Results stability in textual analysis: Its application to the study of the Spanish investiture speeches (1979-2000)

Ramón Álvarez[1], Mónica Bécue[2], Juan José Lanero[1], Olga Valencia[3]

[1]Universidad de León – Spain – dderae@unileon.es

[2] Universidad Politécnica de Cataluña – Spain – monica@eio.upc.es

[3] Universidad de Burgos – Spain – oval@ubu.es

Abstract

The bootstrap technique makes feasible the study of the results stability in correspondence analysis applied to textual data. The kind of results obtained that way are shown through their application to the corpus of investiture speeches of the Spanish Prime Ministers since the Constitution was passed in 1978.

Résumé

La technique du bootstrap fait faisable l'étude de la stabilité des résultats dans l'analyse des correspondances appliquée aux données textuelles. Le genre de résultats obtenu est montré à travers leur application au corpus des discours d'Investiture des Premiers Ministres espagnols depuis que la Constitution est passée en 1978.

Keywords: Bootstrap, Correspondence analysis, Lexical tables, Confidence regions.

1. Introduction

The simulation methods such as the bootstrap are a useful instrument in the validation of the results obtained in the correspondence analysis applied to textual data. They allow to complete the graphic representations obtained through the determination of the confidence regions (Lebart, 1996). In so doing, they make feasible the selection of stable elements which is of particular interest when it is referred to words. Furthermore, the systematic comparison of the correspondence analysis carried out on the tables built from simulated corpora is very useful in order to determine the dimension of the stable subspace.

In this paper these tools are applied to the corpus made up of *the Investiture speeches of the heads of the Spanish government (1979-2000)*, corpus which is presented in section 2. In section 3 it is shown how the bootstrap tables are built and in section 4 the coordinates of speeches and words are obtained. The next step is the analysis of the eigenvalues of correspondence analysis of the simulated tables to finish off in section 6 with the study of coordinates and contributions of words and speeches.

2. Investiture speeches since the democratic restoration in Spain (1979-2000)

The analyzed corpus is made up of eight investiture speeches delivered by the four candidates to the Head of Government in Spain since Constitution of 1978.

The eight speeches are normalized but they are not lemmatized. The total length of this corpus is 75504 occurrences, made of 7801 different words. The average length of a speech is of 9438 words.

In the correspondence analysis of the table *Words x Speeches*, only the words said at least five times are preserved: 65839 occurrences and 1754 different words.

Tables 1 and 2 show the eigenvalues obtained in the analysis and the contribution of the speeches to the formation of the axis.

	Eigenvalue	%	%AC
1	0,08981595	24,19	24,19
2	0,06313169	17,00	41,19
3	0,04863081	13,10	54,29
4	0,04798980	12,93	67,22
5	0,04581298	12,34	79,56
6	0,03928819	10,58	90,14
7	0,03661108	9,86	100,00

Table 1. Eigenvalues

DATE	**Speech**	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7
March 30, 1979	**Suárez-79**	9,87	26,78	2,99	28,79	13,73	0,40	1,67
February 19, 1981	**Calvo-Sotelo-81**	0,32	7,68	3,54	42,64	0,93	29,63	4,45
November 30, 1982	**González-82**	0,20	16,50	18,81	0,05	36,79	9,65	5,91
July 23, 1986	**González-86**	53,36	0,45	23,56	2,99	2,73	1,38	0,00
December 5, 1989	**González-89**	14,84	0,98	37,59	0,11	24,52	6,19	5,54
July 8, 1993	**González-93**	0,67	5,40	2,66	5,25	2,92	28,43	43,87
May 3, 1996	**Aznar-96**	13,75	9,24	9,65	6,65	3,86	6,51	36,61
April 25, 2000	**Aznar-2000**	6,99	32,97	1,20	13,53	14,52	17,80	1,94

Table 2. Absolute contributions of the speeches

3. Validation of results by means of simulation

3.1. Bootstrap methodology

The non parametric bootstrap resampling method does not need the formulation of previous hypothesis about the distribution of estimators.

