

“In reality there are as many religions as there are papers” – First Steps Towards the Generation of Internet Memes

Diogo Costa, Hugo Gonçalo Oliveira, Alexandre Miguel Pinto

CISUC, Department of Informatics Engineering

University of Coimbra, Portugal

dcosta@student.dei.uc.pt, hroliv@dei.uc.pt, ampinto@dei.uc.pt

Abstract

We report on the first steps towards the automatic generation of Internet memes starring public figures. Their images are retrieved from the Web and combined with famous quotes, altered according to recent information on the figures. Current implementation, in Portuguese, exploits several computational resources and aims to produce artifacts with coherent text, image, and some humor value. A preliminary evaluation survey confirmed a strong relation between generated memes and present events. Results on humor were also positive.

Introduction

The term meme originally denotes *an idea, behavior, or style that spreads from person to person within a culture* (Dawkins, 1976; Blackmore, 2000). On the Internet domain, memes became a popular and effective way of transmitting an idea. They are a product of human creativity that typically take the form of an image, often combined with a short phrase. They tend to be funny, make people laugh, and aim to be spread throughout the World Wide Web by sharing and re-sharing in social media.

We present the first steps towards the development of MemeGera, a system for the automatic generation of Internet memes – or better, protomemes¹ – starring public figures (hereafter, characters). MemeGera uses famous quotes, altered as follows: one word is replaced by another that is semantically related to the character and its current information². These sentences, presented together with a character’s image, should convey a simple and effective idea, make sense for the character, even if only for a short period of time after generation, and exhibit some novelty. To deal with the latter, the system exploits fresh information on the character, such as that in recent news or tweets. The produced text+image combinations have thus a transient flavor which, together with their humor potential, may qualify them as “jokes *du jour*”. Long-term knowledge on the character, from its Wikipedia page, is also explored, but so far only used to favor fresh information.

¹The definition of meme implies social sharing, which will only occur if people actually spread the protomeme.

²As the title of the paper, a twist of Mahatma Gandhi’s quote: “In reality there are as many religions as there are individuals”.

We see the generation of meme sentences as a kind of linguistic creativity, a topic that covers tasks such as the generation of: poetry (Toivanen, Gross, and Toivonen, 2014; Gonçalo Oliveira and Cardoso, 2015); metaphors (Veale and Hao, 2008); neologisms (Smith, Hintze, and Ventura, 2014); or verbally-expressed humor (Binsted and Ritchie, 1994; Valitutti et al., 2013). Given the funny aspect inherent to memes, our work is probably closer to the latter. Yet, although not essential, the character and its image also play an important role in the success of our memes.

In the remaining of this paper, we provide some background knowledge on the study of humor, together with computational approaches to this topic. We then present the automatic method for generating memes and list each of the steps involved. The current implementation targeted Portuguese, our native language, and is described right after. Although the method may seem quite straightforward, our effort involves the combination of several knowledge sources. Before concluding, we describe an illustrative example and report on the results of an online survey, which suggests that we are heading in the right direction. All the memes used in the survey are shown in the end of the paper, together with information about their generation and evaluation.

Background and Related Work

This section addresses the topic of humor from a theoretical point of view, followed by an enumeration of computational approaches for humor generation and recognition.

Theoretical Study of Humor

Humor has been studied from a variety of perspectives ranging from psychology and philosophy (Morreall (2013)), to its sociological aspects in literature (Kuipers (2010)) and, more recently, via the computational approach (e.g. Suslov (1992); Ritchie (2014)). Theoretical accounts of humor encompass the superiority theory, endorsed by Descartes, where “*our laughter expresses feelings of superiority over other people or over a former state of ourselves*”; the relief theory, a hydraulic model proposed by Shaftesbury, and later refined by Sigmund Freud, according to which laughter acts as a mechanism for releasing accumulated nervous energy built up from many possible emotionally-charged situations; and the incongruity theory, proposed by Beattie and sponsored by Kant, Schopenhauer, and Kierkegaard, among oth-

ers, which claims “laughter is the perception of something incongruous – something that violates our mental patterns and expectations”, which is now the dominant theory.

Socioliterary studies (e.g. Kuipers (2010)) explore the mechanisms through which humor is related to social boundaries, and how it differs between groups; whereas computational approaches address the building of formal theories of humor (Ritchie (2014)), the synthesis of a sense of humour via specific algorithms (Suslov (1992)), and the generation of humorous text and jokes (Ritchie (2009)).

