

A fight in words: electoral speeches during the Vice Chancellor election in an Italian university

Lorenzo Bernardi¹ and Arjuna Tuzzi²

¹ Dip. Scienze Statistiche,
Università di Padova
lorenzo.bernardi@unipd.it

² Dip. Studi Storici e Politici,
Università di Padova
arjtuuzzi@stat.unipd.it

Abstract. Do ideas and the language used to express them always assure the successful outcome of an electoral competition? In order to answer this question we analysed the candidates' electoral speeches during the campaign for the election of the Vice Chancellor of a university. The main results show that the language and the style chosen to introduce the few pertinent subjects are very important.

1 Introduction

Which is the role of language in determining an electoral success? This is an uncomfortable question if the aim is to examine the context of a National electoral position, because the subjects on debate are many as well as the medias and the actors in the arena. In order to obtain some indications, we decided to analyse a simpler context: the candidates' electoral speeches during the campaign for the election of the Vice Chancellor (*Magnifico Rettore*) of an important Italian university¹. In fact in this case the electoral speeches concern few pertinent subjects and a very limited context, hence it should be easier to recognize the fundamental subjects of the competition, the importance of the authors' style and of the language used to express ideas.

As prescribed by the local rule, every three years the investigated university elects the new Vice Chancellor. During the electoral campaign assemblies *ad hoc* are organized and documents containing the candidates' programmes are presented. The variety of relevant subjects depends on the political and historical period of the competition, though there is always a set of core-subjects typical for an academic context: teaching work, students' services, resources for the development of scientific research, *etc.* Each candidate weights tradition and innovation in different ways and has his own opinions about how theoretical principles have to be arranged in a real context.

¹ Applications of textual techniques to electoral speeches can be found in Bolasco (1996), Carbone (2000), Alvarez *et al.* (2002), Aureli and Bartoletti (2002).

2 The corpus examined

The speeches of the four candidates (in this work we will refer to these texts using the letters: A, B, C and D) are written using different styles, logics and structures. A, B and D are written as letters, the candidates speak directly to the electors (teachers, staff-personnel, students' delegations) and, in order to simplify the reading, the texts are organized in steps. These three speeches contain in the last part a brief candidate's *curriculum vitae*. Differently, C is a schematic programme, directly written in fifteen steps. Moreover, it presents no *curriculum vitae* at the end. In this study we have decided to consider only texts given to electors at the official assembly and to exclude the *curriculum vitae* and all other forms of propaganda used during the electoral campaign (newspapers articles, web pages, *etc.*).

Candidate-A opens his own speech with a slogan as a title and ends with another strong, though banal, slogan: the historical motto of the object-university. Candidate-B writes a slogan at the start of his own speech as a dedication. Just like Candidate-C, but the latter chooses a shorter and more effective opening than the preceding and this *incipit* goes on with the title of the programme. Candidate-D is the only one who doesn't produce a slogan. In order to make up for this lack he writes a preliminary remark: a second brief letter attached to his own speech in which the three main statements of the programme are summarized. This first observation concerning the presence and the substantial character of a slogan gives information on the authors' style. Surely A and B are more rhetorical texts and C and D more pragmatic. Candidate-C bets more on a sensational image and brief presentation while Candidate-D opens slowly and with a modest low tone.

Before proceeding further with statistical analysis we have performed a normalization of the texts in order to obtain the candidates' word types as homogeneous as possible. This means we have worked preliminary on the use of capital letters, punctuation marks, accents, acronyms, *etc.* At the end of this first cleaning work our corpus is composed of 15 thousand words (14,855 word tokens²), hence it is a little corpus (Bolasco 1999). D is the longest text (4,184 word tokens) and the shortest one is C (3,014 word tokens). Using the lexical richness index obtained as ratio of word types to word tokens, we can see that word types are the 19% of the corpus. Due to limited dimensions³, this relative weight is higher in each single speech (around 30%). The richest

² In textual analysis problems the term word token means an occurrence of a textual unit and word type a textual unit recognized as a string of letters taken from the alphabet of the language isolated by means of separators (blanks and interpunctuation characters). This definition is useful to build new forms: if the underscore '_' is not a separator new word types can be constructed: *'carta_di_credito / credit card'*.