Given an original sample of *n* observations, big *n*, *m* bootstrap sample of the same size are worked out through the procedure of random sampling with replacement. The *m* values of a particular statistic, worked out for each of the simulated samples, make up its "bootstrap sampling distribution". This empirical distribution is used in order to estimate the different characteristics of the statistics, their variance in particular, what allows to work out the confidence intervals for the estimated parameters.

3.2. Construction of the replica of the Investiture corpus

In order to apply the bootstrap method to the correspondence analysis of the *Investiture* corpus, *m* replica of the corpus are built with the same length, that is, 65839 occurrences. The said replica are built through the random extraction with replacement of *k*=65839 occurrences among the initial occurrences. To each occurrence the word and the text they belong to are associated.

5000 replica of corpus have been made (*m*=5000), a sufficient big target number for the aimed target of analyzing the results stability.

The simulated tables (rows and columns) can be used as supplementary (partial bootstrap); another possibility could be a correspondence analysis for each table (total bootstrap) and in so doing to the study the stability of subspaces.

4. Construction of confidence regions for texts and words

In order to study the stability of the speeches-columns and words-rows, the rows of the simulated tables are considered as supplementary rows, and the columns of the simulated tables are supplementary columns in the correspondence analysis of the initial table. Thus, the confidence ellipsoids which are obtained from the distribution of simulated coordinates, can be represented on factorial planes.

4.1. Confidence ellipsoids for the speeches

Figure 1 shows the confidence ellipsoids (level 95%) worked out for the speeches on the first factorial plane of the correspondence analysis of the table *Words x Speeches*.

In order to obtain the ellipsoids the coordinates of the 5000 replica of the speeches are worked out and percentiles 5 and 95 of the distribution of the said coordinates are determined.

In our study, the simulations of the eight speeches posed as illustrative show very stable configurations for them all in the first factorial axis (Figure 1).

4.2. Stability analysis of words

The representation of the ellipsoids of confidence of all the words on a factorial plane implies to include an excessively high number of points.

Therefore, it seems more appropriate to select the most stable words without representing them graphically. Starting from these stable words the axes are interpreted.

In the example given, on the positive side of the first axis the words are located with bigger absolute contributions. Among the hundred which are the most stable ones, there are the words that correspond to the entry in the European Economic Community (*comunitaria*, *fecha*, *Acta Única*, *Adhesión*), internal questions (*españoles*, *socialista*, *ETA*) and characteristic words of González's speech (*pagan, decía, punto, vista, extraordinaria, haya, intentar, que, consiguiente, esa, cualitativo, desde*). One can notice that among the thousand words which are the most unstable ones there is not any with high absolute contribution.

On the negative side of the first axis there are the less contributive words, but a great number of them are highly stable (*presida*, *emprender*, *moderno*, *Popular*, *regiones*, *efectivo*, *1*, *pluralismo*, *solamente*, *clarificación*, *Territoriales*, *efectiva*, *familia*, *municipal*, *logro*, *imperio*...)

On the following axes, the stability of words (and speeches) is diminished, therefore the bootstrap techniques can constitute another approach for the election of the axes to preserve, those on which the element-rows and columns remain sufficiently stable.

