Humor in the Blogosphere: First Clues for a Verbal Humor Taxonomy

Antonio Reyes, Paolo Rosso, and Davide Buscaldi

Natural Language Engineering Lab –EliRF, Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, Spain

ABSTRACT

The importance of analyzing processes related to cognitive phenomena through Natural Language Processing techniques is acquiring a greater relevance every day. Areas such as Opinion Mining, Sentiment Analysis or Automatic Humor Recognition are samples about how this kind of research gradually grows. In this paper, we focused on studying the features that define a corpus of humorous data (one-liners) to assess whether they may be used as elements for building a verbal Humor taxonomy. We analyzed, through several experiments, a set of well-known features described in the literature, besides a set of new ones, to determine the importance of each. An evaluation of all features was performed by means of an automatic classification task over a collection of humorous blogs. Our results suggest that some features may represent basic information for creating the taxonomy.

KEYWORDS: humor recognition, natural language processing, semantic ambiguity, computational humor

1. INTRODUCTION

The analysis of phenomena related to cognitive processes is a very important trend in Natural Language Processing research. The study of characteristics linked to

---

Reprint requests to: Antonio Reyes, Natural Language Engineering Lab–EliRF, Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, Spain; {areyes,prosso,dbuscaldi}@dsic.upv.es

human behavior such as emotions or mood (Balog et al., 2006) is a sample about the importance of this kind of research, which leads to exploring more abstract spheres that acquire representation at linguistic level. On that, the investigations in areas such as Opinion Mining (Ghose et al., 2007), Sentiment Analysis (Pang et al., 2002) or Computational Humor (Mihalcea, 2007) have shown how to address these challenging tasks through the use of machine learning or pattern recognition techniques, besides the use of linguistic resources.

In this framework, we present a research work that focuses on the analysis of a set of features that define a corpus of humorous one-liners (Mihalcea & Strapparava, 2006a,b) to obtain elements that allow us to build a Humor taxonomy automatically. On this subject, we aim at investigating how the features that have been considered as descriptors of these one-liners, besides a set of new features, can be employed for depicting the concepts that underlie the phenomenon of Humor. We evaluate the hypothesis that this objective implies going through a classification task over a collection of blogs whose main topic is Humor.

The outline of this paper is organized as follows. In Section 2, we describe the research works on Computational Humor, focusing on Automatic Humor Recognition. In Section 3, we present our initial assumptions and aim. In Section 4, we detail all the experiments. In Section 5, we present the evaluation and the discussion of the results. Finally, in Section 6, we draw some conclusions and address future work.

2. COMPUTATIONAL HUMOR

HUMOR is one of the most amazing and fuzzy aspects of the human behavior that despite its common practice, is still not clearly defined (Ritchie, 2003; Attardo, 2001). Cognitive features as well as cultural knowledge, for instance, are some of the variables that must be analyzed to obtain some answers about how HUMOR works. Such factors turn HUMOR in a subjective and fuzzy entity that changes according to cultures, societies, persons or mood. Nonetheless, HUMOR's automatic processing seems promising. For instance, as part of the Affective Natural Processing tasks, the Computational HUMOR area has demonstrated that this characteristic can be automatically handled from two angles: generation and recognition. The first builds models from recurrent templates taken into account linguistic patterns. For instance, the research work performed by Binsted and Ritchie (1997, 2001) showed the importance of phonetic and semantic patterns as features for automatically generating punch lines. Likewise, the results obtained from the European project HaHacronym (Stock & Strapparava, 2005) demonstrated how the incongruity and the opposite senses are relevant triggers for generating humorous meanings.

The recognition task, which implies detection and extraction of HUMOR descriptors by means of analyzing textual information, has shown that it is possible to learn the discriminating features that define the humorous samples. The research works in (Mihalcea & Strapparava, 2006a,b; Mihalcea & Pulman, 2007; Sjöbergh & Araki, 2007; Buscaldi & Rosso, 2007) have focused on analyzing *verbal HUMOR*, i.e. HUMOR expressed through language, contributing to provide different features for defining their data as humorous. Some of the features are ambiguity, irony, adult slang, antonymy, human centric vocabulary, negative orientation, bag of words, n-grams or professional communities. Although Mihalcea & Pulman (2007) have experimented with humorous news articles, others works have focused on the analysis of one-liners, which are short humorous structures that produce their comic effect with few words. The results reported by these authors are encouraging, despite the one-liners, whose properties suppose to learn more complex features to recognize whether an input is humorous. For instance, let us consider the following example:

(a) Children in the back seats of cars cause accidents, but accidents in the back seats of cars cause children.

The humorous effect in this sentence is caused by the interrelation of opposite concepts given the focused elements in the syntax, i.e. the subjects: children and accidents, respectively. This information is not at surface level and it is necessary to find strategies and methodologies that extract and represent the knowledge that is not given a priori and that determines the relations which turn a sentence into humorous or serious. Thus, to obtain the knowledge for identifying what are the features that best describe the patterns that produce HUMOR, Automatic HUMOR Recognition relies on models and resources that take advantage of linguistic information for describing features such as antonym, alliteration or ambiguity.