Humor expressed in Portuguese has also been studied from a theoretical point of view. While presenting linguistic mechanisms for achieving humor in this language, Tagnin (2005) states that, since humor breaks conventionality in language, understanding it is a sign of fluency.

Humor generation

The automatic generation of humor has been a research topic for more than two decades. In early work by Binsted and Ritchie (1994), a model, implemented under the name of JAPE, was proposed for generating punning riddles. The generated puns (e.g. *What do you call a murderer that has fiber? A cereal killer*) took advantage of spelling or word sense ambiguities. STANDUP (Manurung et al., 2008) follows the lines of JAPE, but is more robust, user friendly, and was developed with the purpose of allowing young children, especially those with linguistic disabilities, to explore language and improve their skills.

Given a concept and an attribute, HAHAcronym (Stock and Strapparava, 2005) rewrites existing acronyms and generates new ones with a humor intent. It relies on an incongruity detector and generator that selects opposing domains and opposing adjectives, while considering also rhythm and rhymes. For instance, the acronym FBI may become *Fantastic Bureau of Intimidation*. Or given the concept of ‘processor’ and the attribute ‘fast’, it generates the acronym OPEN – *Online Processor for Effervescent Net*.

Valitutti et al. (2013) explored the generation of adult humor based on the replacement of a word in a short message. The word should introduce incongruity and lead to a humorous interpretation, achieved by applying three constraints. It must: (i) be of the same form as the original word, i.e. match the part-of-speech and either rhyme or be orthographically similar to the original word; (ii) convey a taboo meaning, e.g. an insult or something related to sex; (iii) take place at the end of the message and keep the coherence of the original sentence. An example of an output is: *I’ve sent you my fart.. I mean ‘part’ not ‘fart’....*

Besides English, there were attempts for generating puns in Japanese (e.g. Sjöbergh and Araki (2007a)). We are not aware of any work of this kind for Portuguese.

Humor recognition

In the scope of natural understanding, there has been work on the automatic recognition of verbally-expressed humor. Researchers typically focus on a specific kind of jokes, such as *knock-knock* (Taylor and Mazlack, 2004) and *That’s what she said* (Kiddon and Brun, 2011), or on a less specific kind of humor but transmitted in bounded kinds of text, such as

single sentences (Mihalcea and Strapparava, 2006; Sjöbergh and Araki, 2007b), or tweets (Barbieri and Saggion, 2014). Humor recognition is generally seen as a text classification problem and relies on a set of humor relevant features to train a classifier, given their presence in humorous and non-humorous text. For instance, Barbieri and Saggion (2014) exploit hashtags, such as #humuor or #irony, to collect positive examples. Selected features generally include the occurrence of antonymous or ambiguous words, alliteration, and other words or expressions typically used in jokes, such as slang or idiomatic expressions.

For Portuguese, the closest works to humor recognition we are aware of include the automatic detection of irony (Carvalho et al., 2009) or proverbs (Rassi, Baptista, and Vale, 2014) in text.

Internet Memes

Internet memes are a current trend in social media. They are typically a reusable combination of text and graphics. Popular memes include Boromir from the Lord of the Rings with the template “*One does not simply X*”, Morpheus from the Matrix with “*What if I told you Y*”, or Batman slapping Robin, with a personalized text in their speech balloons. There is however a subtype of Internet memes related to current events, where new images, text, or both, can be used – if successful enough, they might be reused. Events that triggered several memes include the football player Luis Suárez biting his opponent in a World Cup 2014 match (e.g. “*If you can’t beat them, eat them*”), or when the pop singer Madonna fell on stage, while wearing a cape, during a performance in the BritAwards 2015 ceremony (e.g. “*56 years old, still does her own stunts*”, “*Has a cape, can’t fly*”). While most memes show a break of conventionality (e.g. unexpected situation, confusing interpretation, taboo meaning), we address the previous subtype, which, as suggested by the superiority theory, makes fun of the portrayed character. In fact, the image is sometimes enough to make people laugh (e.g. when it displays a funny person or situation).

