ISBN 978-3-9802659- -

Studies
in Quantitative Linguistics
6

Arjuna Tuzzi
Ioan-Iovitz Popescu
Gabriel Altmann

Quantitative Analysis of Italian Texts

RAM - Verlag

Quantitative Aspects

of

Italian Texts

by

Arjuna Tuzzi
Ioan-Iovitz Popescu
Gabriel Altmann

2010
RAM-Verlag

Studies in quantitative linguistics

Editors
Fengxiang Fan (fanfengxiang@yahoo.com)
Emmerich Kelih (emmerich.kelih@uni-graz.at)
Reinhard Köhler (koehler@uni-trier.de)
Ján Mačutek (jmacutek@yahoo.com)
Eric S. Wheeler (wheeler@ericwheeler.ca)

1. U. Strauss, F. Fan, G. Altmann, *Problems in quantitative linguistics 1*. 2008, VIII + 134 pp.
2. V. Altmann, G. Altmann, *Anleitung zu quantitativen Textanalysen. Methoden und Anwendungen.* 2008, IV+193 pp.
3. I.-I. Popescu, J. Mačutek, G. Altmann, *Aspects of word frequencies*. 2009, IV +198 pp.
4. R. Köhler, G. Altmann, *Problems in quantitative linguistics 2.* 2009, VII + 142 pp.
5. R. Köhler (ed.), *Issues in Quantitative Linguistics.* 2009, VI + 205 pp.
6. A. Tuzzi, I.-I. Popescu, G.Altmann, *Quantitative aspects of Italian texts*. 2010, IV+160 pp.

ISBN: 978-3-9802659-….

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
RAM-Verlag@t-online.de
http://ram-verlag.de

Preface

The present book can be considered a continuation of two streams in textology initiated by different groups of scholars. The first group consisting of Italian scientists analyzed the 57 end-of-year speeches of 9 Italian presidents since 1949 to 2005 from very different points of view. The investigation resulted in the book edited by M.A. Cortelazzo and A. Tuzzi (2007) containing some statistical vistas but concentrated mostly on political, social, linguistic and cultural matter. The second group, lead by I.-I. Popescu consisting of linguists and mathematicians, focused on general aspects associated with texts, namely text characterization, fixed points, vocabulary richness, thematic concentration, compactness, study of autosemantics, the impact of the position in sentence; this group also focused on the possibility of drawing conclusions from text to the morphology of language, etc. This investigation resulted in two books authored by I.-I. Popescu et al. (2009) and I.-I. Popescu, J. Mačutek and G. Altmann (2009).

The present book is a further development of methods introduced by the second stream but fully concentrated on the end-of-year speeches of Italian presidents. Moreover, the updated version of the corpus includes 60 end-of-year speeches of 10 presidents since 1949 to 2008. Our aim was to look at the highest floor of the linguistically perceived reality, namely the sets of word associations, compare these worlds and pursue a possible development in this thematically homogeneous corpus. Of course, not everything could be made but besides associations we added some chapters concerning the golden section, Zipf´s law, parts-of-speech analysis and some issues concerning vocabulary richness, continuing the examinations resulting from earlier common publications.

Acknowledgement

We want to express our gratitude especially to Mrs. Mihaela Vasilescu for her kind help in computing affairs in places where our abilities seemed to be exhausted. She followed our motto: *viribus unitis*.

A.T., I.-I.P., G.A.

Contents

1. Introduction

The number of methods that can be used for text examination is infinite. Their exploitation depends on the aim of the study. And the number of problems associated with every possible aim is not a decreasing but rather an expanding set. Every problem has many facets and every solution opens a new world of problems. This is caused by the fact that in science we always strive for better explanations. Either we "explain" a phenomenon by telling verbally "why" it behaves in the given way or we subsume it under a set of hypotheses (in the best case laws) having a formal character. But whatever we do, every answered why-question evokes the next one because explanations are hierarchic and even laws must be later on explained by more general ones. The situation seems to be hopeless but the opposite case is still more un-imaginable: what would a scientific discipline do after it has captured all phenomena of its universe of discourse?

In text analysis one always must take into account three kinds of restrictions:

(1) One must drastically reduce the objectives of the investigation,

(2) one must restrict the choice of data and

(3) one must restrict the possible set of procedures applied to the analysis.

In the first case, there is no possibility to explain or even capture everything by one approach. Science is a mosaic whose tessellation proceeds over centuries and changes incessantly. One can follow only one trace simultaneously but consequently and hope that somewhere one meets other tessellators who came from different directions and followed a quite different trace. Thus linguists may meet on their way psychologists, neurologists, biologists, physicists and other kinds of specialized scientists. The objectives are neither inherent in nor given by the object of investigation; they are the results of our striving for orientation, understanding/explanation and control.

In the second case the situation is two-sided: in a case study one takes the complete data that are at one´s disposal. This is, e.g., the situation in which the data are exactly limited as it is in the present book. On the other hand, if one strives for general statements about texts or genres or epochs, one can never obtain a satisfactory sample or a set of samples even if today the number of corpora increases daily. Corpora are no data, they are only collections of material frequently suitable for *creating* data for special purposes. One should not entertain the hope that a corpus consisting of billions of words can yield something else than grammatical or graphematic information which is relatively stable. In all other than grammatical or graphematic aspects it is a non-homogeneous set of unique states which must be examined separately and the resulting statements must be generalized very cautiously. A certain homogeneity of a set of texts may be achieved by restricting the aspect of investigation, by concentrating on a

common genre, small time interval, selected set of writers, etc. However, even one single text may contain non-homogeneities, as has been shown elsewhere.

In the third case the problem is even more overwhelming: at the lowest level one must decide for qualitative or quantitative proceeding. In the latter case, one can choose the deterministic or the probabilistic way, crisp or fuzzy reasoning, but the difference is not based on the properties of reality but on our concept formation. They are translatable into one another. Both ways are only approximations to something we call truth, a fairy we try to pursue. Sometimes we see it from distance but it is not relevant which way we have chosen to approach it. The only important thing is to test in some way that our result agrees with the starting hypothesis. And though statisticians did a lot of work to furnish us with appropriate methods, sometimes we feel like Heracles on the crossroad. Which of the tests is appropriate? This is not so much a question but rather a nightmare. In samples containing millions of entities (words, sentences, texts,…) the classical tests fail to work because even the smallest difference can be made significant. But the nightmare disappears if we begin to realize that a test does not tell anything about truth, it only gives us a motivation to decide in this or that way. If the test is positive, should we follow this way or not? But if we follow it, there is no guarantee that one day we shall see the fairy dancing in the night and meet all other scientists at the other side of the mosaic.

For all these reasons we restricted the scope of our investigation to a small set of texts, problems and methods. The texts consist of the collection of 60 end-of-year speeches of Italian presidents (Luigi Einaudi 1949-1954, Giovanni Gronchi 1955-1961, Antonio Segni 1962-1963, Giuseppe Saragat 1964-1970, Giovanni Leone 1971-1977, Sandro Pertini 1978-1984, Francesco Cossiga 1985-1991, Oscar Luigi Scalfaro 1992-1998, Carlo Azeglio Ciampi 1999-2005, Giorgio Napolitano 2006-2008). In Italy, the Presidency of the Republic is the first Office of the State and is the most important, most authoritative, and most beloved Institution of the Italian leadership. From a symbolic point of view the president represents the values and the essence of the nation; for the Constitution the president of the Italian Republic is the Head of State and represents the unity of the nation (Italian Constitution, par. 87). The duration of presidential term is seven years and each President usually delivered seven addresses (with three exceptions: Einaudi delivered the first address during the second year of his office, Segni resigned from his position after two years, Napolitano is the present President). Enrico De Nicola, Provisional Head of State and first President of the Republic (1946-48), is not considered in this analysis because the tradition of the end-of-year speech began with his successor, Luigi Einaudi. The speeches are available on the web site of the presidential office, i.e. the *Quirinale* (www.quirinale.it). Generally, the textual and the audio-visual (when available) versions differ slightly, and so manual corrections of the written texts downloaded from the site were needed in order to obtain the texts actually delivered by the presidents. For example, Pertini's speeches are often different from the

available written versions because of this president's habit of giving extemporaneous speeches. The corpus is available also in a lemmatized version. Lemmatization was conducted through a partly manual and partly automatic process associating each token form with a pair including a lemma and a grammatical category (a part of speech tag). In the traditional presidential end-of-year speech the president addresses directly the Italian citizens, which is a civil ritual and a media event. Besides the obvious contents of good wishes for the New year, these speeches are an important source of information on the last sixty years of Italian history.

For solving some problems the data are homogeneous, for solving other ones they are heterogeneous. A corpus having this character is a gate to an infinite paradise for textologists. The methods of investigation were adapted to solve the given problem or at least to show a perspective for further research and application to other languages.

In the second chapter we scrutinize the autosemantic word associations in texts, i.e. we study the compactness of the world view of the individual presidents, their similarities and differences and study the dynamics of the development. We necessarily chose one special way of investigation but did not exclude other ones which may lead, as we hope, to similar results.

In the third chapter we pursue the peculiar convergence of some text characteristics to the golden section known from many other sciences and arts. It seems to be present in texts in two different forms, but words seem to be the only entities displaying this behaviour. The given chapter applied to the Presidential speeches is only a stimulus for scrutinizing other textological entities. A first trial is made in the fifth chapter.

In the fourth chapter we test the validity of Zipf´s law on Presidential speeches and propose an alternative which seems not only to give better fits but in addition it yields an insight into some text phenomena which can be useful for future explanations.

In the fifth chapter the restricted classes of parts of speech are studied from the viewpoint of Zipf´s law. The results are similar to those in the previous chapter, i.e. the new approach yields better fitting functions but the convergence to the golden section is not given and its cause will be scrutinized in the future.

In the sixth chapter the well known problem of vocabulary richness is examined using a new method and the results are compared both to other Italian texts and to various texts in 20 languages. As shown earlier, this way of investigation can be useful in other domains of textology and typology, too.

In the seventh chapter we show a third method for computing similarity of texts without any recourse to individual words and text content, but useful for the study of style, text type or variation in a corpus of homogeneous texts.

The results are in no case final, even if Chapter 2 is rather extensive. There are a number of other mathematical possibilities that can lead – as we hope – to the same result, or in our terminology, to the same decisions. Our approach

is classical in natural sciences but not very beloved in linguistics. We adhere to Bunge´s (1967: 469) dictum: “A start in mathematical modelling, however unrealistic, is better than either a prolix but unenlightening description or a grandiose verbal sketch.”

2. Associative analysis

2.1. Introduction

Associative analysis is based on co-occurrences of some exactly defined basic entities in a priori defined environments. The basic entities can be syllables, morphemes, words, phrases or even non-identical units referring to the same thing in the reality or in the text. As a matter of fact, this was the starting point of denotative analysis. One tried to find the associations between signs denoting the same thing or referring to the same thing. However, the approach is general enough to be applied to any kind of associations. Here we shall use lemmas whose environment will be the sentence defined traditionally. In order to concentrate on a well defined set of texts we analyze all the end-of-year speeches of the Presidents of the Italian Republic beginning with Luigi Einaudi (1949) and ending with Giorgio Napolitano (2008). Our analysis tries to find overt or covert associations which may but need not be evident from the text. Each text (i.e. each end-of-year speech) and each President (i.e. his end-of-year speeches as a whole subcorpus) will be analyzed separately. The corpus is fully lemmatized, so we work with lemmas.

There is a number of different approaches by means of which one tries to find associations, e.g. conditional probability, conditional entropy, different indicators, etc. (cf. e.g. Pereira, Tishby, Lee 2008). Here we shall use exact probabilities computed as follows. The end of sentences is marked by a dot, but in written language other signs may also be interpreted as end of sentence (e.g. “!”, “?”, “;”, “:”). Let lemma *A* occur *m*-times in the text, lemma *B* *n*-times ans *N* is the number of sentences. Then the probability that they occur simultaneously in *x* sentences is given by the formula

$$P_x = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}, \tag{2.1}$$

a result that can easily be shown by combinatorial argument. It is irrelevant how many times the given word occurs in the same sentence. In order to compute the associativity, we need also the more extreme probabilities, i.e. if the common occurrence in *x* sentences is greater than the expectation, that is, if $x > mn/N$, we compute

$$(2.2) \quad P(X \geq x) = \sum_{j=x}^{\min(m,n)} \frac{\binom{m}{j}\binom{N-m}{n-j}}{\binom{N}{n}}$$

where X is the variable "number of common occurrences". An association is given, if $P(X \geq x)$ is smaller than a predetermined α. In that case (i.e. number of sentences that contain both *A* and *B*) we say that two lemmas are associated at level α. Two significantly associated lemmas can be considered as vertices of a graph. Hence computing all probabilities for pairs of lemmas we obtain the graph of associations – both latent and manifest – in the given text.

The evaluation of such graphs can be made categorically or one can take into account also the weights of the edges in order to obtain weighted graphs. We adhere to the former, which is simpler. Let us illustrate the procedure using the end-of-year speech of Napolitano in 2008. There are 63 (= *N*) sentences, the word *italiano*/Italian occurs 4 times (*m = 4*), the word *augurio*/wish three times (*n* = 3). Together they occur in 2 (= *x*) sentences. Since 2 > 2(2)/63, we compute (2.2) consisting of two terms

$$P(X \geq 2) = \frac{\binom{4}{2}\binom{62-4}{3-2}}{\binom{62}{3}} + \frac{\binom{4}{3}\binom{62-4}{3-3}}{\binom{62}{3}} = 0.0053$$

Setting our decision level e.g. at α = 0.01 we can consider this association significant.

The level of association can be determined conventionally but only graphs based on the same level can be compared.[1]

The program used for this computation (*Adjazenz* by R. Köhler) prints first all words (lemmas) and their respective frequencies, then the individual head words with the probability of their association with other words. Since the probabilities are given in increasing order, it is easy to see the significant associations. An example of the output can be seen in the following column

```
 1 (vigilia/eve) with:
---------
 11 1 0.01639 (Medio-Oriente/Middle East)
 10 1 0.01639 (cuore/heart)
 4  1 0.01639 (dominare/to dominate)
```

[1] For shorter texts in German, Ziegler and Altmann (2002) use α = 0.1.

```
8 1 0.01639 (immagine/picture)
6 1 0.01639 (sgomento/dismayed)
7 1 0.01639 (notizia/piece of news)
5 1 0.03279 (animo/spirit)
9 1 0.04918 (giungere/to come)
2 1 0.08197 (nuovo/new)
3 1 0.19672 (anno/year)
```

In the first column is the order number of the word, in the second column the number of co-occurrences of the given word with *vigilia*/eve, in the third column the probability of such a co-occurrence computed according to (2.2), and in the last column the associated word.

However, in this investigation, we require that all associated words occur at least twice. Otherwise, under special circumstances a random cooccurrence could often yield a significant result. Consider any two words such that $m = n = x = 1$. In that case the associative status depends only on the number of sentences N because according to (2.2)

$$P(X \geq 1) = \frac{\binom{1}{1}\binom{N-1}{1-1}}{\binom{N}{1}} = \frac{1}{N}.$$

Hence, if $N = 20$, we obtain $1/20 = 0.05$. That means that if there are 20 or more sentences in the text, every co-occurrence of hapax legomena in the same sentence signalizes a significant association at the $\alpha = 0.05$ level.

2.2. Associative concentration

The presentation of these associations in form of a matrix would be rather obscure (the matrices are larger than a printed page), we show in Figure 2.1 the association graph of Leone´s speech 1971, containing 106 lemmas. It is to be noted that using this way of computation one can take into account all lemmas; whether autosemantics or synsemantics, the computations do not change. One can include synsemantics in the final evaluation or omit them. In denotative or referential analysis they must be taken into account; in associative analysis they can be omitted. We preferred the latter way.

Figure 2.1. Associations at α = 0.05 in Leone´s 1971 speech

The text consisting of r = 106 non-hapax autosemantics (with frequency higher than or equal to two) did not contain many significant associations as can be seen in Figure 2.1. That means, it is not strongly associatively concentrated. There are many different themes which must be touched upon in an end-of-year speech. In order to express the **associative concentration**, we compute the relative number of edges in the graph. If all r vertices (i.e. autosemantics) would be joined with one another, we would have $\binom{r}{2} = \frac{r(r-1)}{2}$ edges, here 106(105)/2 = 5,565, and the graph would be *connected*. But in Figure 2.1 we find only k = 7 edges. We define the *relative associative concentration* using the indicator

$$(2.3) \quad C_{rel} = \frac{2k}{r(r-1)}$$

yielding in our case C_{rel} = 2(7)/[106(105)] = 0.0013. It can be assumed that generally this small concentration is a typical feature of speeches. Indicator (2.3) is at the same time the expression for the relative graph degree.

Looking at Figure 2.1 where we did not plot the isolated vertices, there are only 3 connected subgraphs called components (cf. West 2001:22) containing 8 lemmas. The number of all components is: number of lemmas minus the number of lemmas in connected (nontrivial) components plus the number of connected components. Here it is κ = 106 – 8 + 3 = 101. To avoid misunderstanding: k = number of associated vertices, κ (kappa) = number of components of a graph. Theoretically, there is at least 1 component in the graph (if all vertices are reachable from all the others) and at most r components (no vertex is connected with any other). Hence the measure of relative connectedness can be defined also as

$$(2.4) \quad \kappa_{rel} = \frac{r-\kappa}{r-1}.$$

In our case, we have

$$\kappa_{rel} = \frac{106-101}{105} = 0.0476,$$

showing that the associative concentration in Leone 1971 at the 0.05 level is very small. Other different indicators are known from the graph theory; for our purposes these two are sufficient.

In Table 2.1 we present the C_{rel} and κ_{rel} of all speeches. We set the significance level at $\alpha = 0.05$, i.e. we consider an association as given if $P(X \geq x) \leq 0.05$.

Table 2.1
C_{rel} and κ_{rel} in the presidential speeches

Year	**No of sentences N**	**No of auto-semantics r**	**No of. edges k**	**No of components κ**	C_{rel}	κ_{rel}
1949	5	83	0	83	0	0
1950	4	70	0	70	0	0
1951	7	109	3	107	0.0005	0.0185
1952	6	87	0	87	0	0
1953	6	88	0	88	0	0
1954	6	118	0	118	0	0
1955	11	161	4	157	0.0002	0.0250
1956	16	269	27	249	0.0007	0.0746
1957	25	381	67	330	0.0009	0.1342
1958	22	312	46	279	0.0009	0.1061
1959	15	276	27	256	0.0007	0.0727
1960	18	306	34	278	0.0007	0.0918
1961	25	427	126	353	0.0014	0.1737
1962	17	280	31	252	0.0008	0.1004
1963	24	382	40	350	0.0005	0.0840
1964	12	179	17	172	0.0011	0.0393
1965	34	359	31	333	0.0005	0.0726
1966	31	434	99	363	0.0011	0.1617
1967	34	379	38	346	0.0005	0.0873
1968	33	388	68	336	0.0009	0.1344
1969	60	481	129	391	0.0011	0.1917
1970	62	571	151	467	0.0009	0.1825
1971	11	106	7	101	0.0013	0.0476
1972	37	262	35	234	0.0010	0.1073

1973	41	429	76	370	0.0008	0.1379
1974	26	287	39	259	0.0010	0.0979
1975	41	456	92	395	0.0009	0.1341
1976	48	446	75	383	0.0008	0.1416
1977	50	484	108	403	0.0009	0.1677
1978	48	388	137	308	0.0018	0.2067
1979	115	513	177	390	0.0013	0.2402
1980	71	329	105	263	0.0019	0.2012
1981	135	546	250	401	0.0017	0.2660
1982	113	513	209	392	0.0016	0.2363
1983	197	701	348	507	0.0014	0.2771
1984	78	314	118	226	0.0024	0.2811
1985	70	581	315	424	0.0019	0.2707
1986	48	389	80	329	0.0011	0.1546
1987	49	662	203	538	0.0009	0.1876
1988	71	610	292	458	0.0002	0.2496
1989	52	546	201	435	0.0013	0.2037
1990	82	835	509	614	0.0015	0.2650
1991	16	155	10	146	0.0008	0.0584
1992	123	657	247	514	0.0011	0.2150
1993	138	716	215	579	0.0008	0.1916
1994	192	780	396	582	0.0013	0.2542
1995	235	858	413	644	0.0009	0.2497
1996	98	576	112	483	0.0007	0.1617
1997	343	860	317	662	0.0009	0.2305
1998	254	750	273	573	0.0010	0.2363
1999	107	552	106	461	0.0007	0.1652
2000	109	544	59	493	0.0004	0.0939
2001	114	604	108	518	0.0006	0.1426
2002	109	633	113	542	0.0006	0.1440
2003	97	492	66	437	0.0005	0.1120
2004	88	544	86	472	0.0006	0.1326
2005	58	342	49	301	0.0008	0.1202
2006	85	651	91	567	0.0004	0.1292
2007	53	562	83	490	0.0005	0.1283
2008	62	532	107	461	0.0008	0.1337

Concerning C_{rel} and κ_{rel} we have three options: either they are historically developing entities, or they depend on text length or they are stylistic text- or author dependent indicators. Let us scrutinize these problems step by step.

2.3. The indicator C_{rel}

If we multiply C_{rel} by 10,000 in order to obtain greater numbers, we can state that it is not correlated with text length N as can be seen in Figure 2.2. Hence it has nothing to do with text length.

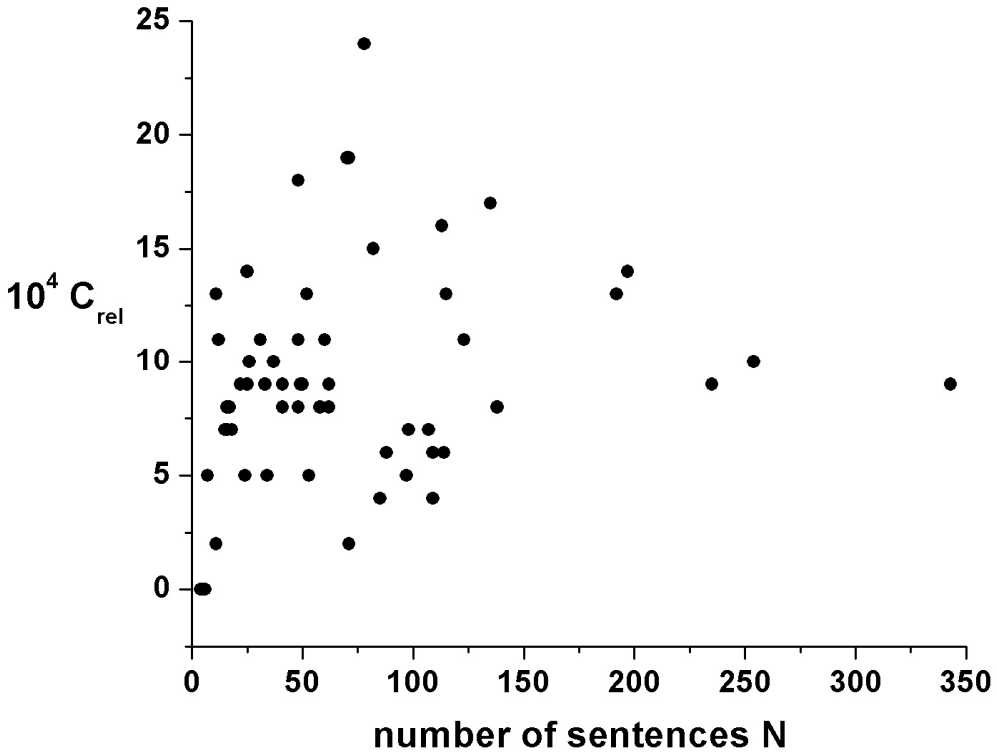

Figure 2.2. The relation of C_{rel} to N

The historical development of C_{rel} displays a peculiar movement as can be seen in Figure 2.3. The values seem to lie around a concave arc but not even polynomials of order 20 are able to capture the movement. In order to see the trend, we pool every ten classes (i.e. 1-10,11-20,…), multiply the C_{rel} values by 10000 and get their means; in the same way we simply call the years 1,2,3,… and take their means. In this way we obtain the result in Table 2.2. Here, the bow is evident, as can be seen in Figure 2.4 and we can capture it for example by means of the beta-function

$$y = 0.0192x^{1.1215}(60 - x)^{0.7815}$$

yielding $R^2 = 0.94$ which is rather a preliminary result. There is a number of functions yielding still better fits.

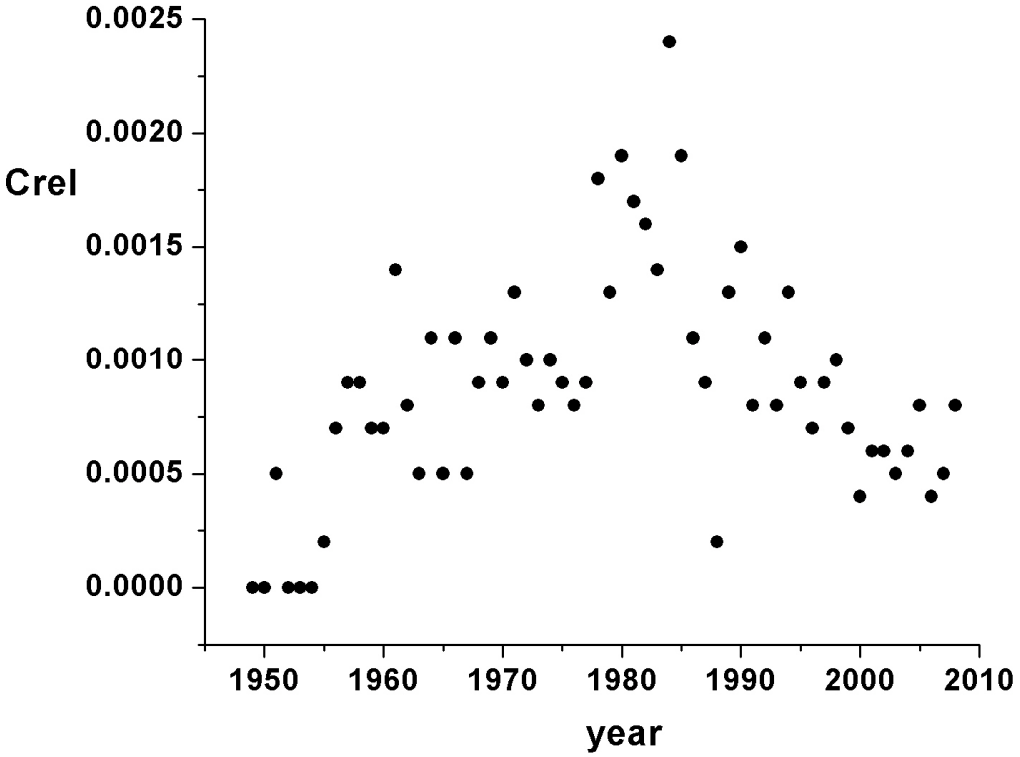

Figure 2.3. Historical development of C_{rel}

Table 2.2
C_{rel} means in 10 year intervals

Ten year period	Centroid	C_{rel}	$C_{rel,comp}$
1949-1958	5.5	3.2	2.96
1959-1968	15.5	8.2	8.08
1969-1978	25.5	10.5	11.58
1979-1988	35.5	14.4	12.84
1989-1998	45.5	10.3	11.26
1999-2008	55.5	5.9	5.64

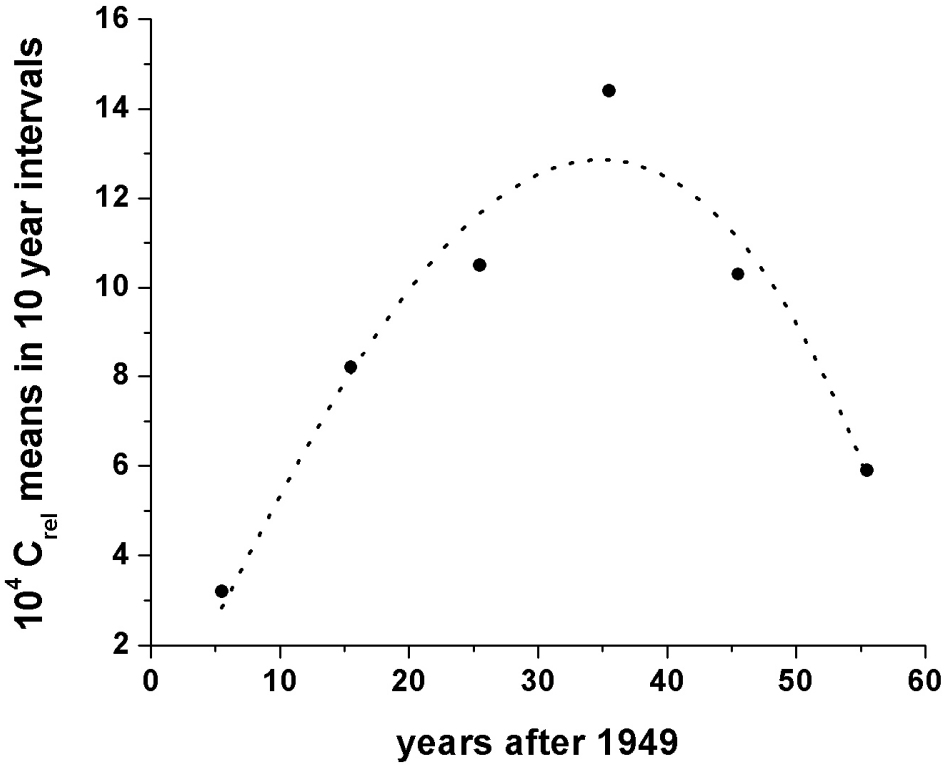

Figure 2.4. Historical development of C_{rel}

Perhaps non-presidential texts from the same years would strengthen the general historical trend, but the possibility may exist that there is no trend. The extrapolation on the basis of the above formula is impossible because we set 60 as the right boundary of the domain of x.

Thus C_{rel} is conditioned historically but only a thorough qualitative analysis of the given texts could reveal the very circumstances. It can be used as a style characteristic for other texts.

2.4. The indicator κ_{rel}

The indicator κ_{rel} in its relation to N also displays a very dispersed picture, though a quasi increasing concave movement can be visually discerned. Again, only a polynomial of high order could attain $R^2 \approx 0.87$, while simpler curves yield maximally $R^2 \approx 0.71$.

In its relation to years, κ_{rel} displays visually a dispersed bow as can be seen in Figure 2.5, which cannot be captured in a simple way. Hence we use the same procedure as above with C_{rel}.

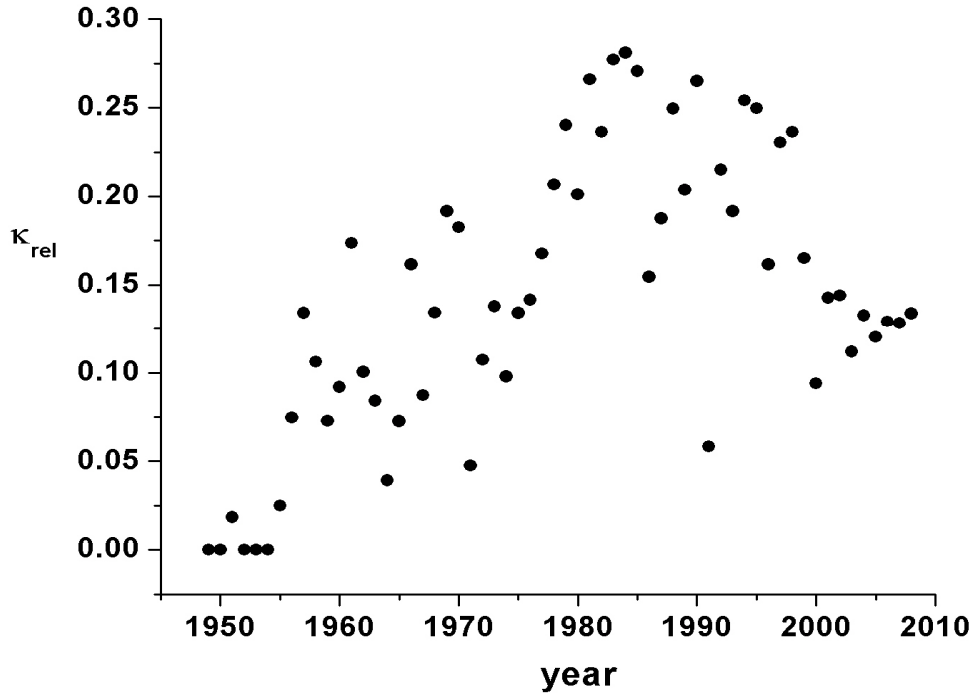

Figure 2.5. κ_{rel} in its relation to years

We divide the sixty years in intervals of 10, transform also x as above and obtain the result in Table 2.3.

Table 2.3
Time-dependence of κ_{rel}

Ten-year period	**Time x**	**$\kappa_{rel} = y$**	**$\kappa_{rel,comp}$**
1949-1958	5.5	0.0358	0.0248
1959-1968	15.5	0.1018	0.0962
1969-1978	25.5	0.1415	0.1663
1979-1988	35.5	0.2364	0.2130
1989-1998	45.5	0.2066	0.2143
1999-2008	55.5	0.1302	0.1302

Trying to use again the beta function we obtain

$$y = 0.000145x^{1.4417}(60 - x)^{0.6706}$$

yielding $R^2 = 0.95$. The result can be seen in Figure 2.6.

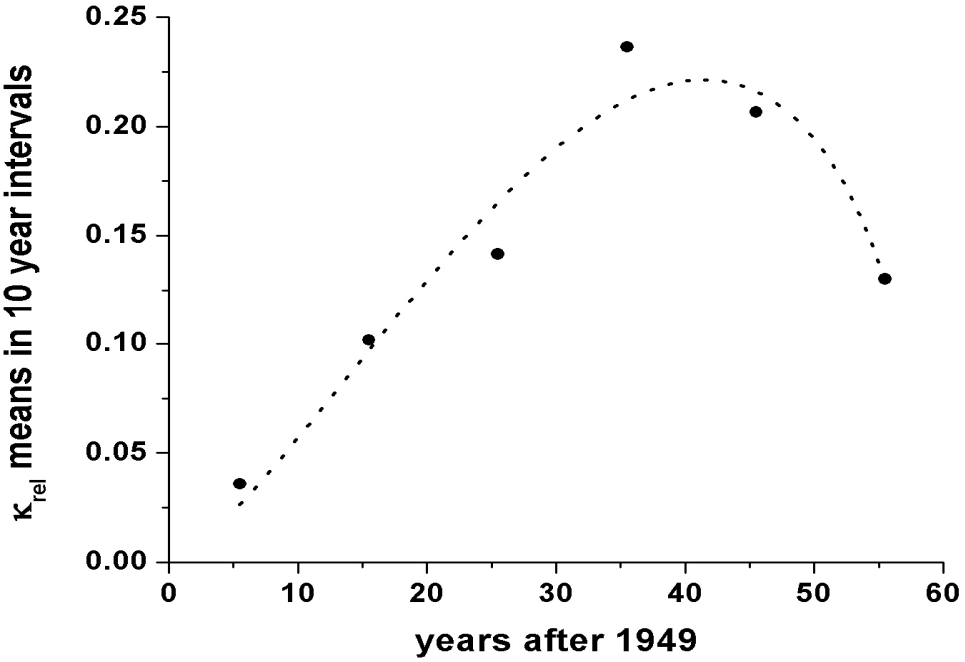

Figure 2.6. Historical development of κ_{rel}

Table 2.4
The relation of C_{rel} to κ_{rel}

C_{rel}	κ_{rel}	$\kappa_{rel,comp}$
0.0	0.31	0.00
0.2	13.73	9.46
0.4	11.16	11.64
0.5	9.68	12.56
0.6	13.97	13.42
0.7	11.32	14.26
0.8	12.63	15.09
0.9	16.96	15.91
1.0	14.72	16.72
1.1	15.25	17.54
1.3	18.64	19.21
1.4	22.54	20.07
1.5	26.50	20.94
1.6	23.63	21.83
1.7	26.60	22.74
1.8	20.67	23.68
1.9	23.55	24.63
2.4	28.11	29.85

The relationship between C_{rel} and κ_{rel} is not quite simple. In order to get a more lucid picture, we multiply C_{rel} by 1,000 and κ_{rel} by 100 and take means of κ_{rel} for each C_{rel} value. In this way we obtain the results in Table 2.4.

Though the dependence is relatively smooth, it is not simple. The theoretical course of κ_{rel} can be captured by the function

$$\kappa_{rel} = 613620288\ C_{rel}^{\ 0.2252}\ (22.5 - C_{rel})^{\ -5.6772},$$

whose first parameter is somewhat unreal. Nevertheless, $R^2 = 0.86$. The relationship is presented graphically in Figure 2.7.

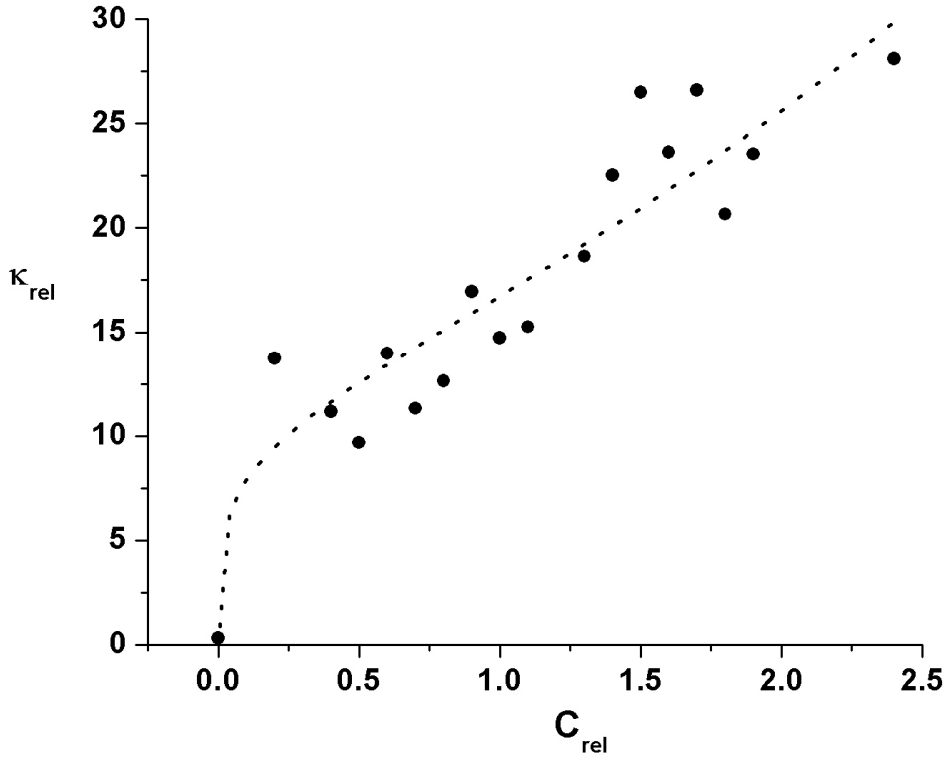

Figure 2.7. The relationship <C_{rel}, κ_{rel}>

2.5. The levels

Needless to say, we can determine α in several ways. It is a convention. However, one could determine different significance levels and characterize the associativity of the text by means of these levels. For example (cf. Ziegler, Altmann 2002):

α = that level at which no two words are associated according to (2.2), i.e. the probability slightly smaller than the minimum probability found;

β = that level at which there are no isolated vertices, i.e. all components consist of at least two vertices;

γ = that level at which the text represents one unique component, i.e. all vertices are mutually reachable.

