

Opinion and Factivity Analysis of Italian political discourse

Rodolfo Delmonte¹, Daniela Gifu², Rocco Tripodi¹

¹ Ca' Foscari University, Department Language Science,
Ca' Bembo, dd. 1075, 30123, Venice
delmont@unive.it, rocco.trip@gmail.com

² „Alexandru Ioan Cuza“ University, Faculty of Computer Science,
16, General Berthelot St., 700483, Iași
daniela.gifu@info.uaic.ro

Abstract. The success of a newspaper article for the public opinion can be measured by the degree in which the journalist is able to report and modify (if needed) attitudes, opinions, feelings and political beliefs. We present a symbolic system for Italian, derived from GETARUNS, which integrates a range of natural language processing tools with the intent to characterise the print press discourse from a semantic and pragmatic point of view. This has been done on some 500K words of text, extracted from three Italian newspapers in order to characterize their stance on a deep political crisis situation. We tried two different approaches: a lexicon-based approach for semantic polarity using off-the-shelf dictionaries with the addition of manually supervised domain related concepts; another one is a feature-based semantic and pragmatic approach, which computes propositional level analysis with the intent to better characterize important component like factuality and subjectivity. Results are quite revealing and confirm the otherwise common knowledge about the political stance of each newspaper on such topic as the change of government that took place at the end of last year, 2011.

Keywords: journalist opinion, sentiment analysis, political discourse, lexical-semantic, syntax, print press, Government of Italy.

1 Introduction

In this paper, we discuss paradigms for evaluating linguistic interpretation of discourses as applied by a light scaled version the system for text understanding called GETARUNS. We focus on three aspects critical to a successful evaluation: creation of large quantities of reasonably good training data, lexical-semantic and syntactic analysis. Measuring the polarity of a text is usually done by text categorization methods which rely on freely available resources. However, we assume that in order to properly capture opinion and sentiment [6,10,11,17] expressed in a text or dialog, any system needs a linguistic text processing approach that aims at producing semantically viable representation at propositional level. In particular, the idea that the task may be solved by the use of Information Retrieval tools like Bag of

Words Approaches (BOWs) is insufficient. BOWs approaches are sometimes also camouflaged by a keyword based Ontology matching and Concept search [10], based on SentiWordNet (*Sentiment Analysis and Opinion Mining with WordNet*) [2]– more on this resource below –, by simply stemming a text and using content words to match its entries and produce some result [16]. Any search based on keywords and BOWs is fatally flawed by the impossibility to cope with such fundamental issues as the following ones, which Polanyi and Zaenen [12] named contextual valence shifters:

- presence of negation at different levels of syntactic constituency;
- presence of lexicalized negation in the verb or in adverbs;
- presence of conditional, counterfactual subordinators;
- double negations with copulative verbs;
- presence of modals and other modality operators.

It is important to remember that both PMI and LSA analysis [16] systematically omit function or stop words from their classification set of words and only consider content words. In order to cope with these linguistic elements we propose to build a propositional level analysis directly from a syntactic constituency or chunk-based representation. We implemented these additions on our system called GETARUNS (*General Text And Reference Understanding System*) which has been used for semantic evaluation purposes in the challenge called RTE and other semantically heavy tasks [1,4]. The output of the system is an xml representation where each sentence of a text or dialog is a list of attribute-value pairs. In order to produce this output, the system makes use of a flat syntactic structure and a vector of semantic attributes associated to the verb compound at propositional level and memorized. Important notions required by the computation of opinion and sentiment are also the distinction of the semantic content of each proposition into two separate categories: objective vs. subjective.

This distinction is obtained by searching for factivity markers again at propositional level [14]. In particular we take into account: modality operators like intensifiers and diminishes, modal verbs, modifiers and attributes adjuncts at sentence level, lexical type of the verb (from ItalWordNet classification, and our own), subject's person (if 3rd or not), and so on.

As will become clear below, we are using a lexicon-based [9,15] rather than a classifier-based approach, i.e. we make a fully supervised analysis where semantic features are associated to lemma and concept of the domain by creating a lexicon out of frequency lists. In this way the semantically labelled lexicon is produced in an empirical manner and fits perfectly the classification needs.