We are not aware of any published work on the automatic generation of Internet memes. Existing web services for meme generation rely on the user input of both images and text. There is work however on the automatic combination of images and text, such as Grafik Dynamo (2005) and Why Some Dolls Are Bad (2008), by Kate Armstrong³. In those projects, a narrative is dynamically generated by combining sequences of images, retrieved from social networks, with speech balloons. The result is often non-sense.

Method

This section provides a high-level description of our proposed method for meme generation. Specific details of its current implementation are given in the next section.

Among other parameters, our algorithm for the generation of memes (see figure 1) uses the name of a public figure, our character, currently provided by the user. Informally, it starts by retrieving n recent messages (e.g. tweets) mentioning the character, from where the top- k frequent nouns

³<http://katearmstrong.com/>

are collected. Then, it selects a random quote from a pool of famous quotes, pairs it with one of the top- k nouns, and generates a sentence, more precisely, an altered quote where the last noun of the original quote is replaced by one of the top- k nouns – similarly to Valitutti et al. (2013), replacing the last noun will increase surprise and humor potential. After repeating this process for a predefined number of times, generated sentences are ranked by a dedicated scoring function. The highest-ranked sentence is pasted on an image of the character, automatically retrieved from the Web, and the combination is finally returned as the generated meme.

The scoring function considers the humor value of the sentence, the frequency of the replacement noun, and its presence in a more stable long-term information source on the character. Words without previous associations to the character are considered novel and are thus favored in the ranking. We may find some parallelism between this and the work of Toivanen, Gross, and Toivonen (2014), where novel associations in documents are identified by their overlap with known associations from a background corpus.

Require:

```

charName: name of character
n : # of messages to retrieve
k : # of top frequent common nouns to consider
m : # of < quote, frequent noun > pairs to generate
1: procedure MEMEGERA
2:   messages ← {msg : msg mentions charName}
      : #messages = n
3:   freqNouns ← top-k most frequent nouns in messages
4:   quotes ← {quote : quote is a famous quote}
5:   pairs ← {< quote, freqNoun > randomly generated}
      : #pairs = m
6:   maxEval ← 0
7:   bestQuote ← ∅
8:   for each < quote, freqNoun > ∈ pairs do
9:     nq ← replace last noun in quote with freqNoun
10:    ne ← score(nq, freqNoun, charName)
11:    if ne > maxEval then
12:      maxEval ← ne
13:      bestQuote ← nq
14:   image ← get image of charName from the Web
15:   resultingMeme ← paste bestQuote in image
16:   return resultingMeme

```

Figure 1: Meme generation algorithm

Implementation

Although our method is language-independent, its current implementation targets Portuguese. MemeGera was implemented in Java and exploits several available resources, for different purposes, including a classifier for Portuguese humor, currently in development. We also describe the function that currently ranks the generated sentences.

Tools and Resources

Famous quotes used in this work were acquired from the Portuguese edition of Wikiquote⁴, a collaborative repository of quotes, run by the Wikimedia Foundation. For the current

⁴<http://pt.wikiquote.org/>

version of the system, we selected quotes from three well-known thinkers – Mahatma Gandhi, Aristotle and Confucius – who were the authors of many quotes, most of them timeless and generic enough for our purpose. We soon realized that long quotes would not produce the desired effect, so we only used quotes with up to 15 words, totaling 90.

We use the social network Twitter⁵ and Twitter4J⁶, a Java API, to retrieve tweets mentioning the names of the selected characters. While we could have used a news site or aggregator, the choice of Twitter relied on the fact that its messages are shorter, up-to-date, and mix different and less controlled opinions. In recent years, Twitter has been widely exploited by computer programs, not only for text mining, but also in computational creativity research (e.g. Veale (2014); Cook, Colton, and Gow (2014) or the recent PROSECCO Code Camp⁷, focused on the development of creative Twitterbots).

Natural language processing is made by the OpenNLP toolkit⁸ and its models trained for Portuguese tokenization and part-of-speech tagging. Since the models were not trained with tweets, a few annotation errors are expected. But this is not severe because we end up using only words in a morphological lexicon, LABEL-Lex⁹, in which we rely to perform inflection, so that words agree with the sentence they are put in. Also, we count the lemmas frequency in the tweets, and not the words frequency. Lemmatization is performed by LemPort (Rodrigues, Gonçalo Oliveira, and Gomes, 2014), a Portuguese lemmatizer.

Nouns long-associated to famous people were collected from the abstracts of their articles in the Portuguese Wikipedia, retrieved directly from the DBpedia¹⁰ entries under the category of *Person*.