The three given numbers are characteristic for every text and could be used in text analysis for different kinds of classification, as a background of content analysis and for historical studies. Above, we considered associations only if the concerned words occurred at least twice. Here we shall take into account also words occurring *once*, otherwise in each text many words would remain isolated and the results would be quite different. On the other hand, we must pay this allowance by the fact that we do not obtain any plain relationship between the levels, years and text length (in terms of sentence numbers). In each relation we find outliers disturbing the smooth course of a proposed curve. For example, in text 1976 the word *irresponsabile/*irresponsible occurring only once is associated only with *essere/*to be with $P = 0.54167$ which represents all levels for this word. For the whole text it is its β, yielding an outlier. Evidently, this technique should be used only with long texts and the evaluation of the levels must be delayed. For our texts the results of computation are presented in Table 2.5.

Table 2.5
Alpha, beta and gamma values of 60 texts

year	alpha	beta	gamma	year	alpha	beta	gamma
1949	0.10000	0.40000	0.90000	**1979**	0	0.07826	0.96999
1950	0.16667	0.50000	0.83333	**1980**	0	0.07042	0.97119
1951	0.04762	0.28571	0.71429	**1981**	0	0.03704	0.99378
1952	0.06667	0.33333	0.93333	**1982**	0	0.0531	0.99399
1953	0.06667	0.33333	0.60000	**1983**	0.00508	0.4467	0.95027
1954	0.06667	0.33333	0.93333	**1984**	0.01282	0.26923	0.61538
1955	0.01818	0.18182	0.93939	**1985**	0.00007	0.07143	0.97747
1956	0.00179	0.12500	0.98951	**1986**	0.00023	0.08333	0.99131
1957	0.00043	0.09565	0.95359	**1987**	0.00022	0.06806	0.96327
1958	0.00065	0.11688	0.90226	**1988**	0.0004	0.04225	0.97997
1959	0.00879	0.13333	0.99301	**1989**	0.00009	0.17308	0.97798
1960	0.00490	0.18301	0.97059	**1990**	0.00001	0.04878	0.97897
1961	0.00174	0.1200	0.9762	**1991**	0.00833	0.1875	0.8500
1962	0.0063	0.19118	0.96267	**1992**	0	0.09756	0.99254
1963	0.00362	0.16667	0.96281	**1993**	0	0.38406	0.96954
1964	0.01515	0.42424	0.93333	**1994**	0	0.06771	0.98706
1965	0.00178	0.11765	0.97118	**1995**	0	0.33617	0.97233
1966	0.00089	0.13447	0.98822	**1996**	0.00001	0.04113	0.91894
1967	0.00167	0.08824	0.98620	**1997**	0	0.40233	0.99318

1968	0.00073	0.12121	0.98185	**1998**	0	0.18110	0.98760
1969	0	0.10000	0.98529	**1999**	0.00002	0.07477	0.99309
1970	0.00003	0.06452	0.96832	**2000**	0.00017	0.05505	0.97081
1971	0.01818	0.27879	0.89394	**2001**	0.00011	0.08772	0.99239
1972	0.00013	0.08752	0.92847	**2002**	0.00001	0.06422	0.99775
1973	0.00038	0.12195	0.96486	**2003**	0.00021	0.07216	0.95449
1974	0.00234	0.08462	0.95435	**2004**	0.00026	0.05682	0.97080
1975	0.00094	0.09756	0.99397	**2005**	0.00051	0.08621	0.95813
1976	0.00058	0.54167	0.98462	**2006**	0.00004	0.07059	0.99588
1977	0.00020	0.08000	0.97629	**2007**	0.00017	0.09254	0.95855
1978	0	0.06250	0.95143	**2008**	0.00031	0.06452	0.95729

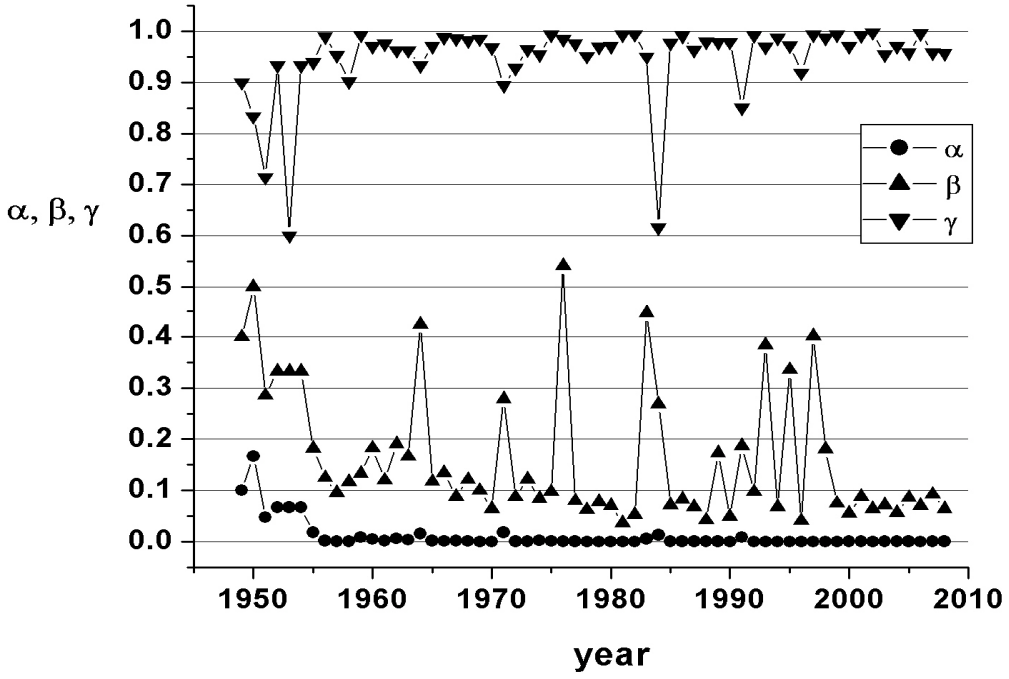

Figure 2.8. Alpha, beta, gamma sorted according to years

If we order the values according to years, then the convergences are evident as can be seen in Figure 2.8, but the results do not yield a historically relevant picture. The number of outliers or the oscillation in beta and gamma is too great and too irregular to say that the speeches have an associative tradition. As mentioned above, this is caused by the admission of words occurring once.

The dependence of levels on *N* (number of sentences) is, again, associated with the same problem. Nevertheless, α yields a quite clear dependence, as can

be seen in Figure 2.9. It could be captured by a simple monotonously decreasing function.

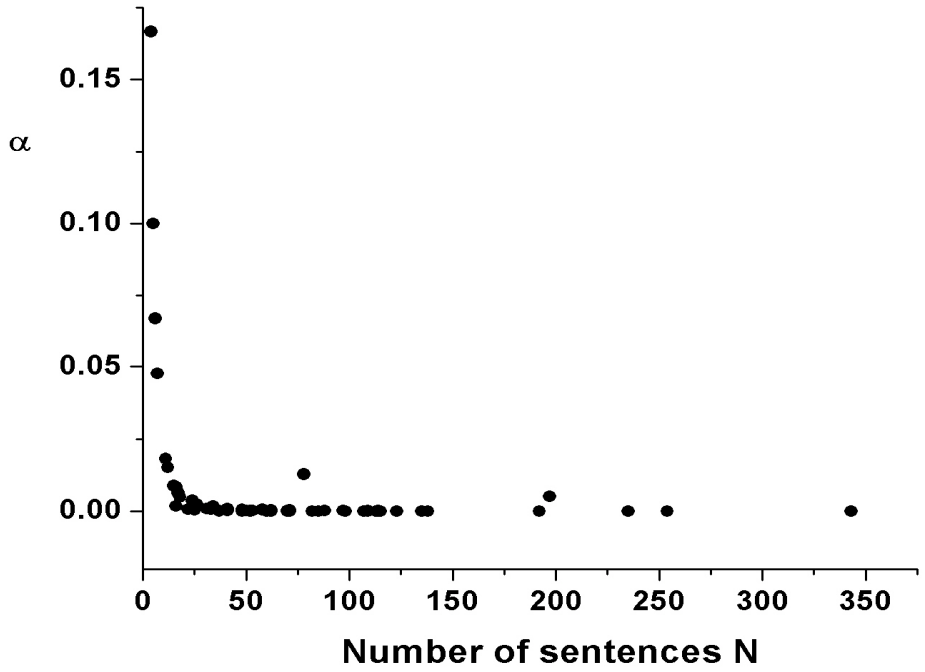

Figure 2.9. Dependence of alpha on *N* (sentence number)

For beta and gamma this rule does not hold. Trials to combine β and γ with some complex indicators led to the same picture: the outliers disturbed the smooth course of curves.

2.6. Cliques

There are groups of words in text all co-occuring with each other with the same probability and with no other word of the text with probability smaller than 0.05. Such groups are called here *cliques*. The criterion of equal probability is not necessary in unweighted graphs but here we set it as a condition. A further condition is that there must be at least three words in the clique, otherwise all associated pairs of words would form cliques. And, as mentioned above, associated words must occur at least twice. In any case, cliques present a very compact pattern of thinking which could, perhaps, be found also in other texts of the given authors. For example the words

{*migliorare, culturale, favorire*}/{to improve, cultural, to foster},

form a clique in Gronchi 1958, all words being mutually associated with P = 0.0043. In order to present a qualitative picture of thinking patterns we show all components (marked with "{ }") and cliques (marked with "< >") in Table 2.6. The texts 1949, 1950, 1952, 1953, 1954 do not contain associated words.

Table 2.6
Components and cliques in individual texts
Cliques are placed in <> with the given *P*

Text	**Components and cliques**
1951	<elevare, nuovo, via, P = 0.0476> 1
1955	{augurale, voto}, {altro, avvenire}, {proprio, sociale}, {legittimo, ricordare} 4
1956	{anno, italiano}, {avere, corso}, {arma, nazione, parte, unito}, {coscienza, forza, prendere}, {rapporto, violenza}, {iniziativa, responsabilità}, {attività, lavoro, più}, {libertà, spirito}, {fiducia, opera}, {esigenza, intento, libero, pace, peso, potere, proprio, solidarietà} 10
1957	{anno, civile, italiano}, {evento, vicenda}, {possibilità, soltanto}, {comune, raggiungere}, {esigenza, essenziale, istituzione, libertà, parte, popolo, potere, problema, prospettare, rinnovato, sempre, soluzione, volontà}, {andare, democrazia, organo, porre, positivo, solo}, {interesse, nazionale, società}, {articolare, crescente, già, giuridico, morale, politica, sociale, struttura, valido, vita}, {amministrativo, assunzione, capacità, cittadino, consapevole, costume, democratico, discussione, dovere, limite, pubblico, responsabilità}, {costituzione, giusto, repubblicano}, {carta, libero, pieno, proprio}, {nuovo, valore}, {uomo, dirittto} 13
1958	<culturale, favorire, migliorare, P = 0.0043>, {avere, clima, compressione, esigenza, giustizia, italiano, mutuo, popolo, progresso, solidarietà}, {avvenire, fare, fiducia, potere, sicuro} {comune, nazione, porre, spirito}, {coscienza, lavoro, maggiore, pieno}, {economico, politico, produttivo, senso, sociale}, {Italia, sforzo, Stato}, {capacità, valore, volere}, {ancora, libertà}, {atto, dovere, nazionale}10
1959	{anno, mondo}, {così, soltanto}, {migliore, sempre}, {coscienza, rapporto}, {arretratezza, miseria}, {andare, odio, paese}, {libertà, morale, progresso, rispetto}, {campo, convergenza, ormai, problema, soluzione, volontà}, <internazionale, grave, benessere, consolidare, P = 0.0095>, <riconoscimento, fare, valore, P = 0.0095> 10
1960	{anno, avvenire}, {iniziativa, tecnico}, {rimanere, unità}, {governo, italiano, potere, problema, secondo}, {capacità, dovere, grado, insieme, libero, mettere, nazione, sociale}, {accento, atto, civile, collaborazione, convivenza, creare, libertà, mondo, pace, parte, popolo, porre, tendere, uomo, volere} 6
1961	{auspicio, sermone}, {controllo, rapporto}, {meno, trovare}, {fare, or-

	dine}, {benessere, equo}, {progresso, sviluppo}, {avvenire, tempo}, {dare, tale}, {nazione, pace, pure}, {anno, assicurare, avere, campo, capace, capacità, civile, comune, condizione, considerare, continuo, cosa, così, costituzione, deficienza, democrazia, dire, economico, esperienza, favorevole, forza, insieme, interesse, internazionale, istituzione, Italia, italiano, lasciare, lavoro, libertà, lieto, lungo, necessario, oggi, organizzato, parte, particolarmente, partito, più, politico, porre, possibilità, potere, problema, progredire, progressivo, prosperità, pubblico, raggiungere, realizzare, rendere, ricchezza, riconoscere, ripresa, risultato, segno, sempre, sentire, sistema, sociale, solido, termine, tranquilità, trarre, vita} 10
1962	{anno, italiano}, {opera, problema}, {mondo, nazione}, {grande, svolgere}, {disposizione, mezzo}, {rapporto, sviluppo}, {civile, progresso}, {coscienza, potere}, {aumento, reddito}, {continuare, pace, stesso}, {civiltà, collaborazione, internazionale, Italia, popolo, volere}, {assicurare, bene, cittadino, classe, largo, partecipazione}, {costituzione, economico, eguaglianza, libertà, ordine, paese, persona, pieno} 13
1963	<conseguire, internazionale, ovunque, riconoscimento, sapere, P = 0.00362>, {avvenimento, profondo}, {animo, consentire}, {grande, uomo}, {indimenticabile, mobile}, {adoperare, ideale}, {apportatre, augurare}, {fiducia, seno}, {giustizia, libertà, via}, {commosso, pensiero, raccogliere}, {azione, ispirare, volere}, {concerto, Italia, tanto, testimonianza}, {dare, fare, più, potere, sempre, solidarietà}, {anno, avvenire, cuore, famiglia, italiano, lavoro, nuovo, popolo} 14
1964	{bene, comune}, {affettuoso, animo, augurio, caro, cordiale, famiglia, sereno} 2
1965	{affetto, atmosfera}, {colpire, disoccupazione}, {ente, sindaco}, {avere, caldo},{pace, politica}, {popolazione, sempre}, {contributo, Italia}, {produttivo, settore, sforzo}, {economico, lavoratore, operatore}, {chiudere, potere, ricordare}, {benessere, confortante, constatare, fare}, {anno, augurio, esprimere, favorevole, fine, inizio, più, produzione, rendere, rivolgere, vedere} 12
1966	{giovane, lavoratore}, {amministrazione, coordinamento}, {conto, tenere}, {difesa, politica}, {migliore, nuovo}, {governo, situazione}, {costo, prezzo}, {corso, già, solo}, {patrimonio, sofferenza, tanto}, <colpire, doloroso, perdita, P = 0.0022>, <ancora, progresso, pure, P = 0.0022>, {atto, produttivo, pubblico, responsabile}, {anno, chiudere, grazia, italiano, popolo, sorgere}, {attività, aumento, cittadino, collocare, contribuire, contributo, creare, dare, dovere, economia, economico, famiglia, figlio, forza, grande, industriale, ispirare, Italia, lavoro, levare, materiale, moderno, mondo, morale, nazionale, nazione, occorrere, opera, pace, paese, Parlamento, partito, più, popolazione, progredito, realtà, ripresa, sempre, settore, sindacato, sociale, spesa, sviluppo, tutto,

	umanità, umano, vero, vita} 14
1967	{italiano, Stato}, {più, umano}, {amministrazione, pubblico}, {adesione, giorno}, {affermare, repubblica}, {fiduciario, affermare}, {istituzione, libero}, {figlio, lavoro}, {paese, patria, secondo}, {augurio, bene, caro, concittadino}, {Italia, mondo, nazione, trovare}, {apertura, attacco, comunità, europeo, integrazione, mercato, processo}, {democrazia, dovere, economico, fermezza, libertà, pace, popolo, potere, progresso, solitario, soltanto, soprattutto} 13
1968	{capacità, struttura}, {Italia, strada}, {principio, pure}, {moderno, precedente}, {corso, dovere, economia}, {anno, augurio, finire, migliore, sorgere}, {assicurare, democrazia, giustizia, indipendenza, libertà, mondo, pace, patria}, {avere, capace, caro, categoria, ceto, comune, crisi, dirigente, diritto, economico, europeo, fare, grado, modo, morale, nascere, nazione, particolare, più, politica, popolare, potere, problema, proveniente, pubblico, raggiungere, rapporto, responsabilità, riforma, risolutivo, scuola, sentire, sviluppo, trattare, unto, vita} 8
1969	{coraggio, invece}, {legato, nuovo}, <contestazione, forma, generoso, P = 0.0006>, {affrontare, alto, anno, articolo, aspetto, avere, cercare, cittadino, collaborazione, compito, condizione, conseguenza, conseguire, coscienza, costituzione, creare, creazione, crisi, democratico, democrazia, derivare, diritto, dovere, economico, effettivo, equilibrio, equo, esistere, essere, Europa, fare, garantire, gettare, giustizia, grande, grave, impegno, insieme, istituzione, Italia, italiano, lasciare, lavoratore, lavoro, libero, libertà, livello, lutto, minoranza, modo, mondo, morale, occasione, pace, paese, pericolo, persona, pieno, più, politico, popolo, posto, potere, problema, progresso, raggiungere, realtà, repubblica, responsabilità, risolvere, risparmio, risposta, risultato, senso, sociale, soltanto, sopravvivenza, speranza, stesso, sviluppo, traguardo, trovare, tutto, umanità, umano, vedere, violento, violenza, vita} 4
1970	{lavoro, ritmo}, {ancora, produttivo}, {dare, significato}, {sapere, volere}, {innanzi, passo}, {atto, credere}, {quarto, secolo}, {cosa, democratico, sovranità}, {consiglio, regionale, statuto}, {dire, diverso, oggi, potere, principio, valere}, {altro, cittadino, fondamentale, già, nuovo, pace, presidente, regione, repubblica}, {affrontare, alto, ammonimento, andamento, anno, aumento, avere, bilancia, casa, congiuntura, considerare, corso, costituzione, democrazia, diritto, diventare, dovere, economico, equo, essere, esterno, famiglia, fare, fine, forma, giovane, giustizia, grande, grave, guerra, importante, infatti, inflazionistico, insegnante, intenso, internazionale, Italia, italiano, lavoratore, libertà, livello, mese, mezzo, misura, mondo, paese, pagamento, parte, più, politica, politico, porre, privato, problema, proprio, quanto, rappresentare, realtà, riforma, riserva, risolvere, rispetto, salvaguardare, sanità, scuola, segnare, sensibilità, sociale, solo, soluzione, studente, studio, sviluppo,

	tale, tensione, tradurre, trovare, ultimo, umano, vedere, vivere} 12
1971	{concordia, sentimento}, <posto, responsabilità, uomo, P = 0.0182 >, <mettere, più, realizzare, P = 0.0182> 3
1972	{anno, italiano, lavoro, proprio}, {ammirazione, lavoratore}, {maggiore, pace, popolo, sempre}, {azione, comune, comunità, dovere, economico, europeo, generale, nazione, paese, politica, potere, problema, provvedimento, realizzare, registrare, ripresa, sforzo, sintomo, soltanto, tempo}, {Italia, lavorare, progresso} 5
1973	<delitto, rapina, strage, P = 0.0012>, {crisi, vastità}, {affrontare, dubbio, fermo, paese, vero}, {augurio, difficile, essere, italiano, momento, problema}, {diritto, libertà}, {consapevole, necessario}, {salvaguardare, sociale}, {consentire, diretto, fare, forza, occorrere, ordine, potere, pubblico, sentire, sicurezza, società}, {bene, già, prezzo}, {avere, capacità, cittadino, civile, comprensione, credere, cuore, dare, difficoltà, dovere, governo, impegno, impulso, indicazione, livello, maggiore, operare, Parlamento, pieno, più, presente, provvedimento, quadro, riconoscimento, sempre, senso, sforzo, situazione, solidarietà, superare, ulteriore, vita, vivere, vivo} 10
1974	{convinto, Stato}, {istituzione, legge}, {giustizia, imporre, rappresentare, sala}, {colpire, concludere, dire, fine, grande, italiano, maggioranza, mondo, stare}, {capacità, chiedere, corso, dare, dovere, governo, momento, morale, paese, più, popolo, responsabilità, risorsa, senso, sentire, stanchezza} 5
1975	{delinquenza, forma}, {sviluppo, valido}, {economia, recessione}, {economico, sociale}, {ambito, problema}, {azione, giovane}, {attività, necessario}, {dire, messaggio}, {constatare, determinare, dovere, lavoratore, positivo, settore}, {avere, bene, imprenditore, porre}, {cuore, irresponsabile, spesso}, {eccitare, potere, purtroppo, solo}, {dare, forza, impegno, politica, solidale}, {anno, augurio, avvenire, capacità, civile, comune, consentire, crisi, difficoltà, disagio, esigenza, famiglia, fiducia, fondato, giustizia, grave, infatti, libero, maggiore, momento, morale, paese, più, popolo, portare, pubblico, realizzare, resistenza, ricco, rispondere, scorso, sistema, speranza, stesso, superare, vero, volontà} 14
1976	{consentire, responsabile}, {invece, occorrere}, {certo, dare, parte}, {inflazione, più}, {dovere, preoccupazione, struttura} {abbattere, colpire}, {difficoltà, lavoro, paese, tecnico}, {crisi, formulare, industriale, nuovo, problema, realtà}, {capo, difficile, fare, intatto, mantenere, momento, potere, produttivo, rendere, ricco, rischiare, sentire, serietà, solo, spesso, Stato, ulteriore}, {alimentare, causa, condizione, confermare, cosa, coscienza, criminalità, difficile, dire, economico, esprimere, fiducia, forza, funzionamento, giustizia, governo, invito, lotta, magistrato, momento, ordine, Parlamento, profondamente, proprio, pubblico, rendere, sembrare, senso, sentire, sforzo, sicurezza, situazione, sociale,

	spesso, spirito, venire, vero} 10
1977	{prospettiva, trovare}, {bene, quadro}, {crisi, grave}, {coscienza, pesante}, {determinare, predisporre}, {decisione, diritto, rendere}, {cittadino, ripresa}, {necessario, sforzo, specie}, {fiducia, immagine, paese, profondo, sala, solidarietà, stesso}, {spesso, troppo, vario}, {affermare, affrontare, anno, assurdo, avere, base, certamente, chiedere, civile, classe, collegamento, compiere, condizione, costituzione, democratico, difficile, dire, dovere, esigenza, fare, fondo, funzione, generale, giustizia, grande, impegno, lavoro, libero, mondo, nazionale, nuovo, oggi, operante, pensare, più, politica, politico, potere, previsione, problema, produttivo, programma, proprio, qui, rimanere, risolvere, ritenere, scuola, sempre, servizio, società, sostenere, speranza, Stato, strumento, uomo, valore, vivere, volere} 11
1978	{indice, ripresa}, {ondata, terrorismo}, {pubblico, sicurezza}, {credere, gioventù, frangia}, {augurio, popolo, presentare}, {adesso, animo, anno, armare, attrezzare, avere, carcere, celebrare, cittadino, contemporaneamente, creatura, difesa, dire, diritto, dovere, economico, esempio, estero, europeo, fame, fare, fede, forza, giovane, grande, interferenza, invece, Italia, italiano, legato, liberamente, libertà, lottare, mano, milione, modo, molto, nazione, ordine, Parlamento, particolare, patria, pensare, personale, più, potere, preoccupare, presidente, protestare, qui, realizzare, repubblica, ricevere, ripetere, ritenere, riuscire, rivolgere, sapere, scelta, scuola, sempre, sentire, sicuro, sociale, soffrire, sorgere, stare, Stato, superpotenza, umano, unità, uomo, vedere, vita, volere} 7
1979	{carta, costituzionale}, {camminare, insieme}, {soglia, varcare}, {invece, rovina}, {disoccupazione, giovanile}, <costruire, miliardo, ordigno, spendere, P = 0.0015>, {amico, andare, animo, anno, anziano, arma, ascoltare, atto, augurio, avere, carabiniere, caro, chiedere, connazionale, conoscere, conversazione, credere, dannato, danno, dare, depresso, destabilizzare, difendere, differenza, dire, disperare, dittatura, dovere, educatore, esempio, esperienza, essere, estero, Europa, fare, fertile, fervido, fine, forza, genitore, giorno, giovane, gioventù, guerra, illuminare, intero, ipotesi, Italia, italiano, lavorare, libertà, linea, lotta, luce, mai, mano, marca, mentalità, modo, molto, momento, mondo, morto, ordine, pace, paese, particolare, patria, pensare, perdizione, personale, pianeta, più, politica, poliziotto, popolo, potere, presidente, prima, qui, recare, rendere, repubblica, ripetere, rivolgere, sano, sapere, scatenare, scelta, sentire, situazione, solco, speranza, stabilimento, superare, terrorismo, trovare, umanità, uscire, vedere, vero, violenza, visione, vita, volere, volontà} 7
1980	{augurale, saluto}, {casa, paese}, {assistere, subito}, {avere, cuore}, {esperienza, speranza}, {camminare, insieme}, {fame, mondo, umano}, <invece, morte, umanità, P = 0.0004>, <amicizia, partito, solidarietà, P

	= 0.0004>, {forza, ordine, tanto}, {molto, sapere, serio}, <democratico, destabilizzare, regime, volere, P = 0.0004>, {amico, caro, sempre, uomo}, {anno, milione, morire, stare}, {affrontare, andare, antisismico, anziano, appartenere, cataclisma, cercare, colmo, contatto, coraggio, corrotto, criterio, dimostrare, disonesto, dovere, estero, fare, fermezza, gente, giovane, impegno, Italia, italiano, lavoro, meridione, pensare, persona, piazza, più, popolo, posto, prendere, protesta, qui, scatenare, sentire, sismico, terrorismo, trovare, venire, vittima, zona} 15
1981	{carcere, giovinezza}, {presidenza, repubblica}, {attentato, papa}, {adesso, manifestare}, {assistere, funerale}, {afghano, partigiano}, {domanda, porre}, {morte, ordigno}, {avvenire, elevare, protesta}, {americano, generale, sequestro}, {artista, grande, mano, merito, solo}, {cadere, codice, massoneria, morale, pidue}, {accadere, amare, andare, angosciato, animo, anno, appena, arabo, ascoltare, avere, battere, caro, città, civile, combattere, conoscere, consumare, cosa, coscienza, dire, dirigente, diritto, dittatura, domani, dovere, entrare, esortare, esplodere, essere, famiglia, fare, fatto, fianco, finalmente, forse, forza, fronte, già, giovane, godere, grave, guerra, indifferente, internazionale, intesa, invece, inviare, italiano, lavoro, legame, liaison, liberare, libero, libertà, lotta, mai, materiale, matrice, mila, misfatto, molto, mondiale, mondo, nazione, nominare, ordine, orrore, pace, paese, parlare, parte, patria, pericolo, più, polacco, politico, popolo, posto, potere, preoccupato, preoccupazione, proprio, protestare, qui, ragione, rendere, restare, ricevere, riguardare, rimanere, ripetere, sapere, scolaresca, scorso, sempre, senatore, sentire, situazione, soffrire, solidarietà, sospettare, spegnere, speranza, stare, Stato, studente, svolgere, talvolta, tanto, territorio, terrorismo, terzo, tormentare, tragico, tristezza, trovare, umano, uomo, usare, venire, veramente, vero, violenza, vita, volere, volta} 13
1982	{civile, diritto}, {assassinare, feretro, intorno}, {fine, ordigno, umanità, usare}, {arrivare, atto, dare, posto}, {amico, andare, anno, appello, armato, avere, avvenire, avversario, barriera, bisognare, calabrese, camorra, carcere, caro, casa, cercare, ceto, classe, combattere, compagno, così, dichiarare, dintorno, dire, domanda, dovere, emigrato, entrare, essere, fabbrica, famiglia, fare, figlio, forza, funerale, generale, già, giovane, governo, grande, grave, inflazione, internazionale, invece, Italia, italiano, lasciare, lavoro, legame, leva, libanese, liberamente, liberare, madre, mafia, malcontento, male, mattina, mila, molto, mondo, movimento, napoletano, nuovo, oggi, operaio, ordine, parte, patria, persona, più, poi, popolo, porre, potere, preoccupare, presidente, problema, proprio, protestare, provare, pubblico, quesito, qui, recente, restituire, ricevere, risolvere, riuscire, saluto, sapere, sconfitto, sempre, sentire, siciliano, silenzio, situazione, sociale, sorgere, sperare, stare, strada, studente, tanto, terrorismo, terrorista, tornare, trascinare, trovare, ultimo,

	veramente, vita, volta} 5
1983	{esuberante, età}, {angoscia, dolore}, {capo, Stato}, {internazionale, situazione}, {cuore, sorte,}, {bravo, soldato}, {presidente, strada, tanto}, {agente, carabiniere, pubblico, sicurezza,}, {adesso, affacciare, americano, amico, animo, anno, anziano, arrivare, ascoltare, avere, avvenire, avversario, bando, barriera, battere, bene, bisognare, buono, cagnesco, camorra, campo, carcere, carta, civile, competente, confondere, connazionale, considerare, contingente, continuo, controllare, controllato, corpo, cosa, così, costruire, curare, dare, democrazia, difendere, differrenza, dimostrare, dire, diritto, disarmo, disoccupazione, domanda, domani, dovere, ebreo, eroe, escluso, esistenza, esprimere, essere, Europa, europeo, fame, fare, fine, finora, forte, forza, francese, generoso, già, giornale, giovane, giovinezza, girare, giustizia, governo, grande, guardare, guerra, idea, intelligenza, invece, Italia, italiano, largo, lasciare, lavoro, libero, libertà, lottare, lungo, mafia, mai, male, massacro, mattina, merito, mese, miliardo, militare, minoranza, modo, molto, mondo, morire, morte, napoletano, nazione, nucleare, nuovo, oggi, opera, ordigno, ordine, orrendo, ospedale, pace, palestinese, parlare, parte, patria, peoccupazione, piazza, più, politico, popolazione, popolo, post, posto, potere, prima, pronto, punto, qui, ragazzo, ragione, regione, ricevere, riconoscere, rimanere, ripetere, sabato, sano, sapere, scatenare, scendere, scienziato, scorso, sempre, sentire, siciliano, sociale, società, solo, soprattutto, sperare, stare, studente, superpotenza, terminare, terra, terrorismo, tonnellata, totale, tranquilità, tremendo, troppo, trovare, turbare, ultimo, umanità, umano, unità, unito, uomo, usare, vedere, venire, veramente, vincere, vita, vittima, vivere, volere, volontà, volta} 9
1984	{pagare, prezzo}, {segreto, servizio}, {occupazione, soprattutto}, {incontrare, studente}, {economia, pieno, ripresa}, {campo, fertile, trasformare}, {domanda, fare, sentire, studiare}, {accordo, altro, ambasciatore, andare, animo, anno, ascoltare, assicurare, atomico, augurio, avere, bambino, caro, casa, cercare, chiedere, combattere, compatriota, conoscere, cosa, deserto, dire, disoccupazione, domani, dovere, emigrato, esperienza, essere, estero, fame, familiare, fervido, fine, giorno, giustizia, guerra, indagare, infamia, Italia, italiano, lavoro, male, marchio, mila, moderno, molto, morire, nuovo, paese, parente, passare, pianto, più, popolo, portare, posto, potere, rendere, ripetere, saluto, servire, sorgere, stare, stesso, strage, tecnologia, tonnellata, tremendo, triste, trovare, ultimo, umanita, valere, vedere, venire, visitare, vittima, viveri} 8
1985	{atto, inaudito}, {condanna, comprensione}, <aeroporto, ferire, morte, terrorista, uccidere, P = 0.00041>, {amministrazione, anno, anzitutto, aperto, armato, augurio, autorità, avere, battere, bene, buono, casa, causa, città, cittadino, civile, coerenza, comprendere, comune, comunità,

	conferire, conflitto, consapevole, consapevolezza, Costituzione, convivenza, coraggio, cultura, culturale, democratico, democrazia, difesa, difficile, dignità, diritto, dolore, domani, donna, dovere, economico, eguaglianza, esperienza, essenziale, essere, europeo, famiglia, fantasia, fare, festa, fondamentale, forma, forza, funzione, giustizia, grande, impegno, importante, imprenditore, indipendenza, iniziativa, innovazione, insieme, intelligenza, internazionale, interno, istituzione, Italia, italiano, lavoratore, lavoro, legge, libero, liberta, limite, lotta, lottare, messaggio, moderno, mondo, morale, nazionale, nazione, nuovo, operare, operativo, ordine, pace, pacifico, paese, Parlamento, parte, particolare, patria, patrimonio, personale, pienamente, più, poco, politica, politico, popolo, porre, potere, presenza, prima, problema, progresso, proprio, pubblico, purtroppo, realismo, realizzare, religioso, repubblica, responsabile, responsabilità, rivolgere, sacrificio, sapere, sconfiggere, scuola, sempre, sera, serenità, servizio, servire, sicurezza, significare, sindacato, sociale, società, soggetto, solo, soprattutto, speranza, spirito, tecnico, tenacia, terrorismo, trovare, uguale, umano, unità, unitario, unito, uomo, venire, vigore, violenza, vita, vivere, volere} 4
1986	{anno, ora}, {morale, responsabilità}, {senso, stare}, {economico, capacità}, {giovane, sociale}, {internazionale, pace}, {impegno, personale, sapere}, {comune, problema, questione, sempre}, {ascoltare, carenza, cittadino, esistere, pubblico, vitalità}, {crescere, democratico, dire, dirittto, dovere, società, soltanto, Stato}, {area, avere, bene, bisogno, comunita, conflitto, consapevolezza, convivenza, costante, essere, famiglia, fare, garantire, grande, insieme, istituzione, lavoro, legato, minaccia, necessario, nuovo, occorrere, ordinato, paese, più, potere, regola, responsabile, sicurezza, significativo, solo, stesso, superare, sviluppo, tensione, vedere, vita} 10
1987	{percorrere, strada}, {destino, vicenda}, {ruolo, svolgere}, {giusto, pubblico, sempre, utilità}, {incisivo, istituzione, soprattutto}, {anno, appuntamento, augurio, italiano, serenità}, {accentuato, adeguamento, affrontare, amministrativo, avere, bene, cammino, cittadino, civile, complesso, comportare, comune, condizione, consapevole, continuare, costituire, costituzione, credere, credibile, crescita, criminalità, culturale, democrazia, difficile, dimostrare, diritto, dovere, economico, efficace, efficiente, esercizio, esigenza, essenziale, essere, fare, fondamentale, formulare, forza, funzione, garantire, governo, grande, grave, idea, impegno, importante, indispensabile, industriale, inquietante, interesse, internazionale, interno, istituzionale, Italia, legge, legislativo, liberta, mercato, migliore, missione, moderno, momento, mondo, nazionale, nuovo, oggi, opera, operare, ormai, pace, paese, Parlamento, parte, pieno, più, politica, politico, popolo, potere, presenza, presidente, principio, profilo, progresso, promuovere, proprio, punto, puntuale, regione,

	rendere, repubblica, richiedere, riforma, riguardare, rilievo, rispetto, sapere, senso, servizio, sforzo, sicuro, sistema, sociale, soltanto, specifico, spirito, Stato, strumento, struttura, sviluppo, tale, vedere, volere} 7
1988	{fine, repubblica}, {presidente, rapporto}, {cittadino, diritto}, {entrata, vigore}, {internazionale, processo, sviluppare}, {affrontare, alto, amministrazione, anno, antico, avere, base, benessere, bisogno, centrale, certamente, certo, civile, codice, collettivo, comune, comunità, concreto, condizione, continuare, convivenza, costituzione, credere, crescente, crescita, criminalità, cultura, culturale, democrazia, dialogo, dimenticare, dovere, economia, economico, efficace, esempio, esigenza, essere, europeo, famiglia, fare, forse, forza, fronte, governo, grande, grazie, impegno, integrazione, intero, istituzione, Italia, lavoratore, legalita, legittimo, libertà, linea, livello, lotta, maggiore, mai, maniera, mercato, mezzogiorno, misura, mobilitazione, modo, mondo, morale, mutamento, natura, necessità, nuovo, operare, ora, ordine, organizzato, pace, paese, Parlamento, parte, particolare, penale, pensare, persona, pienamente, più, politico, popolo, potere, prevalere, primavera, principio, problema, procedura, prodotto, profondo, pronto, proprio, pubblico, questione, rapido, realizzare, regola, relazione, responsabilità, richiedere, riconoscere, riforma, rigore, rinnovare, segno, sembrare, sempre, serie, servizio, sforzo, sistema, sociale, società, solitudine, solo, soltanto, soluzione, soprattutto, specie, stare, Stato, stesso, straniero, stretto, struttura, sviluppo, tanto, tempo, toccare, traguardo, umano, unico, valore, vedere, venire, vista, vita, vivere, volta} 6
1989	{cadere, muro}, {civiltà, vita}, {accadere, agire, alleanza, anno, aperto, assumere, avanzato, avere, avvenimento, cammino, certo, cittadino, civile, comunità, condurre, consapevolezza, continuare, convivenza, così, credere, crescente, cultura, culturale, dare, democratico, democrazia, dovere, economico, edificazione, esercizio, essere, Europa, europeo, fare, forza, giusto, governo, grande, guardare, immane, impegnato, impegno, incontro, infatti, iniziare, inportante, internazionale, interno, Italia, italiano, liberazione, libero, libertà, materiale, milione, moderno, molto, mondo, morale, nazionale, nazione, necessario, nuovo, occidente, pace, pacifico, paese, parte, pieno, più, politica, politico, popolo, potere, progresso, giustamente, religioso, rendere, riconoscere, ricordare, riferire, rinnovamento, rispetto, segno, sembrare, sempre, sereno, sistema, società, solidarietà, solo, speranza, spirito, stagione, Stato, storia, struttura, sviluppare, sviluppo, tanto, termine, tirannico, umano, uomo, valore, vasto, venire, vento, vivere, volgere} 3
1990	{presagio, umanità}, {individuale, sempre}, {contributo, positivo}, {previsto, sorta}, {adeguare, affermare, alimentare, alleanza, ammodernamento, andare, anno, antico, anzitutto, aperto, apparire, armato, atlantico, attuale, augurare, augurio, autorità, avere, avvenire, avviare,