³ The more the quantity of word tokens increases the more the quantity of new word types increases with a decreasing rate. If a text is short we usually have less repetitions of the existing word types.

speech is the shortest one (C with 34%), while A is the one in which less word types are used (26%) and it is one of the longest texts indeed (with its 3,969 word tokens is only second to D); B (3,688 word tokens) and D (30%) are in an average position. Candidate-A positioning is interesting because his text is not the longest but is the poorest from the lexical richness point of view. This suggests us that his speech is a “letter to electors” written in a generic and colloquial language (further analysis show he uses many common multi-words like *‘per quanto riguarda / as regards’*, *‘in questo momento / in this moment’*, *‘di solito / as usual’*). *Vice versa*, short texts with several word types are typical of a more technical, brief, schematic or over-refined language.

3 Method

Textual analysis always begins with the observation of the most frequent word types (Lebart *et al.* 1998). Since it is not possible to work with all the word types (usually they are several thousand) it is necessary to set a consistent threshold and focus the analysis on word types showing a frequency higher than this threshold. If we rank the word types on a list in a frequency decreasing order (the corpus ordered vocabulary) we find at the top of the list a lot of grammatical words (articles, conjunctions, *etc.*) and topic words (*‘Ateneo / University’*, *‘ricerca / research’*, *‘Università / University’*, *‘Facoltà / Faculty’*, *‘didattica / teaching’*, *‘attività / activity’*, *‘risorse / resources’*, *‘personale / staff’*, *‘studenti / students’*, *‘sviluppo / development’*, *‘Dipartimenti / Departments’*, *‘Nomecittà / Nameofthetown’*, *etc.*). Most of these high frequency topic words need further investigation because they are little informative (and often multi-meaning) if they are isolated from the usage context (word types involved in this test are several: *‘ricerca / research’*, *‘risorse / resources’*, *‘autonomia / autonomy’*, *‘università / university’*, *‘Ateneo / University’*, *‘didattica / teaching’*, *etc.*). For example the word *‘università / university’* was initially distinguished in plural *‘università plur / universities’* and singular, then, only for singular, distinguished again in *‘Università ist / University’* when referred to the university as an institutional body and *‘università Siglacittà / university Citycode’* when referred to the specific university of the town under investigation. In order to frame a high frequency word type such as *‘Ateneo / University’*, which makes no sense if considered in itself because of the high usage, words preceding and following can be used to construct new multi-words: *‘governo dell Ateneo / university management’*, *‘nostro Ateneo / our university’*, *‘Ateneo di Nomecittà / university of Nameofthetown’*. It is not always useful to distinguish a word in its different meaning like in the above mentioned case, because when we specify it we obtain word types with a too low frequency. We can consider for instance the word type *‘autonomia / autonomy’*, that occurs 22 times in the corpus but with different meanings.

By means of TALTAC software (Bolasco *et al.* 2000) we have run a further normalization based on lists in order to find in the texts sequences of words (multi-words) which gain meaning or change meaning if jointly considered as a block. This way some common sequences such as ‘*pubblico impiego / public employment*’, ‘*classe dirigente / ruling class*’, ‘*titolo di studio / studies certificate*’, ‘*salto di qualità / quality improvement*’, ‘*punto di riferimento / landmark*’, ‘*in base / on the ground that*’, ‘*per quanto riguarda / as regards*’, *etc.* become part of our vocabulary. After all these operations the texts need further recoding useful to put together words that can be considered as synonymous with others according to the researcher point of view. This merge allows us to recover low frequency word types putting them together with synonyms (even in a broader sense). Even some words used in singular and plural form by the same author can be merged to spare word types in the analysis, like ‘*meccanismo# / mechanism*’ (‘*meccanismo sing.*’ + ‘*meccanismi plur.*’) used only by Candidate-D or ‘*sanitario# / medical*’ (‘*sanitario m.sing.*’ + ‘*sanitaria f.sing.*’ + ‘*sanitarie f.plur.*’) by Candidate-B, *etc.* At the end of this long normalization process we have decided to use for analysis all the word types showing frequency higher than 5⁴. Then we add the word types showing frequency between two and five that are characteristic units (Lebart *et al.* 1998) for at least one author⁵ and further 38 word types of interest. This way we have dropped out all hapax (words with frequency 1 that are the 55% of the vocabulary) and all the word types that are neither characteristic nor interesting with frequency between two and five. With these criteria setted we have obtained 474 word types to submit to correspondence analysis. We have built a two-ways contingency table with 474 rows named with the word types and 4 columns named with the candidates⁶, where for each word type we have in the cell how many times each author has repeated it in his own text. In order to obtain a representation of this table

⁴ This frequency includes words taken from the first tenth of low frequency vocabulary area and it is a recommended threshold for analysis (Bolasco 1999).