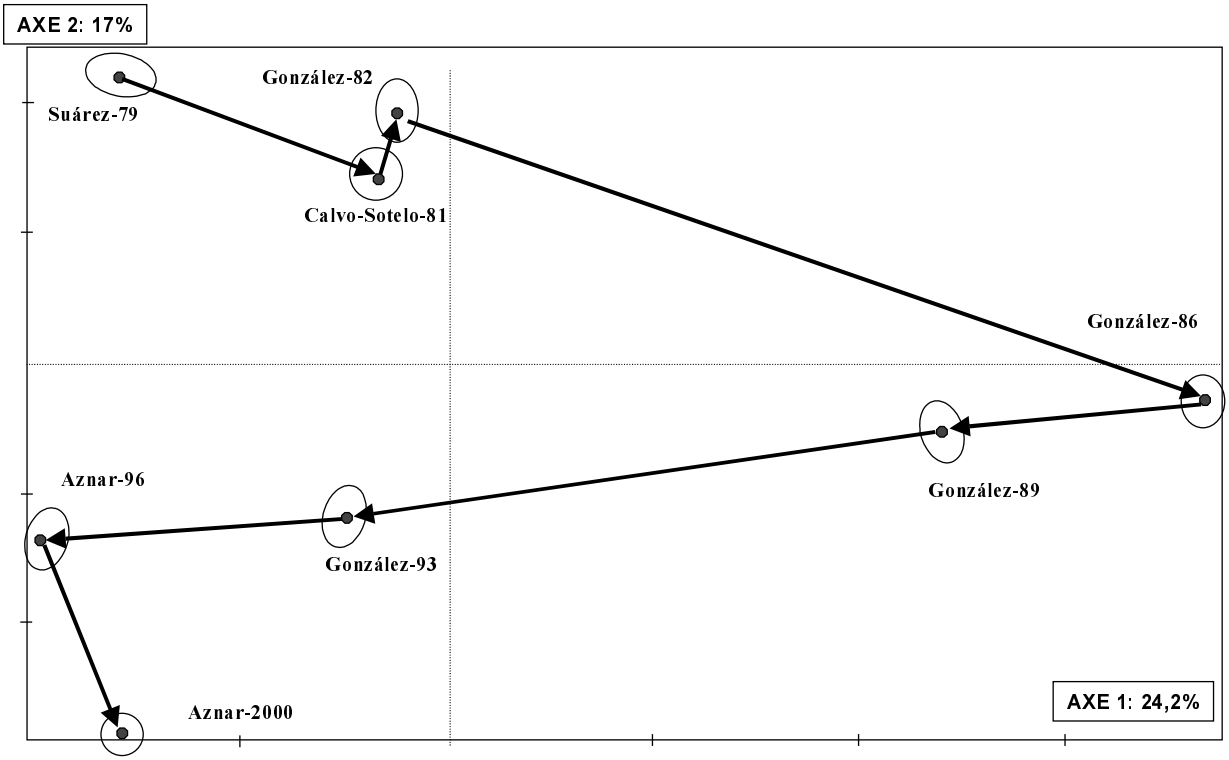

Figure 1. Factorial plane of the first two axes and stability of the coordinates of the speeches

5. Eigenvalues stability in the correspondence analysis of the replied tables

5.1. Non normality of the eigenvalues distribution

The decision about the number of axes or factors that should be preserved in a correspondence analysis can be based on four different approaches : empirical rules, external procedures, asymptotic studies and studies of stability by means of simulation (Lebart, 1996).

In this study we center ourselves in the *non parametric bootstrap*, without formulating previous hypothesis, as a tool to determine the confidence intervals for the eigenvalues. Indeed, the empiric distribution of the eigenvalues obtained by bootstrap simulation shows that, for the axes 4 and 5 (Table 3), you cannot assume the hypothesis of normality.

	Axis 1	Axis 2	Axis 3	Axis 4	Axis 5	Axis 6	Axis 7
Mean	.1096	.08528	.07276	.07007	.06737	.06205	.05866
Std.Deviation	.00218	.00180	.00138	.00119	.00137	.00144	.00154
Skewness	-.040	.073	.237	.073	-.173	.160	-.107
Kurtosis	-.190	-.278	.179	-.186	-.178	.080	.033
Minimum	.10	.08	.07	.07	.06	.06	.05
Maximum	.12	.09	.08	.07	.07	.07	.06
Percentiles 5	.1061	.08238	.07056	.06814	.06510	.05974	.05602
Percentiles 95	.1130	.08829	.07509	.07205	.06954	.06453	.06108
K-S Normality	.675	1.248	1.257	1.416	1.982	1.053	1.177
p value K-S	.753	.089	.085	.036	.001	.218	.125

Table 3. Main statistics of the eigenvalues

5.2. Graphic representation of the eigenvalues of the correspondence analysis of the simulated tables

Carried out the 5000 bootstrap samples generated, the obtained eigenvalues can be represented jointly in the following Figure :

