ITALIAN POLITICAL DISCOURSE ANALYSIS. OPINION AND FEELING

Daniela GIFU[1]

1. Postdoctoral Researcher, "Al.I. Cuza" University of Iasi, Romania
Corresponding author: daniela.gifu@info.uaic.ro

Abstract

From the public opinion's perspective, one can assess the success of an article by the degree to which a journalist is able to change political attitudes, opinions, feelings and beliefs. We present a computational system – GETARUNS – which integrates a wide range of natural language processing tools intended to identify some characteristics of journalistic discourse (here, political issues from the written press). The method can be applied to different languages (Italian, English etc.) to the end of deeply understanding and comparing texts in terms of linguistics (lexical-semantic richness and so on). The system is able to produce tables, graphs, pragmatic marks by means of which one can compare two or more discourses, distinct fields, or to analyse the run of the same subject in different contexts. Our approach is based on the assumption that linguistics can configure a complete analysis of written press discourse. The system can be useful for specialists in various fields, including media, communication, economics and politics, psychology, education etc.

Keywords: *Italian written press, lexical-semantic analysis, syntax, opinion and feeling, rhetoric, journalist, natural language processing.*

1. INTRODUCTION

The goal of an interdisciplinary approach as political discourse analysis, with natural language processing tools (NLP), is to define and explain various discourse contexts, in this case, Silvio Berlusconi's resignation of (November 11th, 2011) and Mario Monti's appointment as Prime Minister of the Italian Government (November 16th, 2011) – two events reflected in the written press, during October-December 2011. Most studies in this regard have focused mainly on three analytical aspects: first one focusing on the emotional dimension identified by how people use language (the discourse means people use to express their feelings); the second one, centering upon understanding the relationship between speech in a certain area and outside it and the third – the role of linguistic structure (here, the Italian language) as a means of communication.

Linguists most often treat language as an abstract object that can be measured without reference to social and political concerns, whatever may their nature be.

Content analysis based on the theories of Osgood *et al.* – but also Taboada *et al.* – calls for a very laborious methodology in obtaining objective results that can be validated by other scientific methods. In this paper, we discuss communication paradigms, especially those that support linguistic discourse interpretation based on natural language processing by means of the computational system GETARUNS (*General Text And Reference Understanding System*), which can complete access to extra-linguistic knowledge. We are considering three critical aspects for a successful evaluation of the political text from the written press: the creation of large amounts of data to train the proposed system as a method of analysis, obtaining important results in terms of semiotics, semantics and lexicology, syntax and pragmatics, and improving the so far implemented heuristics.

A set of criteria for texts selection complete our assumption, *i.e.* that this corpus (Italian political articles from the written press) will be sufficient for extracting opinion and feeling[1] with GETARUNS system. After the texts processing, at the end of the analysis, we get some data output, viable at the sentence level. In particular, this analytical approach can be checked by using tools such as BOWs (*Bag of Words Approaches*) to the end of recovering information – an aspect still insufficiently dealt with. Such approaches, like BOWs, are based on the matching of several keywords, the existence of certain ontologies, and on the search of concepts by means of SentiWordNet (*Sentiment Analysis and Opinion Mining with WordNet*), starting from a text. The system uses the input words, analyzing them

syntactically and semantically. Any search based on keywords and Bows is vitiated by the failure in understanding precisely the processed word, set next to other words – fact that many times changes its meaning[2]: negation at different syntactic levels; verb or adverb negation; the conditional mood; the double negation with copulative verbs, the use of modal verbs etc.

In order to deal with these linguistic elements, we propose a sentence level analysis, starting from a tense political context, with syntactic elements in a tree-like pattern representation. Within the GETARUNS system there have been implemented various computational functions, necessary for the semantic evaluation called RTW, but also for other semantic tasks[3]. The system provides as output texts .xml format, where each sentence is a list of attribute-value pairs. To the intended end of using these data output, the system makes use of a linear syntactic structure and an vector of semantic attributes associated to the main verb at the sentence level.

Evaluation of opinion and feeling is closely related to the semantic content of each sentence, assigned to two separate categories: objective vs. subjective. This differentiation is achieved through the search of some labels (*eg. factivity*), again at the sentence level[4]. In particular, we talk about measuring modal operators such as stress, modal verbs, modifiers and sentence-level attributes, the lexical type of the verb (after the existing classes within ItalWordNet and those in GETARUNS), the person of the subject (if it is at the third person or not) a.s.o.

The paper is structured as follows: chapter 2 contains a brief description of the current state of computational technologies implemented in the purpose of linguistic interpretation. Chapter 3 presents some information about the role of the written media discourse in the public space; chapter 4 describes the GETARUNS system as a multi-dimensional analytical method of political discourse. Chapter 5 presents a case study, the analysis of Italian political speeches from October 12th till December 12th, 2011, collected, stored and processed by the system. The study concludes with Chapter 6, which highlights some conclusions drawn from this analysis, plus new directions of development of the proposed method.

2. PREVIOUS CONCERNS

As we shall see, an important aspect of GETARUNS platform presented here is that of lexical-semantic functionality, found on a large number of computing systems from the field literature. However, the most important similarities are to be found particularly in TACITUS (*The Abduction Commonsense Inference Text Understanding Systems*) and KERNEL (*Knowledge Extraction and Refinement by Neural Learning*).

Tacitus system design can be broadly summarized to the fact that we should not give up any information that might be relevant from a semantic or pragmatic point of view, if we want a complete and correct interpretation. This system performs a syntactic analysis of sentences of a text, using a grammar of the English language, focusing on the main predicates logic. Pragmatics problems are solved by inference[5] (identifying inference with meaning still leaves us with many questions; it would be enough that inference be associated only with specification). Kernel architecture is closest to that of the GETARUNS system, in which the syntactic, semantic and pragmatic functionalities are separated into different processing modules, which are allowed to communicate. Kernel carries out analysis in two stages: first, syntactic parsing, where we identify a "limited access to semantic constraints for disambiguation"[6] and, second, the integration of semantic and pragmatic processing that limited the access to external sources.