⁵ In order to recognize the authors characteristic units we can use the hypergeometric model (Lafon 1980, Lebart *et al.* 1998). It allows us to compute the probability of the word over-usage by an author (Tuzzi 2003, Tuzzi and Tweedie 2000). In this work we have selected over-usage characteristic words fixing 3% as probability threshold.

⁶ From this contingency table (474×4) we obtain a three-dimensional graph in which each author and each word type is a point in the space by means of three coordinates. Unfortunately this kind of graph is not easy to read. It is therefore better to observe one axis per time (unidimensional point of view) or two axes a time (bidimensional point of view). Further complications derive from the number of word types to draw. All the word types can not be drawn on the Cartesian plane at the same time. For this reason we have decided to represent the topic words (concerning politics and university) separately from the rest (non topic words, articles, conjunctions, *etc.*), furthermore we have decided to represent only the most important words for the solution reading (loadings).

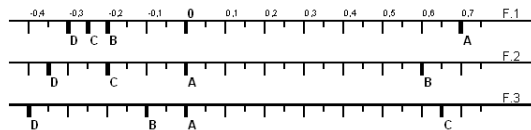


Fig. 1. Candidates' positions on axes.

we can use correspondence analysis (Greenacre 1984) because this statistical technique allows to represent the system of relations author-word type on a Cartesian plan where each word type and each author can be placed by means of coordinates (Figure 1).

4 The first and second dimension: Candidate-A and Candidate-B

The first axis is determined by Candidate-A's speech. This first speech is peculiar because it contains words that are typical of a conservative speech; we find the topic words '*imprenditoria / business*', '*privata# / private*', '*impresa / company*', '*pubblico impiego / public employment*', '*personale sost / staff*', '*personale TA / technical and administrative staff*', '*stato giuridico / legal status*', '*livelli / levels*', '*risorse finanziarie / financial resources*', '*classe dirigente / ruling class*', '*opinione pubblica / public opinion*', '*candidato / candidate*', '*politico / politic*', '*competizione / competition*', '*di carriera / of career*', '*Università Siglacittà / University city-code*', '*identità / identity*', '*Università ist / university as an institutional body*', '*istituzione# / institutional body*', '*universitaria / university*', '*di prestigio / prestige*', '*punto di riferimento / landmark*', '*Italia / Italy*', '*nazionale / national*', '*internazionale / international*', '*primato / record*'. Candidate-A speaks using the first person and a lot of plural forms (non topic and grammatical word types): '*io / I*', '*noi / we*', '*nostro / our m.*', '*nostra / our f.*', '*credo / hope*', '*vorrei / wish*', '*stimo / estimate*'. We can notice an emphasised style: '*grande / great sing.*', '*grandi / great pl.*', '*importanza / importance*', '*importante# / important*', '*super*', '*enorme / enormous*', '*mondo / world*', '*molto / much*'; a generic discourse: '*generale / general*', '*in generale / in general*', '*in questo momento / in this moment*', '*oggi / today*', '*anni / years*', '*un momento / a moment*' and a discourse on doubt: '*dubbio / doubt*', '*debba / should*', '*forse / perhaps*', '*ma / but*', '*non solo / not only*', '*però / but*', '*crisi / crisis*', '*potrebbe / could*', '*preoccupante / worrying*', '*problematica# / problematical*'. Synthetically, Candidate-A characteristics are based on rhetorical emphasis, generic lines and high attention payed to the political and economic world.

