

# **A textual analysis of the lexicon of general policy in an election of the Rector through the electoral speeches of the five candidates**

Enrica Aureli<sup>1</sup>, Silvia Bartoletti<sup>2</sup>

<sup>1</sup>Università di Roma "La Sapienza" – Dip. Statistica , Probabilità e Statistiche Applicate –  
Roma – Italia

<sup>2</sup>Via Tropea 41 – 00178 Roma – Italia

## **Abstract**

In this paper we aim to provide a brief description of the lexical universe of electoral discourse, presenting a method of textual analysis. The corpus contains the policy statements of the five candidates aspiring to become Rector of 'La Sapienza' University, Rome. We aim to use automatic techniques to pick up the themes of the texts, and the positions that the candidates hold on these subjects, by means of "forms-texts" tables.

**Keywords:** vocabulary analysis, lexical factor analysis, textual statistics.

## **1. The events**

In the months of September-October 2000, 'La Sapienza' University, Rome, was busy with the elections of the Rector who was to guide the University over the next four-year period 2000-2004.

The candidates standing for election were the outgoing Rector, OR, and four competing candidates whom we shall denote with the initials AS, BM, CI and DL.

This was a hard-fought election which saw four determined challengers united against OR (in office since '97), forming alliances in which the policy statements of CI and BM appeared to converge in sharper contrast with OR, but which split up in the light of the results of the subsequent rounds of elections, at which point CI joined forces with OR, effectively securing the latter's re-election.<sup>3</sup>

---

<sup>3</sup> The elections also took place in an atmosphere of total uncertainty from a procedural point of view, regarding the electoral body. Initially it seemed that the active electorate was to consist solely of teaching staff, but OR defended the new, but suspended, statute, whereby the whole population of Rome's largest university was entitled to vote.

Just when the outgoing candidate looked beaten in the first round of elections, news arrived that researchers, non-teaching staff and student representatives should also have voted, which led to a complete alteration of the size and preferences of the electorate.

## 2. The methods

Faced with this shifting ground of alliances, our aim was to test whether a textual analysis of the policy statements might disclose or help predict the final results of the elections.

We therefore began with a lexical study of the language used in the formulation of the policy statements, identifying each of the candidates OR, AS, BM, CI and DL with a lexical profile.

When an author formulates his own text, he uses not only his own particular language of reference, but also his own particular capacities of expression; for the purposes of linguistic statistics, each part of such a text will therefore be regarded as a sample of the entire corpus, and each text will be a sample of the language of its author. Moreover, since the statistical calculations are carried out on data taken from texts, every time a calculation suggests a conclusion about the language, we reason by inference (Rizzi 1992).

This evidently pre-supposes a strong hypothesis: electoral discourse is a theoretical and empirical object, possessing its own characteristics, which implies the possibility of carrying out a formal description of the genre by means of its universes and domains of variation.

It will be necessary to dwell upon certain points:

- intertextual stability - this has to do with the basic vocabulary, the written forms which are present in all the discourse profiles observed with a relative frequency, which is seen to be constant;
- intertextual discrimination - the words which are distinctive and characteristic precisely of each executive-policy profile;
- intratextual stability - this makes it possible to obtain an image of the "lexical habits" of the speaker, with the recurring factors within his discourse;
- intratextual discrimination - these are "localized" terms, adopted by the orator exclusively in relation to a certain written form.

## 3. Processing of textual materials and characteristics of the corpus

The body of texts (the corpus) consists of the electoral speeches of the candidates standing for election to the office of Rector, as they were circulated around 'La Sapienza' University in printed form; they are taken from the websites on which they were published.

We deemed it appropriate to analyse the entire corpus in order to get an overall picture of the subjects constituting the basis for discussion regarding university policy.

For the processing of the data, the SPAD-T software was used, a package designed for the statistical analysis of a text by means of a lexicometric approach; this recognizes every word as a written form, ie as an item the main feature of which is its written appearance ("ne" being different from "né", for example) and bases its calculations:

- on the entire lexical table for the observed corpus;
- on the frequency of occurrence of the words within the corpus;
- on the comparison between the lexical "profiles" of the various parts of the corpus.