	capo, centrale, certamente, certo, chiedere, cittadino, civile, collaborazione, competere, comune, comunismo, comunità, concreto, confronto, consiglio, cooperazione, coraggio, così, costituzione, costruzione, credere, crescente, crisi, crollo, culturale, dare, delicato, democratico, democrazia, difensivo, difesa, dignità, dire, diritto, diverso, domanda, dovere, doveroso, duro, economico, edificio, effettivo, egoismo, esigenza, esprimere, essenziale, essere, Europa, europeo, fare, fatto, fermo, fine, forte, forza, garanzia, già, gioia, giustizia, giusto, governo, grande, grave, ideale, impegno, importante, inaccettabile, individuare, influire, indipendenza, iniziativa, istituzionale, internazionale, interno, intero, interrompere, invece, istituzione, Italia, italiano, legge, legittimo, libero, libertà, limitato, lotta, magistrato, mercato, milione, militare, miseria, misura, moderno, molto, momento, mondo, morale, nazionale, nazione, negativamente, nuovo, occorrere, oltre, operare, ordine, organizzazione, ormai, pace, pacifico, paese, parlamentare, Parlamento, parola, parte, persona, pieno, più, pluralistico, poi, polemica, politica, politico, popolare, popolo, porre, portare, potere, presente, presidente, pretesto, principio, privare, problema, processo, profondamente, proprio, provincia, pubblico, quadro, radicato, ragione, reale, realizzare, recare, regime, regione, relativo, religioso, repubblica, resistenza, responsebilità, ricordare, riforma, rischiare, risolvere, rispetto, rispondere, ritenere, sapere, servizio, sforzo, sicurezza, sincero, sistema, situazione, sociale, socialismo, società, soggetto, solo, sovranità, Stato, stesso, storia, storico, straordinario, strumento, struttura, sviluppo, tema, tempo, termine, trasformazione, trovare, tutela, unire, unità, unito, uomo, vasto, venire, veramente, verificare, vero, vittoria, volere, volontà} 5
1991	{anno, repubblica}, {fine, nazione}, {potere, presente}, {circostanza, dovere}, {coraggio, mandato}, {sembrare, tacere}, {proprio, uomo}, {legge, sempre, tradizione} 8
1992	{accogliere, bontà}, {parere, toccare}, {allargare, sguardo}, {disoccupazione, sottoccupazione}, {essenziale, proprio}, {impegno, presenza}, {chiamare, problema}, {ripetere, stesso}, {amare, amore, anno, anzitutto, apporto, aprire, articolo, attendere, attenzione, attuale, augurio, avere, bene, bisogno, buono, capace, capacità, capo, casa, certo, cittadino, colpevole, compiere, comune, controllo, coraggio, costituzionale, credere, dare, debito, debole, democratico, democrazia, denominatore, difendere, dignità, dire, diritto, disposto, diverso, dovere, economico, egualmente, esempio, essere, Europa, famiglia, fare, fede, fiducia, fondamentale, forza, garanzia, gettare, giungere, giudizio, giustizia, governo, grado, grave, grazie, incertezza, indifeso, indispensabile, Italia, italiano, lavorare, lavoro, libero, libertà, lontano, lottare, magistratura, mai, male, modo, mondo, nascere, nuovo, operare, ordine, pace, paese, Parlamento, particolare, partito, patria, penale, pensare, persona, più,

	politica, politico, popolo, potere, procedura, pubblico, pulizia, punto, qui, racogliere, ragione, rendere, responsabilità, risentire, risorgere, rispettare, rivolgere, saluto, sapere, sentenza, servire, sicurezza, soffrire, solo, spegnere, speranza, spugna, stanco, stare, Stato, tanto, tema, umano, uomo, valido, valore, vedere, verità, vero, vita, vitale, vivere, volere, volontà, volta} 9
1993	{cercare, fatica}, {instancabile, poi}, {responsabile, sicuro}, {lasciare, spazio}, {amministrazione, pubblico}, {sete, umanità}, {aiutare, ammirevole, amore, andare, anno, anzitutto, atto, augurio, avere, bisogno, buono, camminare, capace, certo, chiedere, chiesa, cittadino, comunità, conoscere, coraggio, credere, cuore, dare, delicato, democrazia, difficoltà, dimenticare, dire, diritto, disperati, dittatura, donare, dovere, eccezionale, economico, efficace, essenziale, essere, Europa, exdrogato, fare, fede, fermo, fiducia, forte, forza, generosità, già, giovane, giustizia, governo, grande, grazie, guardare, guida, guidare, immigrato, incontro, italiano, lavoro, libero, libertà, magistratura, mai, male, momento, mondo, nulla, nuovo, occorrere, onore, opera, operare, pace, paese, pagina, Parlamento, parte, passato, patria, pensare, pensiero, perdere, pessimista, più, politica, popolo, porre, potere, presenza, primario, proprio, qui, responsabilità, ripetere, risorgere, riuscire, sapere, scelta, scorso, sempre, sentire, sera, servire, sete, sofferente, solidarietà, solo, soprattutto, speranza, sperare, spes desperantium, stare, stesso, superare, tanto, tema, tenere, trovare, umanità, umano, uomo, valido, valore, venire, veramente, vicino, vita, vivere, volere, volontà, volta, voltare}
1994	{generale, interesse}, {libero, sentire}, {elettorale, poi, risultato}, {fermare, qui}, {manifestazione, pensiero}, {affrontare, alto, amore, andare, anno, anzitutto, appello, attendere, augurio, avere, bisogno, buono, Camera, capo, casa, chiamare, chiaro, chiedere, chiudere, cittadino, compito, comune, condizione, conoscere, consentire, considerazione, consiglio, constatare, consultazione, consumare, corretto, cosa, costituire, costituzionale, credere, crisi, dare, delicato, democratico, democrazia, difficile, dire, diritto, disoccupazione, diverso, doloroso, domani, dovere, eccezionale, elenco, elezione, eroico, essere, estero, fare, fiducia, finanziario, finta, forse, forza, fronte, giovane, giungere, giusto, governo, grande, grave, grazie, guardare, ignoto, immediato, impegno, incontro, innocente, insieme, intelligente, intenso, internazionale, interrogativo, Italia, italiano, italico, laboriosità, lavoro, libertà, maggioranza, mai, massimo, meno, meta, mettere, molteplice, momento, mondo, necessario, occorrere, offrire, oggi, ordine, organo, ovunque, pace, pagina, par condicio, parere, Parlamento, parola, parte, patria, pensare, permettere, persona, personale, piano, più, poco, politica, politico, popolo, porta, portatore, positivo, posizione, potere, presente, presenza, presidente, prevalere, problema, promesso, proprio, questione, quoti-

	diano, realtà, registrare, regola, religione, repubblica, responsabilità, richiamare, ricordare, ricorso, riforma, riguardare, ritenere, sacrificio, sapere, scusa, secondo, sempre, Senato, senso, sera, servire, sguardo, silenzioso, sociale, sofferenza, soffrire, solidarietà, solo, soluzione, soprattutto, sostenere, speranza, stare, Stato, stesso, strada, tanto, tema, tempo, terra, tesi, tornare, trovare, ultimo, umano, uomo, urna, valore, vano, vario, vedere, venire, vero, vertice, vita, vivere, vivo, volere, volontà, volta} 6
1995	{arte, cultura}, {accendere, accordo, addietro, aiuto, alto, andare, anno, anticipato, anzitutto, assumere, attendere, attenzione, attesa, attivo, atto, augurio, autonomia, avanti, avere, bene, bilancio, bimbo, buono, Camera, camminare, capace, capacità, capo, carta, casa, certamente, chiamare, chiarezza, chiaro, chiedere, chiudere, cittadino, collaborazione, comprendere, comune, conoscere, constatare, conto, convinto, corte, cosa, così, costituzionale, costo, credere, dare, democratico, democrazia, dettato, dicembre, difesa, dimostrare, dire, diritto, diverso, divisione, dovere, dubbio, elettorale, entità, esaminare, essenziale, essere, Europa, famiglia, fare, fiducia, forse, forte, forza, garanzia, genere, gente, giorno, giudice, giugno, giungere, giustificare, giustizia, giusto, governo, grande, gruppo, idoneo, impegno, indicazione, insieme, interesse, istituzione, Italia, italiano, largamente, lasciare, lavoro, legge, linea, maggioranza, maggioritario, magistrato, magistratura, mai, mandare, meritare, mese, mettere, mezzo, militare, modo, molto, mondo, motivato, movimento, nascere, nascosto, nome, oggi, operare, pace, paese, pagare, pagina, parlamentare, Parlamento, parlare, partecipazione, partire, partito, pensare, pensiero, persona, più, poco, polemica, politica, politico, polo, portare, potere, preciso, preghiera, prendere, preoccupazione, presentare, presente, presenza, presidente, prima, processo, proprio, proseguire, pubblico, qualificato, qui, ragione, rendere, responsabilita, ridurre, rispondere, risultato, sacrificio, sapere, scadenza, scioglimento, scrivere, semestre, semplice, sempre, Senato, sentire, sereno, servire, silenzio, situazione, soffrire, solo, soprattutto, specie, spegnere, speranza, sperare, stancare, stare, Stato, stesso, tanto, tema, tempo, tenere, tregua, ultimo, umanesimo, umano, umile, uomo, utile, vedere, venire, verità, vero, visione, vita, vivere, volare, volere, volontà, volta, voltare, voto} 2
1996	{giorno, portare, soldato}, {entusiasmo, sempre}, {intenso, valore}, {bambino, soffrire}, {elezione, politico}, {impegno, nuovo}, {dipendere, responsabilita}, {caduto, guerra}, {legato, ovunque}, {anno, buono, momento}, {aumentare, senso, timore}, {ampio, attento, bene, eccessivo, essere, famiglia, idoneo, incredibile, maggioranza, opposizione, più, silenzioso, volta}, {faticoso, rendere}, {sforzo, vincere}, {cuore, giustizia, mondo}, {intervento, nazione, unito}, {dare, pensiero,

	progetto, tempo}, <asurdo, così, inumano, P = 0.00021>, {affare, affettuoso, animare, assemblea, attesa, augurio, avere, azione, caro, città, compito, comune, conoscere, dignità, diritto, diventare, dovere, essenziale, estero, fare, formazione, governo, grande, grave, incontrare, indicare, insieme, Italia, italiano, modo, necessario, opera, pace, parlare, parte, pensare, persona, politica, popolo, potere, presente, problema, proprio, rapporto, rispetto, sentire, sofferenza, storia, tanto, terra, tragedia, verità, violenza, vissuto, vita, vivere, volere} 19
1997	{contribuire, fatto}, {bambino, schiavo}, {mai, sognare, spezzare}, {condannare, scorso}, {amnistia, indulto}, {condizione, ritenere}, {maggioranza, opposizione}, {rinunciare, visione}, {mettere, naso}, {problema, vivo}, {adesso, affermare, amare, andare, anno, associazione, assunzione, augurio, avere, bene, binario, bisognare, bisogno, buono, cadere, Camera, capacità, capo, carcerazione, caso, certo, chiamare, chiarimento, chiedere, citare, cittadino, civiltà, coloniale, competenza, compiere, comune, consentire, consiglio, coraggio, cosa, criminalità, cuore, dare, delicato, democrazia, difesa, dimenticare, dire, diritto, diverso, docente, dovere, dubbio, elogio, emozione, entusiasmo, esattamente, essere, estero, Europa, europeo, fare, fermare, fine, firmare, fratellanza, gente, giorno, giovane, giustizia, governo, grande, gratitudine, grazia, grazie, guardare, impegno, importante, imputazione, indefinito, iniziativa, internazionale, intesa, Italia, italiano, lasciare, lavoro, limitato, magistrato, magistratura, marcia, mezzo, ministero, ministro, modo, molto, momento, monco, mondo, morto, nascere, numero, ovunque, pace, paese, parere, Parlamento, parlare, parola, parte, partecipazione, passo, pena, pensare, pensiero, pericolo, persona, piacere, piano, più, poco, poi, polemica, politica, pontefice, popolare, popolo, portare, possibilità, potere, povero, prendere, presidente, preventivo, previsto, prima, proprio, punto, qui, ragione, reinserimento, responsabilità, ricevere, ricordare, riforma, rimanere, ringraziare, riuscire, sapere, scrivere, scuola, segno, sensazione, sentire, servire, settore, sguardo, situazione, sofferenza, solo, soltanto, soprattutto, spazio, specie, stabilità, stare, Stato, strada, studente, superiore, svolgere, tanto, tema, toccare, tornare, trovare, tutelare, umano, unione, unità, unito, uomo, vario, vedere, venire, via, viaggio, vicino, vita, vivere, volere, volotariato, volta} 11
1998	{morte, pena}, {legge, magistrato}, {fede, tenere}, {formazione, giovane, insegnamento}, {democrazia, libertà, vincere}, {alleanza, fedele, mai}, <arma, commercio, droga, P = 0.00003>, {accordo, amore, andare, anno, attendere, attenzione, atto, augurio, aumentare, avere, avvenire, avvicinamento, avvicinare, bambino, bene, bisognare, buono, capacità, capo, certo, chiamare, chiedere, citare, commuovere, compiere, compito, comunità, consiglio, cosa, così, costituzione, credere,

	criminalità, cultura, dare, dicembre, dignità, Dio, dire, diritto, dovere, ecologia, entrare, episodio, esempio, essere, estero, Europa, fare, fatto, fiducia, forza, frutto, gente, già, giorno, giudice, giuridico, grande, grazia, grazie, grido, importante, importanza, incontro, insieme, interesse, Italia, italiano, lasciare, lavoro, legame, lettera, lotta, magistratura, male, messaggio, militare, ministro, modo, moltiplicare, molto, momento, mondo, nascere, nessuno, nuovo, occorrere, ordine, ovunque, pace, paese, Parlamento, parlare, parte, particolare, passare, passato, passo, patria, pensare, pensiero, persona, più, poco, politica, politico, popolo, portare, potere, presente, presenza, presidente, prima, professsore, punto, quotidiano, raccoglimento, regione, ricchezza, ricordare, ricucitura, rilievo, rimanere, ringraziare, rispetto, salvatore, sapere, scorso, scuola, seguire, sempre, sentire, serata, signore, sofferenza, solidarietà, solo, soltanto, soprattutto, spendere, sperare, stare, Stato, statura, strada, studiare, tanto, televisione, tema, temere, tempo, trovare, ultimo, umanistico, umano, unito, vedere, veramente, viaggio, visita, vita, vittoria, voce, volere, volta, votare} 8
1999	{integrazione, processo}, {capitale, umano}, {cammino, confermare}, {patria, unità}, {certo, conflitto}, {caro, giovane}, {esprimere, iniziativa}, {casa, italiano, sentire}, {lasciare, spalla, vedere}, {diritto, cultura, economico, finanziario, sociale}, {anno, crescere, milione, numero, oggi}, {andare, assicurare, avere, bello, benessere, campo, capitale, città, cittadino, cominciare, comune, condizione, consapevole, costituzione, creare, crescente, dare, diventare, dovere, essere, Euro, Europa, europeo, fare, fiducia, fondamentale, fondato, forte, forza, generazione, già, governo, grande, guerra, intero, istituzione, Italia, lavoro, maggiore, male, mantenimento, meta, mondo, moneta, nuovo, occorrere, operare, pace, paese, parte, più, popolo, potere, prestigio, progresso, repubblica, responsabilità, rischiare, sapere, secolo, sempre, sfida, sicurezza, soltanto, soprattutto, sovranazionale, stare, Stato, stesso, tema, umano, unione, valore, vita, vivere} 12
2000	{mente, pensiero}, {dare, forza}, {chiedere, giusto}, {comune, stesso}, {dire, volere}, {affettuoso, augurio}, {presidente, repubblica}, {milione, sentire}, {alto, tenere}, {sempre, volta}, {avere, campagna, elettorale}, {amore, anno, buono, cercare, conoscere, cosa, coscienza, democrazia, dovere, modo, solo, svolgere, votare, voto}, {Europa, europeo, fare, fine, grande, ideale, identità, iniziativa, Italia, legge, meglio, modo, necessario, nuovo, ora, pace, patria, più, progresso, proprio, rafforzare, rispetto, sentimento, singolare, stare, storia, svolgere, valore} 12
2001	{giovane, università}, {dire, significato}, {modo, progresso}, {parlare, potere}, {fare, guerra, orgoglioso}, {acciaio, amore, augurio, avanguardia, avere, avvenire, bene, buono, carbone, caro, circolare, cittadino, civiltà, compito, comune, così, costituzione, costruire, cuore, dare, dia-

	logo, diventare, dovere, esercitare, essere, estero, Euro, Europa, europeo, evento, fiducia, fine, giuramento, governare, governo, grande, idea, ideale, impegno, indipendenza, insieme, integrazione, ispirare, istituzione, Italia, italiano, libertà, mente, mettere, milione, mondo, moneta, nazionale, necessario, oggi, operare, ora, ovunque, pace, paese, patria, pensiero, più, politico, popolo, porre, principio, pubblico, rivolgere, secolo, sempre, stare, Stato, storia, svolgere, terra, unico, unione, unità, vicino, vivere} 6
2002	{augurio, speziale}, {capacità, provincia}, {fiducia, godere}, {coscienza, punto}, {confronto, Parlamento}, {società, vivere}, {bene, fare}, {alto, regione}, {minaccia, terrorismo}, {andare, spirito, vero}, {accrescere, fine, impegno}, {affermare, internazionale, mezzo}, {buono, dare, funzionamento, violenza, vita}, {Europa, generazione, operare, stare, unito}, {comunale, creare, europeo, nuovo, proprio, Stato, unione}, {benessere, caro, cominciare, comunità, forma, forte, gente, incontro, italiano, mondo, naturale, più, raggiungere, vario, vicino}, {anno, assicurare, avere, bisogno, buongoverno, centrale, coesione, collaborazione, costituzione, crescente, dire, disoccupazione, dote, dovere, essere, favorire, forza, futuro, garantire, giustizia, governo, guerra, istituzione, interno, Italia, libertà, livello, locale, magistratura, molto, nazionale, nazione, necessario, pace, parlamentare, parte, politico, popolo, regola, riconoscenza, rispetto, soffrire, solidarietà, solo, stesso, tanto, unità, vivere, vivo, volere} 17
2003	{andare, pensiero}, {famiglia, società}, {forza, parte}, {diritto, rispetto}, {guardare, preoccupazione, scelta}, {economico, intero, ripresa, sistema}, {costituzione, dare, essere, scuola}, {affrontare, amore, animo, anno, bene, coesione, costruire, crescere, diffusa, dovere, Europa, europeo, fare, fiducia, fondamentale, forte, governo, istituzione, Italia, italiano, mancare, nascere, natura, nazionale, nazione, nuovo, oggi, pace, patria, più, politico, reagire, religione, religioso, repubblica, sapere, sempre, sentire, soprattutto, spirito, unione, uomo, valore, vivere} 8
2004	{forte, nazione}, {condizione, popolo}, {ovunque, presente}, {là, Asia}, {combattere, vita}, {contare, potere}, {affrontare, amato, amicizia, anno, attesa, augurio, avanzato, avere, caduto, caro, catastrofe, colpire, cuore, dare, disastro, dovere, essere, estero, Europa, europeo, fare, fiducia, forza, giorno, giovane, grande, immagine, immane, impegno, Italia, italiano, lanciare, libertà, mai, mente, mercato, mese, mezzogiorno, migliaio, molto, mondiale, mondo, naturale, nuovo, oggi, pace, paese, pensiero, più, problema, proprio, realizzare, ricordo, rischiare, rivolgere, sapere, sfida, sofferenza, speranza, stare, storia, tanto, tempo, terra, uomo, vedere, vivere} 7
2005	{incontro, trarre}, {identità, valore}, {dovere, sentire}, {animare, estero}, {intendere, stasera}, {nascere, pace}, {europeo, unione}, {augurio,

	particolare} {fondamentale, società}, {famiglia, patria, stesso}, {anno, avere, caro, cittadino, comune, cuore, dare, esprimere, essere, forza, iniziativa, insieme, Italia, italiano, lungo, mandato, mente, piazza, più, popolo, presidente, provincia, repubblica, sindaco, spirito, tanto, tricolore, vita, vivere, vivo, volere} 11
2006	{anno, buono}, {capo, Stato}, {caso, essenziale}, {migliorare, rinnovare}, {ragazzo, scuola}, {cominciare, sensibilità}, {famiglia, vita}, {essere, giusto}, {dialogo, reciproco}, {dire, mese}, {minacciare, stesso}, {complesso, delicato}, {nuovo, responsabilità}, {giovane, lavoro, molto}, {paese, sviluppo, umano}, {crescere, insieme, momento}, {civiltà, Europa, pace, profondo}, {avere, civile, dare, difficoltà, dovere, esprimere}, {coesione, condizione, costruttivo forte, impegno, maggiore, particolare, più, ruolo, significare}, {europeo, fiducia, forza, futuro, guardare, incontrare, partecipare, pensare, poco, ricerca, rispetto, sapere, superare, tanto}, {allontanare, ascoltare, attuale, autonomia, bene, cittadino, comune, confronto, Costituzione, contare, crisi, decisione, fare, generale, incontro, indirizzare, indispensabile, interesse, internazionale, istituzione, Italia, italiano, mondo, oggi, Parlamento, parte, politica, potere, proprio, radicale, scelta, sentire, solo, tenere, venire, volere} 21
2007	{dedicare, esprimere}, {abbastanza, proporre}, {modo, speciale}, {dire, Italia}, {augurio, sincero}, {dinamismo, dovunque}, {garanzia, reciproco}, {sempre, soluzione}, {azione, sicurezza}, {proprio, ultimo}, {esigente, impegnare, stesso}, {incontro, vedere, volere}, {equilibrio, istituzione, politica, sfera}, {crescere, giovane, impresa, lavoro, malessere}, {apprezzare, avere, condizione, contare, creatività, culturale, dovere, effettivo, esempio, esigere, Europa, fare, fiducia, forza, generale, grande, guardare, insieme, insufficiente, interesse, italiano, libertà, lontano, mai, migliaio, moderno, momento, mondo, motivo, nuovo, paese, parte, particolarismo, patrimonio, paura, più, politico, positivo, potenziale, potere, recente, religioso, richiamare, richiesta, riforma, risposta, sfida, solo, stare, storico, tutela, vita} 15
2008	{andare, sforzo}, {sentire, vicino}, {pace, regione}, {diritto, rispetto}, {parte, Parlamento}, {crescere, diverso}, {pieno, venire}, {grave, porre}, {comune, esprimere}, {quando, riuscire}, {affrontare, problema}, {anno, attendere, iniziare}, {carico, dovere, lungo}, {aperto, augurio, avere, capace, civile, colpire, comunque, condizione, conesquenza, creare, crisi, critico, difficile, economia, economico, esigenza, famiglia, fare, finanziario, fondamentale, forte, forza, futuro, generazione, governo, grande, guardare, impegnare, impegno, Italia, italiano, lavoro, messaggio, missione, modo, mondiale, mostrare, nazionale, nuovo, occasione, operare, paese, più, potere, produttivo, prova, pubblico, reddito, rivolgere, scelta, sistema, sociale, sollecitare, sviluppo, valore, vita, vivo} 14

However, the study of cliques as defined above can turn out to be important only in short texts, because in long texts the cooccurrence probabilities tend to be very different. Or, one sets up different criteria for building cliques.

2.7. The associated vocabulary

The words wired in some way in the head of the writer have the tendency to appear in texts in associated form. These words form the core of the text. Whenever the writer utters a word from this stock, the probability that the associated word will appear in the same sentence increases. Through the analysis of these lists it is possible to individuate values and reconstruct the presidential political plans. Einaudi appears to be the spokesperson of an archaic lexicon and of a country which is not yet fully developed. Pertini shows colloquial and informal features, but also tragic themes discussed with great emphasis. In Ciampi's speeches the concept of homeland finds renewed strength, within a perspective that promotes nationalism from local autonomies to the entire country and as far as Europe. Cossiga seems concerned with innovating in his choice of international and institutional themes. Saragat stands out, who first clearly opens the doors to political lexicon, and so on. This central part of the vocabulary can be approached in different ways. There are, of course, also word pairs clustering in the text on grammatical reasons or because of semantic restrictions (valency) but these can be filtered out if one takes into consideration only autosemantics or removes the pairs of them authoritatively. They can be parts of the phraseology.

2.7.1. Thematic concentration

As has been proposed in Popescu et al. (2009) thematic concentration can be computed using the words or word forms whose ranks are smaller than *h*, which is the *h*-point of the rank-frequency sequence. Since we use only autosemantic lemmas, the computation algorithm must be slightly modified. And since we do not consider frequencies but associations, we treat degrees, i.e. the number of associated words. The procedure is as follows:

1. Using a lemmatized text containing all autosemantic lemmas compute the rank-degree distribution of the text, i.e. arrange the degrees in decreasing order and ascribe them ranks.
2. Compute the *h*-point of the distribution according to

$$h = \begin{cases} r, & \text{if there is an } r = f_r \\ \dfrac{f_i r_j - f_j r_i}{r_j - r_i + f_i - f_j} & \text{if there is no } r = f_r \end{cases} \tag{2.5}$$

i.e., first seek a rank which is equal to its degree (*f*) and use the first part of the formula. If there is no such *r*, then take two (neighbouring) r_i and r_j ($r_i < r_j$) such that $f_i > r_i$ and $f_j < r_j$. Insert them in the second part of the formula.

3. Having the *h*-point, consider all autosemantics whose degree is smaller than or equal to *f*(*h*). These autosemantics occur frequently in the text and determine the theme of the text. To express the thematic concentration, compute

$$TC = 2\sum_{r'=1}^{n} \frac{|h - r'|\ f(r')}{n(n-1)\,f(1)}, \tag{2.6}$$

where *h* is the *h*-point, *r´* are the ranks of autosemantics smaller than or equal to *f*(*h*), and *f*(*r´*) are the degrees of these ranks, *n* is the number of words whose degree is smaller than or equal to *f*(*h*). The expression is normalized using its maximum *n*(*n*-1)*f*(1)/2 where *f*(1) is the degree of the most associated word.

For the sake of illustration we show the rank-degree distribution of autosemantic lemmas in the presidential speech of Gronchi 1961 as presented in Table 2.7.

Table 2.7
The rank-degree distribution of Gronchi 1961

Word	**Rank**	**Degree**	**Word**	**Rank**	**Degree**	**Word**	**Rank**	**Degree**
sentire	1	11	dire	29	3	dare	57	1
problema	2	10	necessario	30	3	tale	58	1
insieme	3	9	tranquillità	31	3	possibilità	59	1
continuo	4	9	ricchezza	32	3	potere	60	1
capace	5	9	assicurare	33	3	termine	61	1
forza	6	9	prosperità	34	2	oggi	62	1
progredire	7	9	pubblico	35	2	tempo	63	1
rendere	8	8	deficienza	36	2	avvenire	64	1
democrazia	9	8	particolarmente	37	2	progresso	65	1
istituzione	10	8	lieto	38	2	sviluppo	66	1
anno	11	6	raggiungere	39	2	parte	67	1
sociale	12	6	lavoro	40	2	fare	68	1
avere	13	5	trarre	41	2	ordine	69	1
comune	14	5	lungo	42	2	equo	70	1
libertà	15	5	Italia	43	2	benessere	71	1
politico	16	5	così	44	2	risultato	72	1
riconoscere	17	4	più	45	2	porre	73	1

realizzare	18	4	cosa	46	2	Costituzione	74	1
progressivo	19	4	esperienza	47	2	campo	75	1
solido	20	4	internazionale	48	2	sistema	76	1
vita	21	4	favorevole	49	2	auspicio	77	1
condizione	22	4	considerare	50	2	sermone	78	1
organizzato	23	4	partito	51	2	pace	79	1
economico	24	3	interesse	52	2	pur	80	1
capacità	25	3	sempre	53	2	meno	81	1
ripresa	26	3	civile	54	2	trovare	82	1
segno	27	3	nazione	55	2	rapporto	83	1
lasciare	28	3	italiano	56	1	controllo	84	1

As can easily be seen, the *h*-point is 8 but there are three words having this degree. Thus we add them all to the set of pre-*f*(*h*) words and obtain the set *f*(*h*) = {*sentire, problema, insieme, continuo, capace, forza, progredire, rendere, democrazia, istituzione*}/{to feel, problem, together, continuous, able, strength, to develop, to render, democracy, institution}, *n* = |*f*(*h*)| = 10. We compute *TC* as follows

$$TC = 2[|8\text{-}1|11 + |8\text{-}2|10 + |8\text{-}3|9 + |8\text{-}4|9 + |8\text{-}5|9 + |8\text{-}6|9 + |8\text{-}7|9 + |8\text{-}8|8 + + |8\text{-}9|8 + |8\text{-}10|8]/[10(9)11] = 0.5980$$

For the individual texts we obtain the thematic concentration as given in Table 2.8.

Table 2.8
Thematic concentration TC for individual texts
(*h* is rounded to integer)

Text	**Pre-h autosemantics**	**h**	**n**	**TC**
1949Einaudi	anno, patria, potere	2	3	
1950Einaudi	anno, paese, uomo, città, borgo, casolare, volere	2	7	
1951Einaudi	via, nuovo, elevare	1	3	1.0000
1952Einaudi	anno, nuovo, italiano, augurio, patria, via	2	6	
1953Einaudi	nuovo, anno, popolo, speranza, richiamo	2	5	
1954Einaudi	anno, nuovo, pensiero	3	3	
1955Gronchi	augurale, avvenire, legittimo, proprio, ricordare, sociale, voto	1	7	1.3333
1956Gronchi	proprio, intento, libero, potere	4	4	0.9000
1957Gronchi	capacita, responsabilità, pubblico, democrazia	4	4	0.9762

1958Gronchi	economico, clima, comprensione, esigenza, mutuo, popolo	4	6	0.6000
1959Gronchi	soluzione, campo, benessere, consolidare, convergenza, grave, internazionale	3	8	0.5000
1960Gronchi	pace, collaborazione, insieme	3	3	0.9444
1961Gronchi	sentire, problema, insieme, continuo, capace, forza, progredire, rendere, democrazia, istituzione	8	10	0.5980
1962Segni	partecipazione, bene, eguaglianza, libertà	3	4	0.5833
1963Segni	conseguire, internazionale, nuovo, ovunque, riconoscimento, sapere	4	6	0.6000
1964Saragat	animo, affettuoso, augurio, caro, cordiale, famiglia	5	6	0.6556
1965Saragat	augurio, più, fare	3	3	0.9333
1966Saragat	Italia, vita, nazione, famiglia, sempre, collocare, economico, industriale, progredito	5	9	0.3722
1967Saragat	europeo, attacco, augurio, dovere, Italia, libertà, pace, processo, progresso	3	9	0.3611
1968Saragat	dirigente, grado, nazione, politica, responsabilità	5	5	0.8286
1969Saragat	libertà, più, umano, cittadino, dovere, effettivo, sociale, sviluppo	7	8	0.7545
1970Saragat	paese, politica, anno, tale, dovere, fine, libertà	7	7	0.6488
1971Leone	mettere, più, posto, realizzare, responsabilità, uomo	2	6	0.7333
1972Leone	potere, dovere, economico, comunita, europeo	4	5	0.5714
1973Leone	dare, sempre, maggiore, cittadino, forza	5	5	0.6769
1974Leone	italiano, dire, maggioranza, colpire, risorsa, capacità	4	6	0.4476
1975Leone	fiducia, popolo, stesso, grave, paese, libero	6	6	0.8267
1976Leone	criminalità, potere, problema, giustizia, dire, esprimere, alimentare, produttivo	4	8	0.4031
1977Leone	forza, anno, più, mondo, società, dire, grande, oggi, volere	5	9	0.4105
1978Pertini	uomo, giovane, potere, armare, molto, presidente, unità	7	7	0.7792
1979Pertini	giovane, guerra, Italia, potere, anziano, difendere, fare, ipotesi, popolo, terrorismo	7	10	0.3937
1980Pertini	fare, dovere, Italia, trovare, cataclisma, giovane, lavoro, meridione	6	8	0.4893
1981Pertini	potere, terrorismo, popolo, dovere, protestare, indifferente, animo, avere	8	8	0.6153
1982Pertini	popolo, fare, operaio, qui, anno, italiano,	7	12	0.3963

	ricevere, avere, persona, più, riuscire, terrorismo			
1983Pertini	popolo, italiano, pace, potere, avere, dire, fare, anno, volere, stare, francese, americano	10	12	0.5970
1984Pertini	giustizia, strage, vittima, familiare, anno, dovere, infamia, marchio, parente, stesso	8	10	0.5778
1985Cossiga	civile, pace, soggetto, dovere, libero, libertà, popolo, uomo, comunita, cultura, vivere	10	11	0.6966
1986Cossiga	comunita, paese, vita, avere, solo, fare, istituzione, sicurezza, occorrere, superare, costante, tensione, minaccia	4	13	0.4029
1987Cossiga	sviluppo, comune, libertà, istituzionale, sistema, fare, economico, Italia, pace, civile, impegno, opera	8	12	0.4874
1988Cossiga	più, nuovo, economico, società, vita, forza, pubblico, tanto, persona, problema, avere	9	11	0.4112
1989Cossiga	europeo, popolo, società, dovere, valore, essere, uomo, culturale, democratico, struttura	8	10	0.6154
1990Cossiga	politico, paese, dovere, internazionale, Europa, europeo, sapere, società, forza, grande, democratico, essere, Stato, sviluppo	13	14	0.6304
1991Cossiga	legge, sempre, tradizione	2	3	0.6667
1992Scalfaro	avere, diritto, dovere, politico, fare, Italia, giudizio, più, essere, libero, responsabilità, sapere	8	12	0.4209
1993Scalfaro	avere, popolo, volontà, dare, responsabilitá, italiano, solo, volere	8	8	0.9536
1994Scalfaro	dovere, tanto, Parlamento, potere, presidente, forza, repubblica, augurio, dire, elezione, governo, mondo, Stato, volta	12	14	0.5348
1995Scalfaro	dovere, diritto, avere, Parlamento, anno, governo, Italia, politico, dire, essere, prendere, atto, politica, potere, responsabilità	12	16	0.5389
1996Scalfaro	potere, tanto, più, politica, azione, diritto	6	6	0.5472
1997Scalfaro	dire, parlare, Stato, fare, avere, capo, consentire, trovare, persona, volta	10	10	0.7101
1998Scalfaro	dire, anno, grande, avere, essere, persona, lavoro, cosa, pace, più, politica, tanto	8	12	0.3954
1999Ciampi	dare, Italia, più, istituzione, Stato, nuovo, secolo	5	7	0.5714
2000Ciampi	Italia, più, elettorale, fare, grande	4	5	0.5000
2001Ciampi	Italia, essere, principio, grande, paese, avere, compito, istituzione, libertà, moneta, unità	5	11	0.4045
2002Ciampi	più, Italia, giustizia, costituzione, pace, avere,	5	8	0.4365

	essere, popolo			
2003Ciampi	pace, più, istituzione, europeo, Italia	5	5	0.9500
2004Ciampi	vivere, mondo, anno, Italia, tanto, avere, disastro, italiano	5	8	0.3750
2005Ciampi	volere, Italia, italiano, repubblica, comune, presidente, provincia, sindaco	4	8	0.4345
2006Napolitano	potere, italiano, dovere, forza, più, politica	5	6	0.6286
2007Napolitano	parte, nuovo, contare, fiducia, forza, potere	5	6	0.5833
2008Napolitano	mondiale, crisi, modo, fare, impegno, nuovo, paese, sociale	6	8	0.5516

2.7.2. Thematic association

Combining the two approaches, namely the associations and the concentration, we want to know how the pre-*f*(*h*) words characterizing the thematic concentration are associated. We use the results of the computation of associativity of all autosemantic pre-*f*(*h*) words with one another and take the mean of their associativity *P*. In this way we obtain an indicator of thematic association which can be defined as follows. Let *n* be the number of autosemantics in the pre-*f*(*h*) domain, then

$$(2.7) \quad TA = \frac{2}{n(n-1)} \sum_{i'<j'} P_{i'j'}$$

where $P_{i'j'}$ is the strength of association between autosemantics *i´* and *j´*. Note that each association is counted only once.

The associations computed in the previous chapter are presented in Table 2.9, which refers to the same vocabulary as column 2 in Table 2.8.