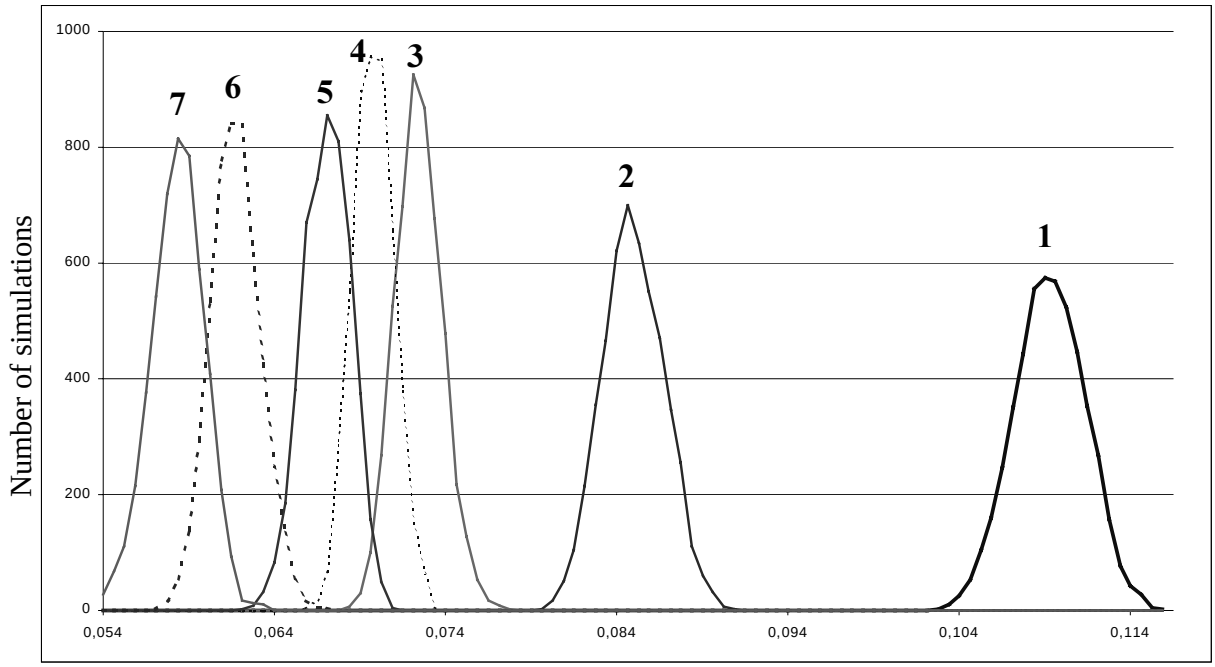

Figure 2. Simulated eigenvalues

The following Figure shows the confidence percentiles for the simulations of the seven eigenvalues :

