INTERNATIONAL JOURNAL OF ADVANCED COMPUTER AND COMMUNICATION TECHNOLOGY (IJACCT)
Rodolfo Delmonte and Daniela Gîfu, Vol. 1, Issue. 1, pp. 24-38, 2013

Opinion and Sentiment Analysis of Italian print press

Rodolfo Delmonte[1], Daniela Gîfu[2]

[1]Ca'Foscari University, Department Language Science,

Ca' Bembo, dd. 1075, 30123, Venice

[2] „Alexandru Ioan Cuza" University, Faculty of Computer Science,

16, General Berthelot St., 700483, Iaşi

*Corresponding Author Email address: daniela.gifu@gmail.ro

ABSTRACT

As it is known, 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. The system is multilingual and can produce deep text understanding. This has been done on some 500K words of text, extracted from three Italian newspaper 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.

INTRODUCTION

The aim of an interdisciplinary approach such as analysing the language of political discourse with NLP tools is to define and explain different discursive contexts, in this case, reflected by the online media. Most studies in this direction have mainly concentrated on three tasks: the first has to do with an emotional side, of how humans feel language. The second aimed at understanding the relationship between the linguistic utterance and the world; and the third, the role of linguistic structure of the language as a communication device. Linguistics has usually treated language as an abstract object which can be accounted for without reference to social or political concerns of any kind (Note 20).

Content analysis, which is based on (Note 14) and (Note 15) – but also (Note 24) -, requires an extremely laborious methodology for objective interpretations. In this paper, we discuss paradigms for evaluating linguistic interpretation of discourses as applied by the system GETARUNS (Note 1),(Note 4), (Note 5), which addresses the needs to restrict access to extra linguistic knowledge of the world by contextual reasoning. We focus on three aspects critical to a successful evaluation: creation of large quantities of reasonably good training data, lexical-semantic and syntactic analysis. We assume that in order to properly capture opinion and sentiment (Note 25) expressed in a text or dialog any system needs a deep 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, based on SentiWordNet (Sentiment Analysis and Opinion Mining with WordNet), by simply stemming a text

and using content words to match its entries and produce some result. 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 (Note 18) 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.

In order to cope with these linguistic elements we propose to build a propositional level analysis directly from a syntactic constituency based representation. We implemented these additions our the 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 (Note 3). 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 (Note 21). In particular we take into account: modality operators like intensifiers and diminishers, 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 rather than a classifier-based approach, i.e. we make a fully supervised analysis where semantic features are associated to lemmata and concept of the domain by creating a lexicon out of frequency lists. In this way the semantically labeled lexicon is produced in an empirical manner and fits perfectly the classification needs.

The paper is structured as follows. Section 2 shortly describes state of the art. Section 3 presents some information about the role of print press discourse, section 4 describes a system for multi-dimensional political discourse analysis. Section 5 discusses an example of 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 6 highlights interpretations anchored in our analysis and presents conclusions.

STATE OF ART

As we will see, one aspect of the platform that we present touches a lexical-semantic functionality, There is a large number of well-documented system in the literature which compare with GETARUNS, in particular TACITUS and KENEL. The design of Tacitus system is that the system, to the maximum extent possible, should not discard any information that might be semantically or pragmatically relevant to a full, correct interpretation. It performs a syntactic analysis of the sentences in the text, using a fairly complete grammar of English, producing a logical form in first-order predicate calculus. Pragmatics problems are solved by abductive inference in a pragmatics, or interpretation, component (Note 10).

Kernel's architecture is closer to ours in that syntactic, semantic and pragmatic tasks are segregated into separate processing modules but they are allowed to communicate. Kernel performs its analysis in two stages: first syntactic parsing "which has limited access to shallow semantic constraints for parse disambiguation" (Note 16) and second integrated semantic and pragmatic processing which has constrained access to external knowledge sources. However, syntactic processing is not itself performed by a context-sensitive semantically guided parser: it is basically a context-free grammar with restrictions, a grammar formalism called restriction grammar.