Table 2.9
Thematic associations
(for TA h is rounded to integer)

Text	h	n	pairs	Sum P	1000*TA
1949Einaudi	2	3	0	0.0000	0.00
1950Einaudi	2	7	0	0.0000	0.00
1951Einaudi	1	3	3	0.1429	47.62
1952Einaudi	2	6	0	0.0000	0.00
1953Einaudi	2	5	0	0.0000	0.00
1954Einaudi	3	3	0	0.0000	0.00

1955Gronchi	1	7	6	0.1091	5.19
1956Gronchi	4	4	6	0.1750	29.17
1957Gronchi	4	4	3	0.0587	9.78
1958Gronchi	4	6	9	0.1818	12.12
1959Gronchi	3	8	9	0.1238	4.42
1960Gronchi	3	3	0	0.0000	0.00
1961Gronchi	8	10	41	0.6581	14.62
1962Segni	3	4	2	0.0662	11.03
1963Segni	4	6	10	0.0362	2.41
1964Saragat	5	6	15	0.3788	25.25
1965Saragat	3	3	0	0.0000	0.00
1966Saragat	5	9	10	0.2047	5.69
1967Saragat	3	9	7	0.1027	2.85
1968Saragat	5	5	3	0.0490	4.90
1969Saragat	7	8	7	0.1267	4.53
1970Saragat	7	7	3	0.0895	4.26
1971Leone	2	6	6	0.0073	7.27
1972Leone	4	5	3	0.0851	8.51
1973Leone	5	5	2	0.0621	6.21
1974Leone	4	6	6	0.0869	5.80
1975Leone	6	6	5	0.1029	6.86
1976Leone	4	8	3	0.1005	3.59
1977Leone	5	9	12	0.2994	8.32
1978Pertini	7	7	3	0.0498	2.37
1979Pertini	7	10	5	0.1316	2.92
1980Pertini	6	8	11	0.2584	9.23
1981Pertini	8	8	4	0.0681	2.43
1982Pertini	7	12	6	0.0350	0.53
1983Pertini	10	12	8	0.0735	1.11
1984Pertini	8	10	25	0.2040	4.53
1985Cossiga	10	11	10	0.2492	4.53
1986Cossiga	4	13	15	0.1698	2.18
1987Cossiga	8	12	10	0.2631	3.99
1988Cossiga	9	11	8	0.1687	3.07
1989Cossiga	8	10	7	0.1759	3.91
1990Cossiga	13	14	22	0.5115	5.62
1991Cossiga	2	3	3	0.0583	19.44
1992Scalfaro	8	12	16	0.2449	3.71
1993Scalfaro	8	8	4	0.0492	1.76

1994Scalfaro	12	14	20	0.3717	4.08
1995Scalfaro	12	16	20	0.3409	2.84
1996Scalfaro	6	6	6	0.0817	5.45
1997Scalfaro	10	10	7	0.1451	3.23
1998Scalfaro	8	12	14	0.2968	4.50
1999Ciampi	5	7	2	0.0116	0.55
2000Ciampi	4	5	0	0.0000	0.00
2001Ciampi	5	11	6	0.1657	3.01
2002Ciampi	5	8	6	0.1452	5.19
2003Ciampi	5	5	2	0.0462	4.62
2004Ciampi	5	8	4	0.0809	2.89
2005Ciampi	4	8	13	0.1422	5.08
2006Napolitano	5	6	0	0.0000	0.00
2007Napolitano	5	6	3	0.0463	3.09
2008Napolitano	6	8	7	0.1708	6.10

If we take the means of 1,000*TA* for each president, we obtain a very seducing picture as can be seen in Figure 2.10. The indicator decreases almost linearly. But this is caused by the fact that Einaudi 1951 displays an absolutely deviant value. If it were zero as the other texts by Einaudi, we would obtain a rather concave function. Thus either Einaudi 1951 is an outlier or his other texts are outliers because of their shortness (the addresses *approximately* show a trend of increasing duration, the two extremes being Einaudi for his conciseness and Scalfaro for his loquacity). The problem cannot be solved without analyzing other texts written by Einaudi. Einaudi's speeches are distinctive also because they were the first end-of-year addresses delivered (on the radio, cf. Zotti Minici, 2007) by a president of the Italian Republic and, from a qualitative point of view, Einaudi shows an archaic lexicon which is not found in his successors.

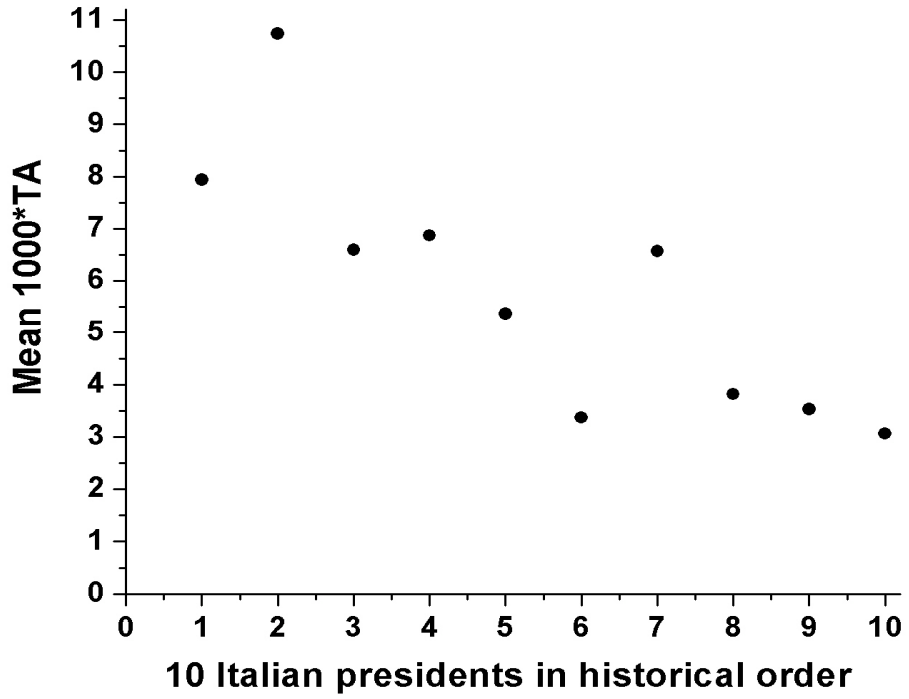

Figure 2.10. The historical order of mean 1000*TA* -values

2.8. Degree distribution

The number of (not directed) edges touching the given vertex designate its *degree*. Looking at Figure 2.1, we see e.g. that *uomo, responsabilità, posto, realizzare, più, mettere* have each 2 degrees, *sentimento* and *concordia* have 1 degree. Those lemmas which do not occur in Figure 2.1 all have degree 0. From graph theory it is known that graphs of a special kind have a special distribution of degrees, hence we may ask what the *degree distribution* is in texts, especially in this kind of texts. Since we accepted only associations of words occurring at least twice at $\alpha = 0.05$, many of the words remained dissociated. That is, our graphs are not Milgram´s *small worlds* (cf. e.g. Buchanan 2002). However, a slight change of the procedure could demonstrate that even here the *small world* idea is prolific.

2.8.1. Degree spectrum

Consider the text Napolitano 2008. Taking into account the degrees of individual autosemantics we obtain the empirical distribution of degrees as presented in Table 2.10

Table 2.10
Degree distribution of autosemantic lemmas in Napolitano 2008

Degree x	**Number f(x)**
0	447
1	39
2	19
3	4
4	6
5	9
6	4
7	1
8	2
9	1

In order to capture the degree distribution formally, we can proceed in two ways.

(1) We consider it a distribution and try to derive and justify it theoretically and test it by a goodness-of-fit test.

(2) We consider the sequence as a function, try to find a curve and judge the result by the determination coefficient.

Our starting point is as follows. From a text we mechanically selected autosemantic lemmas and computed their associations using a procedure creating several isolated components (the α-method). The degrees of individual lemmas are determined on the basis of the component which they are part of. Thus each component has its own degree distribution. The text has consequently as many strata as there are components. The set of isolated words can be considered a further stratum. In Popescu et al. (2009), Popescu, Altmann, Köhler (2009) it was proposed to pool the strata and re-rank the frequencies; here we can do the same with degrees. In the last step we transform the degree distribution in degree spectrum. Here, each $f(x)$ is composed of all x-degrees taken from all components of the text graph. In other words, $f(x)$ is a mixture. As such, it can be modelled either as a mixture of distributions (method 1) or as a superposition of functions (method 2). Using the approach proposed in the above references we suppose that the degree spectrum abides by one of the following functions:

$$f(x) = 1 + \sum_{j=1}^{k} a_j \exp(-x / b_j) \tag{2.8}$$

if we consider it a curve, or by

(2.9) $NP_x = N\sum_{j=1}^{k}\alpha_j p_j q_j^x$, x = 0,1,…,

if it is a distribution. In (2.9), $\sum_{j=1}^{k}\alpha_j = 1$, i.e. if there are only two summands, then $\alpha_2 = 1- \alpha_1$. The formulas tell the same story because exp(-*x*/*b*) can be written as q^x. Approach (2.8) has the advantage of showing us how many strata there are in the text. In the case of overestimating the number of strata, the parameters of two terms turn out to be equal. Approach (2.9) has the disadvantage that the existing software fits only mixtures of geometric distributions consisting of two terms. Nevertheless, the mixed geometric was the only distribution yielding systematically good fittings. For example, for the Napolitano 2008 we obtain the results in Table 2.11 and in Figure 2.11.

Table 2.11
Fitting the mixed geometric distribution
to the degree-spectrum in Napolitano 2008

Degree x	**Number f(x)**	**Mixed geometric**
0	447	442.04
1	39	39.82
2	19	16.94
3	4	10.98
4	6	7.34
5	9	4.91
6	4	3.29
7	1	2.21
8	2	1.48
9	1	3.00
$q_1 = 0.6698$, $q_2 = 0.0380$, $\alpha = 0.2075$, $X^2 = 10.72$, DF = 6, P = 0.10		

It should be noted that pooling some classes in the tails of distributions would yield a slightly better results but the improvement is not essential.

Approach (2.8) yields the results presented in Table 2.12. Here even the one-term approach is sufficient, but the difference of the exponents shows that there are minimally two terms. As can be seen, the parameters b_i are different, hence we may speak about two strata. The results corroborate the discrete approach. A survey of all results is presented in Table 2.13.

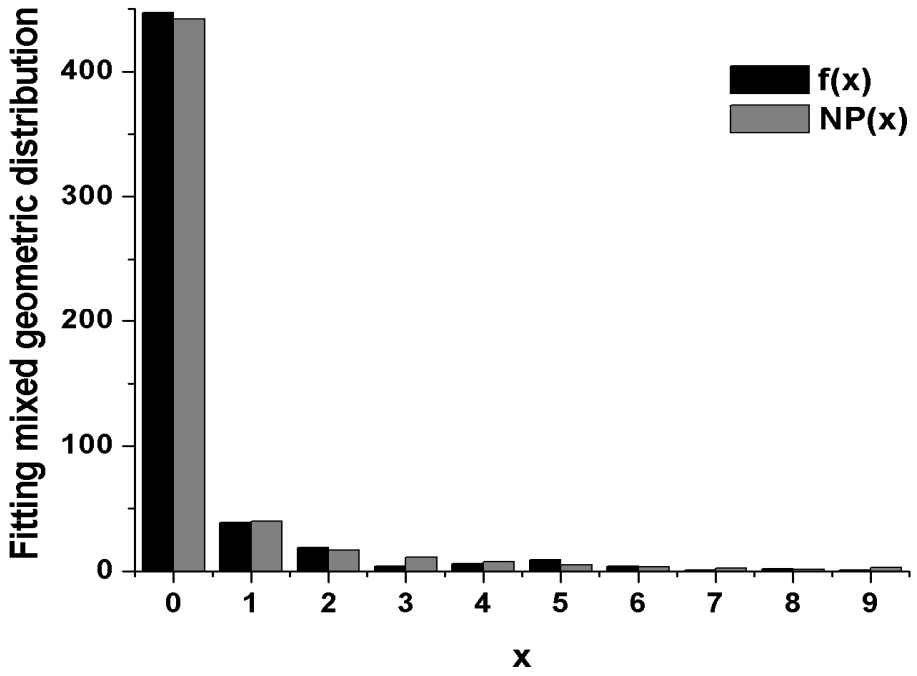

Figure 2.11. Fitting the mixed geometric d. to Napolitano2008

Table 2.12
Fitting the exponential mixture
to the degree-spectrum in Napolitano 2008

Degree x	**Number f(x)**	**One term**	**Two terms**
0	447	446.88	447.00
1	39	41.69	38.15
2	19	4.17	16.39
3	4	1.34	9.80
4	6	1.03	6.15
5	9	1.00	4.02
6	4	1.00	2.77
7	1	1.00	2.04
8	2	1.00	1.61
9	1	1.00	1.36
		a = 445.8757 b = 0.4177 R^2 = 0.998	a_1= 43.6765, b_1= 1.8711 a_2 = 402.3214, b_2 = 0.2884 R^2 = 0.9996

Table 2.13a
Fitting (2.8) to the individual degree-spectra

Text	Continuous (2.8)				
	a_1	b_1	a_2	b_2	R^2
1956	119.0026	0.3943	119.0026	0.3943	0.9995
1957	251.1296	0.0197	64.8737	1.3483	0.9999
1958	219.6216	0.0129	28.7742	2.4507	0.9926
1959	223.3837	0.2120	21.6163	1.6973	0.9998
1960	205.2530	0.0101	23.7059	1.9900	0.9699
1961	297.1777	0.0053	44.8356	2.2703	0.9996
1962	178.7326	0.0234	59.3198	1.1201	0.9997
1963	282.4288	0.1631	52.5712	1.2347	0.9998
1964	152.4188	0.0212	16.5016	2.2109	0.9962
1965	283.4341	0.0194	36.4036	1.9936	0.9986
1966	285.2251	0.0241	61.4363	1.7145	0.9997
1967	301.6089	0.2686	30.3913	1.8491	1.0000
1968	285.1266	0.0145	41.8667	1.8281	1.0000
1969	319.6724	0.0177	64.2998	2.0161	0.9995
1970	368.1022	0.0224	86.2360	1.6573	0.9992
1971	85.5591	0.0219	10.6527	2.0365	0.9946
1972	205.5706	0.2407	22.4295	1.5503	1.0000
1973	316.2515	0.0247	43.4782	1.9087	0.9993
1974	214.3632	0.0183	26.2408	2.1135	0.9959
1975	314.9792	0.0228	65.0208	1.4902	0.9999
1976	337.6561	0.2816	34.3454	2.5512	0.9997
1977	339.9093	0.2748	51.0908	2.0095	1.0000
1978	283.2741	0.3474	16.7355	5.0503	0.9995
1979	349.8897	0.2868	42.1121	3.0019	0.9996
1980	204.6766	0.0125	39.8547	3.2301	0.9961
1981	322.6545	0.2527	64.3453	2.7710	0.9997
1982	339.7524	0.0231	45.1077	3.2475	0.9994
1983	340.4667	0.0197	87.1394	2.8474	0.9773
1984	186.0079	0.0206	31.6541	3.6322	0.9963
1985	380.2267	0.0167	45.2565	3.8522	0.9992
1986	277.4997	0.0115	40.5344	2.2537	0.9997
1987	461.1000	0.0080	68.9000	2.1929	0.9999
1988	400.6732	0.2210	50.3272	3.4273	0.9998
1989	401.2855	0.2860	29.7154	3.7653	0.9998
1990	483.0807	0.0225	106.3265	2.1181	0.9974

1991	114.5653	0.3592	22.4347	0.7507	1.0000
1992	426.9760	0.0229	76.9243	2.3430	0.9999
1993	491.4088	0.0040	79.5287	2.1349	0.9999
1994	465.3785	0.0128	109.6591	2.0441	0.9991
1995	503.5013	0.0214	102.4856	2.0296	0.9962
1996	330.1769	0.1814	132.8231	1.1986	1.0000
1997	540.1509	0.2923	107.8468	2.0681	0.9999
1998	434.7052	0.0216	128.5936	1.7624	0.9993
1999	353.0907	0.0165	94.9561	1.9556	0.9980
2000	434.4487	0.2425	45.5515	1.6604	0.9999
2001	449.9077	0.0133	61.3156	1.9426	0.9998
2002	425.2515	0.2166	98.7486	1.4187	1.0000
2003	363.8970	0.0216	63.9390	1.3560	0.9999
2004	382.5509	0.0162	81.5091	1.3310	0.9999
2005	271.6500	0.3803	17.3491	1.9715	1.0000
2006	363.9951	0.0101	181.0480	0.9464	0.9998
2007	506.3202	0.2817	67.6798	1.4191	1.0000
2008	4.023.214	0.2884	436.764	18.711	0.9996

Table 2.13b
Fitting (2.9) to the individual degree-spectra

Discrete (2.9)						
	q_1	q_2	α	X^2	DF	P
1957	0.5311	0.0091	0.3116	1.19	4	0.88
1958	0.5555	0.0029	0.2685	5.23	1	0.02
1959	0.4692	0.0005	0.2555	4.88	1	0.03
1960	0.5306	0.0004	0.2220	3.71	3	0.29
1961	0.6973	0.0119	0.2847	13.03	7	0.07
1962	0.3903	0.0205	0.3527	2.72	1	0.10
1963	0.4660	0.0076	0.2606	1.59	1	0.21
1964	0.7475	0.0002	0.0943	8.62	2	0.01
1965	0.4172	0.0003	0.2776	7.47	2	0.02
1966	0.5907	0.0013	0.3230	1.11	5	0.95
1967	0.4548	0.0049	0.2671	4.39	2	0.11
1968	0.5251	0.0009	0.2590	3.01	3	0.39
1969	0.6394	0.0030	0.3223	5.69	5	0.34
1970	0.6456	0.0083	0.2879	9.16	7	0.24
1972	0.5661	0.0129	0.2049	0.25	3	0.97

1973	0.6152	0.0044	0.2422	10.91	5	0.05
1974	0.5976	0.0043	0.1902	4.56	3	0.21
1975	0.6247	0.0184	0.2423	3.30	5	0.65
1976	0.5465	0.0029	0.3130	8.82	4	0.07
1977	0.6025	0.0195	0.2992	4.35	5	0.48
1978	0.7245	0.0269	0.2972	12.25	8	0.14
1979	0.6952	0.0212	0.3166	9.26	9	0.41
1980	0.6158	0.0005	0.3828	8.57	5	0.13
1981	0.6948	0.0134	0.4149	10.73	9	0.29
1982	0.7102	0.0004	0.3535	9.48	9	0.39
1983	0.7297	0.0024	0.4002	20.51	12	0.06
1984	0.6991	0.0002	0.4424	10.70	6	0.10
1985	0.7515	0.0005	0.3677	8.80	10	0.55
1986	0.5868	0.0005	0.3250	8.10	4	0.09
1987	0.6922	0.0103	0.2806	6.00	8	0.65
1988	0,7604	0.0281	0.2982	9.32	12	0.68
1989	0.7375	0.0187	0.2694	6.71	9	0.67
1990	0.7938	0.0283	0.3057	20.64	17	0.24
1992	0.7030	0.0164	0.3005	7.61	10	0.67
1993	0.6765	0.0045	0.2996	4.65	7	0.70
1994	0.7889	0.0676	0.2696	28.04	15	0.02
1995	0.7553	0.0200	0.3209	19.11	13	0.12
1996	0.5298	0.0372	0.3152	2.01	4	0.73
1997	0.7378	0.0845	0.2311	9.71	11	0.56
1998	0.7269	0.0467	0.3018	22.68	11	0.02
1999	0.5211	0.0014	0.3638	2.71	4	0.61
2000	0.4930	0.0047	0.2355	3.66	3	0.30
2001	0.5809	0.0006	0.2670	3.06	5	0.69
2002	0.5386	0.0136	0.3024	0.71	5	0.98
2003	0.5472	0.0049	0.2421	5.39	4	0.25
2004	0.5994	0.0378	0.1650	5.08	4	0.28
2005	0.5621	0.0546	0.2038	2.52	3	0.47
2006	0.4308	0.0006	0.3869	5.66	3	0.13
2007	0.5375	0.0308	0.1983	1.06	4	0.90
2008	0.6698	0.0380	0.2075	10.72	6	0.10

The continuous approach seems to be more adequate in all cases. Some texts are too short, the number of degree-classes is too small for fitting the mixed geometric; in some texts the number of associated words is too small and no model can be fitted. In the discrete approach there are several cases at the boundary of

significance but the continuous approach is strongly corroborative. In the text 1956 it even shows that there is only one stratum ($b_1 = b_2$).

2.9. The comparison of vocabularies

The comparison of vocabularies depends on the aim. For forensic research or authorship determination it is something different from stylistic, historical, lexicographical or psycholinguistic research. Sometimes it is sufficient to consider the verbs and draw conclusions about the "activity" of texts; for studying the "ornamentality" of texts a simple ratio of adjectives and nouns is sufficient; the use of synsemantics may be prolific both in forensic and stylistic studies, etc. (cf. e.g. Carroll 1969). But to compare whole vocabularies is reasonable only if the texts concern the same object, i.e. if they have the same content or something in common.

Here we take into account only autosemantic lemmas which are more firmly anchored in text by their associations with other words. That is, we do not compare the whole vocabulary but only the contents of the network established by the writer for expressing the same object. The set of pertinent words or the network itself could be established in many different ways; we restrict ourselves to our above results.

The results will be different if we use the words presented in Table 2.6, i.e. all associated words, or in Table 2.8 containing the cores of the texts. Further, one can compare the individual texts or the individual authors, thus there are four modes of comparison. Other different aspects have been shown in Bernardi, Tuzzi (2007).

A quite different aspect of comparison is taking into account also the frequencies of words, which is currently used in French quantitative linguistics, even if the interpretation of the results, e.g. for authorship attribution, is strongly criticized (cf. Viprey, Ledoux 2006).

2.9.1. Method I

Let us have two sets $A = \{e_{a1},e_{a2},\ldots,e_{aj}\}$ and $B = \{e_{b1},e_{b2},\ldots,e_{bk}\}$ where e are the elements, j and k need not be equal. The similarity will be expressed by means of Levin´s (1964) formula

$$(2.10)\quad S(A,B) = \frac{|A \cap B|^2}{|A| \bullet |B|},$$

where | | is the cardinal number of the set and ∩ is the intersection of sets. $S(A,B) = 0$ when the intersection is empty (i.e. the texts do not contain equal words) and

$S(A,B) = 1$ when the sets are equal ($e_{ai} = e_{bi}$ and $j = k$). Otherwise we obtain some values between 0 and 1. Taking from Table 2.6 the texts 1956 (= A) and 1957 (= B)

A = {anno, arma, attività, avere, corso, coscienza, esigenza, fiducia, forza, iniziativa, intento, italiano, lavoro, libero, libertà, nazione, opera, pace, parte, peso, più, potere, prendere, proprio, rapporto, responsabilità, solidarietà, spirito, unito, violenza}

B = {amministrativo, andare, anno, articolare, assunzione, capacità, carta, cittadino, civile, comune, consapevole, costituzione, costume, crescente, democratico, democrazia, dirittto, discussione, dovere, esigenza, essenziale, evento, già, giuridico, giusto, interesse, istituzione, italiano, libero, libertà, limite, morale, nazionale, nuovo, organo, parte, pieno, politica, popolo, porre, positivo, possibilità, potere, problema, proprio, prospettare, pubblico, raggiungere, repubblicano, responsabilità, rinuovato, sempre, sociale, società, solo, soltanto, soluzione, struttura, uomo, valido, valore, vicenda, vita, volontà}

we can easily state that |A| = 30, |B| = 64, A ∩ B = {anno, esigenza, italiano, libero, libertà, parte, potere, proprio, responsabilità}, | A ∩ B| = 9, hence

$$S(1956, 1957) = 9^2/[30(64)] = 0.0422$$

The results of all comparisons using Table 2.6 are presented in Table 2.14, and those using Table 2.8 in Table 2.15. In order to save space, the similarities are given only for successors, i.e. those of Napolitano must be sought in respective tables of his predecessors.

Table 2.14
Bulk similarities of associated vocabularies from Table 2.6

Table 2.14a
Bulk autosimilarities

Gronchi	**1955**	**1956**	**1957**	**1958**	**1959**	**1960**	**1961**
1955	1	0.0042	0.0078	0.0119	0	0.0147	0.0060
1956	0.0042	1	0.0422	0.0960	0.0178	0.0794	0.0571
1957	0.0078	0.0422	1	0.0629	0.0422	0.0901	0.0744
1958	0.0119	0.0960	0.0629	1	0.0198	0.1008	0.0726
1959	0	0.0178	0.0422	0.0198	1	0.0157	0.0480
1960	0.0147	0.0794	0.0901	0.1008	0.0157	1	0.0686
1961	0.0060	0.0571	0.0744	0.0726	0.0480	0.0686	1

Segni	**1962**	**1963**					
1962	1	0.0429					
1963	0.0429	1					
Saragat	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1964	1	0.0029	0.0013	0.0217	0.0167	0	0.0010
1965	0.0029	1	0.0614	0.0206	0.0439	0.0274	0.0383
1966	0.0013	0.0614	1	0.0663	0.0716	0.0775	0.0863
1967	0.0217	0.0206	0.0663	1	0.0815	0.0817	0.0422
1968	0.0167	0.0439	0.0716	0.0815	1	0.0840	0.0634
1969	0	0.0274	0.0775	0.0817	0.0840	1	0.1437
1970	0.0010	0.0383	0.0863	0.0422	0.0634	0.1437	1
Leone	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1971	1	0.0038	0.0018	0.0152	0.0670	0.0016	0.0057
1972	0.0038	1	0.0277	0.0147	0.0683	0.0249	0.0502
1973	0.0018	0.0277	1	0.0433	0.1008	0.1238	0.0795
1974	0.0152	0.0147	0.0433	1	0.0404	0.0559	0.0348
1975	0.0670	0.0683	0.1008	0.0404	1	0.0754	0.0811
1976	0.0016	0.0249	0.1238	0.0559	0.0754	1	0.0849
1977	0.0057	0.0502	0.0795	0.0348	0.0811	0.0849	1
Pertini	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1978	1	0.1533	0.0959	0.1107	0.1117	0.1263	0.0528
1979	0.1533	1	0.1189	0.1068	0.0765	0.1254	0.0945
1980	0.0959	0.1189	1	0.0957	0.1003	0.0934	0.0869
1981	0.1107	0.1068	0.0957	1	0.1690	0.1632	0.0675
1982	0.1117	0.0765	0.1003	0.1690	1	0.1935	0.0847
1983	0.1263	0.1254	0.0934	0.1632	0.1935	1	0.0829
1984	0.0528	0.0945	0.0869	0.0675	0.0847	0.0829	1
Cossiga	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1985	1	0.1281	0.1595	0.1377	0.1473	0.1767	0.0365
1986	0.1281	1	0.0855	0.1314	0.0847	0.0691	0.0134
1987	0.1595	0.0855	1	0.1636	0.1356	0.1523	0.0220
1988	0.1377	0.1314	0.1636	1	0.1511	0.1550	0.0240
1989	0.1473	0.0847	0.1356	0.1511	1	0.1647	0.0253
1990	0.1767	0.0691	0.1523	0.1550	0.1647	1	0.0376
1991	0.0365	0.0134	0.0220	0.0240	0.0253	0.0376	1
Scalfaro	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1992	1	0.2287	0.1863	0.2058	0.1036	0.1456	0.1502
1993	0.2287	1	0.1902	0.1572	0.0843	0.1526	0.1775

1994	0.1863	0.1902	1	0.2015	0.1276	0.1858	0.1612
1995	0.2058	0.1572	0.2015	1	0.1123	0.1646	0.1732
1996	0.1036	0.0843	0.1276	0.1123	1	0.0984	0.1255
1997	0.1456	0.1526	0.1858	0.1646	0.0984	1	0.2189
1998	0.1502	0.1775	0.1612	0.1732	0.1255	0.2189	1
Ciampi	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
1999	1	0.0844	0.1740	0.1262	0.1852	0.1234	0.1238
2000	0.0844	1	0.1607	0.0882	0.0882	0.0703	0.1305
2001	0.1740	0.1607	1	0.1222	0.1078	0.1498	0.0922
2002	0.1262	0.0882	0.1222	1	0.1311	0.1116	0.1193
2003	0.1852	0.0882	0.1078	0.1311	1	0.0972	0.1346
2004	0.1234	0.0703	0.1498	0.1116	0.0972	1	0.0974
2005	0.1238	0.1305	0.0922	0.1193	0.1346	0.0974	1
Napolitano	**2006**	**2007**	**2008**				
2006	1	0.1341	0.1147				
2007	0.1341	1	0.0596				
2008	0.1147	0.0596	1				

Table 2.14b
Gronchi´s similarities

Gronchi	**Segni**		**Saragat**						
	1962	**1963**	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1955	0	0.0027	0	0.0033	0.0015	0	0	0.0013	0.0097
1956	0.0659	0.0464	0	0.0219	0.0397	0.0464	0.0556	0.05	0.0348
1957	0.0381	0.0217	0.0017	0.0066	0.0419	0.0275	0.0586	0.0718	0.0486
1958	0.0703	0.0626	0.0026	0.0307	0.0479	0.0626	0.048	0.0717	0.0593
1959	0.0659	0.0261	0	0.014	0.0194	0.0181	0.0356	0.042	0.0287
1960	0.1033	0.0409	0	0.007	0.0424	0.0639	0.049	0.06	0.0428
1961	0.0839	0.0373	0.0013	0.0451	0.0512	0.0437	0.105	0.0904	0.0805

Gronchi	**Leone**						
	1971	1972	1973	1974	1975	1976	1977
1955	0	0.0038	0.0018	0	0.0067	0.0064	0.0014
1956	0.0167	0.0495	0.0233	0.0162	0.0360	0.0346	0.0464
1957	0.0078	0.0573	0.0571	0.0232	0.0752	0.0200	0.0582
1958	0	0.1039	0.0490	0.0354	0.0622	0.0598	0.0791
1959	0	0.0364	0.0171	0.0091	0.0111	0.0068	0.0310
1960	0.0147	0.0570	0.0340	0.0321	0.0475	0.0185	0.0338
1961	0.0060	0.0812	0.0551	0.1770	0.0635	0.0441	0.0443

Gronchi	Pertini						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1955	0.0014	0	0.0015	0.0032	0.0089	0.0025	0.0013
1956	0.0245	0.0278	0.0263	0.0540	0.0214	0.0325	0.0125
1957	0.0259	0.0158	0.0123	0.0320	0.0358	0.0343	0.0104
1958	0.0536	0.0198	0.0238	0.0386	0.0272	0.0303	0.0248
1959	0.0061	0.0100	0.0148	0.0255	0.0214	0.0106	0.0087
1960	0.0409	0.0297	0.0232	0.0538	0.0458	0.0528	0.0077
1961	0.0547	0.0358	0.0331	0.0471	0.0689	0.0811	0.0358

Gronchi	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1955	0.0031	0.0018	0.0038	0.0072	0.0011	0.0139	0.0074
1956	0.0598	0.0233	0.0366	0.0306	0.0494	0.0379	0.0314
1957	0.0994	0.0893	0.1003	0.0837	0.0927	0.1056	0.0331
1958	0.0652	0.0412	0.0880	0.0438	0.0754	0.0560	0.0126
1959	0.0168	0.0305	0.0366	0.0416	0.0421	0.0333	0.0078
1960	0.0884	0.0340	0.0323	0.0317	0.0580	0.0633	0.0433
1961	0.0757	0.0750	0.0764	0.0877	0.0878	0.0686	0.0112

Gronchi	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1955	0.0008	0.0009	0.0055	0.0023	0.0011	0.0024	0.0027
1956	0.0371	0.0589	0.0320	0.0188	0.0503	0.0270	0.0353
1957	0.0642	0.0728	0.0600	0.0384	0.0314	0.0330	0.0338
1958	0.0508	0.0475	0.0378	0.0249	0.0478	0.0193	0.0312
1959	0.0140	0.0147	0.0198	0.0076	0.0298	0.0102	0.0180
1960	0.0435	0.0456	0.0467	0.0268	0.0444	0.0407	0.0358
1961	0.0529	0.0434	0.0675	0.0567	0.0562	0.0356	0.0618

Gronchi	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1955	0.0012	0.0077	0.0014	0.0011	0	0.0016	0	0.0012	0.0014	0.0015
1956	0.0457	0.0328	0.0178	0.0783	0.0429	0.0511	0.0410	0.0457	0.0383	0.0392
1957	0.0656	0.0240	0.0333	0.0414	0.0419	0.0239	0.0243	0.0537	0.0404	0.0596
1958	0.0580	0.0297	0.0663	0.0789	0.0457	0.0301	0.0371	0.0383	0.0394	0.0810
1959	0.0384	0.0251	0.0178	0.0248	0.0190	0.0152	0.0026	0.0114	0.0138	0.0251
1960	0.0473	0.0113	0.0540	0.0607	0.0299	0.0450	0.0362	0.0280	0.0216	0.0498
1961	0.0827	0.0469	0.0685	0.0529	0.0682	0.0488	0.0449	0.0709	0.0443	0.0875

Table 2.14c
Segni´s similarities

	1964	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**	**Saragat**
1962	0.0027	0.0160	0.0653	0.0764	0.0585	0.1120	0.0681	
1963	0.0097	0.0206	0.0437	0.0232	0.0130	0.0326	0.0422	
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**Leone**
1962	0	0.0894	0.0348	0.0185	0.0325	0.0113	0.0549	
1963	0.0109	0.0422	0.0376	0.0165	0.0351	0.0178	0.0422	
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**Pertini**
1962	0.0404	0.0246	0.0244	0.0395	0.0327	0.0351	0.0206	
1963	0.0490	0.0261	0.0526	0.0497	0.0499	0.0477	0.0226	
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**Cossiga**
1962	0.0668	0.0502	0.0901	0.0685	0.0856	0.0624	0.0129	
1963	0.0346	0.0376	0.0280	0.0271	0.0551	0.0349	0.0205	
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**Scalfaro**
1962	0.0464	0.0330	0.0306	0.0290	0.0490	0.0421	0.0297	
1963	0.0413	0.0600	0.0516	0.0292	0.0416	0.0518	0.0406	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**Ciampi**
1962	0.0523	0.0304	0.0597	0.0647	0.0248	0.0445	0.0300	
1963	0.0405	0.0683	0.0391	0.0676	0.0676	0.0892	0.0418	
	2006	**2007**	**2008**					**Napolitano**
1962	0.0595	0.0227	0.0485					
1963	0.0530	0.0422	0.0256					

Table 2.14d
Saragat´s similarities

Saragat	**Leone**						
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1964	0	0.0034	0.0063	0	0.0237	0	0.0013
1965	0.0033	0.0646	0.0184	0.0072	0.0351	0.0273	0.0302
1966	0.0015	0.0812	0.0551	0.0519	0.0700	0.0673	0.0268
1967	0.0027	0.1291	0.0252	0.0323	0.0290	0.0178	0.0360
1968	0.0083	0.0611	0.0467	0.0409	0.0642	0.0259	0.0324
1969	0.0117	0.0619	0.0595	0.0455	0.0613	0.0386	0.1007
1970	0.0011	0.0441	0.0493	0.0376	0.0556	0.0399	0.0892

Saragat	Pertini						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1964	0.0051	0.0083	0.0014	0.0063	0.0035	0.0022	0.0104
1965	0.0366	0.0430	0.0117	0.0107	0.0134	0.0221	0.0332
1966	0.0603	0.0321	0.0376	0.0434	0.0639	0.0464	0.0243
1967	0.0640	0.0464	0.0386	0.0446	0.0292	0.0477	0.0383
1968	0.0845	0.0450	0.0167	0.0305	0.0429	0.0366	0.0391
1969	0.0580	0.0543	0.0741	0.0593	0.0603	0.0599	0.0434
1970	0.0571	0.0524	0.0345	0.0631	0.0887	0.0793	0.0517

Saragat	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1964	0.0110	0.0143	0.0076	0.0028	0.0010	0.0020	0
1965	0.0276	0.0305	0.0394	0.0329	0.0332	0.0299	0.0248
1966	0.0958	0.0750	0.0662	0.0594	0.0878	0.0612	0.0112
1967	0.0913	0.0376	0.0664	0.0732	0.0688	0.0604	0.0205
1968	0.0812	0.0610	0.0673	0.0832	0.0585	0.0712	0.0157
1969	0.1141	0.1430	0.0974	0.1170	0.0878	0.0980	0.0153
1970	0.1084	0.0833	0.1161	0.1015	0.0727	0.1076	0.0183

Saragat	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1964	0.0117	0.0008	0.0022	0.0129	0.0357	0.0048	0.0024
1965	0.0339	0.0181	0.0290	0.0240	0.0338	0.0213	0.0205
1966	0.0529	0.0894	0.0491	0.0567	0.0425	0.0415	0.0505
1967	0.0572	0.0661	0.0426	0.0328	0.0328	0.0503	0.0569
1968	0.0860	0.0507	0.0395	0.0375	0.0656	0.0581	0.0477
1969	0.1042	0.0782	0.0901	0.0594	0.0729	0.0576	0.0577
1970	0.8630	0.0771	0.1099	0.0707	0.0647	0.0799	0.0862

Saragat	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1964	0.0042	0.0154	0.0193	0.0092	0.0159	0.0056	0.0342	0.0095	0.0013	0.0118
1965	0.0361	0.0328	0.0286	0.0241	0.0205	0.0403	0.0324	0.0203	0.0302	0.0607
1966	0.0766	0.0310	0.0419	0.0437	0.0484	0.0544	0.0387	0.0600	0.0395	0.0807
1967	0.1002	0.0334	0.0853	0.0646	0.0497	0.0704	0.0506	0.0350	0.0250	0.0309
1968	0.0573	0.0503	0.0587	0.0495	0.0519	0.0540	0.0721	0.0573	0.0375	0.0865
1969	0.1147	0.0519	0.0652	0.0804	0.0478	0.0759	0.0393	0.0723	0.0748	0.0961
1970	0.1006	0.0585	0.0735	0.0760	0.0788	0.0682	0.0424	0.0949	0.0619	0.0795

Table 2.14e
Leone´s similarities

Leone	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1971	0.0129	0.0010	0.0139	0.0071	0.0040	0.0056	0.0052
1972	0.0502	0.0253	0.0239	0.0276	0.0291	0.0182	0.0256
1973	0.0657	0.0476	0.0346	0.0362	0.0500	0.0480	0.0251
1974	0.0282	0.0364	0.0239	0.0276	0.0346	0.0295	0.0382
1975	0.0300	0.0360	0.0199	0.0372	0.0343	0.0293	0.0235
1976	0.0377	0.0386	0.0192	0.0507	0.0407	0.0431	0.0300
1977	0.0477	0.0277	0.0512	0.0612	0.0483	0.0448	0.0346