3. HUMOR FEATURES

As noted in the preceding section, the features obtained through the analysis of one-liners have allowed the automatic discrimination of humorous from non-humorous data with a high percentage of accuracy. That is why we think that these features can be employed for describing another kind of humorous data beyond only one-liners. This means that the underlying concepts that trigger the humorous effect in the one-liners are common to any kind of joke and, consequently, to any kind of verbal HUMOR. For instance, a feature such as adult slang is not a one-liner privative feature. This feature appears in other data, such as in punning riddles or in discussions about HUMOR. Therefore, we think that the set of features that has showed its effectiveness for discriminating humorous from serious data could represent elemental concepts that may be used to build a general HUMOR taxonomy. In this framework, our objective is to assess some of the most relevant features reported in the literature as general descriptors of HUMOR, specifically HUMOR produced by one-liners, to find some hints for conceptualizing a taxonomy about HUMOR. We expect that such features provide information for classifying any kind of verbal HUMOR.

We addressed this issue through a feature extraction task and an automatic classification process. That is, given the one-liners corpus used by Mihalcea & Strapparava (2006a,b) in their experiments, we extracted the main features they reported for characterizing HUMOR. Besides those features, we performed several experiments over the same corpus to retrieve other kind of discriminating elements. Afterward, using the whole set of features, we evaluated the importance of each through an automatic classification process over a test set composed by a collection of blogs whose main topic is HUMOR.

The features we considered in this research work, according to the results depicted in (Mihalcea & Strapparava, 2006a,b; Mihalcea & Pulman, 2007), were as follows:

1. *stylistic features*, focusing on adult slang;
2. *human centric vocabulary*, focusing on personal pronouns;
3. *human centeredness*, focusing on social relationships;
4. *polarity*, focusing on the positive or negative orientation of the data.

Alongside these features, we took into account the following aspects:

1. wh–phrases, focusing on interrogative pronouns;
2. nationalities, focusing on adjectives of nations;
3. keyness, focusing on the extraction of the most representative subjects of the data.
4. discriminating items, focusing on the words that integrate a same cluster;
5. ambiguity, focusing on the sense dispersion of the words.

These features were tested using the Naïve Bayes and the Multinomial Logistic Regression (MLR) classifiers (Witten & Frank, 2005). The data sets and experiments performed are described in the following section.

3. EXPERIMENTS ON FEATURES EXTRACTION

The experiments reported in this section are divided in two phases: (1) in the first, we extracted, from the one-liners corpus, all the features we described in the previous section (see Sections 4.2 to 4.10); (2) in the second, we automatically labeled each blog according to these features (Section 4.11).

4.1 Data Sets

The corpus of one-liners was automatically collected from the web through a bootstrapping process described in (Mihalcea & Strapparava, 2005). This process contains 16,000 one-liners. This corpus, as we have already mentioned, was the main database for extracting the features that define HUMOR. On the other hand, we decided to test the set of features over a collection of blogs because, being a heterogeneous site where HUMOR is represented not only by one-liners but also by jokes, gags, punning riddles, or even by humorous and serious discussions, blogs are a good source to study any type of verbal HUMOR. The collection of blogs we used was retrieved from the web through an automatic request to the Google search engine. Keywords such as *punch line*, *HUMOR*, *joke*, *funny*, *laughter*, *laugh line*, *gag*, *gag line*, *tag line,* and so on, were the seeds for retrieving the results. A total of 200 humorous blogs integrated the collection. Some statistics about the collection are: 23,363 types; 168,100 tokens; tokens/types relation = 7.19. This collection, enhanced with more blogs, will be made soon available.

Given the automatic process, it is possible that the blogs had information not related to HUMOR. Thus, for minimizing the noise, the collection was evaluated according to the measures proposed in (Pinto, 2008) for estimating features on corpora, such as *domain broadness*, *shortness*, *stylometry,* and *structure*. Before measuring these features on the collection, we eliminated the stopwords, enhancing the list with words such as *login, username, copyright, next, top*, etc., to delete information not related to the topic of the request. Using the Watermarking Corpora On-line System (WaCOS)[1], i.e. the tool that implements all the measures mentioned, we obtained the following indicator about the collection:

1. Domain broadness: wide.
2. Shortness: short texts.
3. Stylometry: general language style.
4. Structure: complex.

This information indicates: (1) the collection is not restricted only to one topic (broadness), for instance politics, but several ones. This feature impacts on the fact of having different kinds of discourses expressing humorous information; (2) the blogs we will classify are written without following a standard pattern (stylometry), whereby they do not share a surface similarity (structure) that implies a trend in the way that HUMOR is expressed. According to this information, the collection seems to be wide enough for covering a broad spectrum of HUMOR, and manners in which it is linguistically expressed. Therefore, we considered the collection as valid for our purposes. We provide a sample about the information contained in the blogs in Appendix A.

4.2 Stylistic Features

According to (Mihalcea & Strapparava, 2006b), the sexual information, such as in example (b), represents one of the most relevant features for discriminating HUMOR. Therefore, we reproduced their experiment about *adult slang* extracting all the words labeled with the tag “sexuality” in WordNet Domains (Bentivogli et al., 2004) for getting the first feature.

(b) Artificial Insemination: procreation without recreation.

4.3 Human Centric Vocabulary

One of the most important features reported in the literature is the presence of words that make reference to *human-related scenarios*. For instance, the pronoun *you* appears with a frequency greater than 25% in the one-liners whereas the pronoun *I* occurs 15% (Mihalcea & Strapparava, 2006b). That is why we selected personal pronouns, specifically first, second, and third (masculine and feminine) singular, for integrating the set of elements in this feature. Besides these, we included their correspondent reflexive pronouns for obtaining a broader coverage on this feature.