The paper is structured as follows. Section 2 comments on the role of print press discourse; Section 3 describes the system for multi-dimensional political discourse analysis. Section 4 presents comparative analysis of print press discourses collected during the Berlusconi's resignation in favour of Monti's nominating the President of Italian Government (October 12 – December 12, 2011). Finally, section 5 highlights interpretations anchored in our analysis and presents a conclusion.

2 Print press discourse

Mirror of contemporary society, located in permanent socio-cultural reevaluation, the texts of print press can disrupt or use a momentary political power. In contemporary society, the struggles stake is no longer the social use of technology, but it is the huge production and dissemination of representations, informations and languages.

At present, the legitimacy of competence and credibility or reputation of political authority is increasingly in competition with mediatic credibility and the charisma already confirmed in public space. In political life we see how „heavy” actors are imposed, benefiting preferential treatment in their publicity and/or how insignificant actors, with reduced visibility, are ignored, even marginalized, notwithstanding their possibly higher reputation. Most of the times, launching the new actors is accompanied by changing others, intermediate body, the militants, condemned not only to mediatic silence, but simply silenced: in this way, the role of opinion leaders is drastically reduced.

Print press, in its various forms, assigns political significance to institutional activities and events in their succession; it forms the political life of a nation, from objective information to become the subject of public debate. In this case, the role of print press is double:

1. secure information as a credible discourse to end a rumor;
2. enter politics in language forms, so they become consistently interpretable in a symbolic system of representations.

The press is designed to legitimize the actions of politicians, attending their visibility efforts, confirming or increasing their reputation. Print press includes essentially political discourses, containing both a specific orientation and a political commitment. The reader has the possibility to choose what and when to read, leaving time to reflection, too. Disproportionality is a risk to the reality described.

No wonder why the people in power, if they intend to govern in peace, try to curb the enthusiasm of the media. Most of the times, through excellence in the elections, the print press is focused on topical issues, leading topics of public interest and events of internal and external social life. However, the perception of social reality depends on how it is presented. So the newspaper, like any commercial product, is dependent on aesthetic presentations that may distort any event-selection alternative to news items which are sensational and, often, negative (i.e. our comparative study).

3 The System GETARUNS

In this section we will present a detailed description of the symbolic system for Italian that we used in this experiment. The system is derived from GETARUNS, a multilingual system for deep text understanding with limited domain dependent vocabulary and semantics, that works for English, German and Italian and has been documented in the past 20 years or so with lots of publications and conference presentations[3,5]. The deep version of the system has been scaled down in the last ten years to a version that can be used with unlimited text and vocabulary, again for English and Italian. The two versions can work in sequence in order to prevent

failures of the deep version. Or they work separately to produce less constrained interpretations of the text at hand.

The "shallow" scaled version of GETARUNS has been adapted for the Opinion and Sentiment analysis and results have already been published for English [6]. Now, the current version which is aimed at Italian has been made possible by the creation of the needed semantic resources, in particular a version of SentiWordNed adapted to Italian and heavily corrected and modified. This version (see 3.0) uses weights for the English WordNet and the mapping of sentiment weights has been done automatically starting from the linguistic content of WordNet glosses. However, this process has introduced a lot of noise in the final results, with many entries totally wrong. In addition, there was a need to characterize uniquely only those entries that have a "generic" or "commonplace" positive, or negative meaning associated to them. This was deemed the only possible solution to the problem of semantic ambiguity, which could only be solved by introducing a phase of Word Sense Disambiguation which was not part of the system. So, we decided to erase all entries that had multiple concepts associated to the same lemma, and had conflicting sentiment values. We also created and added an ad hoc lexicon for the majority of concepts (some 3000) contained in the text we analysed, in order to reduce the problem of ambiguity. This was done again with the same approach, i.e. labelling only those concepts which were uniquely intended as one or the other sentiment, restricting reference to the domain of political discourse.