Leone	**Cossiga**						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1971	0.0124	0.0071	0.0010	0.0072	0.0044	0.0089	0.0074
1972	0.0830	0.0524	0.0592	0.0772	0.0598	0.0539	0.0642
1973	0.0746	0.0988	0.0917	0.0713	0.0663	0.0917	0.0134
1974	0.0271	0.0623	0.0391	0.0326	0.0322	0.0486	0.0160
1975	0.0796	0.1008	0.0493	0.0574	0.0516	0.0901	0.0071
1976	0.0581	0.0593	0.0823	0.0552	0.0450	0.0912	0.0121
1977	0.0925	0.0724	0.1014	0.0846	0.0790	0.0945	0.0243

Leone	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1971	0.0074	0.0078	0.0055	0.0052	0.0011	0.0096	0.0007
1972	0.0449	0.0353	0.0380	0.0238	0.0693	0.0326	0.0321
1973	0.0737	0.0947	0.0908	0.0598	0.0735	0.0498	0.0522
1974	0.0391	0.0535	0.0380	0.0509	0.0457	0.0580	0.0419
1975	0.0789	0.0773	0.0588	0.0452	0.0525	0.0399	0.0349
1976	0.0615	0.0693	0.0726	0.0468	0.0413	0.0516	0.0505
1977	0.0926	0.0762	0.1091	0.0655	0.0694	0.0674	0.0487

Leone	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1971	0.0048	0.0077	0.0054	0.0011	0.0079	0.0142	0.0024	0.0048	0.0014	0.0015
1972	0.0739	0.0378	0.0399	0.0278	0.0481	0.0752	0.0373	0.0566	0.0421	0.0513
1973	0.0720	0.0220	0.0685	0.0755	0.0510	0.0653	0.0618	0.0600	0.0322	0.1136
1974	0.0566	0.0466	0.0646	0.0470	0.0236	0.0384	0.0210	0.0416	0.0348	0.0228
1975	0.0367	0.0295	0.0523	0.0442	0.0212	0.0488	0.0369	0.0794	0.0392	0.0904
1976	0.0703	0.0333	0.0169	0.0425	0.0399	0.0318	0.0158	0.0825	0.0477	0.1020
1977	0.0530	0.0453	0.0551	0.0713	0.0730	0.0642	0.0221	0.0985	0.0528	0.0779

Table 2.14f
Pertini´s similarities

Pertini	**Cossiga**						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1978	0.1086	0.0795	0.0640	0.0797	0.0581	0.0777	0.0331
1979	0.0598	0.0386	0.0366	0.0446	0.0387	0.0593	0.0123
1980	0.0645	0.0346	0.0272	0.0315	0.0524	0.0430	0.0182
1981	0.0832	0.0478	0.0406	0.0582	0.0760	0.0675	0.0238
1982	0.0639	0.0500	0.0475	0.0518	0.0401	0.0779	0.0168
1983	0.0935	0.0518	0.0578	0.0642	0.0734	0.0878	0.0143
1984	0.0373	0.0335	0.0287	0.0240	0.0264	0.0338	0.0098

Pertini	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1978	0.1210	0.1027	0.0902	0.0655	0.0497	0.0836	0.0897
1979	0.1014	0.1063	0.0827	0.0856	0.0626	0.0670	0.1038
1980	0.0683	0.0872	0.0741	0.0703	0.0441	0.0605	0.0641
1981	0.0806	0.1006	0.0973	0.0678	0.0615	0.0819	0.0890
1982	0.0640	0.1060	0.0859	0.0651	0.0596	0.0735	0.0948
1983	0.0990	0.1194	0.0878	0.0918	0.0711	0.1101	0.1280
1984	0.0463	0.0563	0.0460	0.0528	0.0492	0.0510	0.0613

Pertini	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1978	0.0985	0.0707	0.0781	0.0558	0.0805	0.0909	0.0798	0.0684	0.0583	0.0654
1979	0.0579	0.0565	0.0660	0.0599	0.0529	0.0887	0.0776	0.0714	0.0552	0.0318
1980	0.0569	0.0321	0.0435	0.0453	0.0331	0.0900	0.0608	0.0569	0.0319	0.0327
1981	0.0617	0.0471	0.0465	0.0711	0.0486	0.0674	0.0352	0.0617	0.0419	0.0394
1982	0.0874	0.0488	0.0539	0.0492	0.0504	0.0679	0.0441	0.0593	0.0442	0.0631
1983	0.0959	0.0558	0.0865	0.0884	0.0809	0.0771	0.0551	0.0836	0.0515	0.0459
1984	0.0620	0.0463	0.0327	0.0463	0.0279	0.0527	0.0339	0.0358	0.0346	0.0354

Table 2.14g
Cossiga´s similarities

Cossiga	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1985	0.1422	0.1296	0.1924	0.0972	0.0801	0.0804	0.0943
1986	0.0962	0.0772	0.0858	0.0598	0.0563	0.0575	0.0522
1987	0.0886	0.0760	0.0973	0.0719	0.0788	0.0644	0.0836
1988	0.1268	0.1012	0.1162	0.0832	0.0657	0.0889	0.0827
1989	0.0878	0.1017	0.0832	0.0686	0.0753	0.0638	0.0759
1990	0.1124	0.1178	0.1098	0.1102	0.0700	0.1164	0.1111
1991	0.0139	0.0199	0.0185	0.0175	0.0257	0.0138	0.0114

Cossiga	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1985	0.1252	0.0550	0.0827	0.0958	0.1074	0.0661	0.0807	0.1094	0.0600	0.0947
1986	0.1307	0.0563	0.0821	0.0755	0.0510	0.0586	0.0464	0.1307	0.0724	0.0741
1987	0.1106	0.0621	0.0850	0.0763	0.1019	0.0604	0.0476	0.1163	0.0640	0.1366
1988	0.1398	0.0662	0.0800	0.0935	0.0793	0.0826	0.0593	0.1020	0.0616	0.0971
1989	0.1023	0.0714	0.0916	0.1224	0.1171	0.1067	0.0675	0.1083	0.0630	0.0868
1990	0.1058	0.0745	0.0812	0.1189	0.0768	0.0689	0.0452	0.1016	0.0738	0.0967
1991	0.0140	0.0443	0.0102	0.0135	0.0336	0.0365	0.0181	0.0090	0.0108	0.0062

Table 2.14h
Scalfaro´s similarities

Scalfaro	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1992	0.1105	0.0684	0.0827	0.1116	0.0761	0.0907	0.0791	0.1159	0.0774	0.0843
1993	0.1390	0.1020	0.0816	0.1012	0.0985	0.0951	0.0642	0.1390	0.0762	0.0884
1994	0.0988	0.0977	0.0810	0.0911	0.0748	0.1043	0.0739	0.1167	0.0902	0.0747
1995	0.0819	0.0927	0.0730	0.0864	0.0664	0.0850	0.0473	0.1152	0.0582	0.0526
1996	0.0667	0.0929	0.1329	0.0947	0.0750	0.1465	0.0989	0.0817	0.0497	0.0710
1997	0.0882	0.0662	0.0874	0.1011	0.0513	0.0875	0.0622	0.1094	0.0463	0.0541
1998	0.0952	0.0906	0.0848	0.0917	0.0722	0.0887	0.0550	0.1089	0.0805	0.0430

Tabel 2.14i
Ciampi`s similarities

Ciampi	**Napolitano**		
	2006	**2007**	**2008**
1999	0.1451	0.0858	0.1220
2000	0.0775	0.0638	0.0523
2001	0.0871	0.0605	0.0564
2002	0.1329	0.0558	0.0787
2003	0.1271	0.0467	0.1262
2004	0.1085	0.0984	0.0788
2005	0.0969	0.0374	0.0733

Table 2.15
Pre-*h*-similarities of associated vocabularies presented in Table 2.8

Table 2.15a
Pre-*h*-selfsimilarities

Einaudi	**1949**	**1950**	**1951**	**1952**	**1953**	**1954**	
1949	1	0.0476	0	0.2222	0.0667	0.1111	
1950	0.0476	1	0	0.0238	0.0286	0.0476	
1951	0	0	1	0.2222	0.0667	0.1111	
1952	0.2222	0.0238	0.2222	1	0.1333	0.2222	
1953	0.0667	0.0286	0.0667	0.1333	1	0.2667	
1954	0.1111	0.0476	0.1111	0.2222	0.2667	1	
Gronchi	**1955**	**1956**	**1957**	**1958**	**1959**	**1960**	**1961**
1955	1	0.0357	0	0	0	0	0
1956	0.0357	1	0	0	0	0	0
1957	0	0	1	0	0	0	0.0250
1958	0	0	0	1	0	0	0
1959	0	0	0	0	1	0	0
1960	0	0	0	0	0	1	0.0333
1961	0	0	0.0250	0	0	0.0333	1
Segni	**1962**	**1963**					
1962	1	0					
1963	0	1					
Saragat	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1964	1	0.0556	0.0185	0.0185	0	0	0

1965	0.0556	1	0	0.0370	0	0	0
1966	0.0185	0	1	0.0123	0.0222	0	0
1967	0.0185	0.0370	0.0123	1	0	0.0556	0.0635
1968	0	0	0.0222	0	1	0	0.0286
1969	0	0	0	0.0556	0	1	0.0714
1970	0	0	0	0.0635	0.0286	0.0714	1
Leone	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1971	1	0	0	0	0	0.0185	0.0238
1972	0	1	0	0	0	0.0250	0
1973	0	0	1	0	0	0	0.0222
1974	0	0	0	1	0	0.0208	0.0185
1975	0	0	0	0	1	0	0
1976	0.0185	0.0250	0	0.0208	0	1	0.0139
1977	0.0238	0	0.0222	0.0185	0	0.0139	1
Pertini	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1978	1	0.0571	0.0179	0.0179	0	0.0119	0
1979	0.0571	1	0.1125	0.1125	0.0750	0.0750	0
1980	0.0179	0.1125	1	0.0156	0.0104	0.0104	0.0125
1981	0.0179	0.1125	0.0156	1	0.0938	0.0938	0.0125
1982	0	0.0750	0.0104	0.0938	1	0.1736	0.0083
1983	0.0119	0.0750	0.0104	0.0938	0.1736	1	0.0083
1984	0	0	0.0125	0.0125	0.0083	0.0083	1
Cossiga	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1985	1	0.0070	0.0682	0	0.0818	0.0065	0
1986	0.0070	1	0.0064	0.0280	0	0.0055	0
1987	0.0682	0.0064	1	0.0076	0	0.0060	0
1988	0	0.0280	0.0076	1	0.0091	0.0260	0
1989	0.0818	0	0	0.0091	1	0.1786	0
1990	0.0065	0.0055	0.0060	0.0260	0.1786	1	0
1991	0	0	0	0	0	0	1
Scalfaro	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1992	1	0.0417	0.0060	0.2722	0.0556	0.0333	0.0625
1993	0.0417	1	0	0.0333	0	0.0125	0.0104
1994	0.0060	0	1	0.1190	0.0476	0.0643	0.0238
1995	0.2722	0.0333	0.1190	1	0.1000	0.0267	0.1389
1996	0.0556	0	0.0476	0.1000	1	0	0.1250
1997	0.0333	0.0125	0.0643	0.0267	0	1	0.0750
1998	0.0625	0.0104	0.0238	0.1389	0.1250	0.0750	1
Ciampi	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
1999	1	0.1143	0.0519	0.0714	0.2571	0.0179	0.0179

2000	0.1143	1	0.0727	0.1000	0.1600	0.0250	0.0250
2001	0.0519	0.0727	1	0.1023	0.0727	0.0455	0.0114
2002	0.0714	0.1000	0.1023	1	0.2250	0.0625	0.0156
2003	0.2571	0.1600	0.0727	0.2250	1	0.0250	0.0250
2004	0.0179	0.0250	0.0455	0.0625	0.0250	1	0.0625
2005	0.0179	0.0250	0.0114	0.0156	0.0250	0.0625	1
Napolitano	**2006**	**2007**	**2008**				
2006	1	0.1111	0				
2007	0.1111	1	0.0208				
2008	0	0.0208	1				

Table 2.15b
Einaudi´s similarities

Einaudi	**Gronchi**							**Segni**	
	1955	**1956**	**1957**	**1958**	**1959**	**1960**	**1961**	**1962**	**1963**
1949	0	0.0833	0	0	0	0	0	0	0
1050	0	0	0	0	0	0	0	0	0
1951	0	0	0	0	0	0	0	0	0.0556
1952	0	0	0	0	0	0	0	0	0.0278
1953	0	0	0	0.0333	0	0	0	0	0.0333
1954	0	0	0	0	0	0	0	0	0.0556

Einaudi	**Saragat**						
	1964	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1949	0	0	0	0	0	0	0.0476
1950	0	0	0	0	0	0	0.0816
1951	0	0	0	0	0	0	0
1952	0.0278	0.0556	0	0.0185	0	0	0.0238
1953	0	0	0	0	0	0	0.0286
1954	0	0	0	0	0	0	0.0476

Einaudi	**Leone**						
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1949	0	0.0667	0	0	0	0.0417	0.0370
1950	0.0238	0	0	0	0.0238	0	0.0635
1951	0	0	0	0	0	0	0

1952	0	0	0	0.0278	0	0	0.0185
1953	0	0	0	0	0.0333	0	0.0222
1954	0	0	0	0	0	0	0.0370

Einaudi	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1949	0.0476	0.0333	0	0.0417	0.0278	0.1111	0.0333
1950	0.0204	0	0	0	0.0119	0.0476	0.0143
1951	0	0	0	0	0	0	0
1952	0	0	0	0	0.0556	0.0556	0.0167
1953	0	0.0200	0	0.0250	0.0667	0.0667	0.0200
1954	0	0	0	0	0.0278	0.0278	0.0333

Einaudi	**Cossiga**						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1949	0	0	0	0	0	0	0
1950	0.0130	0.0110	0	0	0.0143	0.0102	0
1951	0	0	0	0.0303	0	0	0
1952	0	0	0	0.0152	0	0	0
1953	0.0182	0	0	0.0182	0.0200	0	0
1954	0	0	0	0.0303	0	0	0

Einaudi	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1949	0	0	0.0238	0.0889	0.0556	0	0.0278
1950	0	0.0179	0	0.0095	0	0	0.0119
1951	0	0	0	0	0	0	0
1952	0	0.0208	0.0119	0.0111	0	0	0.0139
1953	0	0.0250	0	0.0133	0	0	0.0167
1954	0	0	0	0.0222	0	0	0.0278

Einaudi	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1949	0	0	0	0	0	0.042	0	0.0556	0.0556	0
1950	0	0	0.0130	0	0	0.018	0.0179	0	0	0.0179
1951	0.0476	0	0	0	0	0	0	0	0.0556	0.0417
1952	0.0238	0	0	0	0	0.0833	0.0208	0.0278	0.0278	0.0208
1953	0.0286	0	0	0.0250	0	0.0250	0	0	0.0333	0.0250
1954	0.0476	0	0	0	0	0.0417	0	0	0.0556	0.0417

Table 2.15c
Gronchi´s similarities

Gronchi	Segni		Saragat						
	1962	1963	1964	1965	1966	1967	1968	1969	1970
1955	0	0	0	0	0	0	0	0.0179	0
1956	0	0	0	0	0	0	0	0	0
1957	0	0	0	0	0	0	0.0500	0	0
1958	0	0	0	0	0.0185	0	0	0	0
1959	0	0.0238	0	0	0	0	0	0	0
1960	0	0	0	0	0	0.0370	0	0	0
1961	0	0	0	0	0	0	0	0	0

Gronchi	Leone						
	1971	1972	1973	1974	1975	1976	1977
1955	0	0	0	0	0	0	0
1956	0	0.0500	0	0	0	0.0313	0
1957	0.0417	0	0	0.0417	0	0	0
1958	0	0.0333	0	0	0.0278	0	0
1959	0	0	0	0	0.0238	0	0
1960	0	0	0	0	0	0	0
1961	0	0	0.0200	0	0	0.0125	0.0111

Gronchi	Pertini						
	1978	1979	1980	1981	1982	1983	1984
1955	0	0	0	0	0	0	0
1956	0.0357	0.0250	0	0.0313	0	0.0208	0
1957	0	0	0	0	0	0	0
1958	0	0.0167	0	0.0208	0.0139	0.0139	0
1959	0	0	0	0	0	0	0
1960	0	0	0	0	0	0.0278	0
1961	0	0	0	0	0	0	0
Gronchi	Cossiga						
	1985	1986	1987	1988	1989	1990	1991
1955	0	0	0	0	0	0	0
1956	0.0227	0	0	0	0	0	0
1957	0	0	0	0.0227	0	0	0
1958	0.0152	0	0.0139	0.0152	0.0167	0	0

1959	0	0	0	0	0	0.0102	0
1960	0.0303	0	0.0278	0	0	0	0
1961	0	0.0077	0	0.0364	0	0.0071	0

Gronchi	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1955	0	0	0	0	0	0	0
1956	0.0208	0	0.0179	0.0167	0.0417	0	0
1957	0.0208	0.0313	0	0.0167	0	0	0
1958	0	0.0208	0	0	0	0	0
1959	0	0	0	0	0	0	0
1960	0	0	0	0	0	0	0.0278
1961	0	0	0.0071	0	0	0	0

Gronchi	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1955	0	0	0	0	0	0	0	0	0	0.0179
1956	0	0	0	0	0	0	0	0.0417	0.0417	0
1957	0	0	0	0	0	0	0	0	0	0
1958	0	0	0	0.0208	0	0	0	0	0	0
1959	0	0	0	0	0	0	0	0	0	0
1960	0	0	0	0.0417	0.0667	0	0	0	0	0
1961	0.0143	0	0.0091	0	0.0200	0	0	0.0167	0.0167	0

Table 2.15 d
Segni´s similarities

Segni								
	1964	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**	**Saragat**
1962	0	0	0	0	0	0	0	
1963	0	0	0	0	0	0	0	
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**Leone**
1962	0	0	0	0	0	0	0	
1963	0	0	0	0	0	0	0	
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**Pertini**
1962	0	0	0	0	0	0	0	
1963	0	0	0	0	0	0	0	
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**Cossiga**
1962	0	0	0	0	0	0	0	
1963	0	0	0	0.0152	0	0.0476	0	

	1992	1993	1994	1995	1996	1997	1998	
1962	0	0	0	0	0	0	0	**Scalfaro**
1963	0.0139	0	0	0	0	0	0	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	
1962	0	0	0	0	0	0	0	**Ciampi**
1963	0.0238	0	0	0	0	0	0	
	2006	**2007**	**2008**					
1962	0	0	0					**Napolitano**
1963	0	0.0278	0.0208					

Table 2.15e
Saragat´s similarities

Saragat	**Leone**						
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1964	0	0	0	0	0	0	0
1965	0	0	0	0	0	0	0
1966	0	0.0222	0.0222	0	0	0	0
1967	0	0.0889	0	0	0	0	0
1968	0.0333	0	0	0	0	0	0
1969	0.0208	0.0250	0.0250	0	0	0	0.0139
1970	0	0.0286	0	0	0.0238	0	0.0159

Saragat	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1964	0	0	0	0.0208	0	0	0
1965	0	0.0333	0.0417	0	0.0278	0.0278	0
1966	0	0.0111	0.0139	0	0	0	0
1967	0	0.0111	0.0556	0.0139	0	0.0093	0.0111
1968	0	0	0	0	0	0	0
1969	0	0	0.0156	0.0156	0.0104	0	0.0125
1970	0	0	0.0179	0.0179	0.0119	0.0119	0.0571

Saragat	**Cossiga**						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1964	0	0	0	0	0	0	0
1965	0	0.0256	0.0278	0	0	0	0

1966	0	0.0085	0.0370	0.0404	0	0	0.0370
1967	0.0909	0	0.0833	0	0.0444	0.0317	0
1968	0	0	0	0	0	0	0
1969	0.0455	0	0.0417	0.0114	0.0125	0.0357	0
1970	0.0519	0.0110	0.0119	0	0.0143	0.0408	0

Saragat	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1964	0	0	0.0119	0	0	0	0
1965	0.0278	0	0.0238	0	0	0.0333	0
1966	0.0093	0	0	0.0074	0	0	0
1967	0.0370	0	0.0317	0.0296	0	0	0.0093
1968	0.0167	0.0250	0	0.0533	0.0333	0	0.0167
1969	0.0417	0	0.0089	0.0083	0.0208	0	0.0104
1970	0.0119	0	0.0102	0.0857	0.0238	0	0.0476

Saragat	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1964	0	0	0	0	0	0	0	0	0	0
1965	0	0.0667	0	0	0	0	0	0	0	0.0417
1966	0.0159	0.0222	0.0101	0.0139	0.0222	0.0139	0.0139	0	0	0
1967	0.0159	0.0222	0.0404	0.0556	0.2000	0.0139	0.0139	0.0185	0	0
1968	0	0	0	0	0	0	0	0.0333	0	0
1969	0.0179	0.0250	0.0114	0.0156	0.0250	0	0	0.0833	0	0.0156
1970	0	0	0.0519	0	0	0.0179	0.0000	0.0952	0	0.0179

Table 2.15f
Leone´s similarities

Leone	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1971	0	0	0	0.0139	0	0	0.0152
1972	0.0286	0.0200	0.0250	0.1000	0	0.0167	0.0200
1973	0	0	0	0	0	0	0
1974	0	0	0	0	0.0139	0.0556	0
1975	0	0.0167	0	0.0208	0.0139	0.0139	0.0167
1976	0.0179	0.0125	0	0.0156	0	0.0417	0.0125
1977	0	0	0	0	0.0370	0.0833	0.0111

Leone	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1971	0	0	0.0152	0.0167	0	0	0.0556
1972	0.0727	0.0154	0.0167	0.0182	0.0800	0.0571	0
1973	0	0	0	0.0182	0	0.0143	0.0667
1974	0	0	0	0	0	0	0
1975	0.0152	0.0128	0	0	0.0167	0.0119	0
1976	0	0	0	0.0114	0	0	0
1977	0	0	0	0.0909	0.0111	0.0714	0

Leone	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1971	0.0208	0	0.0111	0.0278	0	0.0139	0.0238
1972	0.0167	0	0.0571	0.0533	0.0333	0	0
1973	0	0.0250	0.0143	0	0	0	0
1974	0	0.0208	0.0119	0.0111	0	0.0167	0.0139
1975	0	0.0208	0	0	0	0	0
1976	0	0	0.0357	0.0333	0.0208	0.0125	0.0104
1977	0.0093	0.0139	0.0714	0.0296	0.0185	0.0111	0.1481

Leone	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1971	0.0333	0	0.0208	0.0333	0	0	0	0.0278	0	0
1972	0	0	0	0	0.0400	0	0	0.1333	0.0333	0
1973	0.0286	0	0	0	0	0	0	0.0333	0.0333	0
1974	0	0	0	0	0	0.0208	0.0208	0.0278	0	0
1975	0	0	0.0152	0.0208	0	0	0	0	0.0278	0.0208
1976	0	0	0	0.0156	0	0	0	0.0208	0.0208	0
1977	0.0159	0.0889	0.0101	0.0139	0.0222	0.0556	0.0139	0.0741	0.0185	0

Table 2.15g
Pertini´s similarities

Pertini	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1978	0.0130	0	0	0	0.0143	0	0
1979	0.0091	0.0077	0.0333	0	0.0100	0	0
1980	0.0114	0.0096	0.0417	0	0.0125	0.0089	0

1981	0.0455	0.0096	0	0.0114	0.0500	0.0089	0
1982	0.0076	0.0256	0.0069	0.0682	0.0083	0	0
1983	0.0303	0.0256	0.0278	0.0076	0.0083	0	0
1984	0.0091	0	0	0	0.0100	0.0071	0

Pertini	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1978	0	0	0.0408	0.0095	0.0238	0	0
1979	0.0333	0.0125	0.0071	0.0267	0.0167	0.0100	0
1980	0.0938	0	0.0089	0.0333	0	0.0500	0.0104
1981	0.0417	0.0625	0.0357	0.0750	0.0208	0.0125	0.0104
1982	0.0625	0.0938	0	0.0222	0.0139	0.0750	0.1111
1983	0.0278	0.1667	0.0238	0.0889	0.0139	0.0750	0.1111
1984	0.0083	0	0.0071	0.0267	0	0	0.0083

Pertini	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1978	0	0	0.0130	0	0	0	0.0179	0.0238	0.0238	0
1979	0.0143	0.0800	0.0091	0.0500	0.0200	0.0125	0.0125	0.0167	0.0167	0.0125
1980	0.0179	0.1000	0.0114	0.0156	0.0250	0.0156	0.0156	0.0208	0	0.0156
1981	0	0	0.0114	0.0625	0	0.0156	0	0.0833	0.0208	0
1982	0.0119	0.0667	0.0076	0.0938	0.0167	0.0938	0.0104	0.0556	0	0.0104
1983	0	0.0167	0.0076	0.0938	0.0167	0.0938	0.0417	0.0556	0.0139	0.0104
1984	0	0	0	0.0125	0	0.0125	0	0.0167	0	0

Table 2.15h
Cossiga´s similarities

Cossiga	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1985	0.0303	0.0114	0.0065	0.0061	0	0	0.0076
1986	0.0256	0.0385	0	0.0051	0	0.0308	0.0064
1987	0.0278	0	0	0.0056	0	0.0083	0.0069
1988	0.0303	0.0114	0.0260	0.0061	0.0606	0.0364	0.1212
1989	0.0333	0.0125	0.0071	0.0267	0	0	0.0083
1990	0.0952	0	0.0459	0.0429	0	0.0071	0.0238
1991	0	0	0	0	0	0	0

Cossiga	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1985	0	0	0.0083	0.0455	0.0182	0.0114	0	0.0152	0	0
1986	0.0110	0.0154	0.0629	0.0096	0.0154	0.0096	0	0	0	0.0385
1987	0.0119	0.0667	0.0303	0.0417	0.0667	0.0104	0.0417	0	0	0.0417
1988	0.0519	0.0182	0.0083	0.0455	0.0182	0.0455	0	0.0606	0.0606	0.0114
1989	0	0	0.0091	0.0500	0.0200	0	0	0.0167	0	0
1990	0.0102	0.0143	0.0584	0.0089	0.0143	0	0	0.0476	0.0119	0.0089
1991	0	0	0	0	0	0	0	0	0	0

Table 2.15i
Scalfaro´s similarities

Scalfaro	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1992	0.0476	0.1500	0.0682	0.1667	0.0667	0.0417	0.0104	0.0556	0	0.0104
1993	0.0179	0	0.0114	0.0625	0	0.0625	0.0625	0.0208	0	0
1994	0.0102	0	0	0	0	0.0357	0.0357	0.1071	0.0476	0
1995	0.0095	0.0133	0.0545	0.0750	0.0133	0.0750	0.0083	0.1000	0.0111	0
1996	0.0238	0.0333	0	0.0208	0.0333	0.0208	0	0.2500	0.0278	0
1997	0.0143	0.0200	0.0091	0.0125	0	0.0125	0	0	0	0.0125
1998	0.0119	0.0667	0.0682	0.1667	0.0667	0.0938	0	0.0556	0	0

Table 2-15j
Ciampi´s similarities

Ciampi	Napolitano		
	2006	**2007**	**2008**
1999	0.0238	0.0238	0.0179
2000	0.0333	0	0.0250
2001	0	0	0.0114
2002	0.0208	0	0
2003	0.0333	0	0
2004	0.0208	0	0
2005	0.0208	0	0

Several hypotheses can be set up and tested using these results. Let us mention only ten of them. The testing can show that a part of them can be answerd negatively, a part positively. We shall formulate them in form of questions.

(i) Is the within-president similarity decreasing or increasing in the course of time?

(ii) Are the within-presidents similarities greater than those between presidents? In other words, are the associations president-specific?

(iii) Are the similarities between direct followers greater than between more distant followers? In other words, does similarity decrease with time-distance? This is an analogue of question (i) in its between-president version.

(iv) Since each president wishes to differ from his predecessor, is it possible that his first speech differs more from the last speech of his predecessor than his speeches in the following years? The hypothesis is plausible: the speech of a new president must(!) differ from the last one of his predecessor but as time goes, a convergence is plausible, too, because a part of what the presidents must tell concerns the same problems and the nation forgets the speech of the predecessor.

(v) If we consider only the greatest component of the α-graph, do the similarities display different results?

(vi) If we obtain some results analyzing Italian presidents, can one generalize them or are they specific for Italian? The examination of this problem is an enormous enterprise but it can be generalized and simplified at the same time, namely

(vii) do consecutive works of one writer display some of the phenomena mentioned above, e.g. in question (ii)?

(viii) Can these phenomena be mutatis mutandis observed in music? Here the influence of predecessors must be evident because musicians learn from them but at least stepwise moving away must be visible.

(ix) Do other similarity measurements, e.g. using the complete dictionary or different similarity indicators yield different results?

(x) What methods are appropriate to answer the individual questions?

The questions can be added ad infinitum because the local solution of one of them creates new ones. Here we shall restrict ourselves to those selected problems that can be answered directly from Table 2.14 or 2.15.

Let us begin with question (i), the course of similarities between individual texts of one president. In other words, does the associative world of a president change in time or is it stable? In order to obtain a result we compute stepwise the means of cells over the diagonal in Table 2.14a. Step 1 consists of those cells which indicate one year difference, i.e. (1955,1956), (1956,1957),… which lie directly over the main diagonal. The mean bulk similarity is obtained dividing the sum by the number of cells. For Gronchi we obtained.

Step 1 = (0.0042 + 0.0422 + 0.0629 + 0.0198 + 0.0157 + 0.0686)/6 =
= 0.0292

Step 2 = (0.0078 + 0.0969 + 0.0422 + 0.1008 + 0.0480)/5 = 0.05896

Step 3 = (0.0119 + 0.0178 + 0.0901 + 0.0726)/4 = 0.0481
Step 4 = (0 + 0.0794 + 0.0744)/3 = 0.0512
Step 5 = (0.0147 + 0.0571)/2 = 0.0359
Step 6 = 0.0060.

If we plot these numbers (cf. Figure 2.12), we obtain the course of bulk similarities which can be interpreted also historically. Here we can do it only formally: the "next year´s" text has small similarity with the previous year, thereafter greater similarity occurs but it decreases almost steadily with increasing time gap. If we perform this computation for all presidents separately, we obtain the results presented in Table 2.16.

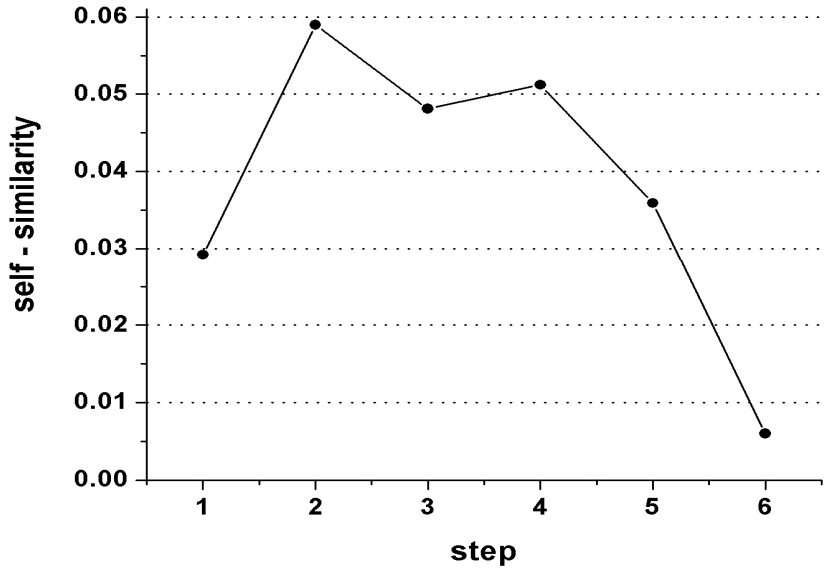

Figure 2.12. Development of Gronchi´s self-similarietes

Table 2.16
Development of self-similarities in the associative vocabulary

Step	Gronchi	Saragat	Leone	Pertini	Cossiga	Scalfaro	Ciampi	Napolitano
1	0.0292	0.0733	0.0459	0.1356	0.1218	0.1750	0.1138	0.0969
2	0.0590	0.0477	0.0527	0.1102	0.1214	0.1522	0.1232	0.1147
3	0.0481	0.0463	0.0605	0.0870	0.0997	0.1623	0.1209	
4	0.0512	0.0435	0.0571	0.1080	0.0795	0.1391	0.1159	
5	0.0359	0.0192	0.0259	0.1104	0.0951	0.1616	0.1270	
6	0.0060	0.0010	0.0057	0.0528	0.0365	0.1502	0.1238	

The graphs for all presidents are in Figure 2.13. Einaudi and Segni were left out because of small number of years or small vocabulary sizes. As can be seen, Gronchi is rather an exception. If one looks at Figure 2.13 one can see quite different courses but all are after six steps smaller than after the first. There is

only one exception, namely Ciampi, whose self-similarity increased after six steps. This fact can be interpreted in different ways: either his associative world is very fixed or he says different things with the same words, or, finally, the possible ghost writers were not flexible enough. This is, of course, rather a problem for sociologists and historians. The course of Napolitano is not yet finished thus the increasing self-semilarity after the second step cannot be evaluated.

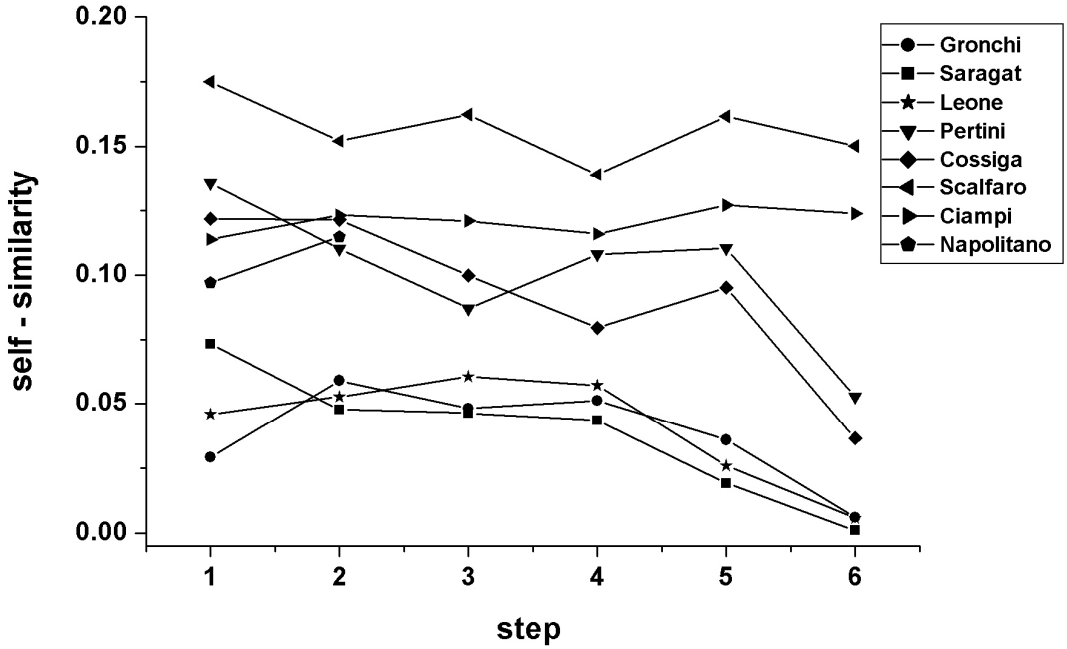

Figure 2.13. Similarities of texts in distances of 1 to 6 years with individual presidents

The general trend can be ascertained if one takes the means of steps (means of rows in Table 2.16) presented in Table 2.17. If one may generalize, then one can say that as years go, our concept formation and "world view" changes, we link the objects of reality in different ways, we set up different relations.

Table 2.17
Means of similarities in 6-year distances

Distance	Mean similarity
1	0.0989
2	0.0976
3	0.0893
4	0.0849
5	0.0822
6	0.0537

Here the monotonous decrease is evident. Thus, other than monotonous decrease with an individual president must be interpreted either historically or individually as a kind of exception. This will be left to specialists.

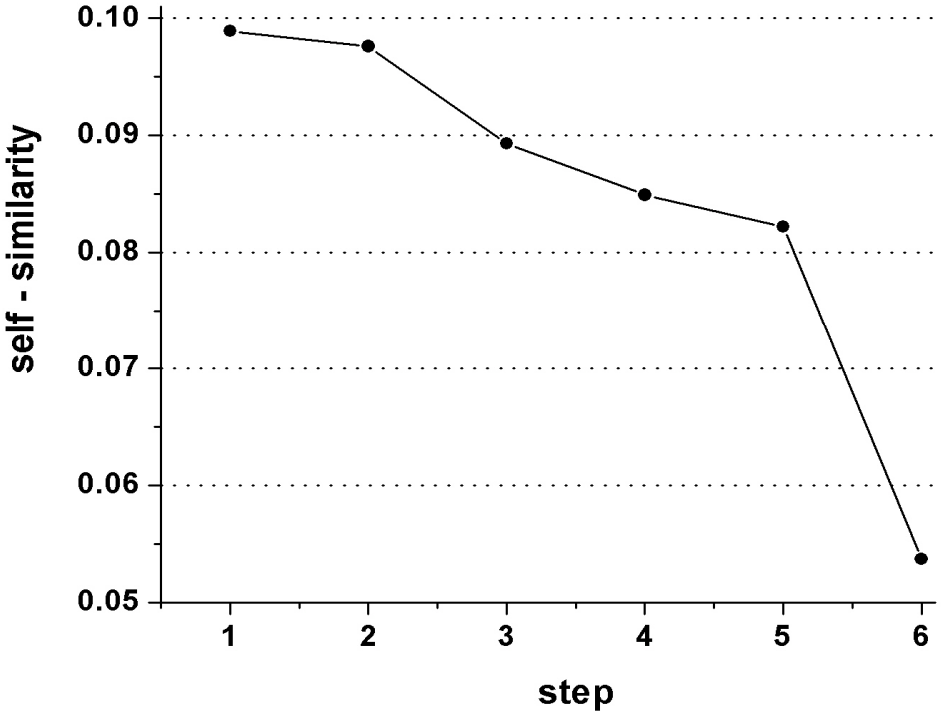

Figure 2.14. The general self-similarity trend of all presidents

Hovever, if we take the means of all similarities of each president, we obtain a not quite historical order as can be seen in Table 2.18. Nevertheless, a kind of increase can be observed.

