

A Technique for Processing of Political Language

Daniela Gifu

Abstract: This paper presents and interprets the context of current political changes in Romania – the new Labor Code - using a new technique in the print media language processing (method PLP-2011). The concept behind this method is that the manner in which individuals speak and write represent a relation between their emotional and cognitive dimensions. Practically, the reader identify with the journalist, who becomes the legitimate voice of common ideals. Our investigation is intended to give support to specialists in social and political sciences, journalists etc., being helpful mainly in measure the reactions of print media in a certain context.

Keywords: discourse, daily, journalist, content analysis, natural language, labor code.

1 Introduction

The motivation for our study relies on the need for objectivity in the interpretation of the political discourse published in newspapers. Among many attributes the political discourse has, we were interested in the lexicon and its interpretation in a range of semantic coordinates. Part of our research, as reported in (Gifu, 2010, 2011), was concentrated on this type of human validation.

The software we developed, PLP (*Political Language Processing*) offers the possibility to analyze efficiently large bodies of text and to characterize them quantitatively and qualitatively, the results having to be as close as possible to the analysis made by a human expert.

The paper is structured as follows. Section 2 shortly describes the functionality of the software and the associated resources for the Romanian language. Then, section 3 discusses one example picked up from the print media about Labor Code 2011 theme, section 4 presents the conclusions.

2 PLP-2011 software. A short description

The PLP-2011 software performs part-of-speech (POS) tagging and lemmatization of words. This is why the lexicon can now be declared as a

The Romanian lexicon contains now approximately 6000 lemmas and 29 semantic classes¹. We plan to populate our lexicon further by importing lemmas from RACAI (Romanian Academy Center for Artificial Intelligence). The semantic classes in AnaDiP are partially placed in a hierarchy. In the future we plan to align this hierarchy with WordNet (Fellbaum, 2001), for languages which support this type of linguistic resource.

3 Application: Labor Code 2011

For the elaboration of preliminary conclusions over the Labor Code 2011 validation process, we collected, stored and processed electronically, in two weeks (one week before and one after March 8, when the censure motion was debated), all texts with this theme, published by three national publications having similar profiles: *Evenimentul zilei*, *Gândul* and *Jurnalul Național*. To exemplify (figure 2), we processed all articles from March 8 about the censure motion filed the opposition parties. We present one type of graphics, "columns", considered for the interpretation relevant for this comparative analysis. Appear only classes that have reached at least 1% frequency, the limit which we taken into account.

The monitored written media corpus was processed directly with PLP-2011. The speech records were previously manually transposed onto text and then they followed the same processing as the written texts. In essence, the program receives the input from one or more text files, and counts occurrences of words belonging to its defined classes. The user can notice directly the mentioned semantic classes (and the corresponding frequencies), as the words belonging to a selected class can be highlighted in the text.

1 Taking into account the criterion of persuasion, which each stage of political article builds, we considered the 29 semantic classes necessary and sufficient for our study: injurii (swear), social (social), familie (family), prieteni (friends), oameni (humans), emoțional (affect), emoțional_pozitiv (positive_emotion), emoțional_negativ (negative_emotion), anxietate (anxiety), furie (anger), supărare (sadness), rațional (rational), intuiție (intuition), determinare (cause), nesiguranță (uncertainty), siguranță (certainty), inhibiție (inhibition), perceptiv (percept), vizual (see), auditiv (hear), tactil (feel), sexual (sexual), muncă (work), realizări (achievements), nerealizări (failures), agrement (leisure), cămin (home), finanțe (money), religie (religion).

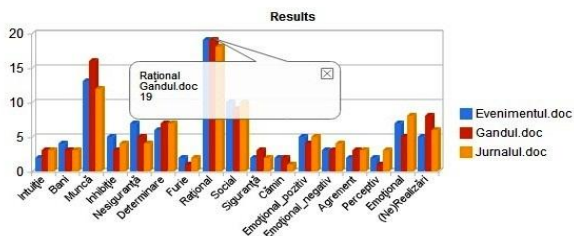


Figure 2. The differences in the frequencies for each class

4 Conclusion

We believe that PLP has a range of features that make it attractive as a tool to assist communication campaigns. It can also be rapidly adapted to new domains and to new languages, while its interface is user-friendly and offers a good range of useful functionalities.

In the future we intend to include a word sense disambiguation module in order to determine the correct senses, in context, of those words which are ambiguous between different semantic classes belonging to the lexicon, or between classes in the lexicon and outside the lexicon.

Acknowledgments. We thank professor Dan Cristea for initiating this research and for supervising the author and his student Mădălina Spătaru which developed the PLP-2011 software, from the Faculty of Computer Science of the “Alexandru Ioan Cuza” of Iași. In order to perform this research the author received financial support from the POSDRU/89/1.5/S/63663 grant.

References

- [1] C. Fellbaum (ed.). *WordNet, An Electronic Lexical Database*, The MIT Press, 2001.
- [2] D. Gîfu and D. Cristea. *Computational Techniques in Political Language Processing: PLP-2011* in J.J. Park, L.T. Yang, and C. Lee (Eds.): FutureTech 2011, Part II, CCIS 185, pp. 188–195, 2011
- [3] D. Gîfu. *The Discourse of the Written Press and the Violence of Symbols*, PhD thesis, Faculty of Philosophy and Political Studies, “Alexandru Ioan Cuza” University of Iași, 2010.

Daniela Gîfu¹

¹Universitatea “Alexandru Ioan Cuza”, E-mail: daniela.gifu@info.uaic.ro