Significant considerations are usually generated by observation of:

- nouns;
- main verbs conjugated in the first person, singular and plural;
- possessives in the first person, singular and plural;
- the main nouns associated with possessive adjectives in the first person, singular and plural.

(Duchastel, Armoni, 1993)

### 3.1. Nouns

From an initial analysis of the corpus in its entirety, it would appear that the nouns used most often are those concerning the main subjects under debate: <ateneo>, <università>, <facoltà>, <medicina>, <policlinico>, <ricerca>, <didattica>, <attività>, <servizi>, <strutture>, <dipartimenti>, <amministrazione>, <personale>. These nouns refer to the main subjects dwelt upon by the candidates: they are written forms which insinuate concrete statements, but which already clash, in an initial, brief analysis, with the low occurrence of <costi>, <budget>, <prezzo>, <deficit>, <retribuzioni>, <tasse>, <debito>. This highlights the general reluctance to tackle subjects which might remind the electorate of unresolved controversies and financial bones of contention. There is also a low occurrence of the following: <infrastrutture>, <parcheggi>, <docenti>, <ricercatori>, <matricole>, <dottorandi/to>, <impianti>, which would highlight the specific subjects of a workable policy statement. There is an attempt to remedy the low occurrence of any such "tangible" nouns, probably all connoted by the generic <problemi>, with the inappropriate and exaggerated use of <sviluppo>, and <progetto>, also generic expressions which do not remotely describe the policies in practical terms. Almost absent is <disoccupazione> (does this represent some kind of alienation from reality? Shouldn't the university interact with the world of work and make at least some attempt to tackle this issue?), along with nouns symptomatic of the desire to modernize and keep up with the times, such as <computer> and <informatizzazione>. There is little mention of the imminent <elezioni> of the Rector, or explicit request for a <voto>. The parsimonious use of <costo> and the absence of <tassazione> are in line with the vote-winning intention of the general-policy statement, as are the low frequencies of nouns with "negative meanings": <disservizi>, <errori> and <abbandono>, referring mainly to the student population, which is not in itself treated as an important audience. From this analysis it would seem that we obtain the image of a typical electoral campaign, conducted on the same, worn subjects and lacking in any critical reflection about the past or in any concrete suggestions for the future.

La Sapienza	211	Sistema	56	Risparmi	1	Progettazione	2
Facoltà	157	Rettore	56	Insuccesso	1	Collettività	2
Ateneo	124	Sviluppo	53	Calcolatore	1	Voto	2
Università	115	Processo	51	Computer	1	Manutenzione	3
Ricerca	104	Medicina	47	Deficit	1	Matricole	3
Strutture	101	Progetto	46	Disoccupazione	1	Impianti	3
Policlinico	98	Esigenza	45	Macchinari	1	Abbandono	4
Didattica	72	Società	45	Musei	1	Elezioni	4
Attività	69	Studente	41	Retribuzioni	1	Errori	4
Problemi	67	Atenei	40	Inefficienza	1	Disservizi	4
Dipartimenti	66	Assetto	39	Cooperazione	2	Costi	5
Stato	66	Amministrazione	37	Equilibrio	2	Debito	5
Risorse	64	Autonomia	36	Tasse	2	Dottorato/ndo	5
Servizi	59	Ruolo	34	Ideologia	2	Prezzi	5
Personale	59	Rispetto	34	Parcheggio	2	Infrastrutture	6

Table 1 - Frequency of written forms (nouns used most and least)

### 3.2. The verbs

The predicates include: verbs of state/condition (be and have: these indicate a property or possession); verbs of action (do, which indicates an action, the realization or production of an object or of an effect); the declarative verbs (say and think, are the declaration of an action, object or sentiment. They make it possible to express a judgment or evaluation (good/bad, true/false)). The analysis of the verbs seems to offer more space to the contents.

The verbs are characterized by size of consensus, winning of trust, assurances of improvement:: <occorre>, <realizzare>, <effettuare>, <fare>, <garantire>, <organizzare>, <potere>, <prevedere>, <presentare>, <programmare>.