Each clause in its turn is then translated and labeled (mapped) in a functional representation, with the attribute-value, called ISR. Semantic interpretation is performed while identifying ISR, fact involving many other features, such as recovery of unexpressed forming elements (we give as examples the subjects or the non-required prepositional phrases), as well as compulsory arguments, given by a verbal predicate. In addition, the noun phrase analysis has a separate analysis mechanism of the clause, in which the previous noun clarifies the relations between the noun subjects. This requires a search for the referent discourse. Here comes a new system limitation not to be neglected: since each forming

element is interpreted in the textual logic order, it is still not yet available for use in the interpretation. The rigidity of the system could be overcome, if it could choose to delay the reference resolution of all nouns, as found in the CANDIDE system[7]. This system implements a similar theoretical approach, the non-compositional interaction of semantic interpretation for the pragmatic context in the reference determining of the noun phrase. The system was designed and implemented to illustrate the acquisition of procedural knowledge, expressed in a combination of diagrams for the ordinary speech in English.

3. WRITTEN PRESS DISCOURSE

Mirror of the contemporary society found in a perpetual socio-cultural revaluation, the written press texts can be uncomfortable or useful for a certain regime at a given moment. "In contemporary post-industrial society, the battle stake is not the social use of technology anymore, but the mass production and dissemination of representations, information and languages"[8]. It explains why the media becomes ever more totality featured: a newspaper summarizes a whole day of activities on the entire planet. It is no exaggeration if we say that the newspaper is both the moral temperature bulletin and notary service of all human acts, thus becoming the most complete archive of history[9]. In other words, "the written press continues to be perceived (at least for some of its components) as an essential instrument for the exposure and free discussion of ideas"[10]. For the newspapers reading to keep inciting the modern people's interest, the *written press* must abandon self-quotation and self-referentiality, as Eco points: "When it is not talking about television, the press talks about itself; it has learnt from the television, that speaks most about television"[11].

In politics, we witness either the prevailing of the "heavy" actors, who benefit from a preferential publicity treatment, effect called *best-seller* by Roland Cayrol, or the ignorance of the low-profile actors, who are marginalized or even ignored. Most times, the propelling of new actors is accompanied by changing others, by changing the intermediate body, the militants, who are condemned not only to media silence, but simply silenced, the role of the opinion leaders being drastically reduced[12].

The written press, in its various forms, assigns political significance to institutional activities and events that, through their succession, form the political life of a nation – object for one's information and topic of public debate. In this case, the role of the written press is twofold: it secures information as a credible *speech* that ends a rumor and frames the political within language forms, this one acquiring the interpretable consistency of a symbolic system of representations.

The press is designed to legitimize the actions of politicians, attending their visibility efforts. The written press includes mainly political speeches, containing both a specific orientation and a political commitment. The reader has the opportunity to choose what and when to read, giving himself/herself time for reflection. Disproportion is a risk to the reality described. "The huge amount of news-printed paper issued on the market had to be filled with attractive texts for an audience that one could no longer force to buy any publication. Something that had never actually been discussed until then had to be correctly applied: the satisfaction of the free people's needs"[13]. No wonder why any authoritarian regime finds it necessary, as a first step to take for governing in peace, to censor the enthusiasm of the press.

Most times, especially during elections, the written press is focused on topical issues of public interest in relation to both internal and external social events. However, the perception of social reality depends on how it is presented. The newspaper, just like any commercial product, depends on its aesthetic presentation, which can distort any alternative of selecting relevant news. It will try, par excellence, to bring to public attention something sensational and, more than often, negative.

4. GETARUNS METHOD

In this section we will present a detailed description of the Italian computational, system, which was used in this study. The system is derived from GETARUNS, a multilingual system

for text analysis of semiotic perspective, applicable for English, German and Italian. Over the past 20 years, the system provided the basis for a lot of research, promoted in various publications and at international conferences in the field.

The present version of the system can be applied to unlimited texts, the vocabulary being constantly improved for both the Italian and English languages. The method was used for different RTE functions (*Rich Text Editor*) and automatic summarization (TAC)[14]. It has also been adapted for opinion and feeling analysis, the results being already published in English.

The current version (3.0), focusing more on Italian, made it possible to create the resources needed for the semantic analysis, particularly by adapting the tool SentiWordNet to Italian. However, this process has introduced a lot of noise in the final results, and therefore a lot of input was totally wrong. In addition, a constant concern was to uniquely characterize only the input that has an associated positive or negative meaning, a stage considered as the only possible solution to the problem of semantic disambiguation, called phase *WSD* (*Word Sense Disambiguation*), which was not part of the GETARUNS. In this case, it was decided to delete all records that had several meanings associated to the same topic, and often with opposing emotional values.

In the summer of 2012 there was created an 'ad hoc' lexicon for most concepts (around 3000 in number) contained in the analyzed text, in order to reduce the ambiguity problem. It is known that a context chosen according to well-established criteria also limits the meaning of words with multiple connotations. Here we have chosen a fervently discussed political context in Italy, in the year of 2011. The stage performed used the same method, *i.e.* the labeling of only the emotionally charged concepts.

The system was recently released through participation of researchers involved in EVALITA (*Evaluation of NLP and Speech Tools for Italian*). In fig. 1 we can see the interface of this computing system, one of the NLP tools (*Natural Language Processing*), where we can identify a number of instruments that can be used in text analysis. This also includes the raw text tokenization operation (sentences segmenting) and the Multi-words option, essential for understanding the specific meanings related to the fields associated to the words sequences. This function is then extended to NER (*Named Entity Recognition*), which can be performed on a larger amount of data, recently provided by the Joint Research Centre, JRC (*Joint Research Centre*)[15]. Of course, we also used a list of texts and multiwords from the data base of the University Ca' Foscari of Venice, created during the Italian national research projects such as ISST (1998-2002), which was furthered by the production of the first Italian treebank (syntactic tree) of the same name.

Words that are not recognized by simple procedures, according to the existing vocabulary within the program (500 K input) are sent to the morphological analyzer. If recognition fails here, too, the guessing option is activated, the first connection of the word with the context of the sentence. We are talking of prefixes, the meaning of other words. The process may continue with the possible suffixes.