The system has been lately documented by our participation in the EVALITA (*Evaluation of NLP and Speech Tools for Italian*) challenge¹. It works in a usual NLP pipeline: the system tokenizes the raw text and then searches for Multiwords. The creation of multiwords is paramount to understanding specific domain related meanings associated to sequences of words. This computation is then extended to NER (*Named Entity Recognition*), which is performed on the basis of a big database of entities, lately released by JRC (*Joint Research Centre*) research centre.² Of course we also use our own list of entities and multiwords.

Words that are not recognized by simple matching procedures in the big wordform dictionary (500K entries), are then passed to the morphological analyser. In case also this may fail, the guesser is activated, which will at first strip the word of its affixes. It will start by stripping possible prefixes and then analysing the remaining portion; then it will continue by stripping possible suffixes. If none of these succeeds, the word will be labelled as foreign word if the final character is not a vowel; a noun otherwise. We then perform tagging and chunking. In order to proceed to the semantic level, each nominal expression is classified at first on the basis of the assigned tag: proper nouns are used in the NER task. The remaining nominal expressions are classified using the classes derived from ItalWordNet (*Italian WordNet*)³. In addition to that, we have compiled specialized terminology databases for a number of common domains including: medical, political, economic, and military. These lexica are used to add a specific class label to the general ones derived from ItalWordNet. And in case the word or multiword is not present there, to uniquely classify them. The output of this

¹ <http://www.evalita.it/>

² <http://irmm.jrc.ec.europa.eu/>

³ http://www.ilc.cnr.it/iwndb/iwndb_php/

semantic classification phase is a vector of features associated to the word and lemma, together with the sentence index and sentence position. These latter indices will then be used to understand semantic relations intervening in the sentence between the main governing verb and the word under analysis. Semantic mapping is then produced by using the output of the shallow parsing and the functional mapping algorithm which produce a simplified labelling of the chunks into constituent structure. These structures are produced in a bottom-up manner and subcategorization information is only used to choose between the assignments of functional labels for argumenthood. In particular, choosing between argument labels like SUBJ, OBJ2, OBL which are used for core arguments, and ADJ which is used for all adjuncts requires some additional information related to the type of governing verb.

The first element for Functional Mapping is the Verbal Complex, which contains all the sequence of linguistic items that may contribute to its semantic interpretation, including all auxiliaries, modals, adverbials, negation, clitics. We then distinguish passive from active diathesis and we use the remaining information available in the feature vector to produce a full-fledged semantic classification at propositional level. The semantic mapping includes, beside diathesis:

- Change in the World; Subjectivity and Point of View; Speech Act; Factitivity; Polarity.

4 A comparative study

Whereas the aims of syntax and semantics in this system are relatively clear, the tasks of pragmatics are still hard to extract automatically. But, we have to recognize the huge relevance of pragmatics in analyzing political texts.

4.1 The corpus

For the elaboration of preliminary conclusions on the process of the change of the Italian government and president of government, we collected, stored and processed - partially manually, partially automatically -, relevant texts published by three national on-line newspapers having similar profiles⁴.

For analytical results to be comparable to those taken so far by second author [20,21], we needed a big corpus, especially considering five rigorous criteria that we list below:

1. Type of message

Selection of newspapers was made taking into account the type of opinions circulated by the Editorial: pro, against Berlusconi and impartial. The following newspapers were thus selected:

- a) Corriere della Sera - www.corriere.it (called The People Newspaper).
- b) Libero - www.liberoquotidiano.it (pro Berlusconi).
- c) La Repubblica - www.repubblica.it (against Berlusconi).

2. Period of time

⁴ www.corriere.it, www.liberoquotidiano.it, www.repubblica.it

The interval time chosen should be large enough to capture the lexical-semantic and syntactic richness found in the Italian press. It was divided into three time periods. We specify them here below with their abbreviations, used during analysis.

A month before the resignation of Berlusconi (12 November 2011), abbreviated to OMBB: October 12 to November 11, 2011

The period between the presentation of Berlusconi's resignation and the appointment of Mario Monti as premier of the Italian Government, abbreviated with PTMB: 12 to 16 November 2011

A month after the resignation of Berlusconi, abbreviated with OMAB: November 17 to December 12, 2011.