As usual, there seems to be no place for verbs referring to anything more realistic, like <fronteggiare> or <concretizzare>. The same is true of another "forceful" verb, <imporre>, and of other verbal predicates with negative connotations: <screditare> e <peggiore>.

Avere	189	Agevolare	1
Essere	181	Arricchire	1
Potere	91	Concretizzare	1
Occorre	88	Screditare	1
Dovere	77	Imporre	3
Realizzare	71	Rispettare	3
Assicurare	41	Peggiorare	3
Garantire	42	Informatizzare	5
Trovare	41	Partecipare	5
Affrontare	38	Interagire	6

Table 2 - Verbs with high and low occurrence.

### 3.3. Adjectives

The adjectives fall into two classes: that of strictly general policy (<potenziale>, <programmatico>, <ampio>, <strutturale>, <pubblico>); and that referring to specific subjects (<universitario>, <sanitario> etc). Conspicuous by their absence are: <multimediale>, <multidisciplinare>, <qualificato>.

There is also significant use of the first person plural of verbs, and the repetition of the pronoun <noi> and of the adjectives <nostro> and <nuovo>, in order to indicate a personalization of commitments, at the same time hinting at the existence of a support group.

Nostro	90	Decisionale	25
Nuovo	89	Scientifico	24
Universitario	73	Internazionale	21
Amministrativo	61	Istituzionale	18
Necessario	54	Multimediale	1
Culturale	47	Multirazziale	1

Table 3- Frequency of the most significant adjectives

From this initial examination it would appear evident that the discourse is designed not so much for the purposes of communication as in order to win consensus. (Bolasco, 1996).

#### 4. Segmentation of the corpus into five areas of general-policy interest

It was therefore decided to test conformity and dissimilarity in the recommendations made in the general policy declarations expressed by the five rectoral candidates.

The entire corpus was therefore divided into five main areas of importance, ie decongestion and construction ("decong" in tab 4), reorganization of the bureaucracy ("bureauc" in tab 4), hospital staff and Faculty of Medicine ("hospital" in tab 4), financial situation ("financ" in tab 4), the role of 'La Sapienza' as a key cultural player ("role cult" in tab 4). (Santini, 1996)

These areas emerged in an explicit manner from the reading of the rectoral speeches, and the corpus was easily divided into the following 5 sub-corpora, which had already been partly identified with the division of the policy statements into paragraphs.

Where such a sub-division was not evident, we proceeded ourselves with the segmentation, with very little risk of error, because the subjects dealt with could easily be inferred from the reading

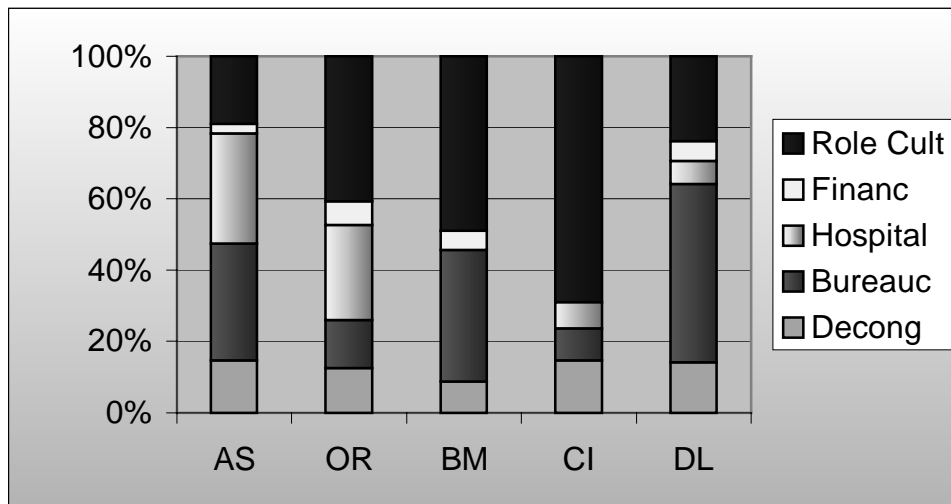
These areas include most of the significant declarations of policy. The analysis of the content of the five main areas has a precise aim: that of highlighting the prescriptive features particular to each candidate in each of the selected areas, by means of the lexical space devoted to each theme and the specific nature of the forms characterizing the policy commitments of each rectoral candidate.