In turn each clause is then translated or mapped into a functional-like representation with attribute-value pairs called ISR. Semantic interpretation is performed while building up ISR and requires among other things recovering unexpressed constituents like subjects or non-obligatory prepositional phrases, as well as implicit but essential and sometimes obligatory arguments of a given verb predicate when used in its nominalised form. Noun phrase analysis in addition has a separate mechanism from clause analysis in that the former but not the latter allows for reference resolution. This requires a search for a likely discourse referent. Here comes another important limitation in the system: since each constituent is interpreted in the order in which it is logically built, there however is not yet available for use in the interpretation process. This rigidity of the system could be overcome in case the system could choose to delay reference resolution of all nouns as for instance in CANDIDE (Note 19). This system implements a theoretically similar approach to non-compositional interactions of semantic interpretation with pragmatic context in determination of noun phrase reference. It was designed and implemented to exemplify the acquisition of procedural knowledge expressed in a combination of flow charts and ordinary English discourse.

PRINT PRESS DISCOURSE

Mirror of contemporary society, located in permanent socio-cultural revaluation, 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, information and languages. This explains why the newspaper is, at the same time, the moral bulletin of temperature, and the notary service of all human acts, thus becoming the most complete archive of history (Note 23).

Print press discourse is perceived as an essential tool for exposure and free to any discussion of ideas (Note 12). For generating interest of modern potential customers in reading newspapers, the print press must renounce to self-citation and self-referentiality. Instead of arousing indignation concerned, this contrasted situation makes the game of a political actor, who finds useful to any statement made before a single medium, to echo to all other media united.

At present, the legitimacy of competence and credibility or reputation of political authority is increasingly in competition with media 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 media 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.

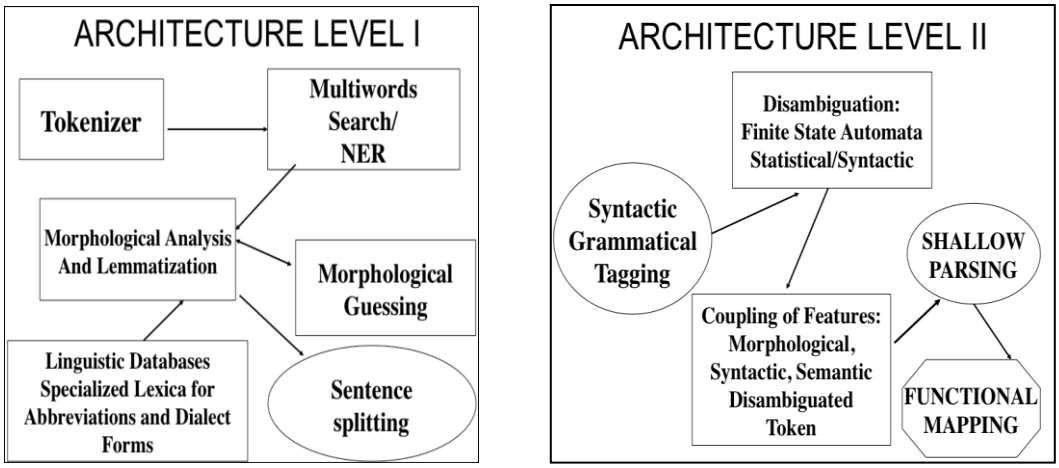
The huge amount of newsprint market removed should be filled with attractive texts for a public that no longer require anyone to buy newspapers (Note 22). No wonder why the people in power, if they intend to govern in peace, try to curb the enthusiasm of the media. Most of the times, through excellence in the elections, the print press is focused on topical issues, leading topics of public interest and events of internal and external social life (Note 8). 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).

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 a nd has been documented in the past 20 years or so with lots of publications and conference presentations, the most recent being (Note 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. This second version has been used for the RTE challenges and for TAC summarization tasks (Note 6) and (Note 7).

The "shallow" version of GETARUNS has been adapted for the Opinion and Sentiment analysis and results have already been published for English. 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, and as a result many entries are totally wrong. In addition, there was a need to characterize uniquely only those entries that have a generic 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, which had conflicting sentiment values. We also created 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, in particular with reference to the domain of political discourse.



a) b)

Figure 1. Italian GETARUNS – fist level one (a) and second level (b).

