Humour and non-humour in religious discourse

Daniela Gifu¹, Liviu-Andrei Scutelnicu¹², Dan Cristea¹²

¹„Alexandru Ioan Cuza” University, Faculty of Computer Science, 16, General Berthelot, 700483, Iași
²Institute of Computer Science, Romanian Academy - Iași branch, „T. Codrescu”, 700481, Iași

E-mail: {daniela.gifu, liviu.scutelnicu, dcristea}@info.uaic.ro

Abstract

The paper describes a pilot study focusing on the exploration of humorous features in religious texts. Although religious discourses have a strong dogmatic tonality, some constructions contain visible humorous features, especially adapted to the audience expectations. Humour facilitates the apprehension of the religious discourse, evidencing not only the oratory dimension of the speaker but also helping the receptor to better perceive the message while also inducing affective effects. A corpus of preaches (which is contrasting with liturgical texts) is collected, in which humour is marked on Adjectival Noun Phrases. We propose a pater-based method for identifying humour in religious discourses in which patterns are lexicalised regular expressions of word categories. Using a religious lexicon, we classified Adjectival Noun Phrases in religious and non-religious. The study is meant to create a tool for automatic detection of humour in religious discourses. Automatically annotated corpora of preaches could become sources for further research that would reveal different valences in the religious texts.

Keywords: detection of humour, religious lexicon, Adjectival Noun Phrases.

1. Introduction

The motivation for our study relies on the need for objectivity in the interpretation of the humour in religious language. In this sense, we will explore the applicability of computational approaches to the recognition of adjectively expressed humour. Our goal is to recognize non-verbal humour (here, in Adjectival Noun Phrases) as other research on this issue so far focused mostly on the recognition of verbal humour (Loehr, 1996; Wolfe, 2011, Mihalcea, 2012).

Marked by a strong inter-textuality (Zafruit, 2010), involving its own discursive practices, religious speech (here, preaches) requires special attention to the construction and adaptation in order to make it easier to understand by people of a very heterogeneous cultural lead (Dinca, 2008). Humorous statements can produce unexpected, sometimes contrary, effects in the hearer – cohesion (Sâftoiu, 2006; Constantinescu, 2006) and exclusion (Koestler, 1964) – and, as such, an objective evaluation becomes essential. We note that humour is a form of communicative behavior for which the transmitter expects an immediate reaction (a certain type of emotion) in the receiver. Humour becomes ironic (Reboul, 1980) when the playful is substituted by offensive intentions and the irony manifestation attracts some limitations (Sălăvăstru, 1995).

Viewed as a set of discursive actions with a religious specificity, religious discourse, with its forms – preach, homily, panegyric, paraenesis, religious conference (Gordon, 2001; Grigoraș, 2000) – can be defined by the intention for producing statements which induce a certain type of emotion in the receiver.

In this study we concentrate on preach, considered to be the most relevant type of discourse in the inherited tradition of Aristotelian rhetoric (Aristotel, 2004) in the European culture and, consequently, in the Romanian culture. A preach becomes persuasive by making use of three components, each of them having specific frequencies: the ethos (speaker individuality), the pathos (use of emotion), the logos (use of rational arguments). Humorous statements investigation identified in the proposed study leads to the definition of new religious contemporary oratory features. Using humour as a mark of oral speech helps perception, understanding and deepening religious message, contribute to group cohesion, moving the rapprochement dominance between the transmitter and the auditor.

The following forms of humour are often present in preaches.

1. Irony is that type of humour through which the speaker expresses the opposite of what the audience expects, implying the words that reflect the discrepancy between appearance and essence. Here is an example:

Sunt unii care îmi spun să mă rog pentru ei şi, când mă duc la Domnul, să nu-i uit. Da! Frăția ta mănâncă și dormi până te saturi, și eu am să mă rog pentru tine!

The speaker mocks the receiver naivety in carrying out Christian’s duties, reminding him that he does violence to the Christian morality values. In some situation, the irony has to penalize the humans’ ignorance and to make people aware of the ethical values.

¹ (EN) - There are people that ask me to pray for them when I go to God, not to forget them. Yes! My brother, eat and sleep till you are fulfilled and I will pray for you!
2. Frames’ humour can be distinguished at the level of serious tonality and at the humorist level of the speech. The joke is based on common knowledge and values, making the audience feel comfortable.