To put it otherwise, the words derivation with suffixes, prefixes and parasynthetic. If none of these tests gives results, the word will be labeled as foreign (only if the word does not end with a vowel). The sentence that includes that unacknowledged word is isolated. The output
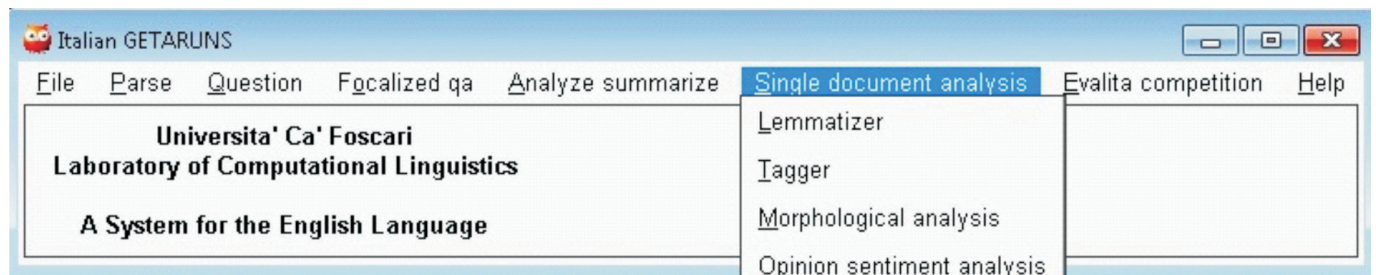

Fig. 1. GETARUNS – work session

is set for the next level, and the text analysis continues.

The second level of analysis has as input (input) one simple/complex sentence, each word being tagged (morpho-syntactically labeled) with a tagset of 91 tags and a set of 31 syntactic tags[16].

Transition to the semantic level is performed as follows: each nominal expression is classified first according to the assigned tag (the proper noun is noted NER). Remaining nominal expressions are classified with the help of the classes derived from ItalWordNet (*Italian WordNet*). In addition, there is a compiled specialized database of terminology for an important number of fields: medical, political, economic, military etc.

Words are assigned to one of the specific classes and to the general classes derived from ItalWordNet. If the word (or multiword) is not present in the lexicons of the two platforms, it is classified as unique. This semantic identification stage provides as output a vector of associated features with word and lemma, together with the index and the position of the sentence. These last indices will be used in understanding the semantic relations that occur in the text (simple/complex sentence) between the main verb and the analyzed word.

Here are a few examples of vectors for the common and proper nouns:

refex(l_omab1-2,n,editoriale,[3,any,sing],
    [artf,lngr])
refex(l_omab1-4,nh,'Maurizio_Belpietro',
    [3,any,sing],[hum])

Verbs, on the contrary, are classified according to specialized semantic lexicon. The features vector includes morphological, syntactic and semantic indices, according to the below example:

refex(l_omab2-1,v,mancare,[mfeats=kl3s],
    [tr,activ])
refex(l_omab3-2,v,cominciare,[mfeats=fl],
    [raisn,achiev,process])
refex(l_omab5-35,v,proseguire,[mfeats=fl],
    [tr,accomp,process])
refex(l_omab6-5,vppt,prendere, [mfeats=tsms],
    [refl,result_state,process])

Semantic labeling is then performed using the output generated from parsing and automatic mapping. In fact, we are talking about a segmented labeling (with the *chunk* attribute) within textual structure. These structures are generated from the bottom up, the subcategories information being used only to choose between the functionally assigned labels for argumentation. In particular, the choice between these labels, such as the argument subject (SUBJ), object2 (obj2), unidentified object (OBL) and so on, are used for additional information with respect to the main verb.

The first element in functional mapping is the verbal complex that contains all language sequences that may contribute to their semantic interpretations, including all auxiliaries, modals, adverbs, negations etc. We also distinguish between the active and passive voice and we use the remaining available information within the feature-vector, in order to generate a complete semantic classification at sentence level (between the elements of multiple sentence parts: multiple subjects, multiple predicate names, multiple attributes).

Semantic mapping, as a consequence of the below described categories, includes voice as well:

– *Change in the World;*
– *Subjectivity and Point of View;*
– *Speech Act;*
– *Factitiviy;*
– *Polarity.*

Here are some examples, the verbal categories, followed by the associated features:

propositional_ semantics(['polarity=pos,
    speech_act=statement, diathesis=active,
    verbal_complexes['
1-[i(7,ostenti,v,ostentare-[cat=verb,pred=
    ostent+are,scat=tr,
    mood=imp,tense=pres,pers=3,num=s])]
2-[i(10,si,clit,si-[sems=nom,mfeats='3spm']),
    i(11,dica,v,dire-[sems=intr,mfeats=hl3s]),
    i(12,convinto,vppt,convincere-[sems=intr,
    mfeats=tsms])]
3-[i(14,arrivare,v,arrivare-[sems=intr,
    mfeats=fl])]
4-[i(24,sa,v,sapere-[sems=intr,mfeats=kl3s])]
5-[i(32,è,vc,essere-[sems=cop,mfeats=kl3s])]]'
'verb_semantics['1-[view=external,factive=
    factive,moodtense=present_subjunct],
2-[view=external,factive=factive,moodtense=
    present_subjunct,view=external,factive=
    factive, moodtense=past_particip],

3-[view=external,factive=factive,moodtense=
infinitival]
4-[view=external,factive=factive,moodtense=
present]
5-[change=null, view=internal_intensional,
factive=nonfactive,moodtense=present]]])'.

5. CASE STUDY

As the syntactic and semantic targets within this system are relatively clear, the heuristics of pragmatic interpretation are still in a test phase. The pragmatic dimension of the text is described by several criteria implemented in the system, the results being compared on the basis of some direct observations: polarity (*polarity*), discursive action (*speech_act*), divided into exclamatory and interrogative, subjectivity (*point_of_view is subjective*) etc.