Two keywords were commonly used to select items from the Italian press, that is the name of the two protagonists: (Silvio) Berlusconi (and appellations found in newspaper articles: Silvio, Il Cavaliere, Il Caimano) and (Mario) Monti.

We tried to select an archive rich enough for each of the three newspapers (meaning dozens of articles per day), the selected period of time as the one of interest, between average values. Text selection was made taking into account the subcriterion *Ordina per rilevanza* (order articles by relevance) that each web page of the corresponding newspapers made available. We then introduced a new subcriterion of selection: storing articles in the first three positions of each web page for every day of the research period. In particular we collected on average 250 articles per newspaper, that is 750 articles overall. Also number of tokens are on average 150K tokens per newspaper, i.e. 450K tokens overall. Computation time on a tower MacPro equipped with 6 Gb RAM and 1 Xeon quad-core was approximately 2 hours.

4.2 The syntactic and semantic analysis

In Fig. 1 below, we present comparative semantic polarity and subjectivity analyses of the texts extracted from the three Italian newspapers. On the graph we show differences in values for four linguistic variables: they are measured as percent value over the total number of semantic linguistic variables selected from the overall analysis and distributed over three time periods on X axis. To display the data we use a simple difference formula, where Difference value is subtracted from the average of the values of the other two newspapers for that class. Differences may appear over or below the 0 line. In particular, values above the 0x axis mean they assume positive or higher than values below the 0x axis, which have a negative import. The classes chosen are respectively: 1. propositional level polarity with NEGATIVE value; 2. factivity or factuality computed at propositional level, which contains values for non factual descriptions; 3. subjectivity again computed at propositional level; 4. passive diathesis. We can now evaluate different attitudes and styles of the three newspapers with respect to the three historical periods: in particular we can now appreciate whether the articles report facts objectively without the use of additional comments documenting the opinion of the journalist. Or if it is rather the case that the subjective opinion of the journalist is present only in certain time spans and not in others.

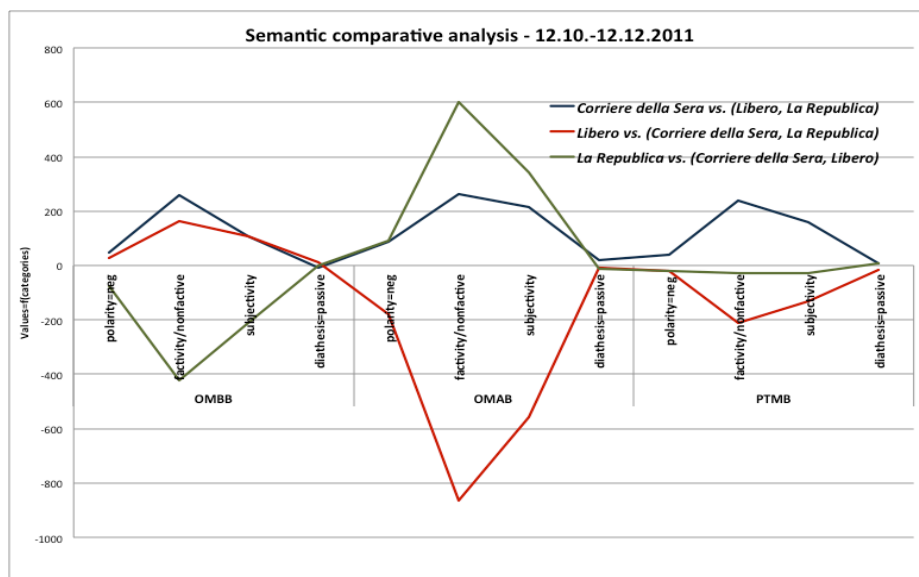


Fig. 1. Comparative semantic polarity analysis of three Italian newspapers.