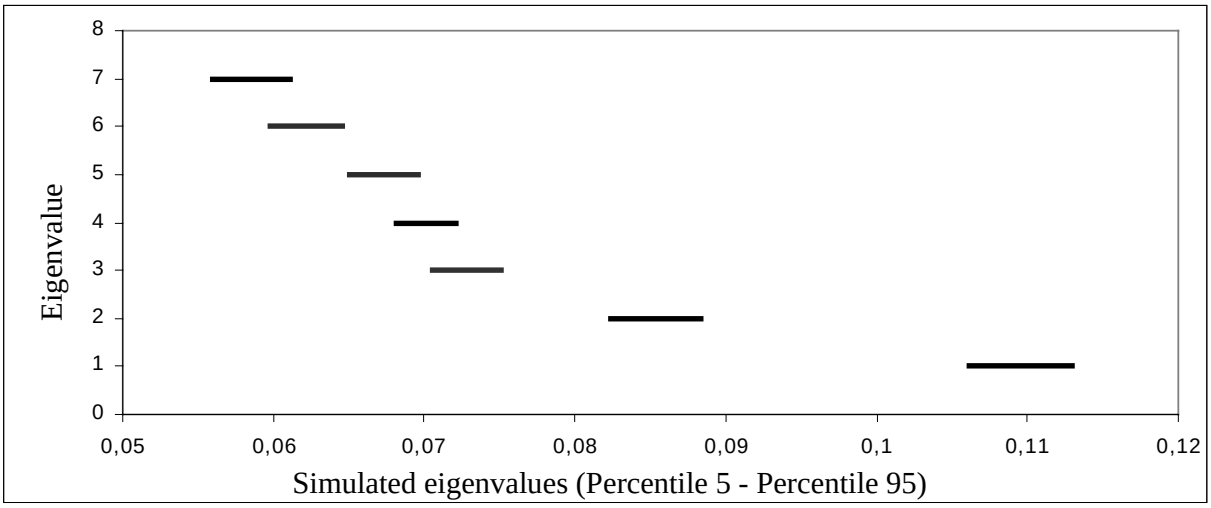

Figure 3. Confidence regions for the 90% of the eigenvalues

The confidence regions worked out in Figure 3 contain the 90% of the observations of the simulated values. The confidence intervals of the first two eigenvalues do not have common areas, whereas the confidence intervals 4, 5 and 6 to one another, and those of the eigenvalues 6 and 7 do.

It should also be noted that it could happen that the axes are exchanged, for example axis 3 in some simulations changes into axis 4 and viceversa.

5.3. Bias of the bootstrap eigenvalues

It should be noted that the eigenvalues of the original table are not inside the confidence intervals of the simulated tables (see Table 1 and Table 3).

In this study it is verified that the inertia of the original table is 0.37128, while the mean of the inertias of the simulated tables is 0.5258 (with a minimum 0.51 and a maximum 0.54).

The increase of the inertia in the simulated tables is due to the bootstrap samples are interferences of the original sample. A cell with zero frequency in the original table will always be a zero in the simulated table, and a cell with low frequency in the original table can be a zero in the simulated tables.

In the literature on bootstrap it is usually admitted that the population eigenvalues are lower to the sample eigenvalues, and these ones in turn are lower to the estimated eigenvalues by bootstrap.

This way, one can establish the following relationship :

$$\lambda_{pop} < \lambda_{samp} < \overline{\lambda}^{*}$$

in which $\overline{\lambda}^{*}$ is the mean of the bootstrap estimates.

This difference among the eigenvalues is due to the sample bootstrap is an interference of the original sample and its estimates are biased in a positive sense, therefore it is necessary to set a correction. This correction of the bias is aiming at obtaining corrected bootstrap eigenvalues subjected to :

$$\lambda_{pop} < \overline{\lambda}_{corr}^{*} < \lambda_{samp} < \overline{\lambda}^{*}$$

The estimate bootstrap of the bias can be worked out as $bias* = \overline{\lambda}* - \lambda_{samp}$.

Wherefore the corrected bootstrap eigenvalues can be settled down as :

$$\lambda*_{(corr)} = \lambda* - bias*$$

The same procedure should be carried out to work out the percentiles of the eigenvalues :

$$P^*_{5(corr)} = P^*_5 - bias*, \qquad P^*_{95(corr)} = P^*_{95} - bias*$$

Van der Burg and De Leeuw (Van of Burg and De Leeuw, 1988) justify the use of this correction theoretically within the analysis of canonical correlations, in those cases where the delta method is applicable, obtaining similar results.

Markus (Markus, 1994a) considers that this correction constitutes a correct procedure to deepen in the external stability of the results of an analysis of homogeneity, providing even better confidence regions than those obtained by the delta method.

Applying the corrections to the previously calculated percentiles, the corresponding corrected eigenvalues by the bias are obtained :

	λ_{samp}	$\overline{\lambda}*$	P^*_5	P^*_{95}	Bias*	$P^*_{5(corr)}$	$P^*_{95(corr)}$
1	0,08982	0,10959	0,10613	0,11295	0,01978	0,08635	0,09317
2	0,06313	0,08528	0,08238	0,08829	0,02215	0,06023	0,06614
3	0,04863	0,07276	0,07056	0,07509	0,02413	0,04643	0,05096
4	0,04799	0,07007	0,06814	0,07205	0,02209	0,04605	0,04997
5	0,04581	0,06737	0,06510	0,06954	0,02156	0,04354	0,04798
6	0,03929	0,06205	0,05974	0,06453	0,02276	0,03698	0,04177
7	0,03661	0,05866	0,05602	0,06108	0,02205	0,03397	0,03903

Table 4. Corrected eigenvalues

The previous graphic representations continue being valid as far as the form of the same ones are concerned, since we have only carried out a change in the origin by subtracting from each bootstrap eigenvalue the corresponding bias.