4.4 Human Centeredness

As reported in (Mihalcea & Pulman, 2007), *human centeredness* tends to find the most discriminating features in humorous data given four a priori semantic classes: *persons*, *social groups*, *social relations* and *personal pronouns*. We selected only the most salient class for representing this feature: *social relations*. The items that integrated this class were chosen as the authors reported, i.e. retrieving all the nouns that deal the synsets *relation*, *relationship,* and *relative* in the WordNet ontology (Miller, 1995).

4.5 Polarity

According to (Mihalcea & Pulman, 2007), the negative orientation is a very important discriminating feature when speaking of HUMOR. Therefore, to verify this assertion, we automatically labeled the corpus with a public tool for Sentiment Analysis: Java Associative Nervous Engine (Jane16)[2]. The underlying algorithm of this tool creates a model of positive and negative words and sentences that are crawled in Internet. Depending on their occurrences, these are ranked, and a weight is assigned to each. In this way, the positive and negative data sets are retrieved. The labeling process matches the information provided with the one in the database and computes the occurrence and weights for assigning the correspondent label.

Furthermore, we used SentiWordNet[3] (Esuli & Sebastiani, 2006) over a set of elements that we found in the experiments reported in Section 4.9 to assess the role that a small set of words could play in the overall objective of this work.

4.6 WH-phrases

A common humorous structure handled in Computational HUMOR is the punning riddle (Binsted, 1996). This kind of jokes takes advantage of syntactic recurrent templates: the *wh−phrases*. These structures are syntactic constituents that are characterized by question words or entire phrases. An example of these structures is given in (c).

(c) *What* are the 3 words you never want to hear while making love? Honey, I'm home!

The quantity of jokes that relies on this template is substantial, at least in the corpus of one-liners. That is why we used the interrogative pronouns for representing a HUMOR feature that may provide elements about the most profiled topics in the jokes.

4.7 Nationalities

Professional communities are common elements that have been associated to HUMOR (Mihalcea & Strapparava, 2006 a,b). For instance, the one-liner that appears in (d) is a sample about this assumption.

(d) *Parliament* fighting inflation is like the *Mafia* fighting crime.

In our study, instead of using this category, we employed a wordlist with adjectives of nationalities for noticing whether or not the information about toponyms is as relevant as the professional communities.

4.8 Keyness

We extracted the most representative items from the one-liners corpus according to their keyness value. This measure estimates the keyness through comparing the frequency of each word in a corpus against the frequency of the same word in a reference corpus. The values are computed taking into account the Log Likelihood test (Dunning, 1993).

To retrieve items with a greater keyness value, we generated a list with all the words from the one-liners corpus, except the stopwords. Likewise, for obtaining a reference corpus, we used the Google n-grams (Brants & Franz, 2006), specifically, the 3-gram section. Given both corpora, we computed the keyness.

Furthermore, we added the items retrieved from the scope module, included in the Jane16 tool, for enhancing the set of keywords. Some representative items according to the keyness value and Jane16 are *mad, paranoid, sick, hell* or *mistake*.

4.9 Discriminating Items

To identify how much similar the items in the one-liners are and be able to determine a set of discriminating features, we carried out five different clustering experiments. We employed two tools: Cluto and SenseClusters. Cluto[4] is a set of algorithms that "operate either directly in the objects feature space or in the objects similarity space" (Karypis, 2003), maximizing or minimizing a criterion function over the solution. On the other hand, SenseClusters[5] is a package that integrates Cluto's algorithms besides a set of tools for identifying similar contexts. SenseClusters works seeking "lexical features to build first and second order representations of contexts" (Kulkarni & Pedersen, 2005).

In the first experiment, we considered Cluto, selecting the following criteria: vector space, direct clustering method, H2 criterion function, and cosine similarity function[6]. The number of requested clusters was 20. Figure 1 shows the distribution of the most discriminating elements in each one of the 20 clusters. The rest of the clustering experiments were carried out employing SenseClusters. In each we varied the parameters and the number of requested clusters. We present in Table 1 the criteria employed in these four experiments[7].

The set of all discriminating items generated with both Cluto and SenseClusters were recorded in a wordlist for removing all duplicates. The remaining items were first labeled with their POS tags using FreeLing[8], described in (Atserias et al., 2006), and then with their positive, neutral or negative polarity tag according to Senti WordNet. This list was used for a second polarity labeling (see Section 4.5) over the collection of blogs. We provide in Appendix B a list with the first 50 discriminative items labeled with their POS and polarity tags.

4.10 Ambiguity

Several research works have pointed out that HUMOR takes advantage of linguistic ambiguity for producing its funny effect (Mihalcea & Strapparava, 2006a,b;

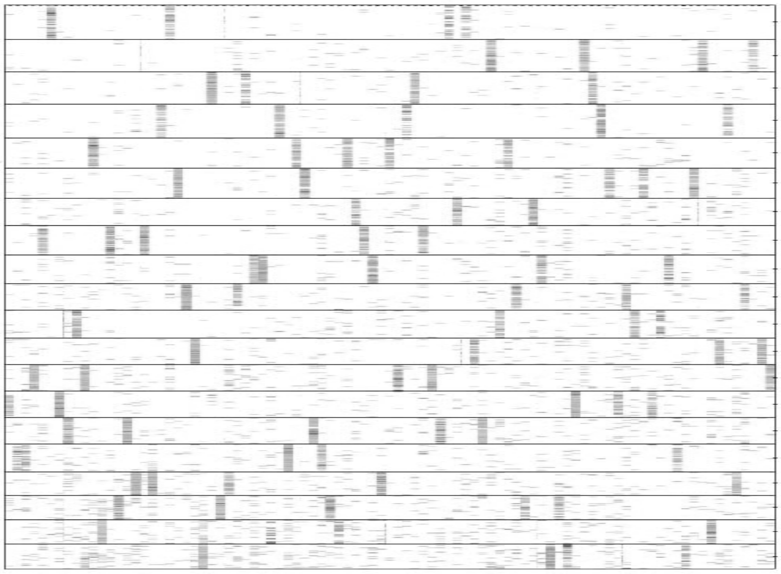

Fig. 1. Distribution of the most discriminating items. The rows represent each cluster and the dots indicate how the items are distributed in the cluster.