The system has been lately documented by our participation in the EVALITA (Evaluation of NLP and Speech

Tools for Italian) challenge (Note 6). It works in a usual NLP pipeline that we show in Figure 1. and 2. below. 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 center.1 Of course we also using our own list of entities and multiwords that we created in the years and that we derived from the participation in National Italian projects like ISST (1998-2002) and that produced the first treebank of Italian with the same name.

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. Other continuations are: inflection suffixes and then derivational suffixes; both prefixes and 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. Then sentence splitting takes place. The output is passed to second level computation that is documented in Figure 2. above.

The second level of computation takes as input one sentence at a time and tags each word with a tagset of 91 tags and a set of 31 syntactic constituency labels. Functional labels are limited to the usual set of 10 which includes: SUBJect, OBJect, OBJ2ect, XCOMPlement, SCOMPlement, OBLique, ADJunct, ADVerbial. Here XCOMP is used for all predicative complements, SCOMP for sentential complements.

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 the NER task. The remaining nominal expressions are classified using the classes derived from ItalWordNet (Italian WordNet). In addition to that, we have compiled specialized terminology databases for a number of common domains including: medical, political, economic, 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 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. Here below are some examples of vectors for common nouns 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 on the basis of our specialized semantic lexicon (Delmonte, 2007;2009). The vector of features includes morphological, syntactic, aspectual and semantic features, as shown below.

```
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 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 assignment 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

[1] http://irmm.jrc.ec.europa.eu/

the feature vector to produce a full-fledge semantic classification at propositional level (Note 2). The semantic mapping is exemplified below and includes, beside diathesis:

- Change in the World
- Subjectivity and Point of View
- Speech Act
- Factitiviy
- Polarity

Here are some examples where we list at first the verbal complexes and then the associated interpretation:

```
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]]])'
```

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 of text. In this paper we presents only syntactic and semantic results for our corpus.

A. 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 profiles[2].

For analytical results to be comparable to those taken so far by second author (Note 9), we needed a big corpus (Note 11), especially considering five rigorous criteria that we list below:

a. 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: *Corriere della Sera* - www.corriere.it (called The People Newspaper), *Libero* - www.liberoquotidiano.it (pro Berlusconi), and *La Republica* – www.repubblica.it (against Berlusconi).

[2] www.corriere.it, www.liberoquotidiano.it, www.repubblica.it

a. Period of time

The interval time chosen should be large enough to be 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.

- October 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 policy: (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.

Of course, this was not sufficient to select an appropriate number of articles for the study. So we had to reduce the number of articles per web page. We introduced a new subcriterion of selection: storing articles in the first three positions of each web page for every day of the researching period.

B. The syntactic and semantic analysis

In Fig. 2 and 3 below, we present comparative syntactic and semantic analyses of the texts extracted from the three Italian newspapers. We show differences on the graph which can assume both positive and negative values. In particular, values above the Ox axis mean they are more positive or higher than values below the Ox axis, which have a negative import. So for instance, Corriere, the blue or darker line, has higher positive values than the other two newspapers, with the exception of one case, the temporal slot we called PTMB where also Republica – green line, or lighter colour – has a higher number of tokens than Libero.

With one month before the Berlusconi's resignation (OMBB), we can highlight the sentences structure of the three dailies as follows: Corriere della Sera has a sentence structure with a rich vocabulary (words, verbal component). There are on average 21 tokens and 3 verbal compounds per sentence; on average 25% of all verbal compounds are subjective. Questions are 492, with only 1 exclamative.

Libero has on average 20 tokens and 2.4 verbal compounds per sentence; on average only 20% of all verbal compounds are subjective. Only 249 sentences are interrogative but 9 are exclamative sentences; more verbs are used with the passive diathesis.

La Republica has on average 18 tokens and 2.2 verbal compounds per sentence; on average only 25% of all verbal compounds are subjective. Only 196 sentences are interrogative and 8 are exclamative sentences; only 22 verbs are used with the passive diathesis.

In other words, Corriere has longer sentences, more questions but less exclamatives than other newspapers. Libero has a style with less subjective verbal compounds, that is it uses more factive verbal compounds and structures, more below on this topic. Repubblica uses shorter sentences than the other newspapers, remarkably less interrogatives.

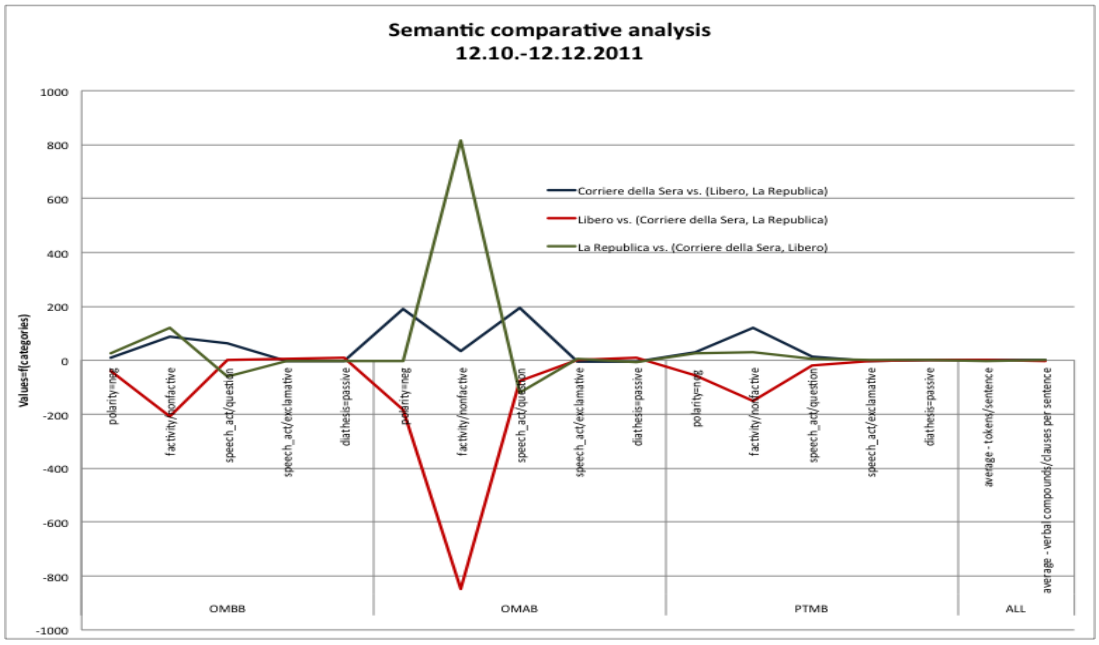


Figure 2. Comparative semantic analysis of three Italian newspapers.

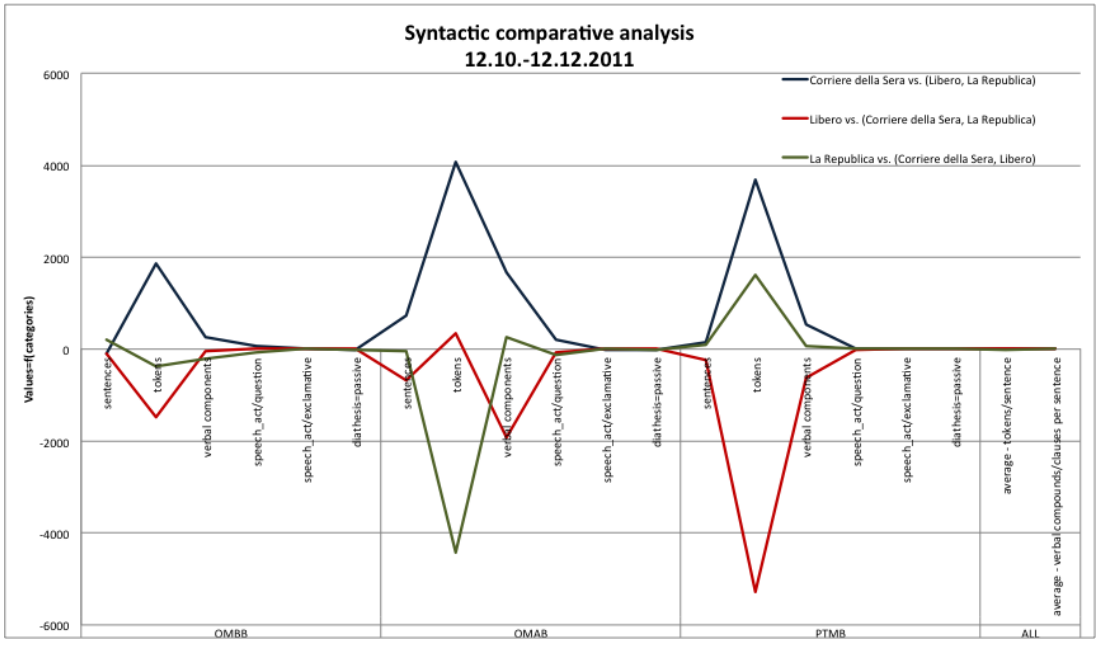


Figure 3. Comparative syntactic analysis of three Italian newspapers.

However subjective and factivity is distributed differently in the three time period.