The area around the "role of 'La Sapienza' as a key cultural player" is the broadest and most significant, in that it includes themes which are not common to all five candidates: not the classic "big issues" of 'La Sapienza' which we have already mentioned, but more person issues, which make it possible to identify and characterize the specific personality and recommendations of each candidate.

	Decong	Bureauc	Hospital	Financ	Role Cult	
AS	2274	5062	4755	410	2933	1543 4
OR	655	698	1397	351	2119	5220
BM	185	787	0	114	1038	2124
CI	439	270	217	0	2061	2987
DL	434	1530	196	172	728	3060
	3987	8347	6565	1047	8879	2882 5

Table 4 - Number of written forms, subdivided by candidate and subject

What is notable in the corpus overall is the greater verbosity of AS, whose policy statement makes up over half the entire corpus, followed by OR, whose a policy statement is expressed in 5220 written forms, equal to less than 20% of the total. CI and DL exhibit around 3000 written forms, while the policy statement of BM is the most synthetic, with only 2124.



Graph 1 - Percentage of written forms devoted by the 5 candidates to each area

Whereas the length of each candidate's speech is a function of their capacity for expression, the relative space devoted to each subject demonstrates the interest and importance attributed to it by them. The interest of the 5 candidates and the focus of their policy statements emerges clearly from this distribution. Indeed, Graph 1 shows clearly that the distribution of the policy statement over the 5 areas is once element which greatly distinguishes the candidates.

The only homogeneous point at which these concur is over decongestion, which registers an almost uniform incidence in the election speeches of all the candidates (with its lowest incidence in that of BM).

The problem of decongestion is most evident, but it is also dangerous ground: while it has now been established by law that 'La Sapienza' must be divided up into a federation of universities, there is no real consensus as to how to implement this legislation. It is therefore necessary to refer to this without entering into any specific details.

In the speeches of OR and AS, it is questions of a financial nature which prevail, compared the other candidates, but whereas these are probably related to the reorganization of 'La Sapienza' for OR, bureaucratic questions are always of central concern to AS.

CI bases his policy statement mainly on the new role to be played by the capital's largest university, relegating the other electoral questions to pure routine discussion; BM seems to be the closest to him, also paying great attention to aspects of a bureaucratic nature. These, on the other hand, constitute the central plank of DL's policy statement.

## 5. Lexical factor analysis

Lexical factor analysis (LFA) provides us with a visualization of the distance between individuals and forms, thus making it possible to identify the forms and/or expressions which differentiate them (Rizzi, 1992).

The textual forms observed are used for the construction of the contingency matrix necessary for any application of simple or multiple correspondences; this matrix may be any one of three different types, depending on the nature of the data and the applications chosen (Lebart, Morineau et al. 1993).

The first type of table we used is obtained by crossing a lexical variable (words) with texts. In this way we obtain a "words by texts" table, the term for which -  $m(i, j)$  - indicates the number of times the word  $i$  is present in the text  $j$ . The analysis of the correspondences of these

matrices makes it possible to obtain a visual representation the proximity/distance between words and texts, through the projection of factor planes, according to the type of analysis carried out.

In particular, we will use the analysis of simple correspondences, applied to lexical tables of the type <<forms-texts>>, generated in correspondence to each area of interest. (Lebart, Salem 1994). The forms represent the vocabulary measured in correspondence to each general-policy area, and the texts represent the five candidates (see LFAs 1-5).

The analysis of the correspondences makes it possible to synthesize the information contained in each lexical table, highlighting associations and contrasts between forms and candidates, and identifying an optimal order of forms, and the relative position of each candidate within this order. For the analysis of the simple correspondences to be applied correctly, it is necessary not only for the corpus to be sufficiently rich in written forms, but above all for the forms to appear with a certain frequency, in order to avoid calculations which are weak, or hard to interpret (Bolasco, 1995). For this reason, it was deemed necessary in subsequent analyses to maintain only those forms exhibiting a frequency of three or more within each general-policy area: the selected threshold thus reduced the size of the lexical tables, but it also increased their informative capacity.