So for instance, *Corriere*, the blue or darker line, has higher nonfactive values in two time spans, OMBB and PTMB; *Repubblica* values soar in OMAB. In the same period *Libero* has the lowest values; whereas in OMBB, *Libero* and *Corriere* have the highest values when compared with *Repubblica*. PTMB clearly shows up as a real intermediate period of turmoil which introduces a change: here *Repubblica* becomes more factual whereas *Libero* does the opposite. Subjectivity is distributed very much in the same way as factuality, in the three time periods even though with lesser intensity. *Libero* is the most factual newspaper, with the least number of subjective clauses. Similar conclusion can be drawn from the use of passive clauses, where we see again that *Libero* has the lowest number. The reasons for *Libero* having the lowest number of nonfactive clauses in OMAB, needs to be connected with the highest number of NEGATIVE polarity clauses, which is related to the nomination of Monti instead of Berlusconi, and is felt and is communicated to its readers as less reliable, trustable, trustworthy. Uncertainty is clearly shown in the intermediate period, PTMB, where *Corriere* has again the highest number of nonfactual clauses.

4.3 The pragmatic analysis

We show in this section the results outputted by GETARUNS when analysing the streams of textual data belonging to the three sections of the corpus (presented in section 4.1). In Fig. 2 we represent comparative differences between the three newspaper in the use of three linguistic variables for each time period. In particular, we plotted the following classes of pragmatic linguistic objects: 1. references to Berlusconi as entity (Silvio, Silvio_Berlusconi, Berlusconi, Cavaliere, Caimano); 2. references to Monti as entity (Monti, prof_Monti, professore, Mario_Monti,

super_Mario); 3. negative words or overall negative content words. To capture coreference mentions to the same entity we built a specialized coreference algorithm.

One month before Berlusconi's resignation (OMBB), we can highlight the opinions of the three dailies as follows: *Corriere della Sera* and *Libero* are concerned mostly with Berlusconi (see *Berlusconi occurrences*), with a remarkable difference however in terms of positive – *Libero* - vs negative – *Corriere* – comments. After Berlusconi resigned (OMAB) *Libero* is more concerned than the other two newspapers on Monti: negative appreciation is always higher with *Libero* and not with the other two. This can clearly be seen from the sudden dip of positive words. Finally in the intermediate period, both *Libero* and *Corriere* seem to be the most concerned with the new government, with the highest number of negative comments.

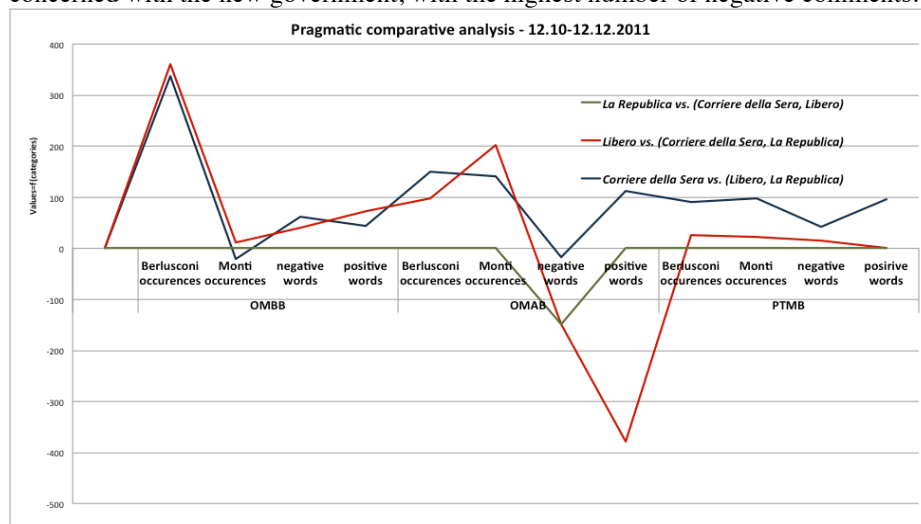


Fig. 2. Comparative pragmatic analysis of three Italian newspapers.