Table 2.18
Mean similarities of individual presidents

President	**mean**	**stdev**
Gronchi	0.0541	0.0183
Segni	0.0595	0.0049
Saragat	0.0643	0.0236
Leone	0.0593	0.0165
Pertini	0.0711	0.0097
Cossiga	0.0823	0.0208
Scalfaro	0.0921	0.0110
Ciampi	0.0826	0.0092
Napolitano	0.0797	0.0131

In Table 2.18 the values of Segni and Napolitano are not quite reliable because of a smaller number of speeches. That of Napolitano can still change but in general a certain regularity can be observed. It seems to be a slowly increasing trend – as can be seen in Figure 2.15 – which must have a maximum and thereafter most probably an irregular oscillation will follow.

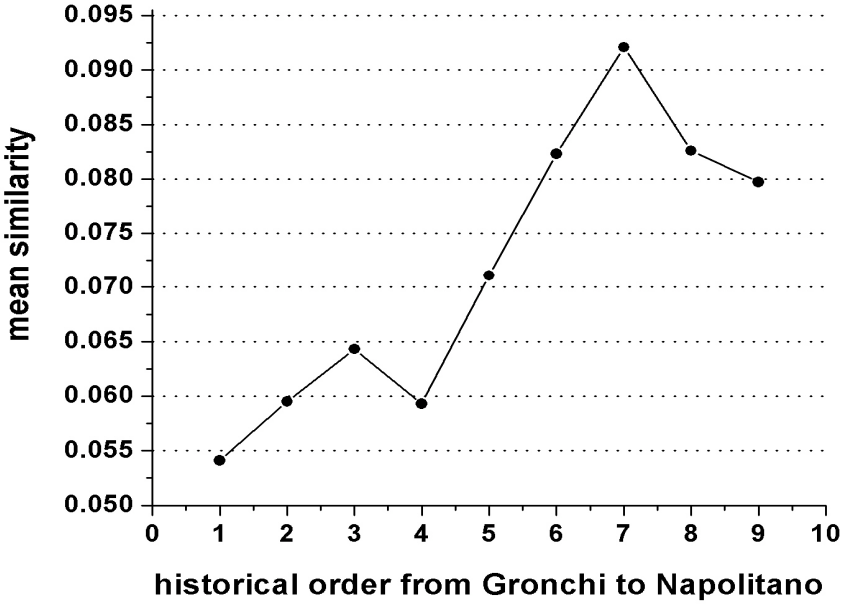

Figure 2.15. Mean similarities of presidents

The next possibility is to take the union of vocabularies of individual Presidents and perform the same procedure. We obtain the results in Table 2.19 and Table 2.20.

Looking at Table 2.19 we try to test again the hypothesis (i) conjecturing the decrease of similarity between presidents with increasing time distance. Even if Segni and Napolitano have a smaller number of texts, the comparison is possible. Step 1 represents those similarities which lie directly over the diagonal, i.e.

(0.0877 + 0.0849 + 0.1635 + 0.1015 + 0.1339 + 0.1856 + 0.1763 + 0.1847)/8 = 0.1398

i.e. the presidents neighbouring in time display a mean similarity of 0.1398. Computing all steps - taking always the next higher diagonal – we obtain the result presented in Table 2.21.

Table 2.19
Bulk similarities among associated vocabularies of Presidents (using Table 2.6) (gray: S > 0.17)

	Gronchi	**Segni**	**Saragat**	**Leone**	**Pertini**	**Cossiga**	**Scalfaro**	**Ciampi**	**Napolitano**
Gronchi 1955-1961	1	0.0877	0.161	0.1634	0.0789	0.1352	0.0947	0.1384	0.1056
Segni 1962-1963	0.0877	1	0.0849	0.0989	0.0553	0.0673	0.0485	0.0989	0.0715
Saragat 1964-1970	0.161	0.0849	1	0.1635	0.1202	0.1678	0.1361	0.1957	0.1415
Leone 1971-1977	0.1634	0.0989	0.1635	1	0.1015	0.19	0.1273	0.1558	0.1502
Pertini 1978-1984	0.0789	0.0553	0.1202	0.1015	1	0.1339	0.1534	0.1318	0.0899
Cossiga 1985-1991	0.1352	0.0673	0.1678	0.19	0.1339	1	0.1856	0.2047	0.1768
Scalfaro 1992-1998	0.0947	0.0485	0.1361	0.1273	0.1534	0.1856	1	0.1763	0.1242
Ciampi 1999-2005	0.1384	0.0989	0.1957	0.1558	0.1318	0.2047	0.1763	1	0.1847
Napolitano 2006-2008	0.1056	0.0715	0.1415	0.1502	0.0899	0.1768	0.1242	0.1847	1

Table 2.20
Pre-h-similarities among associated thematic vocabularies of individual Presidents (using Table 2.8) (gray: S > 0.1)

	Einaudi	**Gronchi**	**Segni**	**Saragat**	**Leone**	**Pertini**	**Cossiga**	**Scal-faro**	**Ciampi**	**Napoli-tano**
Einaudi	1	0.0058	0.0056	0.0135	0.0681	0.0435	0.0165	0.0444	0.0571	0.0523
Gronchi	0.0058	1	0.0026	0.0114	0.0421	0.0051	0.0395	0.0211	0.0068	0.0139
Segni	0.0056	0.0026	1	0	0	0	0.0167	0.0022	0.0029	0.0059
Saragat	0.0135	0.0114	0	1	0.0547	0.0288	0.0721	0.0601	0.0494	0.0572
Leone	0.0681	0.0421	0	0.0547	1	0.0658	0.0782	0.0939	0.0864	0.0721
Pertini	0.0435	0.0051	0	0.0288	0.0658	1	0.0326	0.1237	0.0894	0.0320
Cossiga	0.0165	0.0395	0.0167	0.0721	0.0782	0.0326	1	0.1189	0.1529	0.0534
Scalfaro	0.0444	0.0211	0.0022	0.0601	0.0939	0.1237	0.1189	1	0.1835	0.0641
Ciampi	0.0571	0.0068	0.0029	0.0494	0.0864	0.0894	0.1529	0.1835	1	0.0420
Napoli-tano	0.0523	0.0139	0.0059	0.0572	0.0721	0.032	0.0534	0.0641	0.0420	1

Table 2.21
Mean bulk similarities in time steps

Step	Observed mean similarity
1	0.1398
2	0.1503
3	0.1371
4	0.1056
5	0.1324
6	0.1117
7	0.1050
8	0.1056

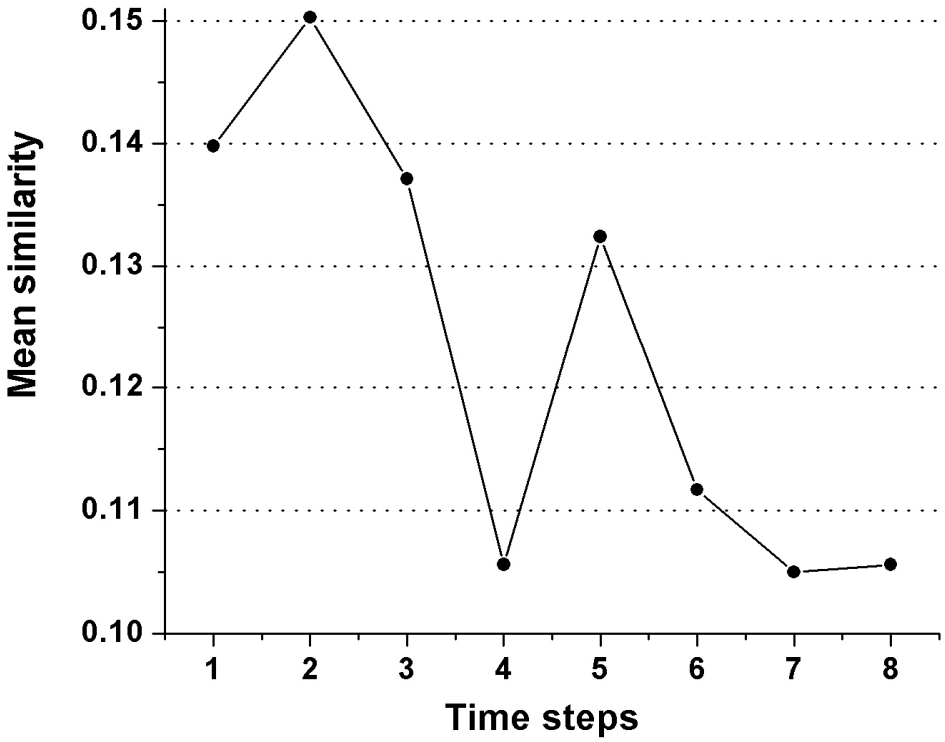

Figure 2.16. Mean bulk similarities

Though the trend is evident, no simple curve can be proposed to capture it. The same has been done with Table 2.20 and we obtained the trend presented in Table 2.22.

Table 2.22
Mean (thematic) pre-h-similarities in time steps in terms of followers (not years)

Step	Similarity
1	0.0562
2	0.0581
3	0.0520
4	0.0447
5	0.0413
6	0.0244
7	0.0190
8	0.0355
9	0.0523

The very peculiar course of similarities with increasing step is presented in Figure 2.17

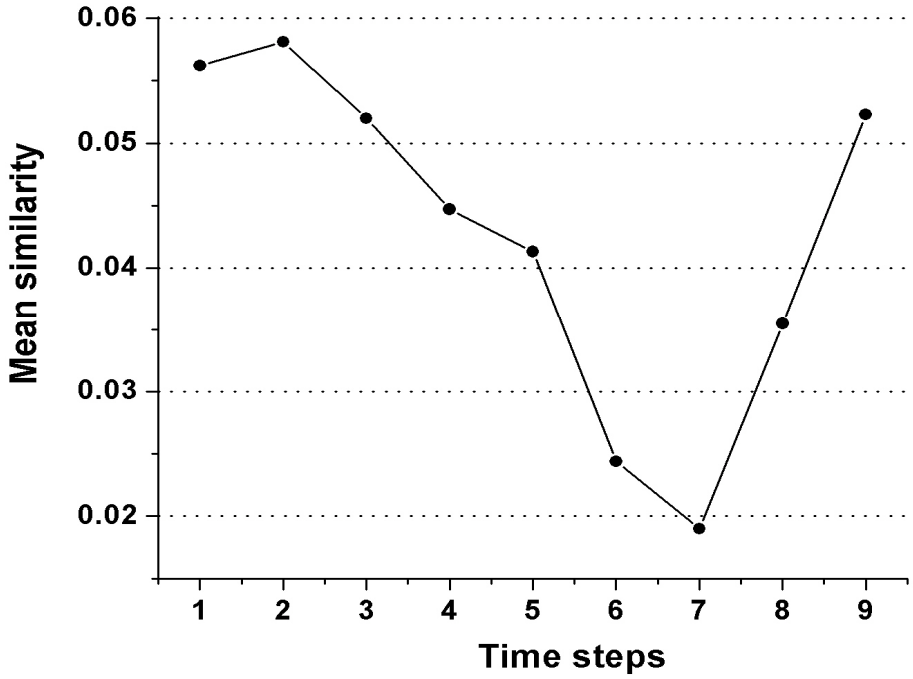

Figure 2.17. Mean (thematic) pre-h-similarities in time steps

Here only historians could give an approximate interpretation of the course of similarity. From the linguistic point of view we register it as a fact and do not try to smuggle in stylistic, historical or political issues. It can be supposed, however, that the course will be oscillating in the future if the texts are long enough. The

mean cannot become zero, though a pair of texts can attain this value. And it cannot become 1 if we already have lower values.

In Tables 2.23 and 2.24 we show only the most conspicuous similarities between texts (Table 23, $S \geq 0.1$) and between presidents (24).

Table 2.23

The most conspicuous pre-h-similarities (thematic) between texts ($S \geq 0.1$)

Thematically highly similar text pairs				**S**
1995	Scalfaro	1992	Scalfaro	0.2722
1953	Einaudi	1954	Einaudi	0.2667
1999	Ciampi	2003	Ciampi	0.2571
1996	Scalfaro	2006	Napolitano	0.2500
2002	Ciampi	2003	Ciampi	0.2250
1949	Einaudi	1952	Einaudi	0.2222
1951	Einaudi	1952	Einaudi	0.2222
1952	Einaudi	1954	Einaudi	0.2222
2003	Ciampi	1967	Saragat	0.2000
1989	Cossiga	1990	Cossiga	0.1786
1982	Pertini	1983	Pertini	0.1736
1983	Pertini	1993	Scalfaro	0.1667
1992	Scalfaro	2002	Ciampi	0.1667
1998	Scalfaro	2002	Ciampi	0.1667
2000	Ciampi	2003	Ciampi	0.1600
1992	Scalfaro	2000	Ciampi	0.1500
1977	Leone	1998	Scalfaro	0.1481
1995	Scalfaro	1998	Scalfaro	0.1389
1952	Einaudi	1953	Einaudi	0.1333
1972	Leone	2008	Napolitano	0.1333
1996	Scalfaro	1998	Scalfaro	0.1250
1988	Cossiga	1998	Scalfaro	0.1212
1994	Scalfaro	1995	Scalfaro	0.1190
1999	Ciampi	2000	Ciampi	0.1143
1979	Pertini	1980	Pertini	0.1125
1979	Pertini	1981	Pertini	0.1125
1949	Einaudi	1954	Einaudi	0.1111
1949	Einaudi	1983	Pertini	0.1111
1951	Einaudi	1954	Einaudi	0.1111
1982	Pertini	1998	Scalfaro	0.1111
1983	Pertini	1998	Scalfaro	0.1111
2006	Napolitano	2007	Napolitano	0.1111
1994	Scalfaro	2006	Napolitano	0.1071

2001	Ciampi	2002	Ciampi	0.1023
1972	Leone	1981	Pertini	0.1000
1980	Pertini	2000	Ciampi	0.1000
1995	Scalfaro	1996	Scalfaro	0.1000
1995	Scalfaro	2006	Napolitano	0.1000
2000	Ciampi	2002	Ciampi	0.1000

Table 2.24
Mean (thematic) pre-h-similarities of presidents

President	**mean**	**stdev**
Einaudi	0.0345	0.0062
Gronchi	0.0207	0.0026
Segni	0.0195	0.0041
Saragat	0.0270	0.0064
Leone	0.0270	0.0066
Pertini	0.0346	0.0094
Cossiga	0.0291	0.0055
Scalfaro	0.0391	0.0084
Ciampi	0.0378	0.0073
Napolitano	0.0351	0.0125

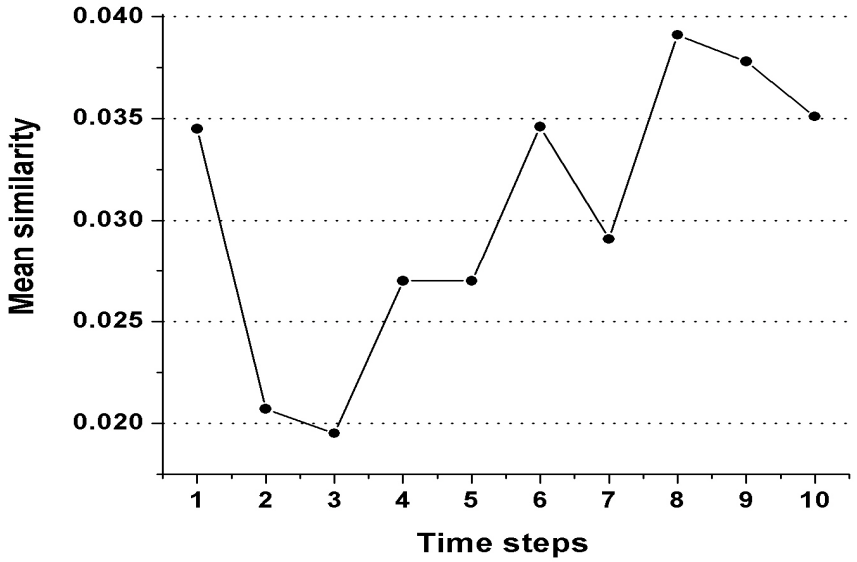

Figure 2.18. Mean similarities between presidents

2.9.2. Method II

If we take into account also the frequencies of occurrence of lemmas of associated words, than perhaps the following procedure is reasonable. Obtain the union of lemmas of all texts and for each text take into account the frequency of the given lemma in the given text. The union of all texts has the size (cardinality) *V*, and in individual texts there will be many zero frequencies.

One can compute the intertextual distance based on lexical connection introduced by Brunet (1988) and recently developed by Labbé (2007), Labbé and Labbé (2001), Merriam (2002).

In this study calculations were made on the basis of lemmas (Pauli and Tuzzi, 2009) rather than forms and, in particular, only on the basis of the "associative-lemmas" (all the lemmas included in components and cliques, Table 2.6).

Given a pair of speeches *A* and *B* of size N_A and N_B with $N_A \le N_B$, the frequency $f_{l,B}$ of each lemma-type *l* in the larger speech *B* is reduced according to the size of the shorter speech *A* in estimating the mathematical expectancy of the frequency of the lemma-type *l* in *A* $f^*_{l,B}$ by means of a simple proportion:

$$f^*_{l,B} : f_{l,B} = N_A : N_B$$

$$f^*_{l,B} = f_{l,B} \frac{N_A}{N_B}$$

hence

$$N^*_B = N_A .$$

The distance *d* between speech *A* and speech *B* is obtained as follows:

$$d(A,B) = \frac{\sum_{l \in L_{A \cup B}} \left| f_{l,A} - f^*_{l,B} \right|}{2N_A}, \tag{2.11}$$

where $L_{A \cup B}$ is the sub-vocabulary containing the "associative-lemmas" of speech *A and* speech *B*, i.e. all the associative-lemmas present in at least one of the two speeches. In brief, intertextual distance is obtained by calculating the difference between the frequency of any associative-lemma in speech *A* and its (estimated) frequency in speech *B*. If two speeches are identical, they contain the same words with the same frequency and their distance amounts to zero. If two speeches have no words in common, they are separated by a distance amounting to 1 (maximum "theoretical" distance).

Intertextual distance is calculated according to the lexical profiles of all possible pair of speeches. The distances between pairs may be easily expressed by a squared matrix of dimensions $n \times n$ ($n = 60$) with rows and columns assigned to speeches. The generic element of matrix *D* is such that:

$$d_{ij} = d_{ji} = d\left(B_i, B_j\right)$$

since the distance between *A* and *B* is the same as the distance between *B* and *A*. The generic element of the main diagonal is:

$$d_{ii} = 0 = d\left(B_i, B_i\right)$$

because the distance between each speech and itself amounts to zero.

The total number of pairs to be considered is therefore 1,770 as expressed through $n(n-1)/2$.

Labbé and Labbé (2001) have provided an intertextual distance standardized scale. According to the authors' comments, an intertextual distance below the 0.20 limit is sufficient to achieve sure authorship attribution; in contrast, beyond 0.30 different authors, text genres and topics may be presumed. Claims on the absolute value of the intertextual distance between two texts were made in a different context and whether they apply to the case at hand is uncertain.

Consequently, relative values were opted for, i.e. we preferred to proceed according to a comparative approach within our matrix. The intertextual distances contained in the matrix provide information on similarities and differences between all speech pairs included in the corpus. The very same distances also allow to represent the 60 speeches in a dendrogram typically used for cluster analysis. A cluster analysis of the 60 presidential speeches was performed by using the squared matrix of distances and an agglomerative hierarchical cluster algorithm with complete linkage, i.e. the distance between pairs of clusters is obtained as the maximum distance among all pairs of elements of the two clusters; pairs of clusters with minimum distance are aggregated. We used complete linkage since we expected to find well separated and tight (convex shaped) clusters. The result is presented in Figure 2.19. Here a word of warning would be in place. Classifications merely reflect a part of our concept formation, they are practical means for ordering the universe of discourse but they are non-theoretical accounts. Each of the immense number of classifications yields a slightly different result and all of them are approximations to the truth. But what kind of truth? A "truth" constructed a posteriori and perhaps a source of future hypotheses, but not a result of testing some a priori hypotheses. Classification is beloved in qualitative linguistics, frequently as the highest aim, but rather shunned in quantitative linguistics. For literary studies it can become a basis of the search for idiosyncrasies, for historians a source of political evaluation, for psychologists a material for studying perseveration, imitation, etc.

Table 2.25 displays all intertextual distances. We shall use it for examination of the increase of distance with years and with followers, in analogy to the analysis performed above. All results were computed to 6 decimal places but rounded to 2 places.

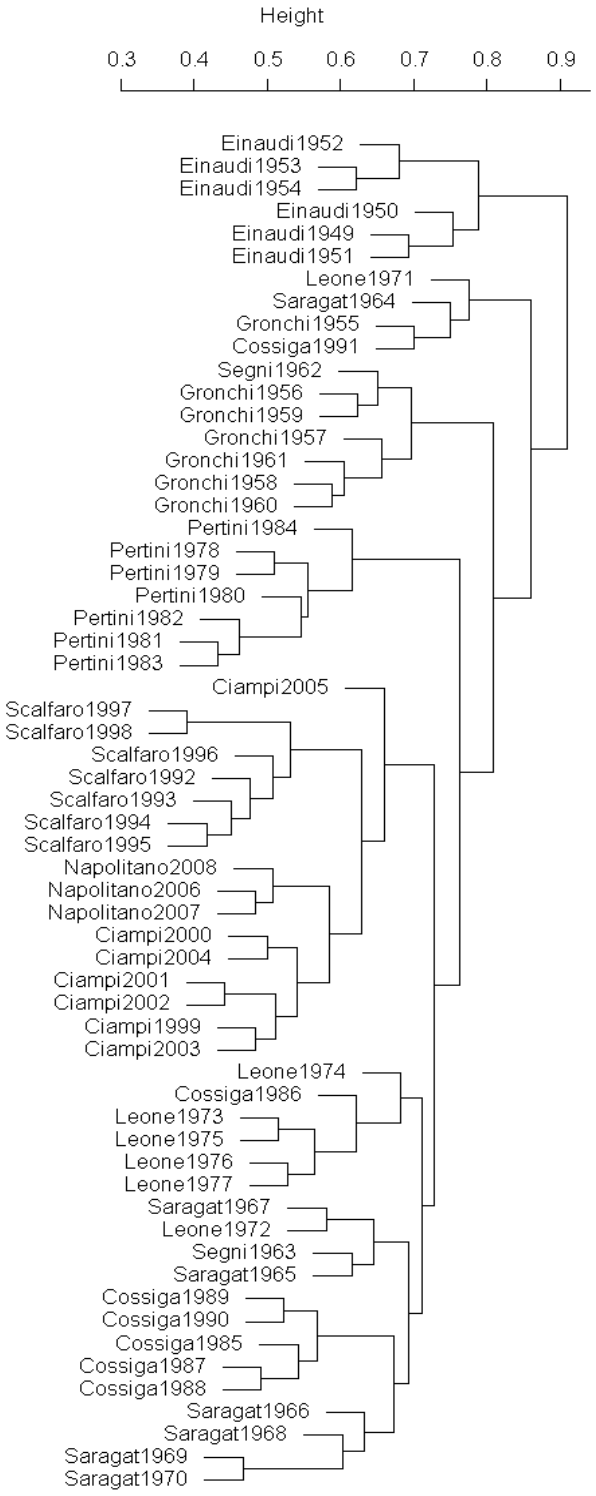

Figure 2.19. Intertextual distances and classification of speeches

Table 2.25
Intertextual distances between speeches

	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
1949	-	0.75	0.69	0.73	0.79	0.71	0.71	0.84	0.87	0.81	0.79	0.81	0.85	0.86	0.83	0.76	0.83	0.90	0.86	0.86	0.89	0.89	0.78	0.84	0.87	0.86	0.87	0.89	0.90
1950	0.75	-	0.72	0.77	0.79	0.73	0.78	0.83	0.87	0.83	0.81	0.82	0.85	0.81	0.80	0.86	0.80	0.86	0.83	0.85	0.84	0.86	0.82	0.84	0.87	0.83	0.88	0.89	0.89
1951	0.69	0.72	-	0.68	0.72	0.68	0.70	0.81	0.83	0.78	0.79	0.81	0.83	0.84	0.75	0.77	0.84	0.86	0.83	0.85	0.83	0.87	0.84	0.78	0.86	0.83	0.83	0.86	0.83
1952	0.73	0.77	0.68	-	0.66	0.68	0.76	0.85	0.84	0.83	0.79	0.83	0.82	0.85	0.81	0.81	0.81	0.82	0.83	0.82	0.86	0.83	0.78	0.83	0.87	0.81	0.83	0.83	0.88
1953	0.79	0.79	0.72	0.66	-	0.62	0.78	0.83	0.82	0.82	0.81	0.83	0.86	0.82	0.81	0.78	0.81	0.85	0.84	0.81	0.84	0.83	0.82	0.80	0.81	0.82	0.84	0.89	0.84
1954	0.71	0.73	0.68	0.68	0.62	-	0.76	0.80	0.82	0.81	0.77	0.83	0.81	0.77	0.79	0.76	0.82	0.78	0.82	0.82	0.80	0.84	0.83	0.77	0.83	0.82	0.84	0.87	0.84
1955	0.71	0.78	0.70	0.76	0.78	0.76	-	0.75	0.74	0.71	0.74	0.77	0.77	0.77	0.73	0.75	0.73	0.78	0.73	0.77	0.77	0.77	0.78	0.73	0.74	0.75	0.76	0.75	0.76
1956	0.84	0.83	0.81	0.85	0.83	0.80	0.75	-	0.65	0.61	0.62	0.62	0.69	0.65	0.65	0.70	0.70	0.69	0.71	0.69	0.64	0.69	0.85	0.69	0.70	0.69	0.68	0.70	0.67
1957	0.87	0.87	0.83	0.84	0.82	0.82	0.74	0.65	-	0.62	0.66	0.66	0.61	0.70	0.68	0.70	0.72	0.68	0.66	0.64	0.61	0.65	0.84	0.67	0.68	0.74	0.67	0.67	0.65
1958	0.81	0.83	0.78	0.83	0.82	0.81	0.71	0.61	0.62	-	0.62	0.59	0.60	0.64	0.65	0.68	0.69	0.68	0.66	0.67	0.64	0.68	0.83	0.62	0.65	0.67	0.68	0.71	0.68
1959	0.79	0.81	0.79	0.79	0.81	0.77	0.74	0.62	0.66	0.62	-	0.60	0.62	0.63	0.65	0.73	0.68	0.72	0.71	0.67	0.66	0.67	0.81	0.67	0.69	0.69	0.70	0.70	0.66
1960	0.81	0.82	0.81	0.83	0.83	0.83	0.77	0.62	0.66	0.59	0.60	-	0.59	0.61	0.68	0.76	0.67	0.72	0.65	0.66	0.66	0.65	0.79	0.67	0.69	0.67	0.68	0.70	0.64
1961	0.85	0.85	0.83	0.82	0.86	0.81	0.77	0.69	0.61	0.60	0.62	0.59	-	0.67	0.69	0.76	0.68	0.67	0.63	0.62	0.61	0.64	0.81	0.66	0.65	0.67	0.63	0.61	0.64
1962	0.86	0.81	0.84	0.85	0.82	0.77	0.77	0.65	0.70	0.64	0.63	0.61	0.67	-	0.67	0.74	0.68	0.66	0.64	0.65	0.65	0.67	0.83	0.66	0.69	0.71	0.74	0.72	0.71
1963	0.83	0.80	0.75	0.81	0.81	0.79	0.73	0.65	0.68	0.65	0.65	0.68	0.69	0.67	-	0.66	0.62	0.66	0.63	0.65	0.65	0.68	0.79	0.65	0.67	0.71	0.67	0.68	0.68
1964	0.76	0.86	0.77	0.81	0.78	0.76	0.75	0.70	0.70	0.68	0.73	0.76	0.76	0.74	0.66	-	0.70	0.73	0.70	0.74	0.70	0.73	0.76	0.65	0.70	0.73	0.68	0.69	0.73
1965	0.83	0.80	0.84	0.81	0.81	0.82	0.73	0.70	0.72	0.69	0.68	0.67	0.68	0.68	0.62	0.70	-	0.63	0.63	0.64	0.64	0.65	0.77	0.64	0.70	0.67	0.70	0.70	0.69
1966	0.90	0.86	0.86	0.82	0.85	0.78	0.78	0.69	0.68	0.68	0.72	0.72	0.67	0.66	0.66	0.73	0.63	-	0.61	0.62	0.63	0.63	0.81	0.67	0.63	0.68	0.65	0.64	0.67
1967	0.86	0.83	0.83	0.83	0.84	0.82	0.73	0.71	0.66	0.66	0.71	0.65	0.63	0.64	0.63	0.70	0.63	0.61	-	0.63	0.61	0.60	0.80	0.58	0.65	0.68	0.64	0.63	0.67
1968	0.86	0.85	0.85	0.82	0.81	0.82	0.77	0.69	0.64	0.67	0.67	0.66	0.62	0.65	0.65	0.74	0.64	0.62	0.63	-	0.57	0.60	0.81	0.64	0.68	0.68	0.69	0.68	0.67
1969	0.89	0.84	0.83	0.86	0.84	0.80	0.77	0.64	0.61	0.64	0.66	0.66	0.61	0.65	0.65	0.70	0.64	0.63	0.61	0.57	-	0.47	0.76	0.61	0.61	0.68	0.63	0.60	0.61
1970	0.89	0.86	0.87	0.83	0.83	0.84	0.77	0.69	0.65	0.68	0.67	0.65	0.64	0.67	0.68	0.73	0.65	0.63	0.60	0.60	0.47	-	0.77	0.63	0.64	0.67	0.60	0.61	0.60
1971	0.78	0.82	0.84	0.78	0.82	0.83	0.78	0.85	0.84	0.83	0.81	0.79	0.81	0.83	0.79	0.76	0.77	0.81	0.80	0.81	0.76	0.77	-	0.74	0.85	0.77	0.86	0.79	0.81
1972	0.84	0.84	0.78	0.83	0.80	0.77	0.73	0.69	0.67	0.62	0.67	0.67	0.66	0.66	0.65	0.65	0.64	0.67	0.58	0.64	0.61	0.63	0.74	-	0.63	0.61	0.64	0.62	0.61
1973	0.87	0.87	0.86	0.87	0.81	0.83	0.74	0.70	0.68	0.65	0.69	0.69	0.65	0.69	0.67	0.70	0.70	0.63	0.65	0.68	0.61	0.64	0.85	0.63	-	0.63	0.52	0.56	0.56
1974	0.86	0.83	0.83	0.81	0.82	0.82	0.75	0.69	0.74	0.67	0.69	0.67	0.67	0.71	0.71	0.73	0.67	0.68	0.68	0.68	0.68	0.67	0.77	0.61	0.63	-	0.63	0.60	0.64
1975	0.87	0.88	0.83	0.83	0.84	0.84	0.76	0.68	0.67	0.68	0.70	0.68	0.63	0.74	0.67	0.68	0.70	0.65	0.64	0.69	0.63	0.60	0.86	0.64	0.52	0.63	-	0.56	0.55
1976	0.89	0.89	0.86	0.83	0.89	0.87	0.75	0.70	0.67	0.71	0.70	0.70	0.61	0.72	0.68	0.69	0.70	0.64	0.63	0.68	0.60	0.61	0.79	0.62	0.56	0.60	0.56	-	0.53
1977	0.90	0.89	0.83	0.88	0.84	0.84	0.76	0.67	0.65	0.68	0.66	0.64	0.64	0.71	0.68	0.73	0.69	0.67	0.67	0.67	0.61	0.60	0.81	0.61	0.56	0.64	0.55	0.53	-

	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1949	0.87	0.88	0.89	0.88	0.88	0.88	0.90	0.88	0.88	0.89	0.88	0.87	0.88	0.88	0.88	0.88	0.90	0.91	0.86	0.91	0.91	0.87	0.86	0.87	0.88	0.85	0.87	0.80	0.88	0.89	0.89
1950	0.84	0.85	0.85	0.84	0.86	0.82	0.87	0.86	0.88	0.89	0.86	0.82	0.86	0.83	0.86	0.81	0.87	0.87	0.84	0.89	0.86	0.85	0.85	0.83	0.85	0.85	0.84	0.80	0.87	0.86	0.87
1951	0.81	0.84	0.84	0.83	0.84	0.82	0.85	0.83	0.82	0.85	0.85	0.83	0.83	0.79	0.85	0.81	0.85	0.86	0.83	0.86	0.85	0.83	0.80	0.81	0.82	0.81	0.81	0.76	0.87	0.84	0.85
1952	0.83	0.88	0.86	0.83	0.84	0.83	0.83	0.85	0.85	0.87	0.84	0.85	0.85	0.82	0.87	0.83	0.85	0.86	0.87	0.85	0.86	0.86	0.85	0.85	0.85	0.83	0.85	0.82	0.87	0.88	0.87
1953	0.82	0.84	0.82	0.83	0.84	0.83	0.85	0.80	0.81	0.82	0.82	0.82	0.85	0.79	0.84	0.81	0.85	0.86	0.84	0.86	0.86	0.82	0.81	0.82	0.85	0.80	0.81	0.80	0.85	0.85	0.82
1954	0.81	0.81	0.80	0.79	0.82	0.79	0.80	0.80	0.84	0.81	0.80	0.79	0.82	0.81	0.81	0.79	0.83	0.84	0.82	0.83	0.84	0.81	0.80	0.81	0.84	0.80	0.82	0.80	0.83	0.82	0.81
1955	0.77	0.78	0.79	0.78	0.78	0.76	0.78	0.75	0.76	0.76	0.73	0.75	0.76	0.70	0.79	0.77	0.78	0.79	0.78	0.79	0.78	0.73	0.74	0.75	0.75	0.75	0.74	0.75	0.77	0.77	0.74
1956	0.75	0.75	0.74	0.72	0.74	0.73	0.78	0.65	0.73	0.71	0.71	0.67	0.68	0.79	0.71	0.72	0.72	0.73	0.69	0.74	0.75	0.67	0.72	0.70	0.67	0.70	0.73	0.68	0.71	0.69	0.72
1957	0.75	0.77	0.77	0.73	0.74	0.74	0.80	0.64	0.65	0.60	0.64	0.67	0.64	0.76	0.69	0.70	0.68	0.71	0.71	0.72	0.72	0.65	0.69	0.66	0.66	0.68	0.69	0.70	0.66	0.68	0.68
1958	0.71	0.71	0.73	0.70	0.72	0.70	0.76	0.66	0.69	0.64	0.67	0.65	0.68	0.78	0.70	0.68	0.71	0.71	0.68	0.73	0.70	0.66	0.70	0.66	0.65	0.66	0.68	0.68	0.65	0.67	0.67
1959	0.72	0.77	0.77	0.72	0.72	0.73	0.79	0.69	0.72	0.68	0.69	0.66	0.67	0.78	0.73	0.73	0.73	0.77	0.74	0.76	0.77	0.68	0.73	0.70	0.70	0.71	0.73	0.75	0.70	0.70	0.71
1960	0.72	0.75	0.76	0.71	0.74	0.70	0.78	0.67	0.71	0.68	0.69	0.67	0.66	0.77	0.74	0.73	0.73	0.74	0.72	0.74	0.76	0.67	0.72	0.70	0.66	0.68	0.72	0.71	0.69	0.70	0.72
1961	0.70	0.71	0.73	0.67	0.71	0.68	0.74	0.65	0.63	0.64	0.64	0.66	0.64	0.79	0.67	0.68	0.66	0.66	0.67	0.68	0.67	0.63	0.65	0.65	0.61	0.64	0.65	0.70	0.62	0.65	0.63
1962	0.76	0.79	0.78	0.74	0.75	0.73	0.81	0.68	0.73	0.67	0.68	0.65	0.67	0.77	0.76	0.72	0.74	0.76	0.72	0.75	0.76	0.69	0.73	0.70	0.69	0.69	0.72	0.75	0.71	0.73	0.73
1963	0.70	0.68	0.67	0.64	0.66	0.65	0.74	0.67	0.71	0.65	0.66	0.64	0.65	0.74	0.68	0.61	0.64	0.69	0.62	0.69	0.66	0.67	0.63	0.63	0.63	0.62	0.62	0.63	0.64	0.66	0.69
1964	0.76	0.78	0.75	0.74	0.77	0.76	0.79	0.71	0.70	0.69	0.72	0.73	0.72	0.74	0.71	0.73	0.72	0.74	0.72	0.76	0.77	0.70	0.71	0.71	0.71	0.68	0.72	0.69	0.72	0.75	0.71
1965	0.68	0.69	0.69	0.66	0.69	0.66	0.70	0.66	0.71	0.67	0.66	0.66	0.69	0.73	0.65	0.64	0.67	0.64	0.63	0.68	0.68	0.63	0.64	0.60	0.63	0.62	0.61	0.67	0.66	0.65	0.68
1966	0.72	0.75	0.74	0.72	0.71	0.70	0.75	0.63	0.67	0.65	0.62	0.67	0.65	0.81	0.67	0.66	0.69	0.69	0.68	0.72	0.70	0.63	0.64	0.66	0.67	0.63	0.66	0.72	0.64	0.66	0.65
1967	0.72	0.72	0.73	0.70	0.72	0.68	0.70	0.64	0.68	0.62	0.64	0.64	0.62	0.73	0.68	0.65	0.67	0.70	0.69	0.68	0.69	0.59	0.61	0.63	0.58	0.63	0.64	0.66	0.63	0.67	0.63
1968	0.70	0.73	0.76	0.70	0.72	0.70	0.74	0.64	0.66	0.64	0.65	0.66	0.63	0.78	0.67	0.68	0.69	0.67	0.65	0.69	0.69	0.65	0.68	0.62	0.63	0.64	0.65	0.69	0.64	0.71	0.65
1969	0.68	0.70	0.68	0.68	0.67	0.66	0.73	0.57	0.60	0.58	0.60	0.63	0.61	0.77	0.59	0.62	0.61	0.64	0.62	0.67	0.66	0.57	0.61	0.62	0.59	0.60	0.63	0.69	0.60	0.58	0.59
1970	0.67	0.69	0.70	0.66	0.67	0.65	0.71	0.60	0.63	0.58	0.59	0.63	0.60	0.78	0.61	0.62	0.60	0.63	0.62	0.63	0.65	0.58	0.60	0.62	0.60	0.61	0.64	0.69	0.56	0.60	0.61
1971	0.80	0.84	0.82	0.84	0.84	0.86	0.80	0.82	0.82	0.79	0.81	0.82	0.82	0.76	0.82	0.82	0.80	0.83	0.83	0.83	0.84	0.79	0.77	0.80	0.82	0.78	0.82	0.77	0.82	0.83	0.82
1972	0.69	0.74	0.74	0.67	0.70	0.69	0.73	0.62	0.64	0.62	0.64	0.68	0.66	0.73	0.65	0.64	0.65	0.69	0.63	0.68	0.68	0.63	0.66	0.65	0.65	0.63	0.65	0.68	0.66	0.65	0.66
1973	0.69	0.70	0.74	0.69	0.71	0.70	0.74	0.61	0.60	0.58	0.60	0.66	0.62	0.78	0.65	0.63	0.62	0.65	0.67	0.68	0.70	0.63	0.66	0.67	0.64	0.62	0.67	0.71	0.61	0.62	0.58
1974	0.69	0.72	0.76	0.71	0.72	0.72	0.74	0.68	0.68	0.69	0.70	0.70	0.68	0.76	0.69	0.68	0.69	0.69	0.69	0.71	0.71	0.68	0.70	0.69	0.65	0.68	0.71	0.73	0.68	0.67	0.66
1975	0.69	0.70	0.68	0.67	0.69	0.69	0.70	0.62	0.61	0.62	0.61	0.69	0.62	0.78	0.64	0.62	0.63	0.65	0.64	0.66	0.67	0.62	0.66	0.67	0.63	0.63	0.66	0.70	0.60	0.64	0.61
1976	0.72	0.73	0.73	0.69	0.71	0.69	0.72	0.65	0.62	0.63	0.62	0.69	0.61	0.78	0.63	0.61	0.60	0.64	0.70	0.65	0.66	0.64	0.65	0.71	0.65	0.63	0.68	0.72	0.61	0.62	0.63
1977	0.67	0.71	0.72	0.67	0.69	0.67	0.74	0.60	0.62	0.57	0.58	0.60	0.56	0.74	0.59	0.60	0.61	0.62	0.63	0.66	0.67	0.61	0.63	0.65	0.62	0.63	0.67	0.70	0.58	0.61	0.59