We therefore proceeded, remembering to take the following precautions:

- elimination of the words of low absolute contribution and of meaning deemed to be too general for the evaluation of the content of the discourse;
- graphical representation of the words on (F<sub>1</sub>, F<sub>2</sub>).

The analysis was performed on the reduced corpora relative to the five thematic areas, so as to establish the lexical vicinity of the policy statements themselves, notwithstanding the different numerical size of the relative texts of each candidate.

ministero prospettive spese AS miliardi ateneo università	settori finanziamento	problema struttura lavoro informatizzazione istituzionale	rettore DL decisioni
	ricerca DL BM finanziaria		strutture amministrativo politica didattica
conto bilancio	risorse	costituzione AS progetti laurea uffici realizzare	personale forte CI impegno accademico stato studenti senato
stato	situazione	esigenze informatico culturale segreterie	interventi efficacia amministrazione governo garantire controllo
obiettivi dipartimenti	OR policlinico	coordinamento	facoltà statuto servizi BM realizzazione risultati adozione gestione OR efficienza risorse azioni legge rafforzamento indirizzo

LFA 4 - financial situation

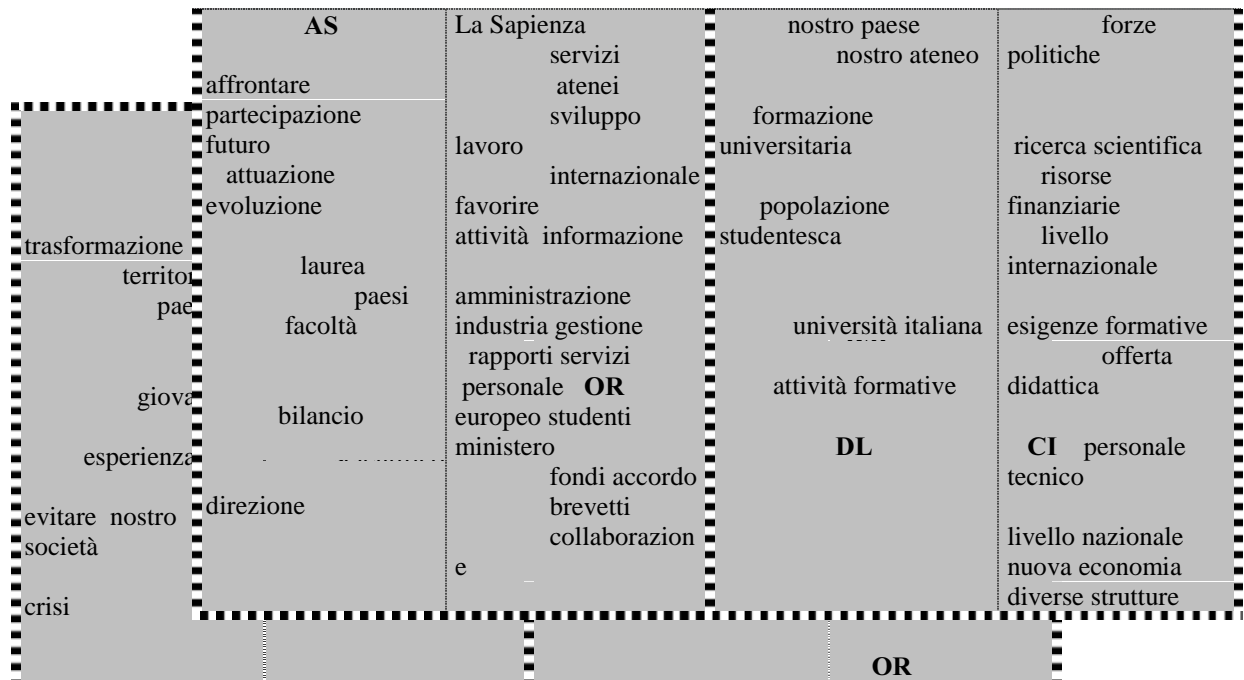
LFA 2 - reorganization of the bureaucracy