TABLE 1

Parameters employed in the experiments performed with SenseClusters.

Exp.	Space	Cl. Method	Cr. Function	LSA	Order	Cluster stop	Clusters
1	Vector	RB/Direct	UPGMA	Yes	Bi/Co	All	2
2	Similarity	Agglo/RBR/Graph	H2	Yes	Uni/Bi	Gap	2
3	Vector	Direct	H2	Not	Co	None	20
4	Vector	RB	I2	Not	Bi	Pk	27

Sjöbergh & Araki, 2007; Reyes et al., 2009a,b). That is why we performed an experiment for verifying how much valuable information could provide a HUMOR characterization through the representation of semantic ambiguity. The experiment consisted in measuring the sense dispersion for each noun of the one-liners. This measure is based on the hypernym distance between synsets, calculated with respect to WordNet. This distance was calculated using the formula depicted in (Reyes et al., 2009a), which appears in Eq. (1):

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

where S is the set of synsets (s_1,...,s_n) for 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). For instance, according to WordNet v. 3.0, the noun *killer* has four synsets. Taking into account only the synsets s_i and s_j, we obtain as first common hypernym *physical entity*. The number of nodes to reach this hypernym is 6 and 2, respectively. Thus, the dispersion of *killer* is the sum of those distances divided by 2. Now, considering all its synsets, we obtain six possible combinations whose distance among them and their first common hypernym generates a dispersion of 6,83. The formula in Eq. (2) shows how the total dispersion per noun is calculated:

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

where *W* is the set of nouns in the collection *N*. The underlying assumption of this measure is to quantify the difference among the senses of a word, meaning that a word with senses that differ significantly is more likely to be used to create HUMOR than a word with senses that differ slightly. The average sense dispersion of the whole set of nouns in the one-liners corpus (calculated as: $\delta_w = \frac{\delta_{TOT}}{|W|}$) was 7.63.

In the last experiment of the features extraction phase, we calculated the total sense dispersion per isolated noun. The results were recorded in a list for measuring the average sense dispersion in the blogs. In Figure 2 we depict the overall distribution of every feature in terms of number of items[9].

4.11 Feature Representativeness

Once we had finished the feature extraction task, we investigated how representative every feature was. To verify such representativeness, we looked for the features in the blogs through a binary distinction: absence/presence. We used the following algorithm:

1. Let (i_1, i_n) be the items that define every feature f_h.
2. Let (b_1, b_m) be the collection of blogs.
3. If any i_n occurs in b_m with frequency $\geq$ 4, then f_h was a representative feature for b_m.

Besides searching the representativeness of each feature, we measured the total sense dispersion for all the blogs according to the formula described in Section 4.10. The results obtained are depicted in Figure 3. The graph (a) shows how representative, in terms of presence in the entire collection, is every feature; whereas the graph (b) displays the total sense dispersion per blog.

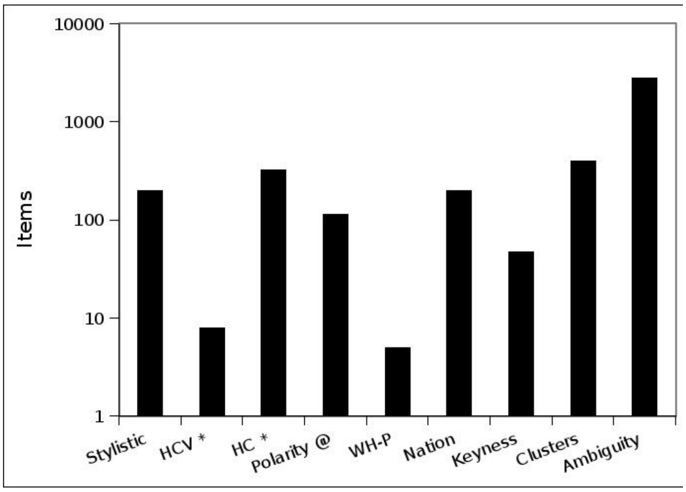

Fig. 2: Items retrieved per feature

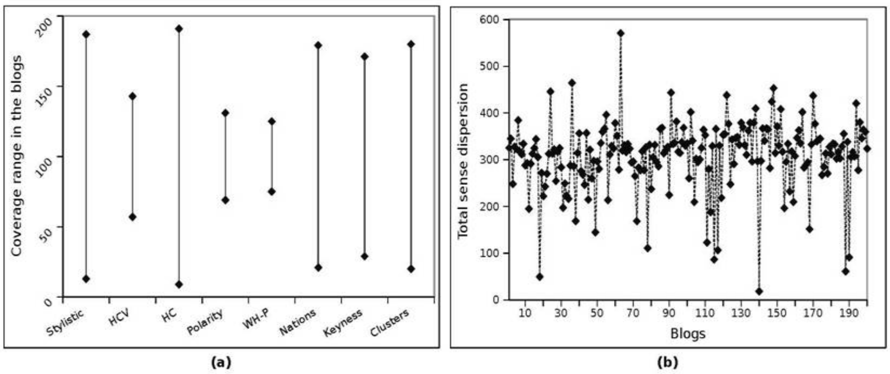

Fig. 3: Feature representativeness in the collection: (a) general features coverage, (b) sense dispersion per blog.