Images of the meme characters are retrieved automatically from Google Images¹¹, at runtime. The first hit for each character is always used.

The Mallet¹² toolkit was used in the development of a humor classifier for Portuguese, presented in the next section. Given a positive and a negative dataset, Mallet automatically converts input text to features, and learns a classifier, using one of the algorithms available out-of-the-box.

Humor Classifier

We have recently started to work on a classifier for recognizing humorous pieces of text, in Portuguese, currently trained with the Mallet toolkit. The first step for its development was the collection of examples of humorous and non-humorous Portuguese documents, labeled respectively as positive or negative. The selected datasets were then imported to Mallet, which was used to train a classifier with the

⁵<https://twitter.com/>

⁶<http://twitter4j.org/>

⁷<http://codecampcc.dei.uc.pt/>

⁸<https://opennlp.apache.org/>

⁹http://label.ist.utl.pt/pt/labellex_pt.php

¹⁰<http://dbpedia.org/>

¹¹<https://images.google.com>

¹²<http://mallet.cs.umass.edu/>

best available learning algorithm. Instead of labeling the examples manually, we collected them from selected sources which we now present.

Positive Dataset: While it is rather easy to collect negative examples, the same does not apply for humorous examples in Portuguese. After searching in the Web, we were able to find the following compilations of Portuguese jokes:

- *Bíblia de Anedotas*¹³ (in English *Bible of Jokes*);
- *O Sagrado Caderno das Piadas Secas*¹⁴ (in English, *The Sacred Book of Dry Jokes*).

To focus on shorter jokes, we discarded all with more than 25 words, and were left with 790 positive examples.

Negative Dataset: The non-humorous dataset should contain text with a similar structure to the positive examples but without a potential humor effect. We thus collected sentences of similar length (≤ 25 words) from non-humorous sources. Since many of the collected jokes have a question-answer structure, we included this kind of text as well. The following resources were used:

- 304,211 sentences from the Portuguese Wikipedia, each collected randomly from a different article;
- Text from Portuguese corpora available through the AC/DC project (Santos and Bick, 2000)¹⁵:
 - 81,478 sentences from CETEMPUBLICO, a corpus with editions of the Portuguese newspaper Público (1991-1998).
 - 25,000 sentences from CONDIVport, a corpus of sports newspapers, fashion and health magazines;
 - 6,767 question-answer pairs from *Museu Da Pessoa*, a corpus of interviews.

In the end, we had a total of 417,456 negative examples.

Validation: After importing the positive and negative datasets, a classifier was trained with the Maximum Entropy algorithm, selected after a 10-fold cross-validation, where it yielded 99.8% accuracy. These numbers look promising, but they were computed in a dataset with mostly negative examples. Although the F_1 for the negative class was 99.9%, it was just 63.7% for the positive, with a recall of 49.4%.

We should stress that the classifier is still in an early stage of development. In the future, instead of relying only in the *black-box* text classification of Mallet, additional features should be integrated, including a subset of those used by others (Sjöbergh and Araki, 2007b; Mihalcea and Strapparava, 2006). Moreover, we are aware that we cannot expect much of the current classifier, at least for the kind of sentences we are generating. While it was trained with classic and timeless jokes, understanding the generated sentences requires not only general world knowledge, but further information that may be valid only on a specific moment in time.

¹³<http://rbep.cm-porto.pt/rbep/upload/downloads/BibliadeAnedotas.doc>

¹⁴<https://www.facebook.com/CadernoDasPiadas>

¹⁵<http://linguateca.pt/ACDC/>



Figure 2: Meme generated for the pop singer Madonna. The text translates to *Keep your thoughts positive, because your thoughts become your falls.*

Ranking function

As referred earlier, MemeGera generates a set of m sentences that combine a known quote with a noun f retrieved from Twitter. Towards the selection of the most promising generated sentences, these are currently ranked by the following linear combination:

$$\begin{aligned} \text{Score} = & \text{humorProb} * \alpha \\ & + \text{wordFrequency} * \beta \\ & + \text{notInWikipedia} * \gamma \end{aligned}$$

There, *humorProb* is the probability returned by the humor classifier; *wordFrequency* is the number of tweets where f occurs, divided by the total number of retrieved tweets, n ; and *notInWikipedia* is a binary function that is 1 if the word is not in the Wikipedia abstract of the character, or 0 otherwise.