	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
1978	0.87	0.84	0.81	0.83	0.82	0.81	0.77	0.75	0.75	0.71	0.72	0.72	0.70	0.76	0.70	0.76	0.68	0.72	0.72	0.70	0.68	0.67	0.80	0.69	0.69	0.69	0.69	0.72	0.67
1979	0.88	0.85	0.84	0.88	0.84	0.81	0.78	0.75	0.77	0.71	0.77	0.75	0.71	0.79	0.68	0.78	0.69	0.75	0.72	0.73	0.70	0.69	0.84	0.74	0.70	0.72	0.70	0.73	0.71
1980	0.89	0.85	0.84	0.86	0.82	0.80	0.79	0.74	0.77	0.73	0.77	0.76	0.73	0.78	0.67	0.75	0.69	0.74	0.73	0.76	0.68	0.70	0.82	0.74	0.74	0.76	0.68	0.73	0.72
1981	0.88	0.84	0.83	0.83	0.83	0.79	0.78	0.72	0.73	0.70	0.72	0.71	0.67	0.74	0.64	0.74	0.66	0.72	0.70	0.70	0.68	0.66	0.84	0.67	0.69	0.71	0.67	0.69	0.67
1982	0.88	0.86	0.84	0.84	0.84	0.82	0.78	0.74	0.74	0.72	0.72	0.74	0.71	0.75	0.66	0.77	0.69	0.71	0.72	0.72	0.67	0.67	0.84	0.70	0.71	0.72	0.69	0.71	0.69
1983	0.88	0.82	0.82	0.83	0.83	0.79	0.76	0.73	0.74	0.70	0.73	0.70	0.68	0.73	0.65	0.76	0.66	0.70	0.68	0.70	0.66	0.65	0.86	0.69	0.70	0.72	0.69	0.69	0.67
1984	0.90	0.87	0.85	0.83	0.85	0.80	0.78	0.78	0.80	0.76	0.79	0.78	0.74	0.81	0.74	0.79	0.70	0.75	0.70	0.74	0.73	0.71	0.80	0.73	0.74	0.74	0.70	0.72	0.74
1985	0.88	0.86	0.83	0.85	0.80	0.80	0.75	0.65	0.64	0.66	0.69	0.67	0.65	0.68	0.67	0.71	0.66	0.63	0.64	0.64	0.57	0.60	0.82	0.62	0.61	0.68	0.62	0.65	0.60
1986	0.88	0.88	0.82	0.85	0.81	0.84	0.76	0.73	0.65	0.69	0.72	0.71	0.63	0.73	0.71	0.70	0.71	0.67	0.68	0.66	0.60	0.63	0.82	0.64	0.60	0.68	0.61	0.62	0.62
1987	0.89	0.89	0.85	0.87	0.82	0.81	0.76	0.71	0.60	0.64	0.68	0.68	0.64	0.67	0.65	0.69	0.67	0.65	0.62	0.64	0.58	0.58	0.79	0.62	0.58	0.69	0.62	0.63	0.57
1988	0.88	0.86	0.85	0.84	0.82	0.80	0.73	0.71	0.64	0.67	0.69	0.69	0.64	0.68	0.66	0.72	0.66	0.62	0.64	0.65	0.60	0.59	0.81	0.64	0.60	0.70	0.61	0.62	0.58
1989	0.87	0.82	0.83	0.85	0.82	0.79	0.75	0.67	0.67	0.65	0.66	0.67	0.66	0.65	0.64	0.73	0.66	0.67	0.64	0.66	0.63	0.63	0.82	0.68	0.66	0.70	0.69	0.69	0.60
1990	0.88	0.86	0.83	0.85	0.85	0.82	0.76	0.68	0.64	0.68	0.67	0.66	0.64	0.67	0.65	0.72	0.69	0.65	0.62	0.63	0.61	0.60	0.82	0.66	0.62	0.68	0.62	0.61	0.56
1991	0.88	0.83	0.79	0.82	0.79	0.81	0.70	0.79	0.76	0.78	0.78	0.77	0.79	0.77	0.74	0.74	0.73	0.81	0.73	0.78	0.77	0.78	0.76	0.73	0.78	0.76	0.78	0.78	0.74
1992	0.88	0.86	0.85	0.87	0.84	0.81	0.79	0.71	0.69	0.70	0.73	0.74	0.67	0.76	0.68	0.71	0.65	0.67	0.68	0.67	0.59	0.61	0.82	0.65	0.65	0.69	0.64	0.63	0.59
1993	0.88	0.81	0.81	0.83	0.81	0.79	0.77	0.72	0.70	0.68	0.73	0.73	0.68	0.72	0.61	0.73	0.64	0.66	0.65	0.68	0.62	0.62	0.82	0.64	0.63	0.68	0.62	0.61	0.60
1994	0.90	0.87	0.85	0.85	0.85	0.83	0.78	0.72	0.68	0.71	0.73	0.73	0.66	0.74	0.64	0.72	0.67	0.69	0.67	0.69	0.61	0.60	0.80	0.65	0.62	0.69	0.63	0.60	0.61
1995	0.91	0.87	0.86	0.86	0.86	0.84	0.79	0.73	0.71	0.71	0.77	0.74	0.66	0.76	0.69	0.74	0.64	0.69	0.70	0.67	0.64	0.63	0.83	0.69	0.65	0.69	0.65	0.64	0.62
1996	0.86	0.84	0.83	0.87	0.84	0.82	0.78	0.69	0.71	0.68	0.74	0.72	0.67	0.72	0.62	0.72	0.63	0.68	0.69	0.65	0.62	0.62	0.83	0.63	0.67	0.69	0.64	0.70	0.63
1997	0.91	0.89	0.86	0.85	0.86	0.83	0.79	0.74	0.72	0.73	0.76	0.74	0.68	0.75	0.69	0.76	0.68	0.72	0.68	0.69	0.67	0.63	0.83	0.68	0.68	0.71	0.66	0.65	0.66
1998	0.91	0.86	0.85	0.86	0.86	0.84	0.78	0.75	0.72	0.70	0.77	0.76	0.67	0.76	0.66	0.77	0.68	0.70	0.69	0.69	0.66	0.65	0.84	0.68	0.70	0.71	0.67	0.66	0.67
1999	0.87	0.85	0.83	0.86	0.82	0.81	0.73	0.67	0.65	0.66	0.68	0.67	0.63	0.69	0.67	0.70	0.63	0.63	0.59	0.65	0.57	0.58	0.79	0.63	0.63	0.68	0.62	0.64	0.61
2000	0.86	0.85	0.80	0.85	0.81	0.80	0.74	0.72	0.69	0.70	0.73	0.72	0.65	0.73	0.63	0.71	0.64	0.64	0.61	0.68	0.61	0.60	0.77	0.66	0.66	0.70	0.66	0.65	0.63
2001	0.87	0.83	0.81	0.85	0.82	0.81	0.75	0.70	0.66	0.66	0.70	0.70	0.65	0.70	0.63	0.71	0.60	0.66	0.63	0.62	0.62	0.62	0.80	0.65	0.67	0.69	0.67	0.71	0.65
2002	0.88	0.85	0.82	0.85	0.85	0.84	0.75	0.67	0.66	0.65	0.70	0.66	0.61	0.69	0.63	0.71	0.63	0.67	0.58	0.63	0.59	0.60	0.82	0.65	0.64	0.65	0.63	0.65	0.62
2003	0.85	0.85	0.81	0.83	0.80	0.80	0.75	0.70	0.68	0.66	0.71	0.68	0.64	0.69	0.62	0.68	0.62	0.63	0.63	0.64	0.60	0.61	0.78	0.63	0.62	0.68	0.63	0.63	0.63
2004	0.87	0.84	0.81	0.85	0.81	0.82	0.74	0.73	0.69	0.68	0.73	0.72	0.65	0.72	0.62	0.72	0.61	0.66	0.64	0.65	0.63	0.64	0.82	0.65	0.67	0.71	0.66	0.68	0.67
2005	0.80	0.80	0.76	0.82	0.80	0.80	0.75	0.68	0.70	0.68	0.75	0.71	0.70	0.75	0.63	0.69	0.67	0.72	0.66	0.69	0.69	0.69	0.77	0.68	0.71	0.73	0.70	0.72	0.70
2006	0.88	0.87	0.87	0.87	0.85	0.83	0.77	0.71	0.66	0.65	0.70	0.69	0.62	0.71	0.64	0.72	0.66	0.64	0.63	0.64	0.60	0.56	0.82	0.66	0.61	0.68	0.60	0.61	0.58
2007	0.89	0.86	0.84	0.88	0.85	0.82	0.77	0.69	0.68	0.67	0.70	0.70	0.65	0.73	0.66	0.75	0.65	0.66	0.67	0.71	0.58	0.60	0.83	0.65	0.62	0.67	0.64	0.62	0.61
2008	0.89	0.87	0.85	0.87	0.82	0.81	0.74	0.72	0.68	0.67	0.71	0.72	0.63	0.73	0.69	0.71	0.68	0.65	0.63	0.65	0.59	0.61	0.82	0.66	0.58	0.66	0.61	0.63	0.59

	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1978	0.00	0.51	0.55	0.51	0.55	0.53	0.61	0.65	0.70	0.69	0.68	0.70	0.67	0.74	0.65	0.60	0.63	0.66	0.66	0.62	0.63	0.65	0.64	0.61	0.64	0.63	0.66	0.67	0.63	0.67	0.66
1979	0.51	0.00	0.52	0.50	0.56	0.51	0.61	0.69	0.72	0.71	0.73	0.71	0.70	0.76	0.65	0.59	0.63	0.63	0.65	0.62	0.62	0.67	0.65	0.63	0.65	0.65	0.66	0.70	0.64	0.66	0.67
1980	0.55	0.52	0.00	0.52	0.54	0.55	0.62	0.69	0.73	0.74	0.73	0.73	0.72	0.77	0.68	0.63	0.66	0.67	0.69	0.64	0.64	0.69	0.66	0.64	0.69	0.66	0.66	0.66	0.68	0.70	0.70
1981	0.51	0.50	0.52	0.00	0.46	0.43	0.58	0.64	0.70	0.69	0.68	0.66	0.65	0.77	0.61	0.56	0.59	0.61	0.62	0.58	0.61	0.66	0.64	0.60	0.63	0.65	0.63	0.68	0.62	0.66	0.68
1982	0.55	0.56	0.54	0.46	0.00	0.46	0.60	0.65	0.70	0.70	0.68	0.69	0.67	0.79	0.65	0.61	0.63	0.65	0.64	0.62	0.62	0.65	0.65	0.62	0.65	0.63	0.65	0.69	0.66	0.66	0.67
1983	0.53	0.51	0.55	0.43	0.46	0.00	0.58	0.64	0.70	0.69	0.67	0.66	0.65	0.77	0.60	0.54	0.58	0.59	0.59	0.55	0.56	0.63	0.60	0.57	0.62	0.61	0.62	0.67	0.62	0.65	0.66
1984	0.61	0.61	0.62	0.58	0.60	0.58	0.00	0.72	0.74	0.75	0.73	0.76	0.74	0.78	0.71	0.66	0.70	0.69	0.69	0.66	0.66	0.69	0.65	0.70	0.69	0.69	0.69	0.73	0.70	0.70	0.71
1985	0.65	0.69	0.69	0.64	0.65	0.64	0.72	0.00	0.56	0.53	0.54	0.56	0.53	0.70	0.59	0.60	0.60	0.63	0.60	0.65	0.64	0.58	0.59	0.55	0.57	0.58	0.61	0.64	0.57	0.60	0.60
1986	0.70	0.72	0.73	0.70	0.70	0.70	0.74	0.56	0.00	0.57	0.57	0.64	0.60	0.74	0.62	0.63	0.61	0.63	0.63	0.66	0.68	0.59	0.64	0.64	0.62	0.59	0.64	0.68	0.60	0.61	0.63
1987	0.69	0.71	0.74	0.69	0.70	0.69	0.75	0.53	0.57	0.00	0.49	0.57	0.54	0.74	0.59	0.62	0.59	0.61	0.60	0.64	0.65	0.55	0.61	0.59	0.57	0.60	0.64	0.68	0.55	0.60	0.57
1988	0.68	0.73	0.73	0.68	0.68	0.67	0.73	0.54	0.57	0.49	0.00	0.55	0.53	0.75	0.60	0.59	0.61	0.61	0.59	0.63	0.65	0.56	0.60	0.58	0.57	0.56	0.60	0.67	0.54	0.56	0.59
1989	0.70	0.71	0.73	0.66	0.69	0.66	0.76	0.56	0.64	0.57	0.55	0.00	0.52	0.76	0.64	0.61	0.64	0.65	0.63	0.67	0.67	0.59	0.60	0.59	0.59	0.60	0.58	0.67	0.62	0.64	0.62
1990	0.67	0.70	0.72	0.65	0.67	0.65	0.74	0.53	0.60	0.54	0.53	0.52	0.00	0.70	0.57	0.60	0.57	0.59	0.62	0.63	0.63	0.58	0.60	0.58	0.55	0.59	0.62	0.66	0.56	0.58	0.61
1991	0.74	0.76	0.77	0.77	0.79	0.77	0.78	0.70	0.74	0.74	0.75	0.76	0.70	0.00	0.75	0.76	0.75	0.76	0.76	0.74	0.75	0.76	0.72	0.72	0.75	0.76	0.76	0.71	0.79	0.77	0.77
1992	0.65	0.65	0.68	0.61	0.65	0.60	0.71	0.59	0.62	0.59	0.60	0.64	0.57	0.75	0.00	0.48	0.47	0.46	0.51	0.53	0.53	0.60	0.62	0.59	0.58	0.61	0.61	0.66	0.57	0.59	0.61
1993	0.60	0.59	0.63	0.56	0.61	0.54	0.66	0.60	0.63	0.62	0.59	0.61	0.60	0.76	0.48	0.00	0.45	0.45	0.49	0.50	0.49	0.58	0.56	0.58	0.58	0.55	0.58	0.62	0.56	0.60	0.59
1994	0.63	0.63	0.66	0.59	0.63	0.58	0.70	0.60	0.61	0.59	0.61	0.64	0.57	0.75	0.47	0.45	0.00	0.42	0.51	0.46	0.49	0.60	0.57	0.59	0.60	0.59	0.60	0.63	0.58	0.58	0.59
1995	0.66	0.63	0.67	0.61	0.65	0.59	0.69	0.63	0.63	0.61	0.61	0.65	0.59	0.76	0.46	0.45	0.42	0.00	0.51	0.44	0.46	0.62	0.59	0.59	0.59	0.60	0.61	0.66	0.56	0.60	0.60
1996	0.66	0.65	0.69	0.62	0.64	0.59	0.69	0.60	0.63	0.60	0.59	0.63	0.62	0.76	0.51	0.49	0.51	0.51	0.00	0.53	0.52	0.57	0.57	0.54	0.56	0.57	0.54	0.61	0.56	0.60	0.61
1997	0.62	0.62	0.64	0.58	0.62	0.55	0.66	0.65	0.66	0.64	0.63	0.67	0.63	0.74	0.53	0.50	0.46	0.44	0.53	0.00	0.39	0.62	0.58	0.58	0.59	0.60	0.62	0.65	0.58	0.62	0.62
1998	0.63	0.62	0.64	0.61	0.62	0.56	0.66	0.64	0.68	0.65	0.65	0.67	0.63	0.75	0.53	0.49	0.49	0.46	0.52	0.39	0.00	0.61	0.58	0.59	0.59	0.60	0.61	0.65	0.58	0.63	0.63
1999	0.65	0.67	0.69	0.66	0.65	0.63	0.69	0.58	0.59	0.55	0.56	0.59	0.58	0.76	0.60	0.58	0.60	0.62	0.57	0.62	0.61	0.00	0.50	0.50	0.47	0.48	0.52	0.59	0.52	0.56	0.56
2000	0.64	0.65	0.66	0.64	0.65	0.60	0.65	0.59	0.64	0.61	0.60	0.60	0.60	0.72	0.62	0.56	0.57	0.59	0.57	0.58	0.58	0.50	0.00	0.54	0.52	0.50	0.50	0.59	0.55	0.59	0.56
2001	0.61	0.63	0.64	0.60	0.62	0.57	0.70	0.55	0.64	0.59	0.58	0.59	0.58	0.72	0.59	0.58	0.59	0.59	0.54	0.58	0.59	0.50	0.54	0.00	0.44	0.51	0.50	0.57	0.53	0.59	0.56
2002	0.64	0.65	0.69	0.63	0.65	0.62	0.69	0.57	0.62	0.57	0.57	0.59	0.55	0.75	0.58	0.58	0.60	0.59	0.56	0.59	0.59	0.47	0.52	0.44	0.00	0.50	0.50	0.55	0.51	0.54	0.55
2003	0.63	0.65	0.66	0.65	0.63	0.61	0.69	0.58	0.59	0.60	0.56	0.60	0.59	0.76	0.61	0.55	0.59	0.60	0.57	0.60	0.60	0.48	0.50	0.51	0.50	0.00	0.49	0.54	0.55	0.55	0.54
2004	0.66	0.66	0.66	0.63	0.65	0.62	0.69	0.61	0.64	0.64	0.60	0.58	0.62	0.76	0.61	0.58	0.60	0.61	0.54	0.62	0.61	0.52	0.50	0.50	0.50	0.49	0.00	0.58	0.58	0.57	0.56
2005	0.67	0.70	0.66	0.68	0.69	0.67	0.73	0.64	0.68	0.68	0.67	0.67	0.66	0.71	0.66	0.62	0.63	0.66	0.61	0.65	0.65	0.59	0.59	0.57	0.55	0.54	0.58	0.00	0.63	0.65	0.63
2006	0.63	0.64	0.68	0.62	0.66	0.62	0.70	0.57	0.60	0.55	0.54	0.62	0.56	0.79	0.57	0.56	0.58	0.56	0.56	0.58	0.58	0.52	0.55	0.53	0.51	0.55	0.58	0.63	0.00	0.48	0.49
2007	0.67	0.66	0.70	0.66	0.66	0.65	0.70	0.60	0.61	0.60	0.56	0.64	0.58	0.77	0.59	0.60	0.58	0.60	0.60	0.62	0.63	0.56	0.59	0.59	0.54	0.55	0.57	0.65	0.48	0.00	0.51
2008	0.66	0.67	0.70	0.68	0.67	0.66	0.71	0.60	0.63	0.57	0.59	0.62	0.61	0.77	0.61	0.59	0.59	0.60	0.61	0.62	0.63	0.56	0.56	0.56	0.55	0.54	0.56	0.63	0.49	0.51	0.00

Our last question concerning similarity is its evolution in time. It is quite natural that any two texts differ but if they are long enough, to what extent they may differ? Independently of the similarity or distance indicators some similarity must be given. But if the associative vocabulary alone is compared, the similarity may become zero.

Here we are not interested in the authorship – which is known – but in the dynamics in the set of texts which can be considered quasi homogeneous concerning the theme. This is the historical variant of the school task for 60 children to write a text about "winter" and test whether the similarity of the associative vocabulary of pupils is a function of their sitting distance. For historical purposes we compute the means of cells on individual diagonals and obtain the results presented in Table 2.26. The result is unequivocal: though for individual text pairs we can find unexpectedly great or small distances, the mean trend is almost linear, as can be seen in Figure 2.20. A slightly quicker increase at the beginning, afterwards a rather weak convex exponential trend which gets its first breaks at the 56th step. The further course of the curve cannot be safely predicted; we can only conjecture that somewhere it begins to converge and will never attain 1 – even if it is linguistically possible.

Though the course of the data can be captured by hundreds of polynomials – which are rather shunned in linguistics – we try to substantiate this course by linguistic argumentation. Writing a speech of this kind consists of two factors: the necessity of imitation, because the theme is equal and the writers live in the same culture; the striving for originality giving political weight to the writer. The imitation gets smaller with time (= increasing steps) and its change can be symbolized as b/t; on the contrary, the originality inreases with time and can by symbolized as $c/(k\text{-}t)$, k being the upper (empirical) boundary of t, and b and c are some parameters. If the change of similarity/distance (y') consists of these two complementary factors, we obtain the differential equation

$$y' = \frac{b}{t} + \frac{c}{(k-t)} \tag{2.12}$$

with the solution

$$y = at^b (k-t)^c \tag{2.13}$$

where a is a form of the integration constant. This is a usual beta-type function. Fitting it to our data we obtain

$$y = 0.8046t^{0.0324}(60-t)^{-0.0657}$$

yielding $R^2 = 0.9693$ and the results presented in Table 2.26.

Table 2.26
Mean distances of associative vocabularies in year steps

Step t	Dist	$Dist_t$	Step t	Dist	$Dist_t$	Step t	Dist	$Dist_t$
1	0.5888	0.6155	21	0.7010	0.6981	41	0.7358	0.7479
2	0.6134	0.6302	22	0.7071	0.7003	42	0.7411	0.7511
3	0.6333	0.6393	23	0.7100	0.7025	43	0.7476	0.7545
4	0.6454	0.6460	24	0.7108	0.7048	44	0.7581	0.7581
5	0.6587	0.6515	25	0.7166	0.7070	45	0.7573	0.7619
6	0.6735	0.6561	26	0.7153	0.7093	46	0.7650	0.7659
7	0.6781	0.6602	27	0.7109	0.7115	47	0.7635	0.7702
8	0.6838	0.6639	28	0.7163	0.7138	48	0.7793	0.7748
9	0.6804	0.6673	29	0.7077	0.7161	49	0.7735	0.7797
10	0.6824	0.6704	30	0.7070	0.7185	50	0.7818	0.7851
11	0.6824	0.6734	31	0.7100	0.7208	51	0.7891	0.7911
12	0.6804	0.6762	32	0.7200	0.7232	52	0.8133	0.7977
13	0.6809	0.6789	33	0.7115	0.7257	53	0.8244	0.8053
14	0.6848	0.6815	34	0.7200	0.7282	54	0.8310	0.8140
15	0.6851	0.6840	35	0.7252	0.7308	55	0.8462	0.8243
16	0.6884	0.6865	36	0.7296	0.7334	56	0.8458	0.8369
17	0.6926	0.6889	37	0.7204	0.7361	57	0.8640	0.8534
18	0.6888	0.6912	38	0.7236	0.7389	58	0.8830	0.8769
19	0.6907	0.6935	39	0.7286	0.7418	59	0.8930	0.9183
20	0.6943	0.6960	40	0.7320	0.7448			

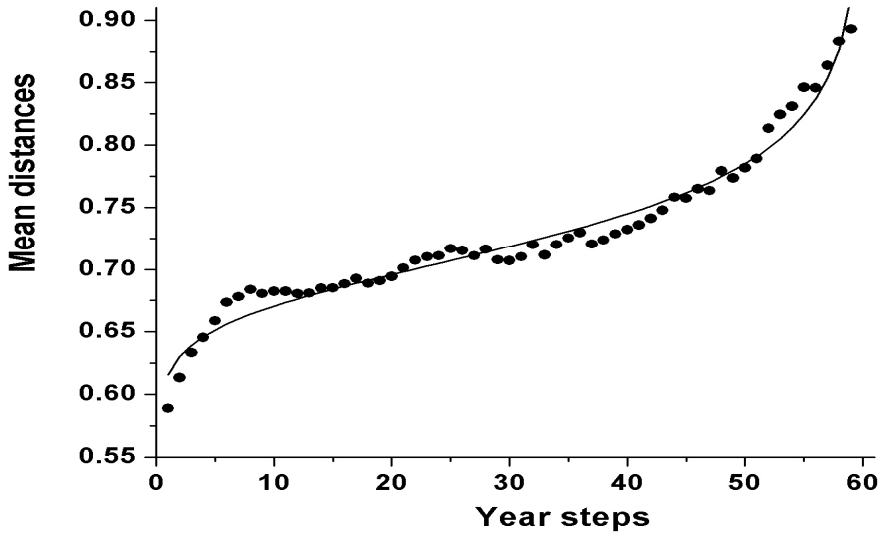

Figure 2.20. Mean distances between speeches in one-year intervals

Or we consider rather the relative rate of change of *y* and write

$$\frac{y'}{y} = c\left(\frac{1}{t} + \frac{1}{k-t}\right). \tag{2.14}$$

Writing the integration constant in form *c ln a* we obtain the solution

$$y = c\ln\left(\frac{at}{k-t}\right). \tag{2.15}$$

Function (2.15) yields

$$y = 0.03628\ln\left(\frac{510153823t}{60-t}\right), \text{ with } R^2 = 0.94$$

The latter result has the advantage of having only two parameters to be estimated but the parameter *a* is not very reasonable. The difference in goodness-of-fit between the two functions is irrelevant. In any case the Zipfian modelling with contrary factors (forces) can be used also in this case. It would be interesting to observe years later whether only the upper boundary (here $k = 60$) changes or other parameters, too.

If we consider now the distances between followers, we first prepare unions of vocabularies of individual presidents and then compare the frequencies in the unions. In this way we obtain the results presented in Table 2.27.

Table 2.27
Mean distances between the individual presidents

DISTANCE	Einaudi	Gronchi	Segni	Saragat	Leone	Pertini	Cossiga	Scalfaro	Ciampi	Napo-litano
Einaudi	0.0000	0.5312	0.5949	0.5676	0.5984	0.6068	0.5748	0.6308	0.5710	0.6134
Gronchi	0.5312	0.0000	0.4686	0.3899	0.4204	0.5601	0.3990	0.5190	0.4435	0.4682
Segni	0.5949	0.4686	0.0000	0.4639	0.5392	0.5925	0.5080	0.5842	0.5215	0.5718
Saragat	0.5676	0.3899	0.4639	0.0000	0.3702	0.5114	0.3771	0.4513	0.3831	0.4162
Leone	0.5984	0.4204	0.5392	0.3702	0.0000	0.5170	0.3779	0.4351	0.4281	0.4092
Pertini	0.6068	0.5601	0.5925	0.5114	0.5170	0.0000	0.5215	0.4348	0.4650	0.5158
Cossiga	0.5748	0.3990	0.5080	0.3771	0.3779	0.5215	0.0000	0.4353	0.3643	0.3969
Scalfaro	0.6308	0.5190	0.5842	0.4513	0.4351	0.4348	0.4353	0.0000	0.4044	0.4363
Ciampi	0.5710	0.4435	0.5215	0.3831	0.4281	0.4650	0.3643	0.4044	0.0000	0.3642
Napolitano	0.6134	0.4682	0.5718	0.4162	0.4092	0.5158	0.3969	0.4363	0.3642	0.0000

Again, we compute the means of means placed on the diagonals, in order to obtain the distances between the direct and indirect followers. For example, the

mean distance from the immediate follower is the sum of cells located directly above the mean diagonal, i.e. 0.5123 + 0.4686 + … + 0.3642 yielding 4.076 and divided by 9, finally 0.453. The individual values are presented in Table 2.28 where x means the x^{th} neighbour.

Table 2.28
Mean distances of associative vocabularies in x^{th} neighbour

Follower x	**Mean Distance D**	**Computed Distance D_t**
1	0.453	0.447
2	0.456	0.464
3	0.465	0.476
4	0.510	0.486
5	0.476	0.497
6	0.508	0.507
7	0.549	0.521
8	0.520	0.541
9	0.613	0.611

The same two functions as above yield an excellent fit, namely

$$y = 0.4953x^{0.0443}(9.1 - x)^{-0.0488} \text{ with } R^2 = 0.89$$

whose values are presented in the third column of Table 2.28, and

$$y = 0.0248 \ln\left(\frac{442728751x}{9.1 - x}\right) \text{ with } R^2 = 0.88$$

whose parameter *b* is somewhat unrealistic. The result is graphically presented in Figure 2.21.

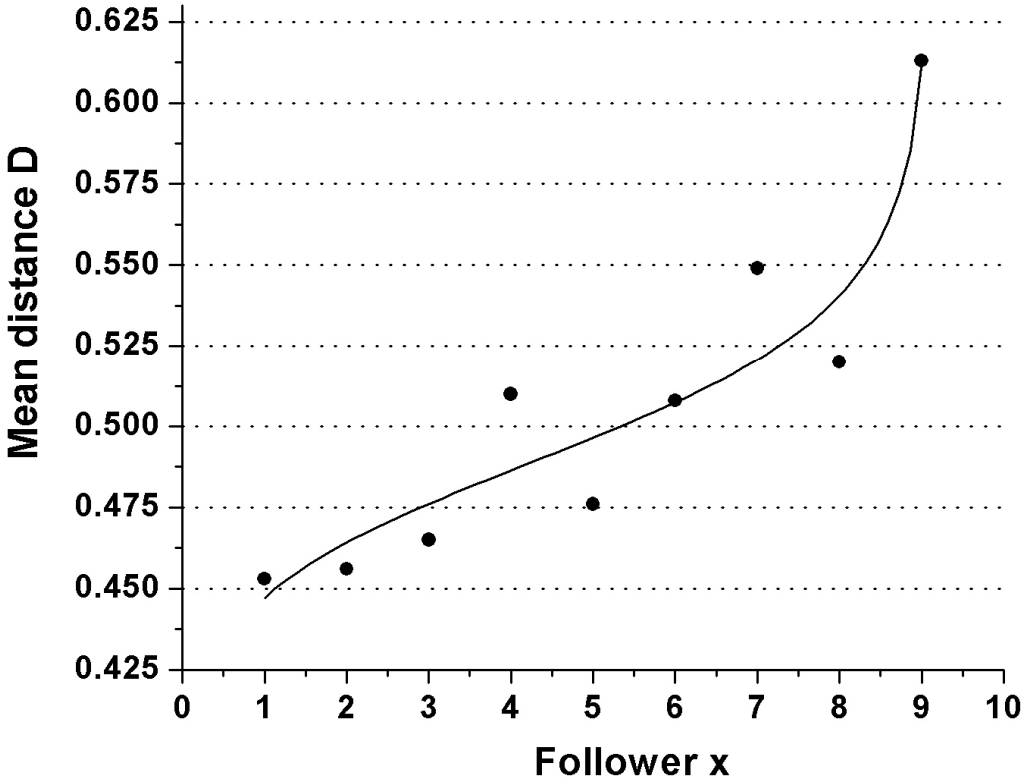

Figure 2.21. The distances between x^{th} followers

Wheter we take the step-by-step distance or that of president-by-president, the result is similar: the associative worlds of presidents change, the distance increases. Even if the theme of the speeches is identical, i.e. thematically we have a very homogeneous basis, the view of the reality changes. This is, of course, caused not only by the personality of the president but above all in the change of the reality which was very fast in the last century.

2.10. Conclusion

The approach of comparing associative vocabularies is rather a view "from above": leaving language accidentals and grammatical details out and looking at the core of the text, taking of it only parts which are more strongly wired in the heads of the text creators. Even in spite of the fact that ghost writers could have helped to draft the text, the president was forced to consider it his own, expressing exactly his view of things. Even the exchange of some words or formulations can change the associative structure without changing the surface form of the

text. There are approaches in forensic linguistics taking into account just the auxiliaries whose distribution may be characteristic for an author, but such an approach strongly depends on language and can be used for a very restricted aim.

The association structure can be presented in form of a graph whose properties can be studied and evaluated but here we kept away from depicting them because complex graphs are clumsy and their processing starting from a graphical picture is for the reader very ineffective. Different subsets of words from a text can be used for special purposes; here we only presented the associated words.

The methods presented here are only a choice of dozens of possibilities. As a matter of fact, all similarity indicators should be applied and those should be *selected* which yield results congruent with our intuition. Comparing aspects of texts we do not attain the whole “truth”; we again obtain only aspects of truth, some parts of the tessellation of the textual reality. A continuation of this study would be necessary.

3. The golden section

The golden section – also called golden ratio, golden mean, Divine proportion etc. – is a geometric peculiarity occurring both in nature, in human artistic products and in many domains of human culture, even in the human body and mind. The literature is overwhelming[1] (see Pacioli 1509; Huntley 1970; Dunlap 1997; Walser 2001; Livio 2002a, 2002b, 2002c; Marquardt 2002; Weiss and Weiss 2003; Hemenway 2005; Doczi 2005; Sahlqvist 2008), for its appearance in texts, see Popescu and Altmann (2007, 2009). In domains where it is a simple geometric matter, its finding is not very tiresome, but in texts it appears only after a series of computations as shown below. It is no fixed number but rather an asymptote to which a special property of texts as a whole converges. The existence of the golden number $\varphi = (1+\sqrt{5})/2 \cong 1.6180$ cannot be substantiated by linguistic arguments, at least not preliminarily.

Consider a text in *statu nascendi* whose words in every moment of writing constitute a rank-frequency distribution. The more the writer approaches the end of the text, the more regular is the distribution. Of course, one can consider it also a simple series or a continuous function; the approach itself is not decisive because all these sequences can be translated into one another. However, the sequence has some basic properties, viz. it is monotonously decreasing (by definition) and convex – properties common to rank-frequency sequences of any linguistic units – and has four cardinal points, namely

1. the origin $O(1,1)$ (not $O(0,0)$ because ranking begins with 1 and there is no smaller frequency than 1)
2. the top point $P_1(1,f_1)$, where f_1 is the greatest frequency
3. the end point $P_2(V,1)$, where V is the vocabulary size and
4. the h-point $H(h,h)$ representing that point at which the line joining two neighbouring frequencies intersects the $y = x$ line. It can be computed according to the formula (cf. Popescu et al. 2009: 18)

$$h = \begin{cases} r & \text{if there is an } r = f(r) \\ \dfrac{f(r_i)r_j - f(r_j)r_i}{r_j - r_i + f_i - f_j} & \text{if there is no } r = f(r) \end{cases} \tag{3.1}$$

where r is the rank and $f(r)$ is the frequency of the word at rank r. If there is no such $r = f(r)$, one takes two neighbouring ranks for which $r_i < f(r_i)$ and $r_j > f(r_j)$. In case that $r_{max} < f(r_{max})$ one transforms the frequency sequence in $f^*(r) = f(r) - f(r_{max}) + 1$. The h-point is a fuzzy boundary separating autosemantics from

[1] cf. *en.wikipedia.org/wiki/Golden_ratio* or *http://www.goldenmuseum.com/1801Refer_engl.html*

synsemantics. In the process of writing one can imagine the writer sitting at this point and controlling the frequency increase of synsemantics to the left and the number of autosemantics to the right and "caring" for the balance in the text. Metaphorically, this view is called writer´s view and is defined as the angle (alpha) subtended between the vectors

HP_1 of Cartesian components $a_x = -(h - 1)$, $a_y = f_1 - h$
HP_2 of Cartesian components $b_x = V - h$, $b_y = -(h - 1)$.

Inserting these components in the general expression

$$\cos\ \alpha\ =\ \frac{a_x b_x\ +\ a_y b_y}{[(a_x^2\ +\ a_y^2)^{1/2}][(b_x^2\ +\ b_y^2)^{1/2}]} \tag{3.2}$$

we obtain the exact formula

$$\cos\alpha\ =\frac{-[(h-1)(f_1-h)+(h-1)(V-h)]}{[(h-1)^2+(f_1-h)^2]^{1/2}[(h-1)^2+(V-h)^2]^{1/2}}\,. \tag{3.3}$$

Now, for $H(1,1)$, i.e. if $h = 1$, one would obtain $\cos\alpha = 0$ and $\alpha = \pi/2$ radians = 1.570796… radians but, peculiar enough, texts never strive to this value. Instead, with increasing vocabulary they tend to 1.618…, i.e. to the golden section. This property is a characteristic of a spontaneously written text, but something lying behind the control of the writer, though we called it metaphorically "writer´s view". Perhaps, it is a kind of text law but still much research is necessary to corroborate its law-like character. In any case, laws cannot be learned by writers; they simply abide by them in the same way as one falls from the window according to the law of gravitation not even knowing that it exists. However, as soon as one begins to consciously adapt the text translating it from one language into another, the background is distorted and the result is not any more a text written spontaneously but something created under a regime foreign to the translator. Let us, for the sake of illustration show the translation of 10 chapters of a Russian novel translated in 11 Slavic languages ("Kak zakaljalas stal'" by N.A. Ostrovskij, 1932-1934), cf. Figure 3.1, taken from E. Kelih´s (2009, 2009a) corpus.

As can be seen, the trend exists and the convergence to the golden section is given but since we have to do with a mixture of translations in 11 languages, and the chapters were considered separately, the dispersion is too great to propose an adequate function capturing the trend. On the other hand, taking 100 texts from 20 languages Popescu and Altmann (2009) obtained a relatively smooth course of the α radians (cf. Figure 3.2), and for 253 texts of 26 German writers the trend is still clearer (cf. Figure 3.3).