In this paper we shall present the corresponding results of the implemented criteria in the system, in three dimensions: syntactic, semantic and pragmatic, that which we may call political discourse semiotic with the help of a natural language processing system.

5.1 CORPUS

For the elaboration of the preliminary conclusions drawn after the changing of the Italian Prime Minister, we collected, stored and processed, both automatically and manually, relevant texts published in three national newspapers available also online and having similar profiles[17].

For the analytical results to be comparable to those performed so far in Romania, we needed a large enough corpus, but, most of all, a corpus selected according to rigorous criteria, given the short time for this research.

We present them below:

1. *Message Type*

Selection of newspapers was made taking into account the type of editorial opinion adopted by each publication. The important thing was to have three types of opinions: for and against Berlusconi, and impartial. Thus, we chose the following newspapers:

a) *Corriere della Sera*, also called *The People's Newspaper* – impartial;
b) *Libero* – pro Berlusconi;
c) *La Repubblica* – against Berlusconi.

2. *Time Period*

A long enough period of time has been chosen so that one could capture the lexical-semantic richness but especially the discursive inflections, consistent with the political change that the Italian people could witness at the end of 2011.

Here were the three periods of time that we will also abbreviate, and the abbreviations will be used throughout this analysis.

a) One month before Berlusconi's resignation (12 November 2011), abbreviated to OMBB: October 12th to November 11th, 2011;

b) The period between Berlusconi's resignation and Monti's appointment as prime minister of the Italian Government, abbreviated to PTMB:
November 12th to November 16th 2011;

c) One month after Berlusconi's resignation, abbreviated to OMAB:
October 17th to December 12th, 2011.

3. *Keyword*

The selection of the Italian written press articles was made according to two keywords, the names of the two political protagonists, along with their appellations: (Silvio) Berlusconi (Silvio, Il Cavaliere, Il Caimano) and (Mario) Monti.

4. *Article Scale*

Given the impressive archive of each monitored publication (*i.e.* dozens of articles per day), the text selection was made taking into account the subcriterion *Ordina per rilevanza* (article's relevance), held by each web page corresponding to the analyzed newspapers.

5. *Article Location*

Of course, the number of articles was greatly reduced, but we had to consider another selection sub-criterion: storage of articles in the first three positions of each web page corresponding to the analyzed days.

In sections 5.2., 5.3 and 5.4, we describe and interpret the results represented by graphics (Fig. 2, Fig. 3 and Fig. 4) given by GETARUNS, and corresponding to the three time periods analyzed for each of the three daily newspapers (the legend is edifying in this sense). Values above the horizontal axis, OX, highlight the criteria that are more relevant for a newspaper than for the other two to which it is compared, while values below the horizontal axis, OX, highlight criteria that are less visible in a newspaper compared to the others.

5.2 SYNTACTIC ANALYSIS

From the chart below, fig. 2, we notice the positive values in the period before Berlusconi's resignation (OMBB), when the newspaper *Corriere della Sera* has the richest speech vocabulary (*tokens*), feature found throughout the entire analyzed period of time. At the same time, in the period between Berlusconi's resignation and the appointment of Monti (PTMB), as well as after the appointment of Monti (OMAB) there is to

be noticed a predominance of the verbal element, in comparison to the other two newspapers. Although this suggests dynamism, it is a pessimistic one (*verbal components* and *polysem = negative*).

Libero has a speech with poor syntactic structure, without predominant features, and in *La Repubblica*, during OMBB, there is a prevalence of negative polysemantic words (*polysem = negative*). Of course, this is only natural, given the type of approach to this topic, against Berlusconi and his government.

5.3 SEMANTIC ANALYSIS

From the chart below, fig. 3, one can notice by following the positive values that in the period before Berlusconi's resignation, OMBB, *Corriere della Sera* pays the most attention to Berlusconi's activity (*Berlusconi occurences*) and to the Italian Government as an institution (*words classified as ORG = organization*), feature found in the period after the appointment of Monti (OMAB) as well. At the same time, during OMAB and between
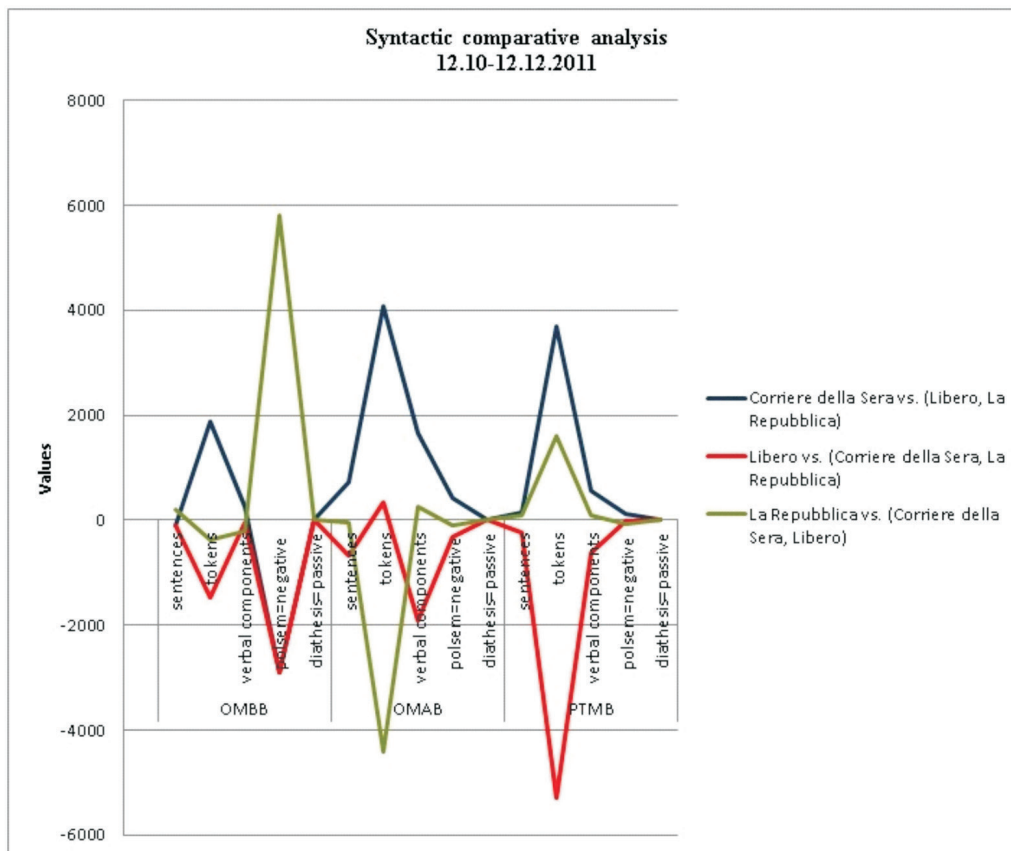