Results

We generated several memes with different configurations. Although not enough experiments were performed to select the best configuration, at a certain point, we started to use fixed parameters, to have a base for comparison. In all reported experiments, generation was based on 200 tweets ($n = 200$), written in Portuguese (according to Twitter), and using the top-5 frequent nouns ($k = 5$). The best sentence was selected from a set of 20 ($m = 20$). The ranking function used the weights: $\alpha = 0.7$, $\beta = 0.25$, $\gamma = 0.05$.

Example

Figure 2 illustrates the output of MemeGera with a meme generated on the 26th February 2015, the day after Madonna fell on stage. The original quote, attributed to Mahatma Gandhi, was *Keep your thoughts positive, because your thoughts become your words.*

Since people were talking about the fall, the most frequent nouns in tweets were: *tombo* (tumble), *queda* (downfall), *palco* (stage), *vídeo* (video) and *madonno*. The last one results from an incorrect part-of-speech tag given to the proper noun Madonna. There is no risk of using it though, because

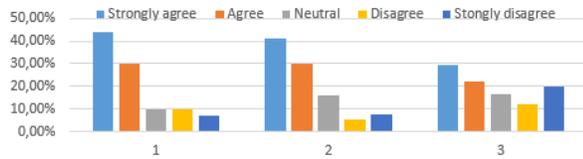


Figure 3: Overall results of the performed evaluation.

it is not in the morphological lexicon. None of the top-5 nouns were in Madonna’s Wikipedia abstract.

In the end of this paper, we present more memes, together with information that will help understanding why they were generated, and their scores according to the online survey where they were used.

Evaluation survey

In order to have a first appreciation of the results produced by MemeGera, we made an online survey, answered by 41 human subjects, all Portuguese native speakers. The survey had the title “Imagens com texto” (Images with text) and it never mentioned the word meme, nor automatic generation.

The survey had 5 memes, for which the name of the character was presented together with three questions, to be answered according to a Likert scale: *strongly agree* (5), *partially agree* (4), *neutral* (3), *partially disagree* (2) and *strongly disagree* (1). The questions were:

1. The text is syntactically and semantically coherent (Does it follow the grammar rules and makes sense?).
2. There is coherence in the combination of text, image and the present time (We suggest to search for breaking news about the character).
3. The combination of the text and the image produce a humorous effect (Did it make you smile?).

The used memes were generated between the 25th February 2015 night and 26th morning. Their characters were manually selected for being mentioned in fresh news in online media. All of these memes used the highest-ranked sentence from the 20 generated. They are presented in the end of the paper, together with information on their generation, as well as their individual scores in the survey questions.

The survey opened just a few hours after generating the last meme, and was opened for about 24 hours. This means that some memes would only be interpreted appropriately by someone following the daily news. Figure 3 presents overall results, which combine the answers to the five memes.

The survey confirmed that it is often safe to replace one word in a sentence by another of the same part-of-speech. If inflection is handled properly, syntax remains coherent, which makes it easier for semantics, especially when using generic quotes. Answers on the coherence between text, image and the present time are also positive. The meme in figure 6 was the one with more negative answers in the first two questions. First, possibly because it is not very easy to find semantic connections between *tumble* and *awake*. Second, because this meme was related to a very recent event and, although we suggested the subjects to search for the

character in the news, most of them probably did not do it, and were not aware of Madonna’s fall.

As for the humor aspect, while we cannot say that the generated memes are very funny and have the ability to make everybody laugh, the overall results are encouraging, as the majority of the answers are positive. It is always subjective to assess the presence of humor, especially in this case, where world knowledge and following recent news was a requirement. A curious fact is that the memes with clearly positive answers in this aspect are those with Portuguese politicians. Given that all our subjects were Portuguese, they are probably better informed about Portuguese characters, who probably play a more relevant role on the subjects lives, and make them more responsive to laugh at. This is related to another issue: the image itself or, sometimes, just the character, might play an important role in the humor value, since there are people for which we are more prone to laugh at than others.