One possibility which is not included in this paper is the construction of bootstrap tables starting from the prospective frequencies instead of from those ones observed, procedures generically named "nonnaive bootstrap methods" (Langeheine and Pannekoek, 1996) with the purpose of putting closer the eigenvalues to the originals.

6. Stability of the axes in the correspondence analysis of replied tables

6.1. Correlations between the coordinates of the original axes and the simulated axes of the same range

The problem of the stability of the axes outlines a more complex problem than the stability of the eigenvalues. If we carry out a correspondence analysis on each one of the simulated tables we have indicated that they can change the eigenvalues, but the eigenvectors can also make it, accordingly an axis can change into unstable. First of all, axes with similar eigenvalues can be

exchanged, but it is also possible that the same subspace comes determined by different axes (by means of rotations, for example).

Compairing in a global way the obtained results of the initial correspondence analysis and those obtained by means of simulation, we have opted for a very simple, but at the same time easy form of interpreting : the calculation of the lineal correlations among the initial vector of coordinates and each one of the vectors of coordinates obtained in the correspondence analysis of the replied tables.

An axis is considered stable if the coordinates of the simulated modalities are highly correlated to the coordinates of the original table. It is necessary to keep in mind that the orientation of the axis (sign of the coordinates) can be changed.

	Axis1	**Axis2**	**Axis3**	**Axis4**	**Axis5**	**Axis6**	**Axis7**
Mean	.9957	.9830	.8256	.8305	.8218	.8779	.9026
Std.Deviation	.003151	.01176	.1025	.1013	.09961	.08479	.08152
Skewness	-2.110	-1.831	-.259	-.293	-.280	-.775	-.980
Kurtosis	7.210	6.038	-1.003	-.936	-.837	-.218	.012
Minimum	.97	.89	.55	.56	.55	.58	.55
Maximum	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Percentiles 5	.9897	.9598	**.6541**	**.6599**	**.6517**	**.7097**	**.7343**
Percentiles 95	.9990	.9958	.9709	.9732	.9671	.9811	.9896

Table 5. Correlations between simulated coordinates and original ones. Speeches

Starting from the results of Table 5, we can conclude that the configurations of the speeches on axis 1 and axis 2 are stable to inclination in all the replica (very high correlations). The smallest stability on the following axes can be explained either because, even being stable the subspaces determined, exchanges exist between the axes or rotations of the same ones, or because starting from a certain dimension the lower range subspaces are unstable. It is evidently important to determine which of the two suppositions is verified and the stable subspaces too.

Once the results for the words are analyzed, we find that the correlations are smaller (it is necessary to keep in mind that the number of words is sensibly bigger than that one of speeches), confirming the conclusions obtained for the speeches.

	Axis1	Axis2	Axis3	Axis4	Axis5	Axis6	Axis7
Mean	. 8927	.8346	.6712	.6715	.6579	.6815	.6885
Std.Deviation	.00527	.01190	.08434	.08266	.07983	.06616	.06189
Skewness	-.481	-1.208	-.247	-.272	-.270	-.758	-.931
Kurtosis	.523	3.229	-.997	-.939	-.828	-.195	-.047
Minimum	.87	.75	.44	.42	.43	.44	.44
Maximum	.91	.86	.84	.81	.81	.80	.78
Percentiles 5	.8837	.8120	**.5304**	**.5330**	**.5212**	**.5527**	**.5614**
Percentiles 95	.9006	.8503	**.7913**	**.7893**	**.7746**	**.7645**	**.7577**

Table 6. Correlations between simulated coordinates and original ones. Words

6.2. Correlations between the coordinates of the original axes and simulated axes of any range

It has also been calculated the correlation between each one of the simulated axes and each one of the original axes. Next the highest correlation is determined between each original axis and the simulated ones, proceeding to calculate the relative frequency with which we associate the simulated axes and the original ones. This way, the simulated axes 1 and 2 have, in all the simulations, the maximum correlation with the axes of originals of the same range (100%); the original axis 3 has a maximum correlation with the axis 3 simulated in 49.6% of the cases, with the axis 4 simulated in 37.4% of the cases and with the axis 5 in 13.0% of the cases, etc.