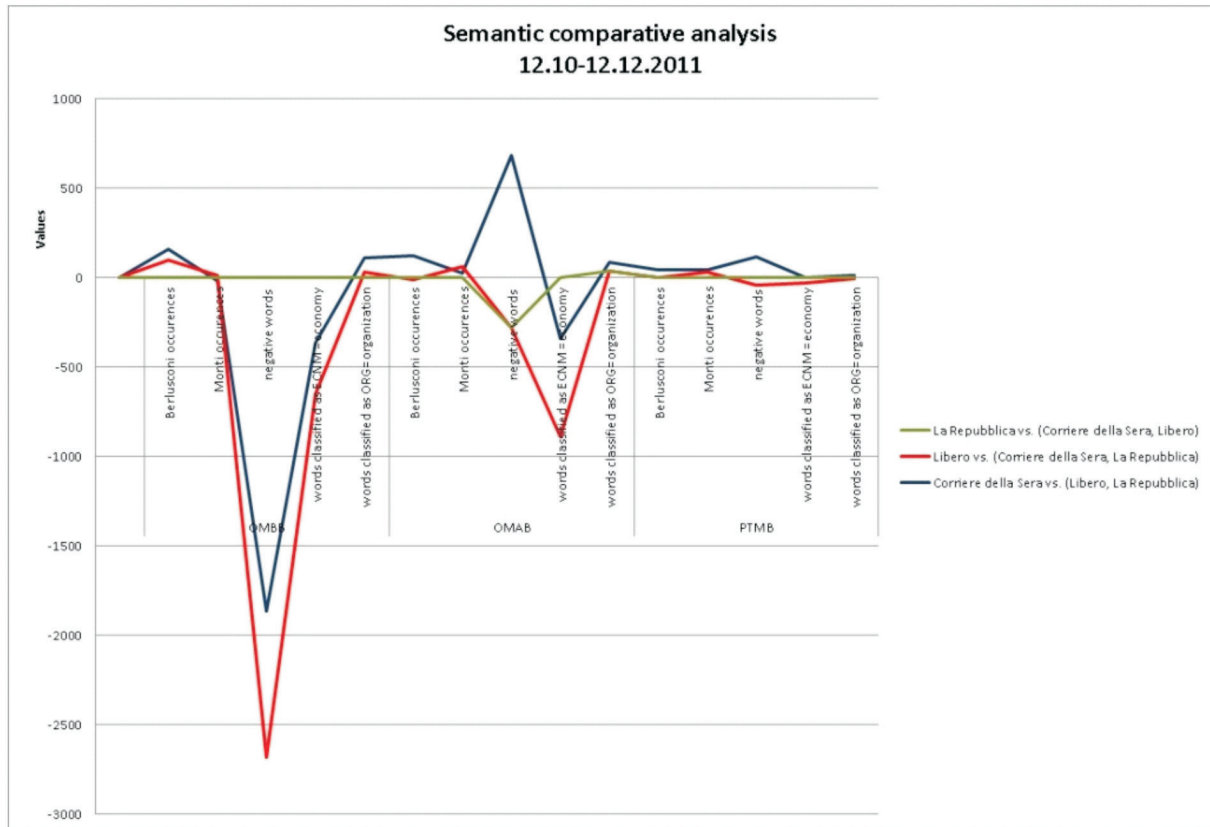**Fig. 2. Comparative Syntactic Analysis of Three Italian Newspapers**

Fig. 3. Comparative Semantic Analysis of Three Italian Newspapers

Berlusconi's resignation and Monti's appointment (PTMB) prevail again the semantically negative loaded words (*negative words*), fact which confirms that none of the two options seemed too good a solution for the Italian people .

After Berlusconi's resignation, *Libero* pays most attention to the new prime minister Monti (*Monti occurences*), and *La Repubblica*, with slight differences after the appointment of the new prime minister, has a linear speech without relevant semantic charges in comparison with the other two publications.

5.4 PRAGMATIC ANALYSIS

From the chart below, fig. 4, one can notice, by following the positive values, that during the period before Berlusconi's resignation (OMBB), *Corriere della Sera* has an interrogative speech (*speech_acts / question*), a feature also found in the period after Monti's appointment (OMAB). Idiomatically and in combination with an exclamatory intonation, some partially positive interrogations have acquired rhetorical value, assuming a negative polarity answer (*negative polarity*), especially after Berlusconi's resignation. During the period between Berlusconi's resignation and the appointment of Monti (PTMB), we note the presence of impersonal constructions and factive verbs (*factivity*)[18].

During OMBB period, *Libero* stands out by negative polarity (*polarity negative*), while in *La Repubblica*, before Berlusconi's resignation and after Monti's appointment as prime minister, we notice the presence of factive verbs (*factivity*) and a high degree of subjectivity (*point_of_view is subjective*).