- In the period Before Berlusconi's resignation, Repubblica has the highest number of nonfactive clauses, 1866, if compared to 1846 of Corriere and 1649 of Libero.

- in the Intermediate period, we see that Corriere has almost the double nonfactive sentences than the other two - 477, if compared to 267 of Repubblica and 145 of Libero;

- in the final time interval, After Berlusconi's Resignation, the data are reversed and converged with the first period: Repubblica has the highest number of nonfactive clauses, 2278, if compared to 1758 of Corriere and 1839 of Libero, which however now has a higher number than Corriere, quite strange, seen that in the previous two time intervals it always had the smallest number.

The reasons for this abrupt change of style in Libero, can only be connected to the uncertainty related to the nomination of Monti instead of Berlusconi, which is felt and is communicated to its readers as less reliable, trustable, trustworthy. More on this issue in the analyses below.

C. 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 5.1). For that, we have used the system feature of performing comparative studies. The values are supposed to reflect correctly the selected categories. In fact, our analysis makes a comparative analysis of the three newspapers mentioned. So, the graphical representation in Figure 5, in which the articles of three newspapers selected should be interpreted as above.

With one month before Berlusconi's resignation (OMBB), we can highlight the opinions of the three dailies as follows: Corriere della Sera is positively concerned mostly with Berlusconi (see Berlusconi occurrences) and it is interested in the organizational system (see words classified as ORG=organization), we also see a big dip in negative words; Libero, on the contrary, is more concerned with Monti (see Monti occurrences) and soaring negative discourse; while La Republica has a rather flat line, with soaring subjective point of view (see point_of_view is subjective), and mentioning economic issues more than others (see words classified as ECNM =economy).

Between Berlusconi's resignation and the nomination of new prime minister, Mario Monti (PTMB), we have the following situation: Corriere della Sera is concerned with both political actors, Berlusconi and Monti (see Berlusconi and Monti occurrences), but, also, the organizational system (see words classified as ORG=organization), rather flat negative tone and subjective point of view. Together with Libero they stand out (see soaring negative words and for Libero decreased subjective point_of_view and Corriere an unusual soaring subjectivity). La Republica has a rather flat attitude.

One month after the nomination of the new prime minister, Mario Monti, (OMAB) the discourses of the three newspapers analyzed return to their initial orientation: Corriere della Sera remains the most established on Berlusconi (see Berlusconi occurrences), with a negative tone (see negative words), interested by organizational system (see words classified as ORG=organization); the unsual aspect concerns the surge of the use of subjectivity. Libero is more concerned by Monti (see Monti occurrences) but has almost no economic orientation. La Republica on the contrary highlights economic issues (see words classified as ECNM =economy), but with higher subjective point of view (see point_of_view is subjective).

As said in the introduction, we were also concerned with measuring the overall attitude with positive words vs. negative words percentage for each newspaper analysed.

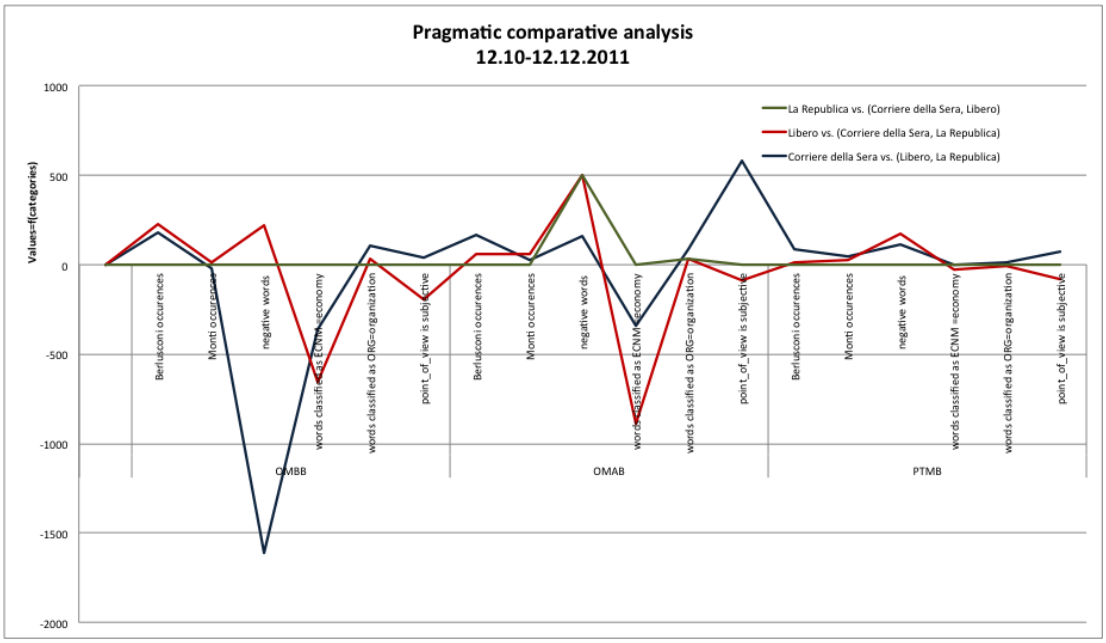


Figure 4. Comparative pragmatic analysis of Italian newspapers.