didattica ricerca culturale	realizzare ateneo	popolazione	<b>DL</b> assicurare docente
<b>BM</b> scientifico	<b>AS</b>	esigenze didat	posti tica letto
autonomia <b>CI</b> facoltà risorse	strutture edilizia territorio	umbertoI policlinico	strutture s.andrea ricerca
dipartimenti	scienza	<b>AS</b>	
sviluppo	interventi <b>DL</b> attività	medicina rispetto	assistenza interventi autonomia direttore <b>OR</b>
progetto	struttura deco	assetto responsabilità	futuro legge murst
università riorganizzazione	ngestionamento	gestione ateneo	problema
	problemi <b>OR</b> accordo		<b>CI</b>

*LFA 1 - decongestion and construction*

*LFA 3 - hospital and Faculty of Medicine*





LFA 5 – role of 'La Sapienza' as key cultural player: written forms.

LFA 6 – role of 'La Sapienza' as key cultural player: repeated segments.

We performed an analysis of lexical type for all the areas, thus based on the individual written forms. In the case of "role of 'La Sapienza' as key cultural player", however, we also used an LFA of the type <<repeated segments by texts>>, bearing in mind that the complexity of the subject might better emerge through the main semantic constructs. The latter are able to convey a much broader autonomous content than what can be obtained from the observations of the frequencies of the individual written forms of which they are composed.

As for the written forms, it is also possible to fix frequency thresholds. In this case all the segments were examined, with a view to using only those featuring greater frequencies.

The first problem is to operate a selection on the segments obtained, without limiting the choice to lexicometric criteria. The real, and perhaps the only problem inherent in using repeated segments is the question of their significance. Frequency on its own is not a suitable criterion for pinpointing the most significant elements. This problem can be overcome by effecting a selection on the basis of the criterion of <<grammatical completeness>> (Morrone, 1993).

This makes use of the distinction between full grammatical forms, treating the function words, ie those more indicative of relations than of contents (articles, prepositions and conjunctions) as merely instrumental. Starting from the consideration that these forms perform the function of connecting the textual fabric, and only have meaning when combined with the main forms, we can drastically reduce the problem of redundancy caused by the numerous sequences which are grammatically complete.

### 5.1. The discriminating capacity of the procedures

Our aim was to extract significant elements from the totality of rectoral-speech messages in order to highlight the specific natures of the individual electoral policy statements, over and above their explicit declarations, using the tools made available by Statistics.

What effectively emerges is that each candidate has his own particular characterization.

OR and AS appear as the most strongly characterized candidates, with clear and sharply defined positions, while BM, CI and DL are harder to define in simple terms. They do not follow the same behavior within the whole rectoral speech; depending on the subject matter, they are sometimes close to each other and sometimes very distant; sometimes they are close to OR, and on other occasions they are close to AS.

That of AS is the most "political" statement, due to its many concrete references to problems of a public nature: a cultural-economic matrix emerges whereby the point of reference for the indication of policy statements always takes account of aspects of funding and management.

OR is diplomatic: he must answer the criticisms of the other candidates, taking account of his past actions during his previous term of office; for the same reason, he is very respectful of the bureaucracy which has set the limits to his management. He is concrete, but in accordance with action that has been initiated.

OR's term of office seems to isolate him from his fellow candidates as regards all the challenges for the future: past management (financial area), decongestion and construction policy, and also the cultural role of 'La Sapienza'. It therefore emerges how opposition regarding these central arguments has been main reason for the proposals for a change of leadership and the existence of an unusually high number of candidates.

CI, on the other hand, fully demonstrates his openness towards civil society, associating himself with <<politica>>, <<progresso>>, <<nazionale>>, <<governo>>, <<valorizzare>>, <<città>>.