The second axis (Figure 1) separates B from C and D. In this picture A is in the axis origin (he is off-side) because all relevant things about this speech are explained by the first axis. As regards to topic word types B and A contain some rhetorical features in common, even if B shows a closer link to

tradition (Figure 2): ‘*nostro Ateneo / our university*’, ‘*Ateneo di Nomecittà / University of nameofthetown*’, ‘*scienza / science*’, ‘*scientifico# / scientific*’, ‘*eccellenza / excellence*’, ‘*attività di ricerca / research activities*’, ‘*attività didattica / teaching activities*’, ‘*didattica agg / teaching*’, ‘*docente# / teacher*’, ‘*universitaria / university*’, ‘*apparato / body*’, ‘*organi / bodies*’, ‘*settori / sectors*’, ‘*sviluppo / development*’, ‘*pubblici / public*’, ‘*reperimento / recovering*’, ‘*risorse finanziarie / financial resources*’, ‘*finanziario# / financial*’. We can notice some specific subjects like ‘*Facoltà di Medicina / Medicine Faculty*’, ‘*assistenziale / relief*’, ‘*sanitario / medical*’ and read a political and polemic essay due to the electoral atmosphere of the competition: ‘*civile / civil*’, ‘*clima / atmosphere*’, ‘*elettorale / electoral*’, ‘*influenza / influence*’, ‘*politici / political*’, ‘*partiti / parties*’, ‘*autonomia / autonomy*’, ‘*schieramenti / positioning*’, ‘*contrapposizioni / oppositions*’. Looking at non topic and grammatical word types (Figure 3) we find a reminder to some classic university aspects: ‘*cultura / culture*’, ‘*società / society*’, ‘*culturale / cultural*’, ‘*collegli / colleagues*’, ‘*creare / to create*’, ‘*energie / energies*’, ‘*adeguata / on purpose*’. Candidate-B is the only one who names ‘*Nomerettoreuscente / Nameofthepastvicecancellor*’ and speaks about what ‘*realizzato / implemented*’, ‘*in passato / in the past*’. We can shortly underline the strong defence of the “academic separation”, the complain for lack of public resources and a minor, but present, rhetorical emphasis.

5 The third dimension: Candidate-C *versus* Candidate-D

The real confrontation comes out on the third axis and we can underline differences between Candidate-C and Candidate-D: the two candidates with wider chances of victory. Concerning topic word types (Figure 2) D is a speech focused on policy and it is a typical management speech given by an evaluator: ‘*gestione / management*’, ‘*organizzazione / organization*’, ‘*progetto / project*’, ‘*bilancio / budget*’, ‘*trasparenza / transparency*’, ‘*interventi / actions*’, ‘*discussione / debate*’, ‘*confronto / confrontation*’, ‘*dati / data*’, ‘*qualità / quality*’, ‘*verifica / check*’, ‘*monitoraggio / monitoring*’, ‘*meccanismo# / mechanism*’ paying attention to the role of university as a public service: ‘*strutture / structures*’, ‘*corsi / courses*’, ‘*offerta formativa / educational offer*’, ‘*servizi / services*’, ‘*didattici / teaching*’, ‘*attrezzature didattiche / teaching environments*’. While Candidate-C focuses his own speech on defence of diversity: ‘*diversità / diversity*’, ‘*complessità / complexity*’, ‘*disciplinari / concerning the subjects*’, ‘*garantendo / assuring*’, ‘*governance*’, ‘*interazione / interaction*’, ‘*regole / rules*’, ‘*accordi / agreements*’, ‘*livelli / levels*’, ‘*autoreferenzialità / self-reference*’, ‘*omogenee / homogeneous*’, ‘*particolare / special*’, ‘*penalizzazione / penalty*’, ‘*subalternità / subordinate*’, ‘*scientifiche / scientific*’, ‘*carico didattico / teaching commitment*’, ‘*didattica sost / teaching*’. Candidate C speaks about ‘*budget*’, ‘*amministrativo / administrative*’, ‘*città*