	Axis1	Axis2	Axis3	Axis4	Axis5	Axis6	Axis7
Axis 1 Original	***100.0%***						
Axis 2 Original		***100.0%***					
Axis 3 Original			***49.6%***	37.4%	13.0%		
Axis 4 Original			42.2%	***41.0%***	16.6%	0.2%	
Axis 5 Original			9.1%	20.2%	***70.4%***	0.3%	
Axis 6 Original				0.1%	0.3%	***86.2%***	13.4%
Axis 7 Original					0.1%	12.3%	***87.6%***

Table 7. Association between simulated axes and original axes for the speeches. Coordinates

The same effect of change of axes is observed when analyzing the correlations among the coordinates of words, as Table 8 indicates.

	Axis1	Axis2	Axis3	Axis4	Axis5	Axis6	Axis7
Axis 1 Original	***100,0%***						
Axis 2 Original		***100,0%***					
Axis 3 Original			***49,8%***	37,4%	12,8%		
Axis 4 Original			42,1%	***41,1%***	16,6%	0,2%	
Axis 5 Original			9,3%	20,2%	***70,2%***	0,3%	
Axis 6 Original					0,4%	***86,2%***	13,4%
Axis 7 Original					0,1%	12,3%	***87,6%***

Table 8. Association between simulated axes and original axes for the words. Coordinates

6.3. Correlations between the contributions of the original axes and simulated axes of any range

Lastly, we outline the analysis starting from the correlations among the simulated absolute contributions of the speeches with the original ones, keeping in mind the grade of existent relationship among coordinates and contributions.

The same study, although it is not included in this paper, has been carried out for the relative contributions. The results are shown in the following Table, observing that good stability exists for the first two axes (small distances among the percentiles 5 and 95), but starting from the third there is not stability. For example, the third one has a confidence interval among 0.0412 0.9595.

	Axis1	Axis2	Axis3	Axis4	Axis5	Axis6	Axis7
Mean	.9950	.9649	.5024	.4783	.6685	.7665	.8449
Std.Deviation	.00534	.03181	.31483	.32025	.23168	.18151	.18187
Skewness	-3.291	-2.411	.068	.111	-.808	-1.471	-2.041
Kurtosis	16.769	9.593	-1.379	-1.396	.053	2.658	4.242
Minimum	.95	.69	.00	.00	.00	.00	.01
Maximum	1.00	1.00	.99	1.00	1.00	1.00	1.00
Percentiles 5	.9853	.9004	**.0412**	**.0297**	**.1907**	**.4032**	**.4352**
Percentiles 95	.9994	.9947	.9595	.9573	.9590	.9690	.9928

Table 9. Correlations between absolute contributions simulated of the speeches with the original ones

Working out the correlations of the absolute contributions of the words on the initial contributions, the results confirm the previous hypothesis, observing that there is good stability for the first axis (0.8762; 0.9170) and for the second one (0.7889; 0.8525).

	Axis1	Axis2	Axis3	Axis4	Axis5	Axis6	Axis7
Mean	.8991	.8255	.4277	.3887	.5257	.5823	.6156
Std.Deviation	.01239	.01983	.24198	.23174	.17488	.13934	.13535
Skewness	-.624	-1.174	-.169	-.020	-.998	-1.726	-1.972
Kurtosis	.384	2.864	-1.290	-1.289	.468	3.235	4.137
Minimum	.85	.69	.00	.00	.00	.00	.00
Maximum	.93	.87	.83	.81	.79	.76	.78
Percentiles 5	.8762	.7889	**.0387**	**.0348**	**.1471**	**.2792**	**.3082**
Percentiles 95	.9170	.8525	.7600	.7333	.7351	.7228	.7372

Table 10. Correlations between absolute contributions simulated of the words with the original ones

7. Conclusions

As it has been indicated along our paper, the lack of normality of the statistics obtained by means of simulation shows, a priori, that it is not advisable the use of the usual asymptotic techniques. The "bootstrap" techniques not only allow to calculate confidence regions for the coordinates of the point-row and for the point-column, but also to test the stability of the subspaces.