In the LFAs on the repeated segments, the vicinity map appears to be different: the <<attività>> and the <<esigenze formative>> of the <<popolazione studentesca>> locate DL and CI in the same area; <<percorsi formativi>>, <<nuovi corsi>> and <<aggregazione culturale>> associate AS and OR, indicating their awareness of the obligation to implement the university reform with reference to the contextualization of these initiatives with respect to <<nostro paese>> and <<nostro ateneo>> (AS) or with respect to the <<forze politiche>> (OR).

BM finally, refers to the <<Unione Europea>> and to the <<realtà professionale>> as points of reference for any reform touching specifically upon the medical profession.

As for the subject of the reorganization of the bureaucracy, textual analysis highlights a closeness of BM and OR around the forms: <<statuto>>, <<realizzare>> and <<risultati>>. This is a result of their joint responsibility in drafting the new university statute, to which BM contributed, while the other three candidates adopted critical positions with regard to this.

One particular feature emerges when we compare the results of the LFAs for the lexical area around the "cultural role of 'La Sapienza'", in terms of both written forms and also repeated segments. In the former, OR always appears in contrast with all the others, overwhelmed by his experience of management, and indeed the written forms close to him are: <<ministero>>, <<studenti>>, <<personale>>, <<rapporti>>, <<servizi>>, and immediately afterwards, <<gestione>>, <<bilancio>> and <<accordo>>, all terms which demonstrate his level of awareness of the objective limits to his formulation of policy.

The written form characterizing AS is <<crisi>>, highlighting a plan critical of the previous management and the idea of making use of <<partecipazione>> in order to <<affrontare>> the <<futuro>>.

BM and DL are strangely close in their expression of cultural recommendations, making use of the key terms <<didattica>>, <<coinvolgimento>> and <<scientifico>>, horizons confined within the university, more open to <<cultura>> in the case of BM, and more concerned with practical details (<<biblioteca>>) in the case of DL.

## 6. Conclusions

All this may be regarded as symptomatic of the current day and age, in which the "politically correct" attitude tends to dominate; this is corroborated by the "neutral" formulation of the speeches, which only adopt personal language at the points of sharpest criticism, and by the certainly intentional omission of themes which it would otherwise be logical to raise, in so far as they fall into the relevant sphere of responsibilities, but which - in contrast - are carefully avoided as their specific nature would arouse a mixed response in the public. This is true of BM, with his omission of "hospital and Faculty of Medicine" (BM, Dean of the Faculty of Medicine, does not expound his point of view in this respect so as not to raise unpopular arguments which might attract opposition; indeed, many are fearful of the excessive power of Medicine, which is already the predominant faculty in every sense, so a rector originating from there might be looked upon with suspicion); it is also true of CI, who avoids any mention of the "financial situation".

## References

- Bolasco S. (1995). Criteri di lemmatizzazione per l'individuazione di coordinate semantiche. In Cipriani R., Bolasco S. (a cura di). *Ricerca qualitativa e computer* Franco Angeli. Milano.
- Bolasco S. (1996). Il lessico del discorso programmatico di governo. In Villone M. Zuliani A. (a cura di). *L'attività dei governi della Repubblica Italiana (1948-1994)*. Il Mulino. Bologna.
- Duchastel J., Armoni V. (1993). Un protocole de description de discours politique. *Jadt 1993*. Montpellier, Paris : Telecom.
- Lebart L., Morineau A. Becue M. (1993). *Spad-T version 1.5. Manuel de l'utilisateur*. CISIA.
- Lebart L., Salem A. (1994). *Statistique textuelle*. Dunod. Paris.
- Morrone A. (1993). Alcuni criteri di valutazione della significatività dei segmenti ripetuti. *Jadt 1993*. Montpellier, Paris: Telecom.
- Rizzi A. (1992). Orientamenti attuali della statistica linguistica. *Statistica*, anno LI, n. 4.
- Santini I. (1996). Caratterizzazione dei governi dal 1947 al 1994 attraverso la lettura dei documenti Programmatici. In Villone M., Zuliani A. (a cura di). *L'attività dei governi della repubblica italiana (1948-1994)*. Il Mulino. Bologna.