As shown in Fig.2, measuring the overall attitude with positive vs. negative affective content for each newspaper allows a clear cut subdivision in the three time periods. Table 1 below shows the same data in a more perspicuous manner. The percentages from Table 1 are organized as follows. Positive values are computed along time line distribution: for each newspaper, we compute the percentage referred to the each time slot. For instance, in OMBB positive values are distributed with the following subdivision in percent values: 33.88 for *Corriere*, 33.75 for *Libero*, and 32.37 for *Repubblica*. In other words, in OMBB, *Corriere* uses the most number of positive words. In fact, as can be easily noticed, *Corriere* is the newspaper that uses most positive keywords in all the three time periods. On the contrary, *Libero* is the newspaper that uses the least number of positive keywords apart from OMBB. *Repubblica* lies in the middle. The second number included in the same cell is needed to account for differences in number of tokens, and this in turn is due to differences in number of days considered for each time period: 31 for OMBB, 5 for PTBM and 26 for OMAB. Average values for each time period for each newspaper in part confirm percent values but also give a deepest idea of the actual numbers at play.

Newspaper / time period	Corriere della Sera		Libero		La Repubblica	
	positive	negative	positive	negative	positive	negative
OMBB	33.95% 52.1	35.49% 21.48	33.74% 51.9	32.6% 19.77	32.34% 49.77	31.91% 18.58
PTMB	42.36% 61.2	44.49% 21.8	24.4% 34.2	25.98% 11.4	33.24% 45.8	29.53% 16
OMAB	35.14% 54.88	32.68% 20.42	25.39% 39.58	28.21% 18	39.47% 49.12	39.12% 19.53

Table 1. Sentiment analysis of three Italian newspapers

Negative opinions are computed in the same way. These data can be interpreted as follow:

One month before Berlusconi's resignation (OMBB), both *Libero* and *Corriere della Sera* have more positive contents than *La Repubblica*, which can be interpreted as follows: Berlusconi's Government is considered a good one; in addition, *Libero*, has the lowest percentage of negative opinions about the current economic situation. In the intermediate period between Berlusconi's resignation and nomination of the new Prime Minister, Mario Monti (PTMB) we see that *Corriere* has by far the highest percentage of positive opinions, whereas *Libero* has the lowest. The other period, one month after the nomination of new prime minister, Mario Monti, (OMAB), we assist to a change of opinions. *Corriere della Sera* becomes more positive than other newspapers and also negative opinions are much higher: the new prime minister seems a good chance for the Italian situation; however, the economic situation is very bad. *Libero* – the newspaper owned by Berlusconi - becomes a lot less positive and less negative than the other two. This situation changes in the following time period, where *Libero* increases in positivity – but remains always the lowest value – and in negativity, but remains below the other two newspaper, on average. This can be regarded as a distinctive stylistic feature of *Libero* newspaper. As a whole, we can see that *Repubblica* is the one that undergoes less changes, if compared to *Libero* and *Corriere* which are the ones that undergo most changes in affective attitude.

We already saw in the Fig. 1 above that *Libero* is the newspaper with the highest number of nonfactual and subjective clauses in the OMAB time period: if we now add this information to the one derived from the use of positive vs. negative words, we see that the dramatic change in the political situation is no longer shown by the presence of a strong affective vocabulary, but by the modality of presenting important concepts related to the current political and economic situation, which becomes vague and less factual after Berlusconi resigned.

Eventually, we were interested in identifying semantic linguistic common area (identification of common words), also called common lexical fields, and their affective import (positive or negative). From previous tables, it can be easily noticed that all three newspapers use words with strong negative import, but with different frequency. Of course, this may require some specification, seeing the political context analyzed. So we decided to focus on a certain number of specialized concepts and

associated keywords that we extracted from the analysis to convey the overall attitude and feeling of the political situation. We collected in Table 2 below all words related to “Crisis Identification” (CIW for short) and noted down their absolute frequency of occurrence for each time interval.