5. EVALUATION AND DISCUSSION

The classification experiments were carried out to understand the difficulty of automatically detecting the features and evaluating whether this set can provide hints about building a HUMOR taxonomy. We automatically classified the blogs using the Bayes and MLR classifiers included in Weka. We evaluated each classifier employing all the features. Given the size of the collection, we used the leave-one-out method for assessing the performance. For the polarity feature, besides using the results obtained with Jane16, we incorporated those of SentiWordNet, dividing the items retrieved in Positive and Negative, according to their polarity tag, and removing all those labeled with neutral tag (see Appendix B). In addition, with respect to sense dispersion, given the differences among ranges, we normalized the values assigning 0 to all the values between 0 and 200; 1 to the values between 201 a 400; and 2 from 401 onward. The results obtained are displayed in Figure 4. The graph (a) shows the classification accuracy for the state-of-the-art reported features, including SentiWordNet polarity results; whereas the graph (b) shows the classification accuracy for the rest of features.

According to the information illustrated in Figures 3 and 4, we can notice that there are features that play a more important role for characterizing the manner in which the bloggers are expressing HUMOR. For instance, the WH-phrases do not seem to be relevant for representing HUMOR. Likewise, the Jane16's polarity results do not reflect the behavior reported in (Mihalcea & Pulman, 2007). This behavior is

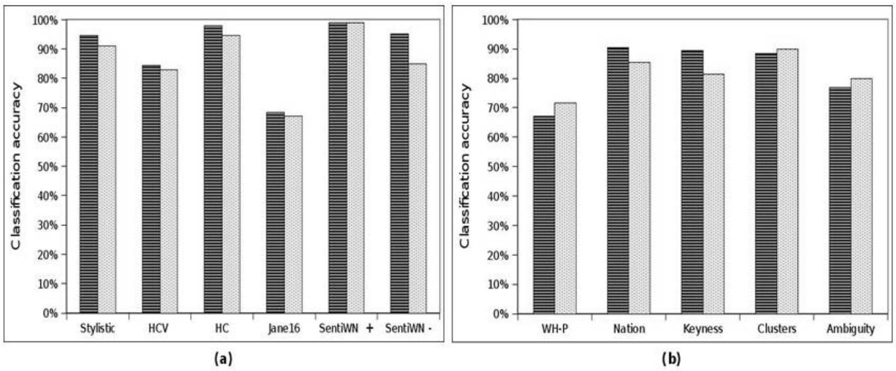

Fig. 4: Classification accuracy for: (a) state-of-the-art and (b) new mined features.

probably due to the polarity data sets employed. Ambiguity also does not seem to play an important role, at least as it is intended as "WordNet ambiguity". However, the experiments must be run over a bigger collection of blogs (or other kind of data: newspapers, political discourse, mass media information, etc.) to verify this behavior. Moreover, it is evident that the state-of-the-art features have overall better learning curves than the new ones. This fact, besides ratifying the results reported in the literature, establishes that the items of these features may represent basic elements for producing a joke, whereas the items in the new features may help to represent background or adjacent information in a humorous process.

Furthermore, after a manual analysis of the results obtained both with the feature extraction task and the classification process, we considered two classes of features for the taxonomy: low level and high level features. The first class integrates a set of features that represent prototypical information for characterizing HUMOR, meaning that several items are used recurrently to promote humorous situations and, consequently, they may be identified as common humorous topics. For instance, Mihalcea and Strapparava (2006a,b) pointed out that sexuality or self-referential elements are present in several jokes. These elements, according to our study, could be represented by the features *slang, personal pronouns, relationship categories, nationalities,* or *keyness*.

With regard to the second class, we consider "high level features" the information that is not clearly related to humorous topics, as the previous ones, but rather is used for producing HUMOR through linguistic strategies. Under this perspective, features such as *polarity, discriminating items*, or *ambiguity* play as the source that represents this class. For instance, in example (e), we can realize how HUMOR is generated by information that is not related to any of the state-of-the-art features, not even by a polarity clue. However, its funny effect relies on information beyond the presence of prototypical items but on the use of linguistic ambiguity as a trigger of the HUMOR effect.

(e) Jesus saves, and at today's prices, that's a miracle!

Now, from the two classes mentioned above and with the available features, we think it possible to build a general structure that roughly represents the underlying HUMOR's topics. In Figure 5 we illustrate how we conceptualize, from the results obtained in this research work, the HUMOR taxonomy.