Concluding remarks

We have presented the first steps towards the development of MemeGera, a system that generates combinations of text and image that may be seen as Internet memes. Famous quotes are altered according to a public figure and complemented with their image, automatically retrieved from the Web. Fresh information on the public figure, in the form of frequent words, is currently obtained from Twitter. Several altered quotes are generated and the best is selected after a ranking that considers the humor value and the novelty of retrieved words, in an attempt to positively discriminate the most promising sentences. The humor value is given by an automatic classifier, trained with positive and negative examples of humor expressed in Portuguese. However, this tool is still far from what we expect from it and gave very low scores to the generated sentences (rarely more than 1%).

On the other hand, we should stress that MemeGera has the ability of generating a different and novel sentence each time, based on fresh news. In fact, the results of an online survey showed that it is not only capable of generating coherent sentences, with some relation to the character, but that the generated combinations have some humor potential.

The work described in this paper lead to the development of the *@memegera* Twitterbot that, from time to time: (i) reads the list of current trends in the Portuguese Twitter; (ii) checks if any of them is the name of a known person; if so, (iii) generates a meme on that person and posts it. The bot is still in a test phase, but we may soon start relying on users feedback (e.g. retweets, favored) for evaluation and adaptation of the weights in the ranking function.

Additional plans include both improvements to the system and to its evaluation. In the scope of this and other projects, the humor classifier shall be improved by: (i) enriching the datasets with humorous text from the Twitter accounts of famous Portuguese humorists; (ii) considering additional features (e.g. ambiguity of words and adult slang, for which there are available Portuguese resources we could use). To increase variation, we will devise adding more quotes to our pool, as long as they are not too specific. Regarding evaluation, in a further survey, we aim at recording the reaction of

the subjects in the moment when the meme is first presented to them, and draw conclusions from their facial expressions.

Acknowledgements

The project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the 7th Framework Programme for Research of the European Commission, under FET grant number 611733.