<i>CIW OMBB</i>	<i>Corriere</i>	<i>Libero</i>	<i>Repub.</i>	<i>CIW OMAB</i>	<i>Corriere</i>	<i>Libero</i>	<i>Repub.</i>
1. crisis	124	71	94	1. crisis	50	21	110
sacrifice	4	14	4	sacrifice	9	23	16
rigour	5	4	4	rigour	23	18	10
austerity	0	6	6	austerity	6	2	0
2. battle	6	12	14	2. battle	14	4	8
dissent	2	8	8	dissent	0	4	0
dictator/ship	2	10	18	dictator/ship	2	6	2
3. fail/ure	8	13	9	3. fail/ure	21	8	15
collapse	10	6	12	collapse	8	2	4
drama/tic	12	14	18	drama/tic	4	0	8
dismiss/al	45	39	20	dismiss/al	3	2	15

Table 2. Crisis Identification words in two time periods

If we look at the list as being divided up into three main conceptualizations, we may regard the first one as denouncing the critical situation, the second one as trying to indicate some causes; and the last one as being related to the reaction to the crisis. It is now evident what the bias of each newspaper is, in relation to the incoming crisis:

- *Corriere della Sera* feels the “crisis” a lot deeper before Berlusconi’s resignation, than afterwards when Monti arrives; the same applies to *Libero*. *La Repubblica* feels the opposite way. However, whereas “austerity” is never used by *La Repubblica* after B.’s resignation and it was used before it, this is the opposite of what *Corriere della Sera* does, the word appears only after B.’s resignation, never before. As to the companion word “sacrifice”, *Libero* is the one that uses it the most, and as expected its appearance increases a lot after B.’s resignation, together with the companion word “rigour” that has the same behaviour. This word confirms *Corriere’s* attitude towards Monti’s nomination: it will bring “austerity, rigour and sacrifice”.

- in the second half, the other interesting couple of concepts is linked to “battle, dissent, dictator”. In particular, “battle” is used in the opposite way by *Corriere della Sera* when compared to the other two newspapers: the word appears more than the double in the second period, giving the impression that the new government will have to fight a lot more than the previous one. As to “dissent”, all three newspapers use it in the same manner: it disappears in both *Corriere della Sera* and *La Repubblica*, and it is halved in *Libero*. Eventually the “dictator/ship” usually related to B. or to B.’s government: it is a critical concept for *La Repubblica* in the first period, and it almost disappears in the second one.

- as to the third part of the list, whereas *Libero* felt the situation “dramatic” before B.’s resignation, the dramaticity disappears afterwards. The same applies in smaller percentage to the other two newspapers. Another companion word, “collapse” has the

same behaviour: Monti's arrival is felt positively. However, the fear and the rumours of "failure" is highly felt by *Corriere della Sera* and *La Repubblica*, less so by *Libero*. This is confirmed by the abrupt disappearance of the concept of "dismiss/al" which dips to the lowest with *Libero*.

5 Conclusion

The analysis we proposed in this paper aims at testing if a linguistic perspective anchored in natural language processing techniques (in this case, the scaled version of GETARUNS system) could be of some use in evaluating political discourse in print press. If this proves to be feasible, then a linguistic approach would become a very relevant to an applicative perspective, with important effects in the optimization of the automatic analysis of political discourse.

However, we are aware that this study only sketches a way to go, and a lot more should be studied until a reliable discourse interpreting technology will become a tool in researcher's hands. We should also be aware of the dangers of false interpretation. For instance, if we take as example the three newspapers we used in our experiments, differences at the level of lexicon and syntax, which we have highlighted as differentiating them, should be attributed only partially to their idiosyncratic rhetorical styles, because these differences could also have editorial roots. Theoretically, at least, *Corriere della Sera*, should embody an impartial opinion, *Libero*, pro Berlusconi and *La Repubblica*, against him. But differences are more subtle, and in fact, in some cases, we could likewise classify *Libero* as being impartial, *Corriere della Sera* as being pro current government and *La Repubblica* as the only one being more critical on the current government disregarding its political stance. It remains yet to be decided the impact that the use of certain syntactic structures could have over a wider audience of political discourse. In other words, this study may show that automatic linguistic processing is able to detect tendencies in the manipulation of the interlocutor with the hidden role of detouring the attention of the audience from the actual communicated content in favor of the speaker's intentions.