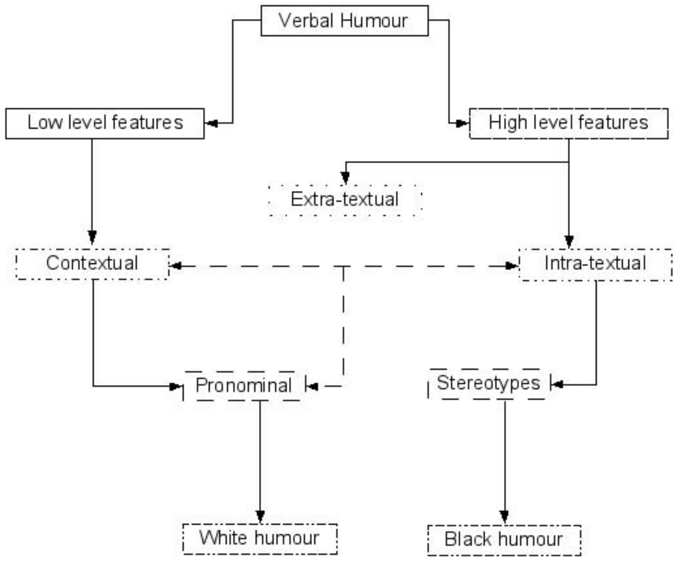

Fig. 5: Toward a verbal humor taxonomy

As noted in this figure, we can identify and extract, according to the items that more recurrently appear, subclasses such as:

1. stereotypes, HUMOR about ethnic groups;
2. pronominal, self-referential HUMOR;
3. white HUMOR, positive polarity orientation;
4. black HUMOR, negative polarity orientation.

And deeper representations such as:

1. contextual, based on items that denote exaggeration, incongruity or absurd;
2. intra-textual, based on linguistic ambiguity;
3. extra-textual, based on pragmatic and cultural information.

6. CONCLUSIONS AND FUTURE WORK

In this study, we evaluated the set of features that have been identified in the main research works on Automatic HUMOR Recognition as discriminating items between humorous and non humorous texts (specifically with respect to one-liners), and a set of new ones obtained on the basis of the study of the *keyness value*, *nationalities*, *discriminating items*, and *ambiguity*, to establish some basic parameters for characterizing any kind of verbal HUMOR. We aimed at assessing this hypothesis through a classification task over a collection of blogs automatically retrieved from the Internet, and whose main topic was HUMOR. The results give us some clues about which features have a greater weight for defining HUMOR. Moreover, it seems probable that, through the items that constitute every feature, some of them may be used for conceptualizing basic information for building a verbal HUMOR taxonomy.

As further work, besides verifying the behavior of these features with more data, we aim at investigating what are the most informativeness features or whether the presence of any of them may change the humorous meaning. Moreover, due to our aim of establishing a verbal HUMOR taxonomy, we plan to verify this behavior in other languages, (besides other kind of data: mass media information, political discourse, etc.), to take benefit of the insight obtained from this research work for tasks such as machine translation or information filtering.

ACKNOWLEDGMENT

The TEXT-ENTERPRISE 2.0 (TIN2009-13391-C04-03) research project has partially funded this work.

REFERENCES

Atserias, J., Casas, B., Comelles, E., González, M., Padró, L., and Padró, M. 2006. FreeLing 1.3: Syntactic and semantic services in an open-source NLP library, in: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006), ELRA*. 48-55.

Attardo, S. 2001. *Humorous Texts: A semantic and pragmatic analysis*. Berlin: Mouton de Gruyter.

Balog, K., Mishne, G., and Rijke, M. 2006. Why Are They Excited? Identifying and Explaining Spikes in Blog Mood Levels, in: *Proceedings of the 11th Meeting of the European Chapter of the Association for Computational Linguistics*. 207-210.

Bentivogli, L., Forner, P., Magnini, B., and Pianta, E. 2004. Revising the Wordnet Domains Hierarchy: semantics, coverage and balancing, in: *COLING 2004 Multilingual Linguistic Resources*. 94-101.

Binsted, K. 1996. *Machine HUMOR: An implemented model of puns*. PhD thesis. University of Edinburgh, Scotland.

Binsted, K., Ritchie, G. 1997. Computational rules for punning riddles, *Humor* **10**, 25-75.

Binsted, K., Ritchie, G. 2001. Towards a model of story puns, *Humor* **14(3)**, 275-292.

Buscaldi, D., Rosso, P. 2007. Some experiments in HUMOR Recognition using the Italian Wikiquote collection, in: *Proceedings of the Workshop on Cross Language Information Processing*, 464-468.

Brants, T., Franz, A. 2006. *Web 1T 5-gram corpus version 1*.

Dunning, T. 1993. Accurate Methods for the Statistics of Surprise and Coincidence, *Computational Linguistics* **19(1)**, 61-74.

Esuli, A. and Sebastiani, F. 2006. SentiWordNet: A publicly available lexical resource for opinion mining, in: *Proceedings of LREC-06, the 5th Conference on Language Resources and Evaluation. (2006)*, 417-422.

Ghose, A., Ipeirotis, P. and Sundararajan, A. 2007. Mining using Econometrics: A Case Study on Reputation Systems, in *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics. (2007)*. 416-423.

Karypis, G. 2003. CLUTO. A Clustering Toolkit. Technical Report 02-017, University of Minnesota, Department of Computer Science.

Kulkarni, A., Pedersen, T. 2005. SenseClusters: Unsupervised Clustering and Labeling of Similar Contexts, in: P*roceedings of the Demonstration and Interactive Poster Session of the 43rd Annual Meeting of the Association for Computational Linguistics. (2005)*, 105-108.

Mihalcea, R. 2007. Multidisciplinary Facets of Research on HUMOR, in: *Proceedings of the Workshop on Cross-Language Information Processing*, 412-421.

Mihalcea, R., Strapparava, C. 2005. Bootstrapping for Fun: Web-based Construction of Large Data Sets for Humor Recognition, in: *Proceedings of the Workshop on Negotiation, Behavior and Language (FINEXIN 2005),* 84-93.