At the same time, our study has also been concerned with the share of feeling (be it negative or positive) during the three periods of time established in section 5.1, thing that completes the discursive picture reflected by the three graphic representations priorly analyzed. We can also establish a relation between the type of opinion and the emotional charge, found on each analyzed newspaper (tab. 1). They are the two dominant tonalities of discourse analysis.
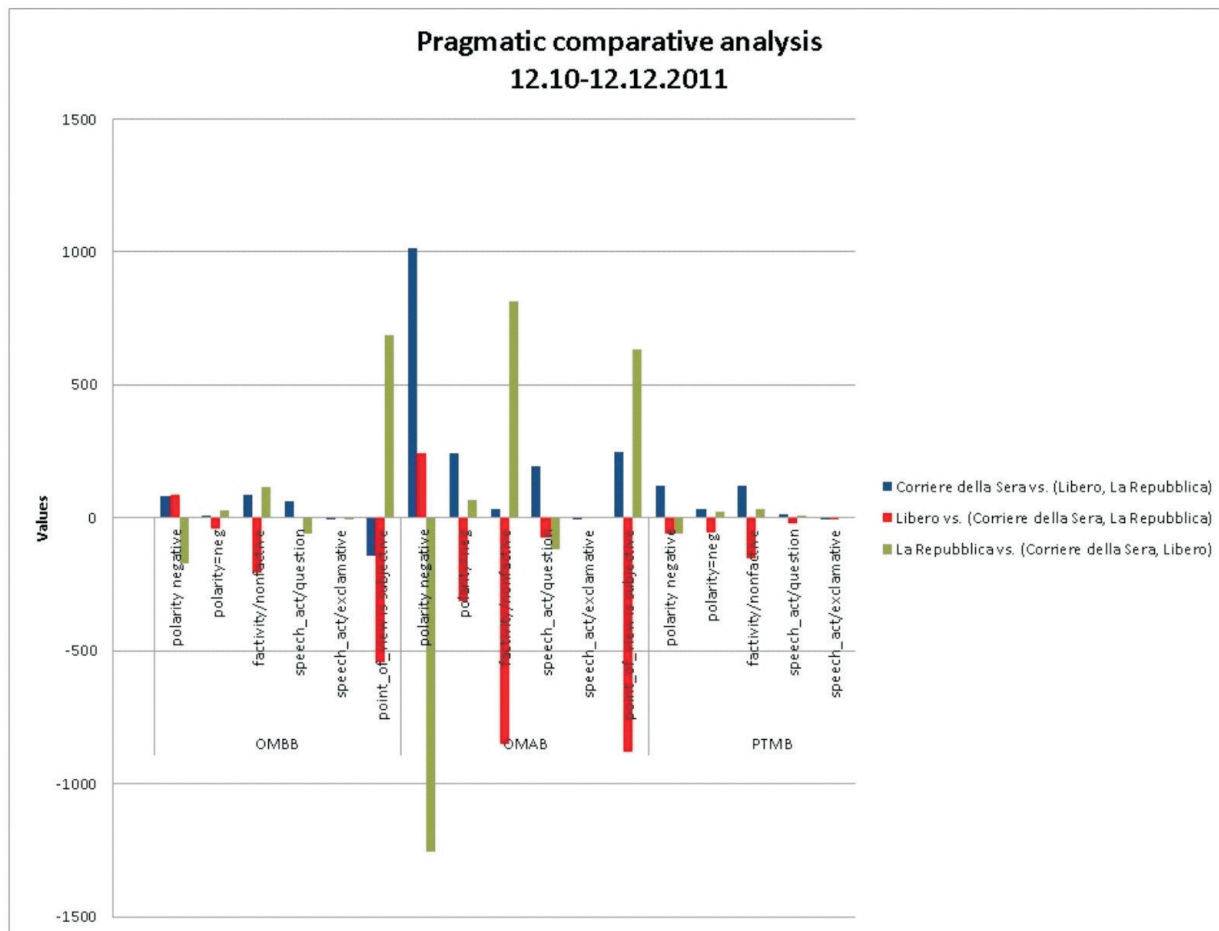
Fig. 4. Pragmatic Comparative Analysis of Three Italian Newspapers

Therefore, one month before Berlusconi's resignation (OMBB), both newspapers, *Libero* and *La Repubblica* have a more positive content than *Corriere della Sera*, values that can be interpreted as follows: *Libero* shows that Berlusconi Government was considered to be good, while *Corriere della Sera* has a strong negative tone.

A month after the appointment of the new prime minister, Mario Monti, (OMAB), we are witnessing a change of opinion. *Corriere della Sera*

becomes more positive than the other two newspapers, and thus decreases the number of negative words; the new prime minister seems to be a good political appointment for the tense situation in Italy. However, *Libero* – a pro Berlusconi newspaper – becomes much more negative than the other two in our attention.

We have also been interested in identifying the linguistic area where we mention the joint words, as well as their emotional import (positive or negative).

Table 1. Feeling Analysis in the Italian Written Press

Newspaper/ period of time	Corriere della Sera		Libero		La Repubblica	
	Positive words	**Negative words**	**Positive words**	**Positive words**	**Positive words**	**Positive words**
OMBB	59,85%	40,15%	62,62%	37.38%	62,93%	37,07%
PTMB	49,76%	50,24%	51,00%	49,00%	51,31%	48,69%
OMAB	49,60%	50,40%	42,16%	57,84%	42,91%	57,09%

Table 2. Joint Critical Words

CORRIERE DELLA SERA	LIBERO	LA REPUBBLICA
OMBB		
crisis	124	71	94
sacrifice	4	14	4
rigor	5	4	4
austerity	0	6	6
fight	6	12	14
collapse	10	6	12
dissidence	2	8	8
failure	8	13	9
dramatic	12	14	18
rejected	45	39	20
dictator	2	10	18
OMAB		
crisis	50	21	10
sacrifice	9	23	16
rigor	23	18	10
austerity	6	2	0
fight	14	4	8
collapse	8	2	4
dissidence	0	4	0
failure	21	8	15
dramatic	4	0	8
rejected	3	2	15
dictator	2	6	2

We note from Table 2 that the three papers made use of a few words with strong negative charge. They are justifiable, if we consider the analyzed political context. We chose a few keywords (11), taking into account their weight, extracted from the lexical-semantic analysis, to identify feeling in the three newspapers texts.