For instance:

Dacă vrei să mă fotografiezi, caută un măgar, fă-i poză şi scrie pe ea Cleopa.2

The donkey is a burden animal and, similarly, the speaker feels saddled with the desire of people to photograph him. In addition, the donkey has big ears and looks funny. The preacher chooses this form of expression in order to put the photographers to shame and to show them the ridiculous situation of the speaker would have when taking photos.

3. Wordplay is achieved by joining disjoint meanings of words, see the example:

Răbdare, răbdare, răbdare…. Nu până la prăşit, că nu spune Sfânta Evanghelie aşa. Cine va răbdă până la sfârşit, acela se mânui teşte, ci nu până la prăşit.3

The speaker makes indirect reference to the way the patience is perceived in a laic sense (temporarily), in total opposition to the biblical sense (for good). The substitution of the word sfârşit with prăşit induces a humorous tone. It is an inspired wordplay that has rhyme and rhythm, both words having two syllables and the same ending – easy to remember.

In this paper we propose a new method of detecting and recognizing the humour in religious discourses. In particular, we investigate whether semi-automatic classification techniques can be seen as a viable approach to distinguish humorous sequences between religious and non-religious texts. We demonstrate with empirical evidence that the humour in religious texts can be detected by automatic means.

The paper is structured as follows. Section 2 shortly describes the background. Section 3 describes the resources used in this paper. Section 4 discusses the methodology applied in the recognition of adjectival humour in religious texts. Section 5 presents statistics and their interpretation. Finally, Section 6 depicts some conclusions and directions for future work.

2. Background

Analysed by humanists (linguists (Attardo & Raskin, 1994) and psychologists (Freud, 1928; Ruch, 2002)), humour has recently raised the interest of computer scientists, who are concerned with the construction of language models for the automatic recognition and categorization of the humorous style (Stock & Strapparava, 2003), or to suggest relevant indicators of humour, authors as Ruch, Bucaria (Ruch, 2002; Bucaria, 2004) focused on alliteration, antonym, and adult slang.

Based on the narrative strategies adopted by the orator, different types of humour often found in discourse, such as anecdote, exaggeration, irony, satire, underestimation, humorous situation, (Solomovici, 2002) are mentioned as sarcasm, exaggeration or minimization, self depreciation, teasing, rhetorical question answers, double meanings, punning, proverbs interpretation (Martin, 2007).

Humour provides oratory depth, even to the religious language, accomplishing also an important social role. The humour theories are seen as complementary (Raskin, 1998; Rutter, 1997; Minsky, 1981). While it is merely considered a way to induce hilarity, humour can have positive effects: it alters attention and memory (Baym, 1995); facilitates social interactions, helping to generate solidarity and group identity (Binsted & Ritchie, 1997; Binsted, et al. 2006); improves communication problems (Bergson, 1980); can establish a common ground between dialogue partners (Hewitt, 2002); enhances the motivation, attention, understanding and capturing of information and gives an affective meaning to the message by bringing into scene an affective sense (Nijholt, 2006). As a primary mechanism for establishing the individuality of humans, humour makes the speaker feel appreciated when the receivers recognize her/his jokes and this improves communication (Black & Forro, 1999), stimulates creativity, memory, and improves the morale and the productivity of speeches (Stock & Strapparava, 2003). But humour can also have negative influences: it may offend, can inhibit the communication when jokes are too harsh, or can create professional stress (Black & Forro, 1999).

In the papers mentioned, we observed that few attempts have been made to develop systems for automatic humour recognition (Mihalcea, 2012). This is obvious, since, from a computational perspective, humour recognition appears to be significantly more subtle and in consequently, difficult to be examined.

In this work, we explore the applicability of computational methods to the recognition of Adjectival Noun Phrases (ANPs) expressed humour in religious speech. Moreover, we investigate whether automatic classification techniques represent a viable approach to distinguish between humorous and non-humorous text especially in preaches.

3. Resources and pre-processing

The study of religious language should necessarily be approached in an interdisciplinary way, in which the rhetoric sciences, communication and doxology cooperate with computational linguistic methods.

To prepare the corpus of humour in religious texts we have processed a collection of texts summing up 16 volumes of preaches in Romanian, authored by the monk priest Ilie Cleopa4, one of the most renowned Romanian Orthodox orators and religious writers. The collection of

2 (EN) - If you want to take me a photo, look for a donkey, take him a picture and write on it Cleopa.
3 (EN) - Have patience, patience, patience... not till hoeing, because The Holly Gospel does not say so. But he who endures to the end will be saved, but not till the time of hoeing.

texts, containing many humorist sequences, was published between 1995 and 1998 by one of his closest disciples. The texts count 588,784 words, over approximately 1,500 pages. The corpus was pre-processed by tokenising it, then tagging it at part-of-speech (POS), with the Romanian POS-tagger web-service (Simionescu, 2011), lemmatising it and then extracting Adjectival Noun Phrases. We considered the adjectives to be semantically relevant for the religious genre.