References

- Barbieri, F., and Saggion, H. 2014. Automatic detection of irony and humour in twitter. In *Procs. of 5th Intl. Conf. on Computational Creativity (ICCC)*.
- Binsted, K., and Ritchie, G. 1994. An implemented model of punning riddles. In *Procs. of 12th National Conf. on Artificial Intelligence (Vol. 1)*, AAAI '94, 633–638. Menlo Park, CA, USA: AAAI Press.
- Blackmore, S. 2000. *The Meme Machine*. Oxford University Press, USA.
- Carvalho, P.; Sarmiento, L.; Silva, M. J.; and de Oliveira, E. 2009. Clues for detecting irony in user-generated contents: Oh...!! it's "so easy" ;-). In *Procs. of 1st Intl. International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion*, TSA '09, 53–56. ACM.
- Cook, M.; Colton, S.; and Gow, J. 2014. Automating game design in three dimensions. In *Procs. of AISB Symposium on AI and Games*.
- Dawkins, R. 1976. *The Selfish Gene*. Oxford University Press, Oxford, UK.
- Gonçalo Oliveira, H., and Cardoso, A. 2015. Poetry generation with PoeTryMe. In Besold, T. R.; Schorlemmer, M.; and Smaill, A., eds., *Computational Creativity Research: Towards Creative Machines*, Atlantis Thinking Machines. Atlantis-Springer. chapter 12, 243–266.
- Kiddon, C., and Brun, Y. 2011. That's what she said: Double entendre identification. In *Procs. of 49th Annual Meeting of the Association for Computational Linguistics*, 89–94. Portland, OR, USA: ACL Press.
- Kuipers, G. 2010. Humor styles and symbolic boundaries. *Journal of Literary Theory* 3:219–239.
- Manurung, R.; Ritchie, G.; Pain, H.; Waller, A.; O'Mara, D.; and Black, R. 2008. The construction of a pun generator for language skills development. *Applied Artificial Intelligence* 22(9):841–869.
- Mihalcea, R., and Strapparava, C. 2006. Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence* 22(2):126–142.
- Morreall, J. 2013. Philosophy of humor. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Spring 2013 edition.
- Rassi, A. P.; Baptista, J.; and Vale, O. 2014. Automatic detection of proverbs and their variants. In *Procs. of 3rd Symposium on Languages, Applications and Technologies (SLATE 2014)*, Bragança, Portugal, OASICS, 235–249. Schloss Dagstuhl.
- Ritchie, G. 2009. Can computers create humor? *AI Magazine* 30(3):71–81.
- Ritchie, G. 2014. Logic and reasoning in jokes. *European Journal of Humour Research* 2(1):50–60.
- Rodrigues, R.; Gonçalo Oliveira, H.; and Gomes, P. 2014. LemPORT: a high-accuracy cross-platform lemmatizer for portuguese. In *Procs. of 3rd Symposium on Languages, Applications and Technologies (SLATE 2014)*, Bragança, Portugal, OASICS, 267–274. Schloss Dagstuhl.
- Santos, D., and Bick, E. 2000. Providing Internet access to Portuguese corpora: the AC/DC project. In *Procs. of 2nd Intl. Conf. on Language Resources and Evaluation*, LREC 2000, 205–210.
- Sjöbergh, J., and Araki, K. 2007a. Automatically creating word-play jokes in japanese. In *Procs. of NL-178*, 91–95.
- Sjöbergh, J., and Araki, K. 2007b. Recognizing humor without recognizing meaning. In *Procs. of Applications of Fuzzy Sets Theory: 7th Intl. Workshop on Fuzzy Logic and Applications (WILF 2007)* Camogli, Italy, volume 4578 of LNCIS, 469–476. Springer.
- Smith, M. R.; Hintze, R. S.; and Ventura, D. 2014. Nehovah: A neologism creator nomen ipsum. In *Procs. of 5th Intl. Conf. on Computational Creativity (ICCC)*.
- Stock, O., and Strapparava, C. 2005. The act of creating humorous acronyms. *Applied Artificial Intelligence* 19(2):137–151.
- Suslov, I. M. 1992. Computer model of a "sense of humour". i. general algorithm. *Biophysics* 37(2):242–248.
- Tagnin, S. E. O. 2005. O humor como quebra da convencionalidade. *Revista Brasileira de Lingüística Aplicada* 5(1):247–257.
- Taylor, J. M., and Mazlack, L. J. 2004. Computationally recognizing wordplay in jokes. In *Procs. of Cognitive Science Conference (CogSci)*, 2166–2171.
- Toivanen, J.; Gross, O.; and Toivonen, H. 2014. The officer is taller than you, who race yourself! In *Procs. of 5th Intl. Conf. on Computational Creativity (ICCC)*.
- Valitutti, A.; Toivonen, H.; Doucet, A.; and Toivanen, J. M. 2013. "Let everything turn well in your wife": Generation of adult humor using lexical constraints. In *Procs. of 51st Annual Meeting of the Association for Computational Linguistics*, volume 2, 243–248. Sofia, Bulgaria: ACL Press.
- Veale, T., and Hao, Y. 2008. A fluid knowledge representation for understanding and generating creative metaphors. In *Procs. of 22nd Intl. Conf. on Computational Linguistics*, volume 1 of COLING '08, 945–952. ACL Press.
- Veale, T. 2014. Coming good and breaking bad: Generating transformative character arcs for use in compelling stories. In *Procs. of 5th Intl. Conf. on Computational Creativity (ICCC)*.

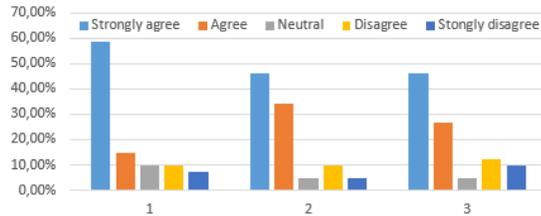


Figure 4: Meme of the Portuguese Deputy Prime-Minister, Paulo Portas.

Character	Paulo Portas (Portuguese Deputy Prime-Minister)
Wikipedia words	<i>jurista, jornalista, conservador, vice-primeiro-ministro, arquitecto, economista, irmão, dirigente, empresário, ...</i> (lawyer, journalist, conservative, deputy prime minister, architect, economist, brother, leader, businessman, ...)
Top-5 nouns	<i>março (29), trajetória (23), alteração (22), dívida (16), ðø (14)</i> (march, trajectory, change, debt, øf)
Context	He had been talking to the media about the downward trend of the Portuguese public debt, which should start on March.
Original quote	<i>Se nós não entendemos a vida, como poderemos entender a morte?</i> (If we cannot understand life, how can we understand death?)
Generated quote	<i>Se nós não entendemos a vida, como poderemos entender a dívida?</i> (If we cannot understand life, how can we understand the debt?)
Humor probability	0.00253
Frequency	0.08
In Wikipedia	No
Score	0.07177

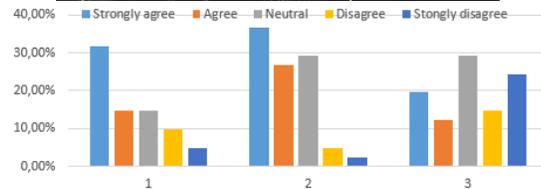


Figure 5: Meme of the Greek Finance Minister, Yanis Varoufakis.