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

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

Mihalcea, R., Pulman, S. 2007. Characterizing HUMOR: An Exploration of Features in Humorous Texts, in: *Proceedings of the Conference on Computational Linguistics and Intelligent Text Processing*, 337-347.

Miller, G. 1995. WordNet: A lexical database, *Communications of the ACM* **38(11)**, 39-41.

Pang, B., Lee, L., and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques, in: *Proceedings of the empirical methods in natural language processing EMNLP. (2002)*, 79-86.

Pinto, D. 2008. *On Clustering and Evaluation of Narrow Domain Short-Text Corpora*. PhD thesis. Universidad Politécnica de Valencia, Spain.

Reyes, A., Buscaldi, D., Rosso, P. 2009a. The Impact of Semantic and Morphosyntactic Ambiguity on Automatic HUMOR Recognition, in: *Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems (NLDB) 2009*.

Reyes, A., Buscaldi, D., Rosso, P. 2009b. An Analysis of the Impact of Ambiguity on Automatic HUMOR Recognition, in: *Proceedings of the 12th International Conference Text, Speech and Dialogue (TSD) 2009*, 162-169.

Ritchie, G. 2003. *The Linguistic Analysis of Jokes*. Routledge.

Sjöbergh, J., Araki, K. 2007. Recognizing Humor without Recognizing Meaning, in: *Proceedings of the Workshop on Cross-Language Information Processing*, 469-476.

Stock, O., Strapparava, C. 2005. Hahacronym: A computational humor system, in: *Demo proc. of the 43rd annual meeting of the Association of Computational Linguistics (ACL05)*, 113-116.

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

APPENDIX A: SAMPLE OF BLOGS

The following fragments represent the kind of information we found in the blogs.

- You don't have to read all the way through. If you just skim read it then you get the general gist of it and it is mediocally funny. The point of the joke (i think) is that it is long and slightly boring (THATS THE POINT!!!!!) and this is one joke on this website that i actually felt was slightly funny. If they made the joke shorter then there wouldn't be a joke at all!!!!
- An Englishman, an American and an Italian are having a conversation, praising their respective countries. The Englishman says: -During the last war we had a ship so large, but so large that for docking maneuvers we needed 24 hours. The American reply: -We had a ship so big that to move on it, there was a bus service. And the Italian: -This is nothing. We had a ship so large that when at bow the war was over, stern even knew that was started.
- A man and his wife were spending the day at the zoo. She was wearing a loose fitting, pink dress, sleeveless with straps. He was wearing his usual jeans and T-shirt. As they walked through the ape exhibit, they passed in front of a large, silverback gorilla. Noticing the wife, the gorilla went crazy. He jumped on the bars, and holding on with one hand and 2 feet he grunted and pounded his chest with his free hand. He was
- Obviously excited at the pretty lady in the pink dress. The husband, noticing the excitement, thought this was funny. He suggested that his wife tease the poor fellow some more by puckering her lips and wiggling her bottom. She played along and the gorilla got even more excited, making noises that would wake the dead. Then the husband suggested that she let one of her straps fall to show a little more skin. She did and the gorilla was about to tear the bars down. Now show your thighs and sort of fan your dress at him, he said. This drove the gorilla absolutely crazy, and he started doing flips. Then the husband grabbed his wife, ripped open the door to the cage, flung her in with the gorilla and slammed the cage door shut. Now. Tell him you have a headache.
- The final piece of advice is writing humor takes time. To excel in humor is a lifetime job, and is not something that you can learn in a day or two. Don't think you can read a joke book and start writing funny stuff an hour later. You will have to teach yourself how to be funny. The process is mostly by trial and error, observing other people's comical situations, mistakes, laughing and applying it

on yourself, etc. No one can teach you exactly how to write something funny, but the possibilities of creating humor on anything and everything are limitless.

- Many companies hold information meetings in the office is not practicing humor, because they do not want to have one of the workers who will be offended. However, at the time the company can cross boundaries on what is acceptable and not acceptable.
- Part of the problem with people telling funny jokes or humor is not acceptable is that if someone can not enjoy the job itself in the workplace will be a drab and unhappy workers.

APPENDIX B: DISCRIMINATING ITEMS

This section shows the 50 most discriminating items, according to their POS and polarity tags. It is necessary to mention that, when an item belongs to different synsets, it was assigned to the first polarity tag according to its POS tag.

Positive:

- *Adjectives*: damned, easy, funny, highly, hot, meek, nice, perfect, positive, real, close, good, weak, fine, wise.
- *Nouns*: art, bag, care, chance, education, energy, eye, fault, freedom, fun, genius, home, ignorance, importance, law, license, line, mind, shareware, strength, word.
- *Verbs*: die, lose, raise, speed, teach, create, see, call, feel, learn, think, understand.

Neutral:

- *Adjectives*: circular, foolish, front, future, green, homosexual, indecisive, Irish, lethal, many, married, middle, own, personal, photographic, proportional, remote, suitable, unanimous, more, hard, little, usually.
- *Nouns*: action, advance, age, air, alcohol, amount, application, arrest, ass, bar, basket, bathroom, bed, beer, being, bite, blood, body, boss, box, brain, bread, bulb, butter, car.

Negative:

- *Adjectives*: bad, common, dark, dead, dull, free, futile, hilarious, impossible, inversely, mad, negative, old, paranoid, sick, silent, stupid, wrong.