This work consisted in two phases:

1. We checked the automatically verbal mapping with GETARUNS system.
2. We mapped manually each verb, which wasn't mapped automatically.

In this sense, we improved our list of verbs from the system, which is essential for any kind of semantic analysis. Here below we propose a table with percentages of positive and negative words for each newspaper in the three periods of time indicated above.

The percentages from table 1, configure the emotional dimension of opinions in the three newspapers, as follow:

One month before Berlusconi's resignation (OMBB), both Libero and Repubblica have more positive contents than Corriere, which can be interpreted as follows: Berlusconi's Government is considered a good one; on the contrary, Corriere della Sera, has the highest percentage of negative opinions. The same evaluation applies for the intermediate period between Berlusconi's resignation and nomination of the new prime minister, Mario Monti (PTMB). 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 words are much less: the new prime minister seems a good chance for the Italian situation; however, Libero – the newspaper owned by Berlusconi - becomes a lot more negative than the others.

Table 1. Sentiment analysis of three Italian newspapers

Newspapers / period of time	Corriere della Sera		Libero		La Republica	
	positive words	negative words	positive words	negative words	positive words	negative 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%

At last, we were interested in identifying semantic linguistic common area (identification of common words) and the affective import of them (positive or negative).

We remark, from table 2, that all three newspapers use some words with strong negative import. 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. They are listed here below with absolute frequency for each time interval.

Table 2. Critical common words in two time peiods

Corriere della Sera	Libero	La Republica
O MBB		
crisis 124	crisis 71	crisis 94
sacrifice 4	sacrifice 14	sacrifice 94
rigour 5	rigour 4	rigour 4
austerity 0	austerity 6	austerity 6
battle 6	battle 12	battle 14
collapse 10	collapse 6	collapse 12
dissent 1	dissent 8	dissent 8
fail/ure 8	fail/ure 13	fail/ure 9
drama/tic 12	drama/tic 14	drama/tic 18
dismiss /al 45	dismiss /al 39	dismiss /al 20
dictator 2	dictator 10	dictator 18
OMAB		
crisis 50	crisis 21	crisis 110
sacrifice 9	sacrifice 23	sacrifice 16
rigour 23	rigour 18	rigour 10
austerity 6	austerity 2	austerity 0
battle 14	battle 4	battle 8
collapse 8	collapse 2	collapse 4
dissent 0	dissent 4	dissent 0
fail/ure 21	fail/ure 8	fail/ure 15
drama/tic 4	drama/tic 0	drama/tic 8
dismiss /al 3	dismiss /al 2	dismiss /al 15
dictator 2	dictator 6	dictator 2

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, and 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. Repubblica feels the opposite way. However, whereas "austerity" is never used by Repubblica after B.'s resignation and it was used before it, this is the opposite of what Corriere 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".

- as to the second half 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 and 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.

- the other interesting couple of concepts is linked to "battle, dissent, dictator". In particular, "battle" is used in the opposite way by "Corriere" 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 fashion: it disappears in both Corriere and Repubblica, and it is halved in Libero. Eventually the "dictator/ship" usually related to B. or to B.'s government, is a critical concept for Repubblica in the first period, and it almost disappears in the second one.