The 11 concepts can be designed in three discursive stages, the first focused on the crisis situation (*crisis*), the second on trying to identify the triggering causes (*austerity*, *fight* and so on), and the last related to response to the crisis (*dictator*). The biased opinion seen in every analyzed newspaper is obvious in terms of the political crisis, triggering the change of Government:

In *Corriere della Sera*, "crisis" is more prevalent before Berlusconi's resignation, (OMBB) than after Monti's appointment, situation similar to that met in *Libero*. *La Repubblica* manifests the opposite extreme of the discourse. However, even if "austerity" is never used by *La Repubblica*, after the resignation of Berlusconi (OMAB), yet it is found used prior to his resignation, an opposite situation to that revealed by *Corriere della Sera*.

Regarding the word "sacrifice", *Libero* is the one using it the most, and, as expected, its interest increased after Berlusconi's resignation, together

with the word "rigor", which confirms the attitude towards Monti's nomination.

In the second half of the list, while *Libero* feels the "dramatic" situation before Berlusconi's resignation, the other two newspapers display the same discursive behaviour, but in a lower percentage. Another word, "collapse", responds to the same pattern: Monti's appointment brings back the positivity of journalistic discourse. However, fear and rumours of "failure" are obvious in *Corriere della Sera* and *La Repubblica*, but less in *Libero*. This is confirmed by the sudden disappearance of the word "reject" the least used by *Libero*.

Another interesting couple of concepts is that related to "fight, dissidence, dictator": especially the word "fight", used in its opposite meaning by *Corriere della Sera*, in comparison with the other two papers (the word appears twice as much in the second period, giving the impression that the new government should fight even more intensely with the social and political problems than the previous one). As far as "dissidence" is concerned, all three newspapers used it in the same way: it disappears in both *Corriere della Sera* and *La Repubblica* and is reduced by half in *Libero*. Finally, the word "dictator" is usually linked to Berlusconi's name or his government, being a clearly preferred concept by journalists from *La Repubblica* in the first period (OMBB), only to disappear after Monti's appointment.

6. CONCLUSIONS

The analysis we proposed in this paper aimed at highlighting linguistic features using political language processing techniques (here, the GETARUNS system) in assessing political discourse in Italian written press. Our endeavor has been shown to be feasible in terms of the semiotic approach that becomes relevant and easily enforceable, with important effects in automatic optimization of political discourse analysis.

The following rhetorical aspects have been reported from the Italian journalists' texts that we stopped upon: lexical diversity and richness of the semantic classes available in the system, syntactic structure, use of metaphorical expressions and persuasive power of the emotionally charged words etc. Our belief is that this type of study confirms the usefulness of language technologies that can complete quickly and objectively, quantitatively and qualitatively, many gaps in the content analysis theory based on computational technologies.

However, we are aware of the fact that this study only outlines an analytical approach, and that there still remains much to be studied till the automatic validation of a discourse interpretation, the technology being yet dependent on the hands and mind of the researcher.

We are also aware of the dangers of false interpretations. For example, if we start from the three newspapers monitored in this study, the differences in vocabulary and syntax reported here can be partially attributed to the rhetorical idiosyncratic styles as well (what we call editorial policy). Theoretically relating ourseves to this analyzed subject, *Corriere della Sera* would have an impartial discursive approach, *Libero*, pro Berlusconi, and *La Repubblica*, against Berlusconi. But the differences are more subtle. In fact, in some cases, we may categorize differently the discursive styles found in the three newspapers: *Libero* – as impartial, Corriere della Sera – as favoring the current government, and *La Repubblica* – as having a critical attitude of the new government, regardless of the political position.

There still remains to be determined the impact that some of the syntactic structures met in the political discourse could have on an audience (*i.e.* extending public hearings, aimed at by all newspapers). Various journalists may use these measures at the level of discursive-rhetorical strategy, exploiting them, sometimes too much, for the benefit of their editorial. In other words, this study may show that by means of automatic processing of natural language there can be detected the trends of recipient's manipulation, designed to divert public attention from the transmitted real content, in favor of the transmitter's hidden intention.

Many interpretation facets are relevant for the specific context in which a speech is expressed. For example, in a critical political context, speeches must be assessed according to the relation between the journalist's agenda and that of the political actor, both in agreement with the public agenda – or so it should be.

Different intensities of the emotional levels have been clearly highlighted, our future efforts being channeled on the improvement and sophistication of the emotional classes. It is well known the fact that the public can be easily manipulated (*e.g.* social and economic classes), by a social actor (journalist, political actor), when their topics are treated with a too emotional tone (in our study, words were predominantly negative).

We are aware that several analytical aspects have not yet been sufficiently refined and improved. One of the most important problems remains that of establishing the meanings of the ambiguous words and expressions in a certain context.

In the future, we intend to include a disambiguation module, to identify the correct meanings of words/sentences in a certain context. In this respect, our focus is on the relationship between the embedded lexicon of the GETARUNS system, assigned in semantic classes, and the external lexicon (Dictionary of the Italian Language). We believe that the GETARUNS system has got a series of features that make it attractive, proving it to be a useful tool for any kind of communication campaign.