Initially we wondered wether a particular technique, in our case the correspondence analysis, could be adapted for the study of the corpus made up of eight investiture speeches. Keeping in mind that sometimes the approach turns around either in the method used or in the chosen analysis, forgetting that frequently the data are those which make possible the application of a certain technique, we start from the study of textual data without establishing previous hypothesis with the purpose of analyzing the quality of the conclusions reached by means of a correspondence analysis in terms of stability.

We have included within the analysis of this corpus the study of the stability, aiming at the determination wether small variations of the departure data can produce big modifications of the obtained results, what implies unstability in the conclusions.

Accordingly, 5000 simulations have been carried out applying on each one a correspondence analysis, having observed the lack of normality of some of the eigenvalues, what takes us to the consideration that it is not advisable the use of the usual asymptotic techniques and that it is possible to apply the "bootstrap" technique, provided it does not establish departure hypothesis about the distribution.

The calculation of the eigenvalues on the simulated tables has implied a considerable increase of the total inertia. The bias of the eigenvalues has been corrected to avoid this effect, checking that the eigenvalues have behaved in a very stable way for the first two factors, but that is not the case for the remaining ones that can pose interpretating problems from this approach.

We have been able to check how the stability of the axes outlines a more complex problem than the stability of the eigenvalues, since when carrying out the correspondence analysis on each one of the simulated tables the eigenvalues can change, even so the eigenvectors, with what an axis changes into unstable. We have observed how axes with closed eigenvalues they are exchanged (axes 3, 4 and 5 on one hand and 6 and 7 on the other).

The study of the absolute contributions of the speeches and words has confirmed the conclusions reached starting from the eigenvalues and the coordinates.

Therefore, in an correspondence analysis the unstability of the configurations should lead us to think about the need to use a certain number of factors only.

We think that this study should be supplemented with the automatic fixation of a threshold of selected words to be sufficiently high as to allow a stability of the results and, at the same time, to be sufficiently small so that an excessive number of words gets lost.

Another developing topic that was not the focus of our paper, is the decrease of the number of bootstrap samples needed to achieve stable results, checking in the generation of each one of these samples if the necessary convergence is reached in the estimators and to interrupt the process at that point.

Note on the software

The computer program created for this study (by the first author Ramón Álvarez) has a first module of generating of the bootstrap samples, another one of correspondence analysis realization and a third one for the study of the simulated estimators.

References

Álvarez R., Bécue M. and Lanero J.J. (2000). Le vocabulaire gouvernemental espagnol (1979-1996). *Mots*, Mars, nº 62: 31-47.

Bécue M., Álvarez R. and Lanero J.J. (1999). Etude statistique des discours d'investiture de la démocratie espagnole (1976-1996). Document de recerca. DR 99/10, noviembre de 1999. Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.

Efron B. and Tibshirani R. (1981). *An Introduction to the Bootstrap*. Chapman & Hall.

Gifi A. (1990). *Nonlinear multivariate analysis*. John Wiley & Sons Ltd.

Chateau, F. and Lebart, L. (1996). Assessing sample variability and stability in the visualization techniques related to principal component analysis : Bootstrap and alternative simulation methods. *Proceedings Computational Statistics. COMPSTAT. 1996*, pages 205-210.

Lebart L., Morineau A. and Piron M. (1995). *Statistique exploratoire multidimensionnnelle*. Dunod.

Langeheine, R. and Pannekoek, J. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods & Research*, vol.(24), Issue 4: 492-516.

Markus M.TH. (1994a). *Bootstrap Confidence Regions in Nonlinear Multivariate Analysis*. DSWO Press.

Markus M.TH. (1994b). Bootstrap confidence regions for homogeneity analysis; the influence of rotation coverage percentages. In Dutter R. and Grossmann W. (eds) *Proceedings Computational Statistics. COMPSTAT. 1994*, pages 337-342.

Reiczigel J. (1996). Bootstrap tests in correspondence analysis. *Applied Stochastic Models and Data Analysis*, vol.(12):107-117.

Van der Burg, E. and De Leeuw, J. (1988). Use of the Multiomial Jackniffe and Bootstrap in the generalized Canonical Correlation Analysis. *Applied Stochastic Models and Data Analysis*, vol.(4):154-172.