2004. Wordnet::similarity - measuring the relatedness of concepts. In *Proceeding of the 9th National Conference on Artificial Intelligence (AAAI-04)*, pages 1024–1025.

D. Pinto, P. Rosso, and H. Jimnez. 2009. On the assessment of text corpora. In *Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems (NLDB) 2009*.

A. Reyes, D. Buscaldi, and P. Rosso. 2009a. The impact of semantic and morphosyntactic ambiguity on automatic humour recognition. In *Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems (NLDB) 2009*.

A. Reyes, P. Rosso, and D. Buscaldi. 2009b. Affect-based features for humour recognition. In *Proceedings of the 7th International Conference on Natural Language Processing ICON-09, (to be published)*.

W. Ruch. 2001. The perception of humor. In World Scientific, editor, *Emotions, Qualia, and Consciousness. Proceedings of the International School of Biocybernetics*, pages 410–425.

J. Sjöbergh and K. Araki. 2007. Recognizing humor without recognizing meaning. In *3rd Workshop on Cross Language Information Processing, CLIP-2007, Int. Conf. WILF-2007*, volume 4578, pages 469–476.

O. Stock and C. Strapparava. 2005. Hahacronym: A computational humor system. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 113–116.

C. Strapparava and A. Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, volume 4, pages 1083–1086.

I. Witten and E. Frank. 2005. *Data Mining. Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Publishers. Elsevier.