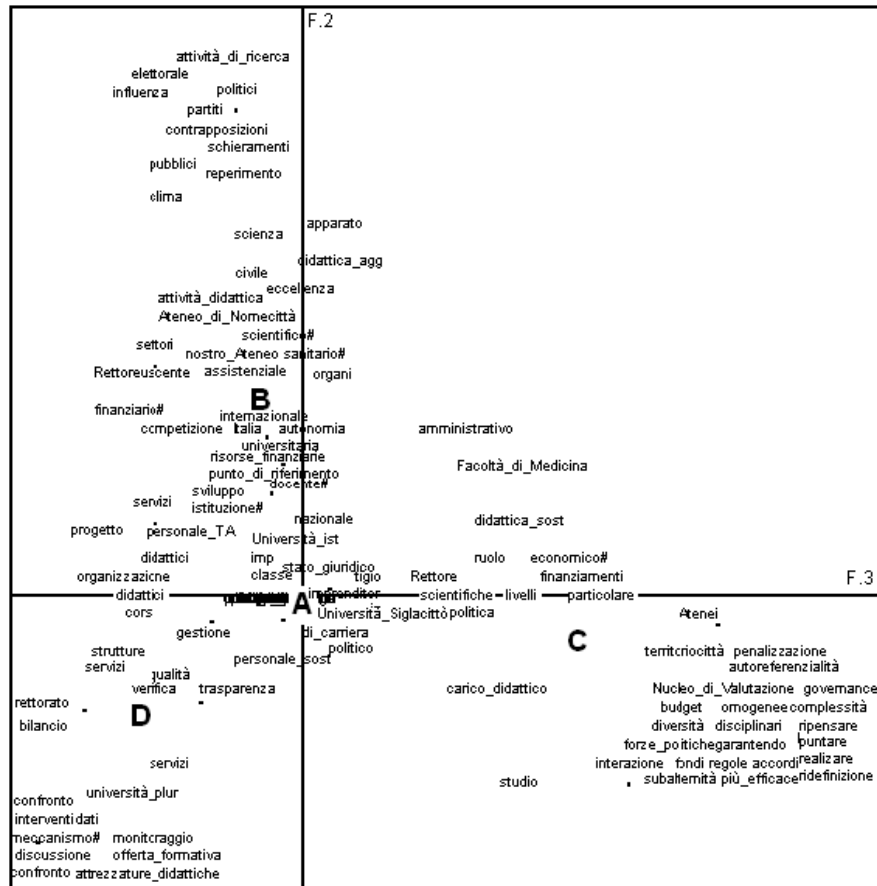


Fig. 2. Topic words. Factor3 (27,6% in.ex.) and factor2 (31,0% in.ex.).

/ city', 'territorio / territory', 'ruolo / role', 'economico# / economic', 'finanziamenti / financing', 'fondi / funds', 'forze politiche / political parties', 'politica / politics' with a different policy approach.

Concerning non topic and grammatical word types (Figure 3), the same third axis underlines the opposition between an optimistic future-oriented speech concerning the project of rebuilding the entire system and a reflective past-oriented speech concerning the improvement of the existing system. In other words Candidate-C was able to present himself as an innovator, ready to change everything: 'ripensare / to think again', 'ridefinizione / redefining', 'realizzare / to make', 'realizzazione / making', 'puntare / to bet', 'giungere / to reach', 'bisogna / need', 'bisognerà / will need', 'converrà / will be necessary', 'deve / must', 'devono / must', 'dovrà / will have to', 'sarà / will be', 'davvero / surely', with good intentions: 'effettiva / actual', 'grande / great',