A lexicon of religious terms was developed using as seeds the lexicon of the semantic class RELIGIOUS used in the research on political discourse analysis in election (Gifu & Cristea, 2012). The RELIGIOUS class is one of the 30 semantic classes, which are considered to optimally cover the necessity of interpreting the political discourse in electoral contexts, in the Discourse Analysis Tool (Gifu & Cristea, 2012). The hierarchy of these categories preserves the structure of a tree. Then, this lexicon was enriched by further importing synonyms from DEX-online5; the greatest public online dictionary for Romanian. DEX has no semantic structure (unlike WordNet, for instance). With this lexicon under eyes, one of the authors went through a tedious process of manual annotation of the corpus for humorous religious and non-religious ANPs. In most of the cases these expressions are a combination of religious word + non-religious word, with a humorous slant: "duh necurarăță; dracul milostivă"; credință strâmbă.

Finally, the lexicon was completed with hyponyms of the Religion synset in the Romanian WordNet (Tufiş et al., 2004)7. When this process was finalised, the religious lexicon contained 2367 entries, considered to cover satisfactorily this type of analysis.

4. The methodology

The research followed the following steps: a). pre-processing the Corpus; b). manual annotation of the humorous Adjectival Noun Phrases in the Corpus; c). automatic detection of humorous ANPs; d). evaluation.

We will call a sequence of one or more adjectives connected by conjunctions and/or commas and modifying a noun an Adjectival Noun Phrase. Whether the semantic category of the content words (nouns and adjectives) is religious or not, we will have Religious Adjectival Noun Phrases (RANP) and Non-Religious Adjectival Noun Phrases (NRANP). These phrases have been extracted by applying lexical-syntactic patterns within the borders of noun phrases. Examples are given below.

In standard Romanian, usually adjectives stay a sequence of one or more adjectives after nouns, and our religious texts statistically follow confidently this rule (for instance, pomi neroditori or ațenie dumnezeiască). Examples of adjectival constructions follows:

nou + article + adjective - Origen cel blestemat11:

<noun + article + adjective - Origen cel blestemat11>

nou + adjective + conjunction + adjective: cărturarilor nebuni și orbii:

<nou + adjective + conjunction + adjective: cărturarilor nebuni și orbii>

nou + (adjective + comma)* + adjective + conjunction + adjective: oameni aleși, drepți și sfinți:

<nou + (adjective + comma)* + adjective + conjunction + adjective: oameni aleși, drepți și sfinți>

5 http://dexonline.ro
6 (EN) - unclean spirit.
7 (EN) - the merciful devil.
8 (EN) - crooked faith.
9 http://www.racai.ro/
As can be seen in the above examples, the corpus includes on XML level basic level annotation that marks morpho-syntactic features and word lemmas attached to each token (<W>W</W>). Above this level, Adjectival Noun Phrases have been marked as XML elements <ANP></ANP>. This way we draw the attention on the adjectives that complement nouns (also including articles, conjunctions, commas and/or other linguistic connectors). The religious/non-religious semantics of the content words belonging to the ANP elements (as expressed by their inclusion in the religious lexicon) is coded in the attribute TYPE of the ANP element. A humorist tonality characterises both types of expressions.

Figure 1 shows a classification of the ANPs in the corpus. They could include or not humorous effects (HANP and NHANP), and the Humorous Adjectival Noun Phrases could be religious and non-religious (RHANP and NRHANP).

5. Statistics and interpretation
Out of the complete set of Adjectival Noun Phrases, the Gold corpus puts in evidence only those including a humorist effect, they being explicitly categorised in 2 classes: religious and non-religious. The fact that the Gold corpus lacks an explicit notiation of humour within an exhaustive set of ANPs has two reasons: first, we relied on the high precision and recall of recognising ANPs (pattern-based) and to the high precision of the tags left behind by the POS-tagger (Tufiş & Dragomirescu, 2004), and second, we wanted to accelerate the process of manual annotation (this way, the annotator paid attention only to sequences displaying different kinds of humour). Identification of types of humour, although of interest, has not been considered in this research. The instances of ANPs extracted from the annotated corpus were used to form the collection of patterns of the recognition. Only variations in word forms were accepted in this phase of the research, therefore we could call our patterns lexicalised regular expressions of word categories. Table 1 compares the automatic detection of humour with that manually annotated in the Corpus with respect to religious/non-religious sequences.