- *Nouns*: animal, bomb, bumper, code, difference, dream, fiction, habit, hell, hurry, hydrogen, matter, mistake, reason, season, shake, stupidity, system, telekinesis, terror, tourist, trouble, worth.
- *Verbs*: clean, forget, keep, succeed, hurt, kill, misquote

ENDNOTES

1 Tool available at: http://users.dsic.upv.es/grupos/nle/demos.html.
2 Tool available at: http://opusintelligence.com/download.jsp.
3 Tool available at: http://sentiwordnet.isti.cnr.it.
4 Tool available at: http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview.
5 Tool available at: http://www.d.umn.edu/ tpederse/senseclusters.html.
6 For a detailed explanation about the meanings of these parameters, see [13].
7 The abbreviations in Table 1 indicate: Exp.: number of experiment. Cl. Method: clustering method employed where RB means repeated bisections; Agglo means agglomerative clustering; RBR means repeated bisections globally optimized; Graph means graph partitioning-based clustering. Cr. Function: criterion function employed. LSA: Latent Semantic Analysis representation. Order: represents contexts where Bi means bigrams; Uni means unigrams; Co means co-occurrences. Cluster stop: cluster stopping measure where Gap means adapted gap statistics; Pk means pk measures. For a detailed explanation consult http://search.cpan.org/dist/Text-SenseClusters.
8 Tool available at: http://garraf.epsevg.upc.es/freeling.
9 The * in Figure 2 refers to Human Centric Vocabulary and Human Centeredness, respectively, whereas the @ indicates the items retrieved from SentiWordNet.

APPENDIX A: SAMPLE OF BLOGS

The following fragments represent the kind of information we found in the blogs.

- You don’t have to read all the way through. If you just skim read it then you get the general gist of it and it is mediocally funny. The point of the joke (i think) is that it is long and slightly boring (THATS THE POINT!!!!!) and this is one joke on this website that i actually felt was slightly funny. If they made the joke shorter then there wouldn’t be a joke at all!!!!
- An Englishman, an American and an Italian are having a conversation, praising their respective countries. The Englishman says: -During the last war we had a ship so large, but so large that for docking maneuvers we needed 24 hours. The American reply: -We had a ship so big that to move on it, there was a bus service. And the Italian: -This is nothing. We had a ship so large that when at bow the war was over, stern even knew that was started.
- A man and his wife were spending the day at the zoo. She was wearing a loose fitting, pink dress, sleeveless with straps. He was wearing his usual jeans and T-shirt. As they walked through the ape exhibit, they passed in front of a large, silverback gorilla. Noticing the wife, the gorilla went crazy. He jumped on the bars, and holding on with one hand and 2 feet he grunted and pounded his chest with his free hand. He was
- Obviously excited at the pretty lady in the pink dress. The husband, noticing the excitement, thought this was funny. He suggested that his wife tease the poor fellow some more by puckering her lips and wiggling her bottom. She played along and the gorilla got even more excited, making noises that would wake the dead. Then the husband suggested that she let one of her straps fall to show a little more skin. She did and the gorilla was about to tear the bars down. Now show your thighs and sort of fan your dress at him, he said. This drove the gorilla absolutely crazy, and he started doing flips. Then the husband grabbed his wife, ripped open the door to the cage, flung her in with the gorilla and slammed the cage door shut. Now. Tell him you have a headache.
- The final piece of advice is writing humor takes time. To excel in humor is a lifetime job, and is not something that you can learn in a day or two. Don’t think you can read a joke book and start writing funny stuff an hour later. You will

have to teach yourself how to be funny. The process is mostly by trial and error, observing other people's comical situations, mistakes, laughing and applying it on yourself, etc. No one can teach you exactly how to write something funny, but the possibilities of creating humor on anything and everything are limitless.

- Many companies hold information meetings in the office is not practicing humor, because they do not want to have one of the workers who will be offended. However, at the time the company can cross boundaries on what is acceptable and not acceptable.
- Part of the problem with people telling funny jokes or humor is not acceptable is that if someone can not enjoy the job itself in the workplace will be a drab and unhappy workers.

APPENDIX B: DISCRIMINATING ITEMS

This section shows the 50 most discriminating items, according to their POS and polarity tags. It is necessary to mention that, when an item belongs to different synsets, it was assigned to the first polarity tag according to its POS tag.

Positive:

- *Adjectives*: damned, easy, funny, highly, hot, meek, nice, perfect, positive, real, close, good, weak, fine, wise.
- *Nouns*: art, bag, care, chance, education, energy, eye, fault, freedom, fun, genius, home, ignorance, importance, law, license, line, mind, shareware, strength, word.
- *Verbs*: die, lose, raise, speed, teach, create, see, call, feel, learn, think, understand.

Neutral:

- *Adjectives*: circular, foolish, front, future, green, homosexual, indecisive, Irish, lethal, many, married, middle, own, personal, photographic, proportional, remote, suitable, unanimous, more, hard, little, usually.
- *Nouns*: action, advance, age, air, alcohol, amount, application, arrest, ass, bar,

basket, bathroom, bed, beer, being, bite, blood, body, boss, box, brain, bread, bulb, butter, car.

Negative:

- *Adjectives*: bad, common, dark, dead, dull, free, futile, hilarious, impossible, inversely, mad, negative, old, paranoid, sick, silent, stupid, wrong.
- *Nouns*: animal, bomb, bumper, code, difference, dream, fiction, habit, hell, hurry, hydrogen, matter, mistake, reason, season, shake, stupidity, system, telekinesis, terror, tourist, trouble, worth.
- *Verbs*: clean, forget, keep, succeed, hurt, kill, misquote