In order to better compare specialized keywords we carefully chose and reclassified a small subset of all lemmata – 100 concepts – using a subset of labels that were suggested by Linguistic Inquiry and Word Count (LIWC), a text analysis software program designed by James W. Pennebaker et al. (Note 17) that we bought and used in a previous experiment on Romania.

Table 3. Selected pragmatic features for three newspapers in three periods of time.

Newspapers/period of time	OMBB	PTMB	OMAB
LIBERO	sadness 11 achievements 24	sadness 11 achievements 24	rational 12 intuition 13
CORRIERE DELLA SERA	intuition 13 negative 8 uncertain 15 failures 25 work 23 social 2 positive 7 emotional 6	rational 12 intuition 13 anxiety 9 financial 28 work 23 social 2	anger 10 failures 25 social 2 positive 7 emotional 6
REPUBBLICA	anxiety 9 anger 10 inhibition 17 financial 28 negative 8	anger 10 inhibition 17 failures 25 positive 7 emotional 6	sadness 11 achievements 24 anxiety 9 inhibition 17 financial 28 negative 8 uncertain 15 work 23

The result of this new classification are highlighted here below, where we list for each newspaper the best

performance in term of number of occurrences, for the first 16 classes in a given time interval: the same conclusion can be now reached by noting that Libero has opposite attitudes to Repubblica, and this has opposite attitude to Corriere.

CONCLUSIONS

The analysis we proposed in this paper aims at testing if a linguistic perspective anchored in natural language processing techniques (in this case, GETARUNS system) could be of some use in evaluating political discourse in print press. If this proves to be feasible, than a linguistic approach would become a very relevant to and applicative perspective, with important effects in the optimization of the automatic analysis of political discourse. Rhetorical aspects (Note 13) to be highlighted in a journalist's prose include: the diversity and richness of the lexicon with a proper mastering of semantic classes (we accentuated the affective dimension, but not only that), syntactic structure, use of metaphoric expressions and the overall style. It is our conviction that the analysis above confirms the ability of current linguistic technology to successfully cover many of these facets.

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 as being pro current government and 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.

Different journalists could raise the use of these measures to the level of a rhetorical strategy, therefore exploiting them perhaps too much to the benefit of some aimed goals. 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.

Many interpretation facets are pertinent to the specific context in which a discourse is being uttered. For instance, in a tense political context, discourses should be evaluated in function of a balance between the journalist's agenda vs. the agenda of the political actor with respect to the public agenda.

Different intensities of emotional levels has been clear 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).

We are aware that many technological aspects have yet to be refined and enhanced. One of the most important is the determining the sense of ambiguous words and expressions in context. In the future, we intend to include a word sense disambiguation module in order to individuate the correct senses, 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 the GETARUNS system 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, J.& Delmonte, R. (eds.), (2008) "Semantics in Text Processing (STEP), Research in Computational Semantics", vol.1, College Publications, London.