Between aRHANP and mRHANP we identified 6932 common Adjectival Noun Phrases and between aNRHANP and mNRHANP we identified 1019 common Adjectival Noun Phrases.

Table 1: Automatic and manual annotation results.

<table>
<thead>
<tr>
<th></th>
<th>Total Number of Words</th>
<th>manual Humorous Adjectival Noun Phrases (mHANP)</th>
<th>automatic Humorous Adjectival Noun Phrases (aHANP)</th>
<th>manual Religious Humorous Adjectival Noun Phrases (mRHANP)</th>
<th>automatic Religious Humorous Adjectival Noun Phrases (aRHANP)</th>
<th>manual Non-Religious Humorous Adjectival Noun Phrases (mNRHANP)</th>
<th>automatic Non-Religious Humorous Adjectival Noun Phrases (aNRHANP)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>588.784</td>
<td>9937</td>
<td>9854</td>
<td>8354</td>
<td>8623</td>
<td>1583</td>
<td>1231</td>
</tr>
</tbody>
</table>

With these values, the Precision (1), Recall (2) and F-measure (3) could be computed, for both Religious and Non-Religious Humorous ANPs.

\[
P = \frac{\#\text{correctly\_identified\_ANP}}{\#\text{automatically\_identified\_ANP}}
\]

\[
R = \frac{\#\text{correctly\_identified\_ANP}}{\#\text{manually\_annotated\_ANP}}
\]

\[
F - \text{measure} = \frac{2*P*R}{P + R}
\]

The values are given in Table 2.
As shown in Table 2 the results for the automatic detection of the adjectival phrases, which have a religious and non-religious nature, are rather high. The fact that the religious humour is scored better than the non-religious humour, could be due to the special attention that we paid on annotating religious terms.

6. Conclusions and future work

This research is a preliminary study in humour recognition in Orthodox preaches and it confirms the hypothesis that humour, defining a specific rhetoric in religious speech, can be depicted by automatic means. However, more efforts are necessary to stabilise these preliminary results, and to look for combined methods of humour detection.

Humour analysis may require multiple interpretations. In preaches, the text approaches colloquial language, without letting aside clarity, accuracy and theological fairness. On the other hand, at the liturgical level a sacred communication act could be evidenced, in which the priest enters into dialogue with God. The preach comes to clarify the liturgical text, which is why there is a permanent adequacy audience expectations and needs. S-a trezi cu total nebun, cu total stricat la minte14, Trupul ăsta este o mână de pământ spre mâncarea viermilor15, or Ai văzut vreodată vreo femeie cu două capete? Nu se poate. Bărbatul este cap şi femeia este trup16, are examples which display a subtle irony of the priest. These are situations for our future work. Although the adjectival phrases are absent, the irony exists in these examples, but our method is yet unable to detect it.

From the perspective of formative function, religious language becomes the key to the receivers’ universe and the language facts require adaptation to specific communication situations.

The results revealed in this study may provide a basis for a new Orthodox oratory identity in the local area. Automatically annotated corpora of preaches could become sources for further research that would reveal different valences in the religious texts. For enlarging our corpus we will take into consideration other priests’ speeches that manifest similar styles with Pr. Cleopa’s (we think of the former Metropolitan Bartolomeu Anania of Cluj, Constantin Necula of Sibiu, Calistrat Chifan of Bârnova Monastery of Iassy).

Another challenge is the classification of adjectival phrases in subclasses that could enhance the accuracy and allow for a finer typification of the humorous sequences.

Finally, we intend also to bring in the research machine learning methods, based on neural networks that would automatically generalise patterns from the annotation. A good neural network for pattern recognition can be feed-forward, the network being trained to associate outputs with the input patterns, in our case different types of humorous Adjectival Noun Phrases.

7. References

Aristotel. (2004). Retorica, I, 2, 1356a, trad. M.-C. Andrieș, Bucharest, IRI.

14 (EN) - He woke up completely insane, completely mind broken.
15 EN - This body is handful of earth that is food for worms.
16 (EN) - Have you ever seen a woman with two heads? You cannot. The man is the head and the woman is the body.