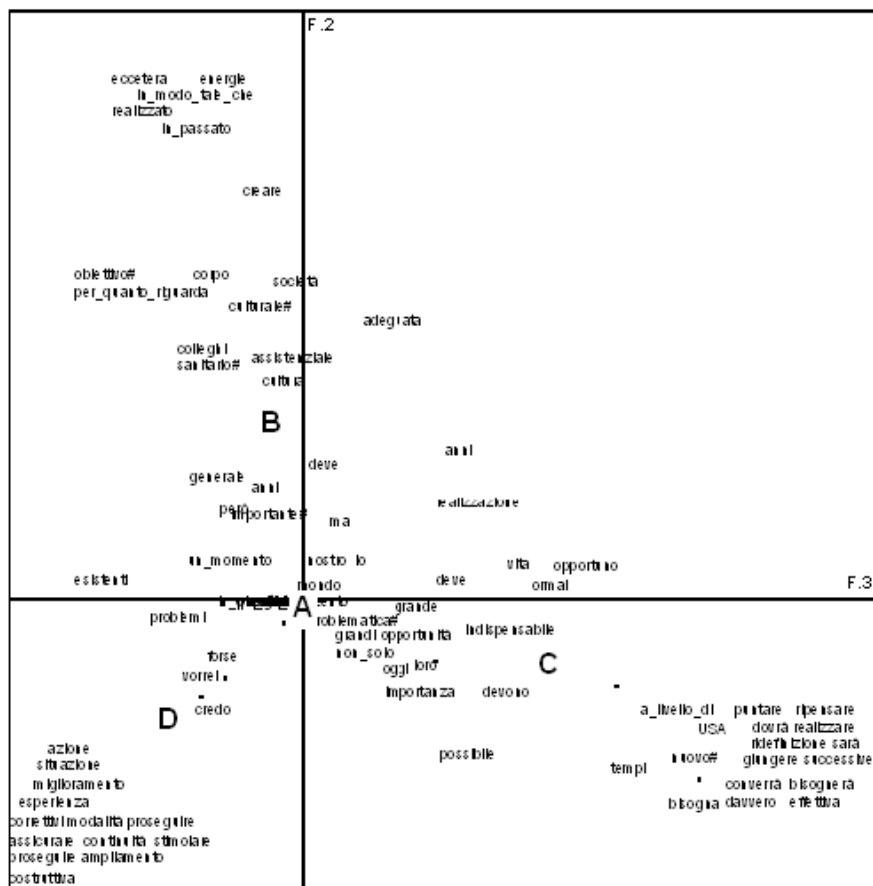


Fig. 3. Non-topic words. Factor3 (27,6% in.ex.) and factor2 (31,0% in.ex.).

'indispensabile / necessary', 'nuovo# / new', 'opportuno / appropriate', 'opportunità / opportunity', 'possibile / possible', 'ormai / it is time', 'tempi / times', 'successive / following'. While Candidate-D wrote in a more modest way, mainly oriented to the past, surely better ready to evaluate the existing situation rather than to make new proposal: 'continuità / continuity', 'esistenti / existing', 'esperienza / experience', 'miglioramento / improvement', 'correttivi / corrective', 'situazione / situation', 'costruttiva / positive', 'ampliamento / enlargement', 'assicurare / to assure', 'azione / action', 'modalità / way', 'proseguire / to go on', 'stimolare / to stimulate'.

Finally Candidate-C won. The elected new Vice Chancellor used few immediate, evocative passwords: features and identity of the university, cultural diversity, optimistic view of the future and need to change. Both modest tones and problematic aspects are avoided. The majority of the electorate has been

probably attracted by this positive and hopeful atmosphere and surely the style of the speeches played a role in determining this result.

References

- ALVAREZ R., BÉCUE M., LANERO J.J. and VALENCIA O. (2002): Results stability in textual analysis: its application to the study of the Spanish investiture speeches (1979-2000), in: *JADT 2002*, Morin A. and Sébillot P. (Eds.), IRISA – INRIA, 1, 1–12.
- AURELI E. and BARTOLETTI S. (2002): A textual analysis of the lexicon of general policy in an election of the Rector through the electoral speeches of the five candidates, in: *JADT 2002*, Morin A. and Sébillot P. (Eds.), IRISA – INRIA, 1, 57–67.
- BOLASCO S. (1996): Il lessico del discorso programmatico di governo, in: *L'attività dei governi della repubblica italiana (1947–1994)*, Villone M. and Zuliani A. (Eds.), Il Mulino, Bologna, 163–349.
- BOLASCO S. (1999): *Analisi multidimensionale dei dati*, Carocci, Roma.
- BOLASCO S., BAIOCCHI F. and MORRONE A. (2000): *TALTAC versione 1.0*, CISU, Roma.
- CARBONE S. (2000): Le parole e le armi, in: *JADT 2000*, Rajman M. and Chappelier J.C. (Eds.), EPFL, 1, 95–102.
- GREENACRE M.J. (1984): *Theory and Application of Correspondence Analysis*, Academic Press, London.
- LAFON P. (1980): Sur la variabilité de la fréquence des formes dans un corpus, *Mots*, I, 127–65.
- LEBART L., SALEM A. and BERRY A. (1998): *Exploring Textual Data*, Kluwer Ac., Boston.
- TUZZI A. (2003): *L'analisi del contenuto. Introduzione ai metodi e alle tecniche di ricerca*, Carocci, Roma.
- TUZZI A. and TWEEDIE F.J. (2000): The Best of Both Worlds: Combining MO-CAR and MCDISP, in: *JADT 2000*, Rajman M. and Chappelier J.C. (Eds.), EPFL, 1, 271–76.