References

1. Delmonte, R. (2004), *Text Understanding with GETARUNS for Q/A and Summarization* in "Proc. ACL 2004 – 2nd Workshop on Text Meaning & Interpretation", Barcelona, Columbia University, 97-104.
2. Delmonte, R. (2007), *Computational Linguistic Text Processing – Logical Form, Logical Form, Semantic Interpretation, Discourse Relations and Question Answering*, Nova Science Publishers, New York.
3. Delmonte, R. (2009), *Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution*, Nova Science Publishers, New York.
4. Delmonte, R., Pallotta, V. (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.
5. Delmonte, R. *et al.* (2006), *Another Evaluation of Anaphora Resolution Algorithms and a Comparison with GETARUNS' Knowledge Rich Approach* in "ROMAND 2006 – 11th EACL", Geneva, 3-10.
6. Derville, Gregory (1997), *Le pouvoir des médias. Mythes et réalités*, PUG, Grenoble.
7. Eco, Umberto (1998), *Cinci scrieri morale*, Pontica Publishing House, Constanta.
8. Hobbs, J. R., Stickel, M., Appelt, D., Martin, P. (1990), *Interpretation as Abduction*, SRI International Artificial Intelligence Center Technical Note 499.
9. Johan, B., Delmonte R. (eds.) (2008), *Semantics in Text Processing (STEP)* in "Research in Computational Semantics", Vol. 1, College Publications, London.
10. Kiparsky, Paul, Kiparsky, Carol (1973 [1970]), *Fact* in Danny D. Steinberg and Leon A. Jakoboviîs (eds.). "Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology". Cambridge: Cambridge University Press, pp. 345-369.
11. Lochard, Guy, Boyer, Henri (1998), *Media Communication*, European Institute Publishing House, Iasi.
12. Osgood, C.E., Suci, G.J., Tannenbaum, P.H. (1957), *The Measurement of Meaning*, University of Illinois Press, Urbana, IL.
13. Osgood, C.E. (1959), *The representational Model and Relevant Research Methods* in I. De Sola Pool (ed.), "Trends in Content Analysis", University of Illinois Press.
14. Palmer, M., Weir, C., Passonneau, R., Finin, T. (1993), *The Kernel Text Understanding System* in "Artificial Intelligence", 63:17-68: Special Issue on Text Understanding.
15. Polanyi, L., Zaenen, A. (2006), *Contextual Valence Shifters* in Janyce Wiebe (ed.), "Computing Attitude and Affect in Text: Theory and Applications", Springer, Dordrecht, 1–10.
16. Romaine, Suzanne (1994), *Language in Society. An Introduction to Sociolinguistics*, Oxford University, Press Inc., New York.
17. Saurì, R., J. Pustejovsky, J. (2012), *Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text* in "Computational Linguistics", 38, 2, 261-299.
18. Schwartz, Gheorghe (2001), *Politics and Press*, European Institute Publishing House, Iasi.
19. Şeicaru, Pamfil (2007), *Press History*, Paralela 45 Publishing House, Pitesti.
20. Taboada, M., Brooke, J., Tofiloski., M., Voll, K., Stede, M. (2011), *Lexicon-based Methods for Sentiment Analysis* in "Computational Linguistics", 37(2): 267-307.
21. Touraine, Alain (1992), *Critique de la modernité*, Fayard, Paris.

22. Wiebe, J., Wilson, T., Cardie, C. (2005), *Annotating Expressions of Opinions and Emotions in Language* in "Language Resources and Evaluation", 39(2):165-210.
23. http://irmm.jrc.ec.europa.eu/
24. www.corriere.it
25. www.liberoquotidiano.it
26. www.repubblica.it.

Endnotes

1. J. Wiebe, T. Wilson and C. Cardie: *Annotating expressions of opinions and emotions in language*, Language Resources and Evaluation, 39(2):165–210 (2005).
2. L. Polanyi and A. Zaenen: *Contextual valence shifters.* In Janyce Wiebe, editor, "Computing Attitude and Affect in Text: Theory and Applications", Springer, Dordrecht, 1–10 (2006).
3. R. Delmonte: *Text Understanding with GETARUNS for Q/A and Summarization*, Proc. ACL 2004 – 2nd Workshop on Text Meaning & Interpretation, Barcelona, Columbia University, 97-104 (2004); R. Delmonte *et al.*: *Another Evaluation of Anaphora Resolution Algorithms and a Comparison with GETARUNS' Knowledge Rich Approach.* In: ROMAND 2006 – 11th EACL. Geneva, 3-10 (2006).
4. R. Saurì, J. Pustejovsky: *Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text*, Computational Linguistics, 38, 2, 261-299 (2012).
5. J.R. Hobbs, M. Stickel, D. Appelt, and P. Martin: *Interpretation as Abduction*, SRI International Artificial Intelligence Center Technical Note 499 (1990).
6. M. Palmer, C. Weir, R. Passonneau, and T. Finin: *The Kernel Text Understanding System.* In "Artificial Intelligence", 63: 17-68: Special Issue on Text Understanding (1993).
7. L. Polanyi and A. Zaenen: *Contextual valence shifters.* In Janyce Wiebe, editor, "Computing Attitude and Affect in Text: Theory and Applications", Springer, Dordrecht, 1–10 (2006).
8. Alain Touraine, *Critique de la modernité*, Fayard, Paris, 1992, p. 184.
9. Pamfil Șeicaru, *Press History*, Paralela 45 Publishing House, Pitesti, 2007, pp. 132-133.
10. Guy Lochard, Henri BOYER, *Media Communication*, European Institute Publishing House, Iasi, 1998, p. 6.
11. Umberto Eco, *Cinque scritti morali*, Bompiani (Romanian translation by Stefania Mincu, *Cinci scrieri morale*, Pontica Publishing House, Constanta, 1998), p. 52.
12. Gregory Derville, *Le pouvoir des médias. Mythes et réalités*, PUG, Grenoble, 1997, pp. 136-137.
13. Gheorghe Schwartz, *Politics and Press*, European Institute Publishing House, Iasi, 2001, p. 50.
14. B. Johan & R. Delmonte (eds.): *Semantics in Text Processing* (STEP), Research in Computational Semantics, Vol. 1, College Publications, London (2008).
15. http://irmm.jrc.ec.europa.eu/
16. Functional labels are limitted to a common set of 10, that includes: subject (SUBJect), object (OBJect), object 2 (OBJ2ect), complement X (XCOMPlement), complement S (SCOMPlement), connector (ADJunct), verb (ADVerbial), spare object (OBLique).
17. They are national newspapers, of general information, they come in tabloid form, with a few ten thousand pieces per issue, and the printed issue of the monitored articles can be found in full, on the Internet: www.corriere.it, www.liberoquotidiano.it, www.repubblica.it.
18. Inclussion of truth or of truth premises within a semantic theory of native languages find their support in the presence of *factive verbs* as well. (see Kiparsky, Paul and Carol Kiparsky (1973 [1970]). *Fact.* In Danny D. Steinberg and Leon A. Jakoboviîs (eds.), "Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology", Cambridge: Cambridge University Press, 345-369, such as: *know, observe*, that imply their objects' truth.