Different intensities of emotional levels have been clearly highlighted, but we intend to organize a much more fine-grained scale of emotional expressions. It is a well-known fact that the audience can be easily manipulated (e.g., the social and economic class) by a social actor (journalist, political actor) when their themes are treated with excessive emotional tonalities (in our study, common negative words). In the future, we intend to extend the specialized lexicon for political discourse in order to individuate more specific uses of words in context, of those words which are ambiguous between different semantic classes, or between classes in the lexicon and outside the lexicon (in which case they would not have to be counted). We believe that GETARUNS has a range of features that make it attractive as a tool to assist any kind of communication campaign. We wish it to be rapidly adapted to new domains and to new languages (i.e. Romanian), and be endowed with a user-friendly web interface that offers a wide range of functionalities. The system helps to outline distinctive features which bring a new and, sometimes, unexpected vision upon the discursive feature of journalists' writing.

Acknowledgments: In performing this research, the second author was supported by the POSDRU/89/1.5/S/63663 grant.

References

1. Bos, Johan & Delmonte, Rodolfo (eds.): "Semantics in Text Processing (STEP), Research in Computational Semantics", Vol.1, College Publications, London (2008).
2. Esuli, A. and F. Sebastiani. Sentiwordnet: a publicly available lexical resource for opinion mining. In Proceedings of the 5th Conference on Language Resources and Evaluation LREC, 6, 2006.
3. Delmonte, R. (2007). Computational Linguistic Text Processing – Logical Form, Logical Form, Semantic Interpretation, Discourse Relations and Question Answering, Nova Science Publishers, New York.
4. Delmonte, R., Tonelli, S., Tripodi, R.: Semantic Processing for Text Entailment with VENSES, published at <http://www.nist.gov/tac/publications/2009/papers.html> in TAC 2009 Proceedings Papers (2010).
5. Delmonte, R. (2009). Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution, Nova Science Publishers, New York.
6. Delmonte R. and Vincenzo Pallotta, 2011. Opinion Mining and Sentiment Analysis Need Text Understanding, in "Advances in Distributed Agent-based Retrieval Tools", "Advances in Intelligent and Soft Computing", Springer, 81-96.
7. Gifu, D. and Cristea, D.: Multi-dimensional analysis of political language, in J. J. (Jong Hyuk) Park, V. Leung, T. Shon, Cho-Li Wang (eds.) In Proc. of 7th FTRA International Conference on Future Information Technology, Application, and Service – FutureTech-2012, Vancouver, vol. 1, Springer (2012).
8. Hobbs, J. R., Stickel, M., Appelt, D., and Martin, P.: "Interpretation as Abduction", SRI International Artificial Intelligence Centre Technical Note 499 (1990).
9. Pennebaker, James W., Booth, Roger J., Francis, Martha E.: "Linguistic Inquiry and Word Count" (LIWC), at <http://www.liwc.net/>.
10. Kim, S.-M. and E. Hovy. Determining the sentiment of opinions. In Proceedings of the 20th international conference on computational linguistics (COLING 2004), page 1367–1373, August 2004.
11. Pang, B. and L. Lee. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the 42nd annual meeting of the Association for Computational Linguistics (ACL), page 271–278, 2004.
12. Polanyi, Livia and Zaenen, Annie: "Contextual valence shifters". In Janyce Wiebe, editor, Computing Attitude and Affect in Text: Theory and Applications. Springer, Dordrecht, 1–10 (2006).
13. Pollack, M., Pereira, F.: "Incremental interpretation". In Artificial Intelligence 50, 37-82 (1991).
14. Saurì R., Pustejovsky, J.: "Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text", Computational Linguistics, 38, 2, 261-299 (2012).
15. Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M.: "Lexicon-based methods for sentiment analysis". In Computational Linguistics 37(2): 267-307 (2011).
16. Turney, P.D. and M.L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. ACM Transactions on Information Systems (TOIS), pages 15–346, 2003.
17. Wiebe, Janyce, Wilson, Theresa, Cardie, Claire: "Annotating expressions of opinions and emotions in language". In Language Resources and Evaluation, 39(2):165–210 (2005).