[2] Delmonte, R. (2004) "Text Understanding with GETARUNS for Q/A and Summarization", Proc. ACL 2004 - 2nd Workshop on Text Meaning & Interpretation, Barcelona, Columbia University, 97-104.

[3] 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, pp. 3-10.

[4] Delmonte, R. (2007) "Computational Linguistic Text Processing – Logical Form, Logical Form, Semantic Interpretation, Discourse Relations and Question Answering", Nova Science Publishers, New York.

[5] Delmonte, R. (2009) "Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution", Nova Science Publishers, New York.

[6] Delmonte, R., S. Tonelli, R. Tripodi (2010) "Semantic Processing for Text Entailment with VENSES", published at http://www.nist.gov/tac/publications/2009/papers.html in TAC 2009 Proceedings Papers.

[7] Delmonte, R. and Vincenzo P. (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.

[8] Gîfu, D. (2011) "Violenţa simbolică în discursul electoral" (Symbolic violence in electoral discourse), Ed. Casa Cărţii de Ştiinţă, Cluj-Napoca.

[9] Gîfu, D. and Cristea, D. (2012) "Multi-dimensional analysis of political language", in James J. (Jong Hyuk) Park, Victor Leung, Taeshik Shon, Cho-Li Wang (eds.) In Proceedings of The 7th FTRA International Conference on Future Information Technology, Application, and Service – FutureTech-2012, Vancouver, 26-28 June, vol. 1, Springer, pp. 213-221.

[10] Hobbs, J. R., Stickel, M., Appelt, D., and Martin, P. (1990) "Interpretation as Abduction", SRI International Artificial Intelligence Center Technical Note 499.

[11] Kennedy, G. (1998) "An Introduction to Corpus Linguistics", London & New York: Longman.

[12] Lochard, G., Boyer, H. (1998) "Comunicarea mediatică" (Communication media), Institutul European, Iaşi.

[13] Mann, W.C., Thompson, S.A. (1988) "Rhetorical Structure Theory: Toward a functional theory of text organization". In Text 8(3), 243-281.

[14] Osgood, C.E., Suci, G. and Tannenbaum, P. (1957) "The Measurement of Meaning", University of Illinois Press, Urbana, IL.

[15] Osgood, C.E. (1959) "The representational model and relevant research methods". In De Sola Pool I. (Ed.), Trends in Content Analysis. University of Illinois Press.

[16] Palmer, M., Weir, C., Passonneau, R., and Finin, T. (1993) "The Kernel Text Understanding System". In Artificial Intelligence 63: 17-68: Special Issue on Text Understanding.

[17] Pennebaker, J.W., Booth, R.J. and Francis, M.E. "Linguistic Inquiry and Word Count (LIWC)", at http://www.liwc.net/.

[18] Polanyi Livia and Annie Z. (2006) "Contextual valence shifters". In Janyce Wiebe, editor, Computing Attitude and Affect in Text: Theory and Applications. Springer, Dordrecht, 1–10.

[19] Pollack, M., Pereira, F. (1991) "Incremental interpretation". In Artificial Intelligence 50, 37-82.

[20] Romaine, S. (1994) "Language in society. An Introduction to Sociolinguistics", Oxford University, Press Inc., New York.

[21] Saurì R., J.Pustejovsky, (2012) "Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text", Computational Linguistics, 38, 2, 261-299.

[22] Schwartz, G. (2001) "Politica şi presa" (Politics and media), Institutul European, Iaşi.

[23] Şeicaru, P. (2007) "Istoria presei" (History of press), Ed. Paralela 45, Piteşti.

[24] Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M. (2011) "Lexicon-based methods for sentiment analysis". In Computational Linguistics 37(2): 267-307.

[25] Wiebe, J., Wilson, T. and Cardie, C. (2005) "Annotating expressions of opinions and emotions in language". Language Resources and Evaluation, 39(2), pp.165–210.