Character	Yanis Varoufakis (Greek Finance Minister)
Wikipedia words	Not in the used version of Wikipedia
Top-5 nouns	<i>salário (37), governo (29), funcionário (22), manutenção (18), ministro (17)</i> (salary, government, federal employee, maintenance, minister)
Context	Since he became a member of the Greek government, he has been all over the news regarding the negotiations between Greece and the Eurogroup. This time, he was negotiating the maintenance of the current number of federal employees and their salaries.
Original quote	<i>Na busca da virtude, não temas superar teu professor.</i> (In the pursuit of virtue, do not fear overcoming your teacher.)
Generated quote	<i>Na busca da virtude, não temas superar teu governo.</i> (In the pursuit of virtue, do not fear overcoming your government.)
Humor probability	0.00102
Frequency	0.145
In Wikipedia	No
Score	0.08696

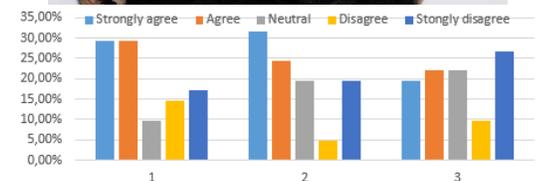


Figure 6: Meme of the pop singer Madonna.

Character	Madonna (pop singer)
Wikipedia words	<i>cantor, compositor, atriz, dançarino, empresário, produtor, álbum ...</i> (singer, songwriter, actress, dancer, businessman, producer, album ...)
Top-5 nouns	<i>tombo (46), queda (21), madonna (29), palco (21), vídeo (13)</i> (tumble, downfall, madonna, stage, video)
Context	A few hours before this generation, she had fell on stage, during a live performance.
Original quote	<i>A esperança é um sonho acordado.</i> (Hope is a waking dream.)
Generated quote	<i>A esperança é um tombo acordado.</i> (Hope is a waking tumble.)
Humor probability	0.00231
Frequency	0.23
In Wikipedia	No
Score	0.10911



Figure 7: Meme of the Russian President, Vladimir Putin.

Character	Vladimir Putin (Russian President)
Wikipedia words	<i>presidente, ex-agente, chefe, serviço, primeiro-ministro, governo, país, ...</i> (president, former agent, chief, service, prime-minister, government, country, ...)
Top-5 nouns	<i>passo</i> (88), <i>risco</i> (36), <i>país</i> (36), <i>intervenção</i> (34), <i>vitória</i> (33) (step, risk, country, intervention, victory)
Context	He had been on the news due to his role in the Ukrainian crisis. While there is not an international intervention, people try to figure out his next step.
Original quote	<i>Em todas as coisas, o sucesso depende de preparação prévia.</i> (In all things, success depends on a previous preparation.)
Generated quote	<i>Em todas as coisas, o sucesso depende de intervenção prévia.</i> (In all things, success depends on a previous intervention.)
Humor probability	7.8×10^{-4}
Frequency	0.17
In Wikipedia	No
Score	0.09305



Figure 8: Meme of the former Portuguese Prime-Minister, José Sócrates.

Character	José Sócrates (former Portuguese Prime Minister)
Wikipedia words	<i>político, secretário-geral, ministério, ordenamento, mestrado ..</i> (politician, secretary general, ministry, planning, masters, ...)
Top-5 nouns	<i>prisão</i> (93), <i>mês</i> (54), <i>ex-premiê</i> (41), <i>retour</i> (34), <i>novembro</i> (33) (jail, month, ex-Prime Minister, comeback, november)
Context	He has been detained into custody for being a suspect in a corruption case. This week he had got out for a few hours to give his testimony in a court.
Original quote	<i>Minha vida é minha mensagem.</i> (My life is my message.)
Generated quote	<i>Minha vida é minha prisão.</i> (My life is my jail.)
Humor probability	0.00975
Frequency	0.465
In Wikipedia	No
Score	0.17308