Even if the original text is translated but long enough, the convergence is clearly visible. The necessary values of translations of very long texts are pre-

sented in Table 3.1 and shown in Figure 3.4. In these texts some of the values are smaller than 1.618 but the convergence is evident. Some of the texts are mixed and other ones are created under different regimes, i.e. containing several chapters which were not written in one go but with pauses in which the disposition of the writer could change.

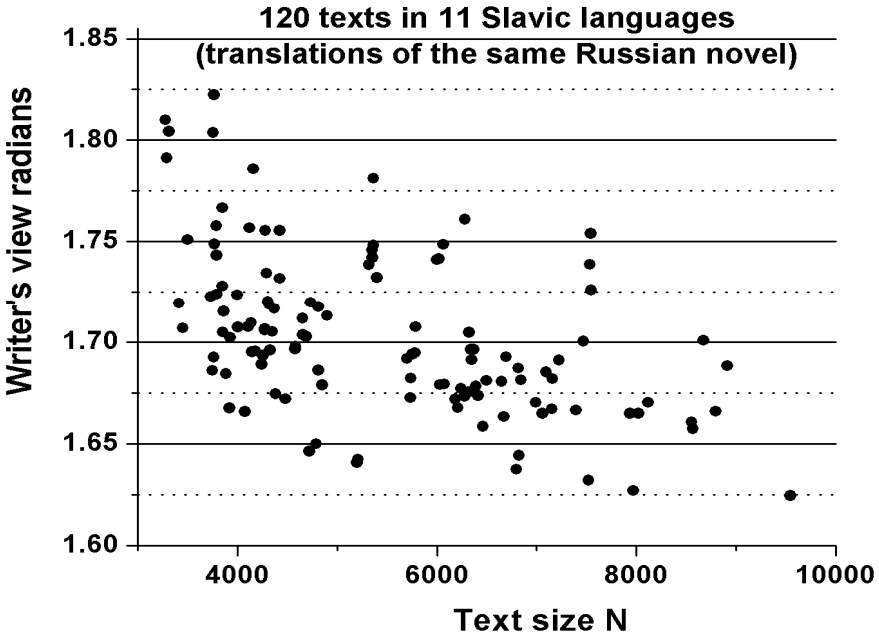

Figure 3.1. α radians of 10 chapters of a Russian novel translated in 11 Slavic languages (data from Emmerich Kelih's project on Slavic languages 2009)

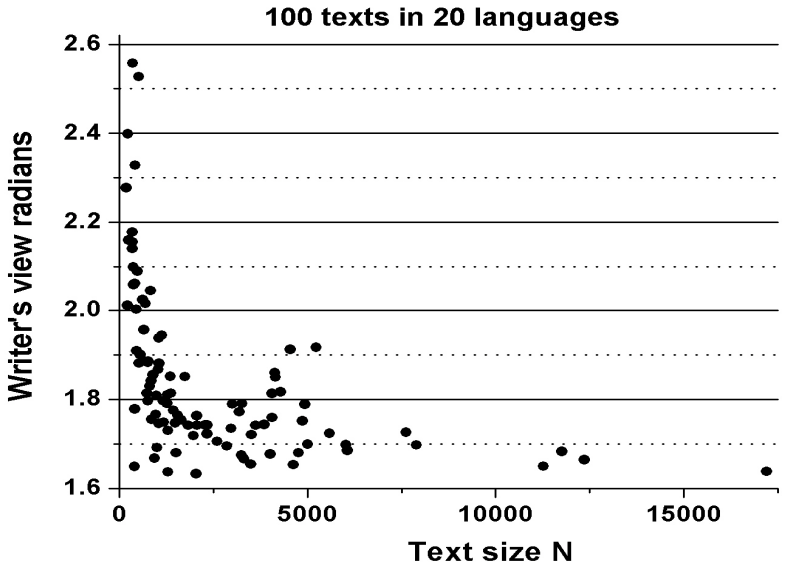

Fig. 3.2. α radians in 100 texts in 20 languages
(data from Popescu, Altmann 2009)

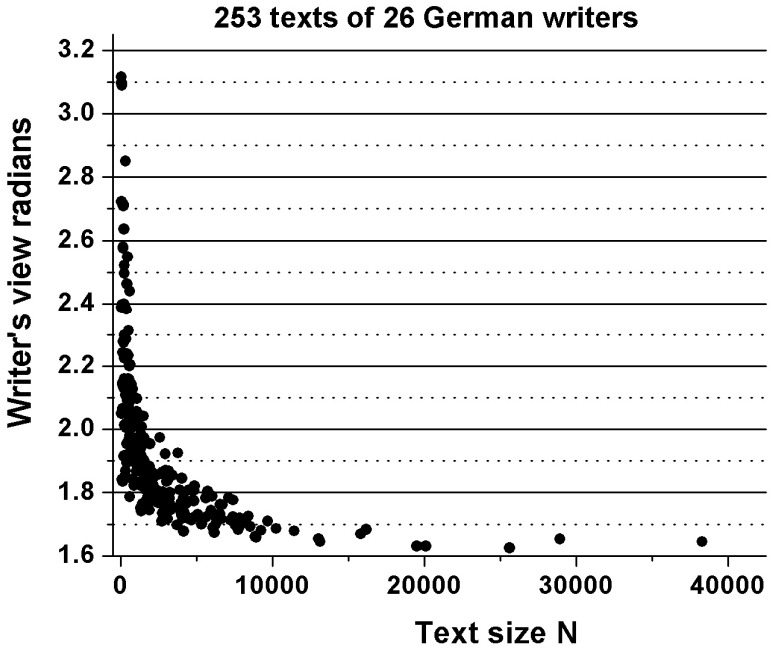

Figure 3.3. α radians in 253 German texts (data from Popescu, Altmann 2009)

Table 3.1
Long texts (Popescu, Altmann 2007)

Author		**Text**	**N**	**V**	**f(1)**	**h**	**cos α**	**α rad**
Goethe	G	Faust 1.	30625	6303	918	64	-0,08496	1,6559
Goethe	G	Faust 1. and 2.	75050	13341	2089	90	-0,05176	1,6226
Milton	E	Paradise Lost	79879	10211	3330	98	-0,03999	1,6108
The Evangelists	E tr.	The Gospels	83932	3501	5669	112	-0,05316	1,6240
Conan Doyle	E	Sherlock Holmes	104230	8324	5601	112	-0,03403	1,6048
Homer	E tr.	The Odyssey	117386	6800	5875	137	-0,04442	1,6152
Homer	E tr.	The Iliad	152455	7776	9945	150	-0,03497	1,6058
Moses	E tr.	The Pentateuch	156872	4797	13667	150	-0,04335	1,6142
The Bible	E tr.	New Testament	180573	6005	10976	160	-0,04215	1,6130
Dickens	E	Great Expectations	185104	11376	8139	161	-0,03453	1,6053
Dostoevsky	E tr.	Crime and Punishment	202853	10728	7768	174	-0,03938	1,6102
Joyce	E	Ulysses	263324	29457	14905	169	-0,01724	1,5880
Dickens	E	David Copperfield	360779	17225	13918	210	-0,02766	1,5985
Tolstoy	E tr.	War and Peace	561723	20094	34391	255	-0,02032	1,5911
The Bible	E tr.	Old Testament	610051	10751	52934	270	-0,03088	1,6017
The Bible	Etr.	Old & New Testament	790624	12698	63910	294	-0,02832	1,5991

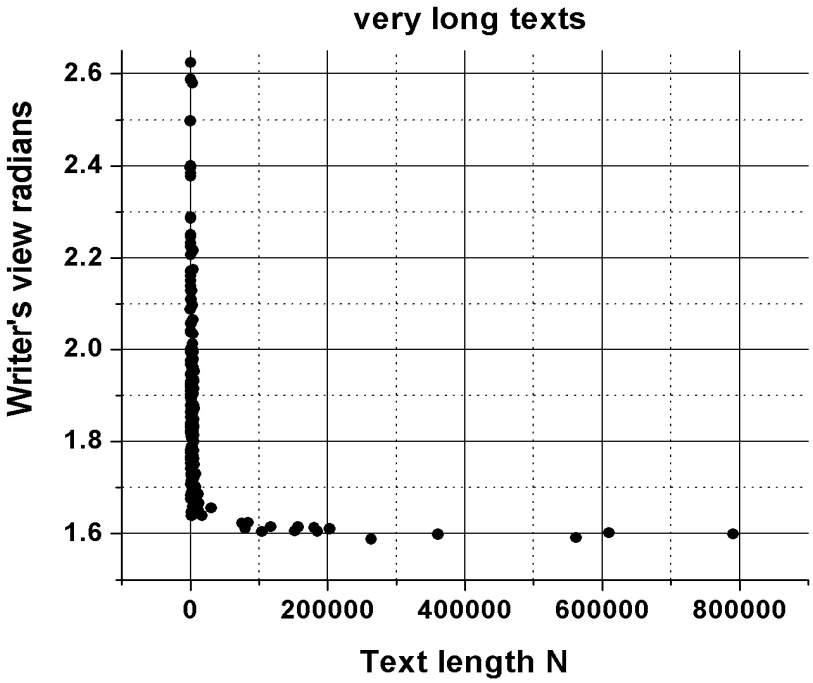

Figure 3.4. α radians for long texts (data from Popescu, Altmann 2007)

Our question is, of course, the form of this regularity in Italian texts. They are of very different length, mostly too short to display convergence. All quantities necessary for computing writer´s view are presented in Table 3.2

Table 3.2
Writer's view of 60 End-of-Year speeches of Italian Presidents (1949 - 2008)

Text	***N***	***V***	***f*(1)**	***h***	**cos *α***	***α* rad**
1949Einaudi	194	140	10	5.00	-0.6475	2.2752
1950Einaudi	150	105	9	4.00	-0.5397	2.1409
1951Einaudi	230	169	9	5.00	-0.7241	2.3806
1952Einaudi	179	145	7	4.00	-0.7220	2.3775
1953Einaudi	190	143	8	4.00	-0.6171	2.2359
1954Einaudi	260	181	12	5.00	-0.5157	2.1127
1955Gronchi	388	248	16	6.66	-0.5382	2.1391
1956Gronchi	665	374	29	8.00	-0.3343	1.9117
1957Gronchi	1130	549	65	12.00	-0.2232	1.7959
1958Gronchi	886	460	41	11.00	-0.3373	1.9148
1959Gronchi	697	388	33	9.00	-0.3362	1.9137
1960Gronchi	804	434	41	10.00	-0.2991	1.8746
1961Gronchi	1252	622	67	13.00	-0.2361	1.8092

1962Segni	738	381	35	10.00	-0.3614	1.9406
1963Segni	1057	527	46	11.66	-0.3162	1.8925
1964Saragat	465	278	21	8.00	-0.4968	2.0907
1965Saragat	1052	510	52	11.66	-0.2761	1.8505
1966Saragat	1200	597	44	12.50	-0.3614	1.9405
1967Saragat	1056	526	51	11.00	-0.2613	1.8352
1968Saragat	1173	562	56	13.00	-0.2898	1.8648
1969Saragat	1583	692	86	15.00	-0.2137	1.7862
1970Saragat	1929	812	85	16.50	-0.2397	1.8128
1971Leone	262	168	12	5.00	-0.5173	2.1145
1972Leone	767	394	32	9.50	-0.3740	1.9541
1973Leone	1250	616	67	12.00	-0.2139	1.7864
1974Leone	801	426	32	9.00	-0.3466	1.9247
1975Leone	1328	632	63	13.00	-0.2522	1.8257
1976Leone	1366	649	52	13.00	-0.3121	1.8882
1977Leone	1604	717	80	14.00	-0.2114	1.7838
1978Pertini	1492	603	53	14.33	-0.3472	1.9254
1979Pertini	2311	800	70	18.00	-0.3313	1.9085
1980Pertini	1360	535	50	13.75	-0.3548	1.9335
1981Pertini	2819	911	96	20.00	-0.2632	1.8371
1982Pertini	2486	854	90	19.00	-0.2666	1.8406
1983Pertini	3746	1149	118	23.66	-0.2531	1.8267
1984Pertini	1340	514	42	13.66	-0.4308	2.0162
1985Cossiga	2359	859	118	17.00	-0.1752	1.7469
1986Cossiga	1348	561	65	14.00	-0.2700	1.8441
1987Cossiga	2092	904	109	15.00	-0.1629	1.7344
1988Cossiga	2384	875	123	19.00	-0.1912	1.7632
1989Cossiga	1912	778	85	17.00	-0.2495	1.8229
1990Cossiga	3345	1222	155	20.00	-0.1550	1.7264
1991Cossiga	418	241	22	7.00	-0.3951	1.9769
1992Scalfaro	2774	978	118	17.50	-0.1789	1.7507
1993Scalfaro	2942	1074	129	18.60	-0.1739	1.7456
1994Scalfaro	3606	1190	171	21.00	-0.1491	1.7205
1995Scalfaro	4233	1341	180	22.66	-0.1526	1.7240
1996Scalfaro	2085	866	88	16.00	-0.2212	1.7938
1997Scalfaro	5012	1405	167	27.50	-0.2055	1.7778
1998Scalfaro	3995	1175	137	23.50	-0.2136	1.7860
1999Ciampi	1941	831	66	16.50	-0.3169	1.8933
2000Ciampi	1844	822	70	16.00	-0.2855	1.8604

2001Ciampi	2098	898	89	18.00	-0.2516	1.8251
2002Ciampi	2129	909	96	17.00	-0.2160	1.7886
2003Ciampi	1565	718	63	14.00	-0.2742	1.8486
2004Ciampi	1807	812	76	15.00	-0.2408	1.8140
2005Ciampi	1193	538	54	12.66	-0.2927	1.8679
2006Napolitano	2204	929	125	16.50	-0.1582	1.7297
2007Napolitano	1792	793	101	16.00	-0.1928	1.7648
2008Napolitano	1713	775	75	15.00	-0.2451	1.8184

The trend is graphically presented in Figure 3.5. In order to capture it formally, we order the texts according to increasing *N* and suppose that in the ideal case the decrease of values can be captured by a power function used frequently in text analysis (Zipf´s or zeta function).

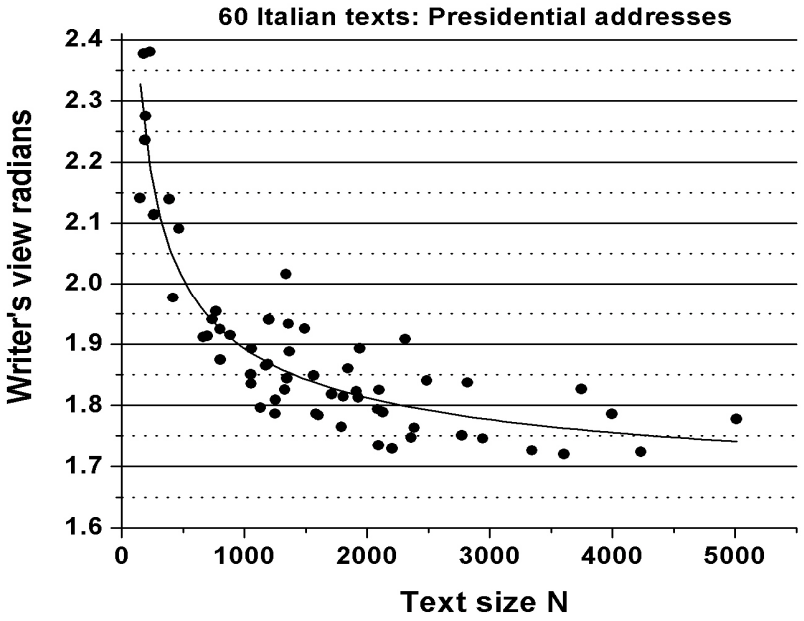

Figure 3.5. α radians in 60 Italian texts: Presidential addresses (data from Tuzzi, Popescu, Altmann 2009c)

However, since the curve should not drop under the golden number, we add it as a constant to the function and obtain

(3.4) $y = 1.618 + bx^c$

where y is the α radian and x is text length N. Fitting this function to our data we obtain

$$y = 1.618 + 8.7094x^{-0.5002}$$

(with $R = 0.833$) which is practically

$$y = 1.618 + 8.7094/\sqrt{x}.$$

The individual values are presented in Table 3.3 and graphically in Figure 3.5.

Table 3.3
The course of writer´s view in Presidential addresses

N	α rad	computed	N	α rad	computed
150	21.409	2.328	1366	18.882	1.853
179	23.775	2.268	1492	19.254	1.843
190	22.359	2.249	1565	18.486	1.838
194	22.752	2.243	1583	17.862	1.837
230	23.806	2.192	1604	17.838	1.835
260	21.127	2.158	1713	18.184	1.828
262	21.145	2.156	1792	17.648	1.823
388	21.391	2.060	1807	18.140	1.823
418	19.769	2.044	1844	18.604	1.821
465	20.907	2.021	1912	18.229	1.817
665	19.117	1.955	1929	18.128	1.816
697	19.137	1.948	1941	18.933	1.815
738	19.406	1.938	2085	17.938	1.808
767	19.541	1.932	2092	17.344	1.808
801	19.247	1.925	2098	18.251	1.808
804	18.746	1.925	2129	17.886	1.807
886	19.148	1.910	2204	17.297	1.803
1052	18.505	1.886	2311	19.085	1.799
1056	18.352	1.886	2359	17.469	1.797
1057	18.925	1.886	2384	17.632	1.796
1130	17.959	1.877	2486	18.406	1.792
1173	18.648	1.872	2774	17.507	1.783
1193	18.679	1.870	2819	18.371	1.782
1200	19.405	1.869	2942	17.456	1.778
1250	17.864	1.864	3345	17.264	1.768
1252	18.092	1.864	3606	17.205	1.763
1328	18.257	1.857	3746	18.267	1.760
1340	20.162	1.856	3995	17.860	1.756
1348	18.441	1.855	4233	17.240	1.752
1360	19.335	1.854	5012	17.778	1.741

Though the given evidence corroborates – but not proves – the existence of the golden section, let us approach the problem from another point of view. It is known that there is a relation between text size N and h. Hirsch (2005) proposed $N = ah^2$ but since the origin of a rank frequency sequence is $O(1,1)$, Popescu and Altmann (2009) proposed $N = b(h\text{-}1)^2$ and using the arc length established an indicator which is characteristic for many language phenomena displaying a rank-frequency sequence, namely

$$p = \frac{L_{\max} - L}{h - 1}, \tag{3.5}$$

where L is the arc length of the rank-frequency sequence computed in form of Euclidean distances between the individual frequencies

$$L = \sum_{r=1}^{V-1} \{[f(r) - f(r+1)]^2 + 1\}^{1/2}, \tag{3.6}$$

where V is the vocabulary (= maximum rank), r is the rank, ,$f(r)$ are the frequencies at rank r and L_{max} is the maximal possible length given as

$$L_{max} = R - 1 + f(1) - 1. \tag{3.7}$$

If the frequency at the highest rank, i.e. $f(V)$ is not 1 – a case that frequently occurs with other linguistic phenomena whose inventory is small – it is recommended to transform all frequencies in $f^*(r) = f(r) - f(V) + 1$, i.e. to "pull them down". Because of the relation between h and N we can set up another indicator which is a counterpart of (3.5), namely

$$q = \frac{L_{\max} - L}{\sqrt{N}}. \tag{3.8}$$

Combining them we obtain

$$p + q = (L_{\max} - L)\left(\frac{1}{\sqrt{N}} + \frac{1}{h-1}\right). \tag{3.9}$$

Computing (3.9) for texts we can state that it converges to the golden section. In order to illustrate this peculiar phenomenon, we take the 60 texts and compute p and q for each of them as can be seen in Table 3.4.

Table 3.4
The (*p*,*q*) indicator pair of 60 End-of-Year speeches of Italian Presidents
(Tuzzi, Popescu, Altmann 2009c)

Text	***N***	***V***	***f*(1)**	***h***	***L***	L_{max}	***p***	***q***	***p+q***
1949Einaudi	194	140	10	5.00	143.54	148	1.1142	0.3200	1.4342
1950Einaudi	150	105	9	4.00	108.78	112	1.0733	0.2629	1.3362
1951Einaudi	230	169	9	5.00	172.23	176	0.9417	0.2484	1.1901
1952Einaudi	179	145	7	4.00	146.89	150	1.0357	0.2322	1.2679
1953Einaudi	190	143	8	4.00	145.82	149	1.0603	0.2308	1.2911
1954Einaudi	260	181	12	5.00	186.29	191	1.1772	0.2920	1.4692
1955Gronchi	388	248	16	6.66	255.36	262	1.1739	0.3373	1.5112
1956Gronchi	665	374	29	8.00	392.75	401	1.1782	0.3198	1.4980
1957Gronchi	1130	549	65	12.00	599.38	612	1.1476	0.3755	1.5231
1958Gronchi	886	460	41	11.00	488.04	499	1.0956	0.3681	1.4637
1959Gronchi	697	388	33	9.00	409.82	419	1.1475	0.3477	1.4952
1960Gronchi	804	434	41	10.00	462.23	473	1.1969	0.3799	1.5768
1961Gronchi	1252	622	67	13.00	674.05	687	1.0789	0.3659	1.4448
1962Segni	738	381	35	10.00	404.01	414	1.1101	0.3678	1.4779
1963Segni	1057	527	46	11.66	559.52	571	1.0771	0.3532	1.4303
1964Saragat	465	278	21	8.00	289.03	297	1.1383	0.3695	1.5078
1965Saragat	1052	510	52	11.66	547.78	560	1.1466	0.3768	1.5234
1966Saragat	1200	597	44	12.50	624.77	639	1.2376	0.4109	1.6485
1967Saragat	1056	526	51	11.00	562.98	575	1.2019	0.3699	1.5718
1968Saragat	1173	562	56	13.00	602.83	616	1.0978	0.3847	1.4825
1969Saragat	1583	692	86	15.00	759.82	776	1.1556	0.4066	1.5622
1970Saragat	1929	812	85	16.50	877.58	895	1.1242	0.3967	1.5209
1971Leone	262	168	12	5.00	173.02	178	1.2443	0.3075	1.5518
1972Leone	767	394	32	9.50	414.71	424	1.0932	0.3355	1.4287
1973Leone	1250	616	67	12.00	669.22	681	1.0710	0.3332	1.4042
1974Leone	801	426	32	9.00	445.78	456	1.2770	0.3610	1.6380
1975Leone	1328	632	63	13.00	678.97	693	1.1688	0.3849	1.5537
1976Leone	1366	649	52	13.00	685.16	699	1.1535	0.3745	1.5280
1977Leone	1604	717	80	14.00	780.72	795	1.0982	0.3565	1.4547
1978Pertini	1492	603	53	14.33	639.45	654	1.0918	0.3768	1.4686
1979Pertini	2311	800	70	18.00	848.35	868	1.1558	0.4087	1.5645
1980Pertini	1360	535	50	13.75	567.95	583	1.1800	0.4080	1.5880
1981Pertini	2819	911	96	20.00	983.94	1005	1.1085	0.3967	1.5052
1982Pertini	2486	854	90	19.00	921.74	942	1.1257	0.4064	1.5321
1983Pertini	3746	1149	118	23.66	1236.65	1265	1.2513	0.4633	1.7146

1984Pertini	1340	514	42	13.66	539.18	554	1.1704	0.4048	1.5752
1985Cossiga	2359	859	118	17.00	955.75	975	1.2033	0.3964	1.5997
1986Cossiga	1348	561	65	14.00	610.09	624	1.0699	0.3788	1.4487
1987Cossiga	2092	904	109	15.00	993.76	1011	1.2312	0.3769	1.6081
1988Cossiga	2384	875	123	19.00	976.91	996	1.0606	0.3910	1.4516
1989Cossiga	1912	778	85	17.00	842.21	861	1.1742	0.4297	1.6039
1990Cossiga	3345	1222	155	20.00	1351.79	1375	1.2214	0.4012	1.6226
1991Cossiga	418	241	22	7.00	254.77	261	1.0384	0.3047	1.3431
1992Scalfaro	2774	978	118	17.50	1072.80	1094	1.2848	0.4025	1.6873
1993Scalfaro	2942	1074	129	18.60	1179.30	1201	1.2327	0.4000	1.6327
1994Scalfaro	3606	1190	171	21.00	1333.26	1359	1.2869	0.4286	1.7155
1995Scalfaro	4233	1341	180	22.66	1492.52	1519	1.2227	0.4071	1.6298
1996Scalfaro	2085	866	88	16.00	934.04	952	1.1975	0.3934	1.5909
1997Scalfaro	5012	1405	167	27.50	1538.44	1570	1.1908	0.4458	1.6366
1998Scalfaro	3995	1175	137	23.50	1281.19	1310	1.2806	0.4559	1.7365
1999Ciampi	1941	831	66	16.50	877.32	895	1.1405	0.4012	1.5417
2000Ciampi	1844	822	70	16.00	871.20	890	1.2531	0.4377	1.6908
2001Ciampi	2098	898	89	18.00	965.54	985	1.1446	0.4248	1.5694
2002Ciampi	2129	909	96	17.00	984.94	1003	1.1287	0.3914	1.5201
2003Ciampi	1565	718	63	14.00	763.50	779	1.1925	0.3919	1.5844
2004Ciampi	1807	812	76	15.00	869.71	886	1.1639	0.3833	1.5472
2005Ciampi	1193	538	54	12.66	576.22	590	1.1815	0.3989	1.5804
2006Napolitano	2204	929	125	16.50	1033.53	1052	1.1918	0.3935	1.5853
2007Napolitano	1792	793	101	16.00	874.57	892	1.1621	0.4118	1.5739
2008Napolitano	1713	775	75	15.00	831.25	848	1.1961	0.4046	1.6007

Here we see, that the $p+q$-values lie in an interval which gets narrower with increasing text size and approaches the golden section. The graphical presentation can be seen in Figure 3.6.

Experiments with many other data in 28 languages brought the same results (cf. Tuzzi, Popescu, Altmann 2009c) but neither the background mechanism nor the mathematical argumentation leading to this surprising phenomenon are known and must be left as problem for the future.

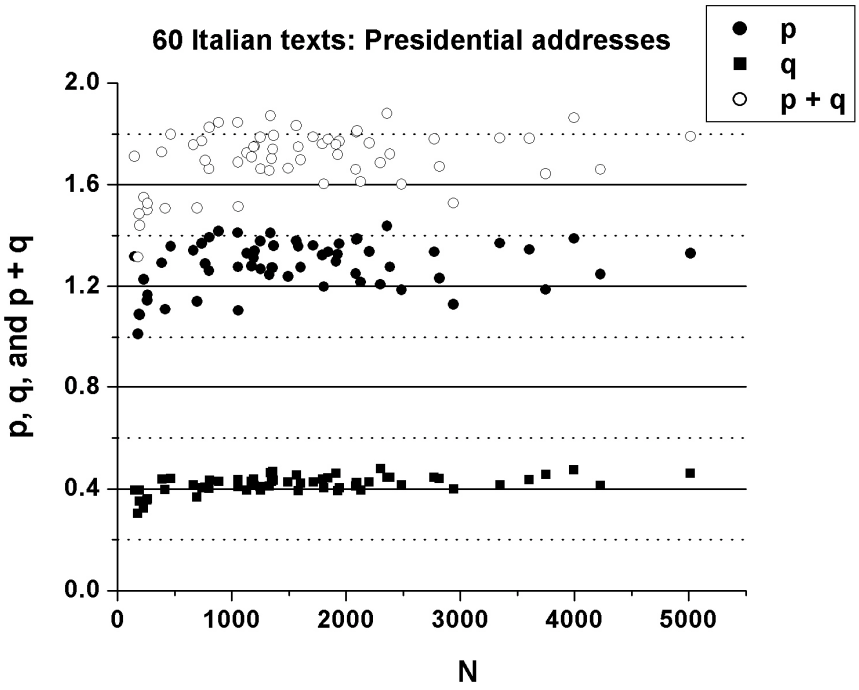

Figure 3.6. Indicators *p*, *q*, and their sum in terms of the text size *N* in 60 Italian texts: Presidential addresses (data from Tuzzi, Popescu, Altmann 2009c)

4. Zipf´s law

4.1. Validity

Usually the first question of all those who treat rank-frequency data is that of the validity of Zipf´s law. Whatever is its history, substantiation, modifications, variations, controversies etc., its fitting to new data seems to be a kind of performance of duty. If the fit is satisfactory, then one considers the analysis as finished; everything is correct. But if it is not satisfactory, three questions arise:

(1) Is the text adequate? A simple answer is that text is text in any case, i.e. always adequate but professionals know that texts can be mixed (e.g. written by two authors), added as in a corpus or representing different speakers as in a stage play and in these cases they need not fulfil the conditions required for the holding of a law. Further a text can be very short, hence the words do not have the possibility to unfold their frequency of occurrence which is necessary for testing a law.

(2) Did the analyst perform the segmentation in words correctly? In the Euro-American tradition one defines the word as a sequence between two blanks which is very practical and may turn out to be correct but it is neither general enough nor necessarily correct. In Japanese there is the problem whether some postpositions are parts of the word or not; in Slovak one does not know whether prepositions taking over the main accent of the following word are preclitics; what is the status of postclitics (e.g. *–lah*, *-kah*) or short preposition (*ke-*, *di-*) in Indonesian; are dependent French personal pronouns affixes (*je t´aime*), or are articles obeying liaison prefixes in French or Italian; is the hyphen to be treated always in the same way, etc. To follow simply the written form is generally possible but not always "correct". Thus text is not "data" because in every language there are fuzzy boundaries between words; text is the result of grammarian´s decisions called criteria which are not present in the data but set up conventionally by us.

(3) Lastly the mathematical question arises: is the model adequate? There are many exceptions to the simple Zipf´s law represented by the power function whose appearance in many scientific disciplines supports its acceptability. Now, it must be decided whether the model should be modified in order to allow for the "peculiarity" in the data caused perhaps by a strange influence on the pertinent entities, or the whole system (e.g. a phonemic) is wandering to another attractor, etc. Either one modifies the old model or proposes new ones. The literature is enormous and not yet systematized. We must always reckon with exceptions because language is in a steady development and sometimes a small change, e.g. a merging of two vowels can cause radical reorganization of different subsystems.

In order to cope with the bewildering situation we reduce the problem to two models: the original Zipf approach, considering the data a continuous sequence, and an approach initiated by Popescu et al. (2009). In both cases we test the goodness-of-fit using the determination coefficient showing the ratio of unexplained deviations to all deviations rather than using the chi-square test for goodness-of-fit, which does not yield unique results (it increases with increasing sample size), and in case of word frequencies strongly depending on the way of pooling small frequencies.

For testing the simple Zipf´s law we use the power function

(4.1) $f(r) = cr^{-a}$

where $f(r)$ is the frequency at rank r, and a and c are parameters. If (4.1) is used as a probability distribution, either a should have a value greater than 1 – otherwise the series does not converge – or the distribution must be truncated at the right hand side. In this case a could attain any positive value. Using (4.1) as a simple function, the parameters seem to be more realistic. The result of fitting (4.1) to the 60 texts is presented in Table 4.1.

Table 4.1
Fitting the power sequence (4.1) to the speeches

Text	***N***	***V***	***f*(1)**	***c***	***a***	$\mathbf{R^2}$
1949Einaudi	194	140	10	10.41	0.5616	0.9381
1950Einaudi	150	105	9	8.49	0.5300	0.9421
1951Einaudi	230	169	9	9.93	0.5226	0.9195
1952Einaudi	179	145	7	6.61	0.4543	0.8825
1953Einaudi	190	143	8	8.44	0.5065	0.8928
1954Einaudi	260	181	12	12.61	0.5679	0.9468
1955Gronchi	388	248	16	18.97	0.6037	0.9340
1956Gronchi	665	374	29	34.71	0.6617	0.9334
1957Gronchi	1130	549	65	63.41	0.7219	0.9889
1958Gronchi	886	460	41	47.32	0.6861	0.9578
1959Gronchi	697	388	33	36.92	0.6717	0.9683
1960Gronchi	804	434	41	44.25	0.6915	0.9786
1961Gronchi	1252	622	67	69.26	0.7306	0.9790
1962Segni	738	381	35	40.14	0.6730	0.9642
1963Segni	1057	527	46	50.47	0.6674	0.9743
1964Saragat	465	278	21	23.57	0.6248	0.9668
1965Saragat	1052	510	52	59.46	0.7052	0.9640
1966Saragat	1200	597	44	55.78	0.6605	0.9430

1967Saragat	1056	526	51	54.24	0.6865	0.9760
1968Saragat	1173	562	56	62.00	0.6957	0.9734
1969Saragat	1583	692	86	89.12	0.7242	0.9759
1970Saragat	1929	812	85	94.95	0.6942	0.9642
1971Leone	262	168	12	12.23	0.5385	0.9642
1972Leone	767	394	32	33.15	0.6141	0.9692
1973Leone	1250	616	67	68.50	0.7287	0.9844
1974Leone	801	426	32	37.61	0.6450	0.9644
1975Leone	1328	632	63	72.41	0.7141	0.9598
1976Leone	1366	649	52	63.68	0.6722	0.9557
1977Leone	1604	717	80	89.54	0.7334	0.9628
1978Pertini	1492	603	53	68.43	0.6536	0.9446
1979Pertini	2311	800	70	98.88	0.6465	0.9132
1980Pertini	1360	535	50	66.74	0.6553	0.9265
1981Pertini	2819	911	96	119.11	0.6564	0.9444
1982Pertini	2486	854	90	110.88	0.6671	0.9477
1983Pertini	3746	1149	118	157.94	0.6610	0.9063
1984Pertini	1340	514	42	53.53	0.6062	0.9457
1985Cossiga	2359	859	118	134.06	0.7381	0.9717
1986Cossiga	1348	561	65	70.15	0.6860	0.9786
1987Cossiga	2092	904	109	121.64	0.7560	0.9664
1988Cossiga	2384	875	123	133.64	0.7349	0.9831
1989Cossiga	1912	778	85	98.10	0.7002	0.9598
1990Cossiga	3345	1222	155	181.59	0.7430	0.9610
1991Cossiga	418	241	22	24.47	0.6546	0.9449
1992Scalfaro	2774	978	118	145.79	0.7118	0.9513
1993Scalfaro	2942	1074	129	158.40	0.7264	0.9522
1994Scalfaro	3606	1190	171	188.32	0.7324	0.9833
1995Scalfaro	4233	1341	180	215.64	0.7248	0.9603
1996Scalfaro	2085	866	88	103.39	0.7029	0.9639
1997Scalfaro	5012	1405	167	227.30	0.6858	0.9198
1998Scalfaro	3995	1175	137	186.04	0.6848	0.9240
1999Ciampi	1941	831	66	81.98	0.6579	0.9420
2000Ciampi	1844	822	70	80.25	0.6669	0.9593
2001Ciampi	2098	898	89	94.82	0.6836	0.9598
2002Ciampi	2129	909	96	104.06	0.7065	0.9801
2003Ciampi	1565	718	63	70.35	0.6680	0.9657
2004Ciampi	1807	812	76	83.24	0.6863	0.9794
2005Ciampi	1193	538	54	57.99	0.6654	0.9745

2006Napolitano	2204	929	125	121.38	0.7445	0.9775
2007Napolitano	1792	793	101	103.55	0.7530	0.9901
2008Napolitano	1713	775	75	89.65	0.7096	0.9611

The parameter *c* could evidently be replaced by the greatest frequency *f*(1) and need not be further interpreted. The parameter *a* is, however, a characteristic of texts. Since it is always smaller than 1, the rank-frequency sequence lies over the harmonic sequence; it is not a merely mathematical construction. The presidents have different *a*'s. Computing the means of *a*'s for individual presidents we obtain the result as displayed in Table 4.2.

Table 4.2
The mean of parameter *a* for individual presidents

Historical order	**President**	**Mean *a***
1	Einaudi	0.5238
2	Gronchi	0.6810
3	Segni	0.6702
4	Saragat	0.6844
5	Leone	0.6637
6	Pertini	0.6494
7	Cossiga	0.7161
8	Scalfaro	0.7098
9	Ciampi	0.6764
10	Napolitano	0.7357

The graphical representation in Figure 4.1 shows that *a* has a slightly increasing character in time but the sequence is not long enough to make predictions.

On the other hand, if we link the parameter *a* with the sample size *N*, then one can see that *a* increases with increasing *N*, and at *ca N* = 1,000 it attains its final value and oscillates around 0.7. This value can be accepted as a local value holding for presidential speeches but not even for Italian as a whole where also *a* = 0.7895 has been found (cf. Popescu, Altmann 2008). The speeches seem to stabilize and after 60 years to obtain an individual form. In Figure 4.2 the time sequence of *a* is presented, in Figure 4.3 its relation to text size *N*. The same picture can be obtained if we link *a* with the vocabulary size *V*.

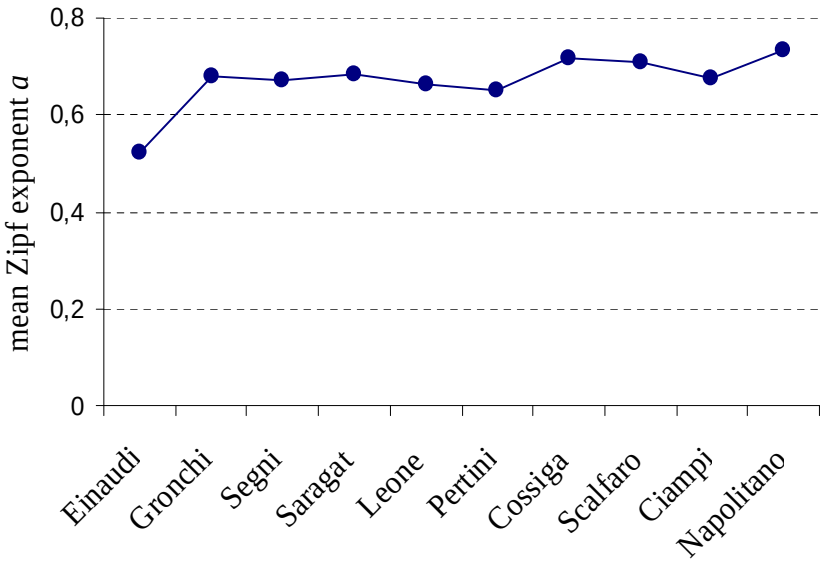

Figure 4.1. The change of parameter *a*

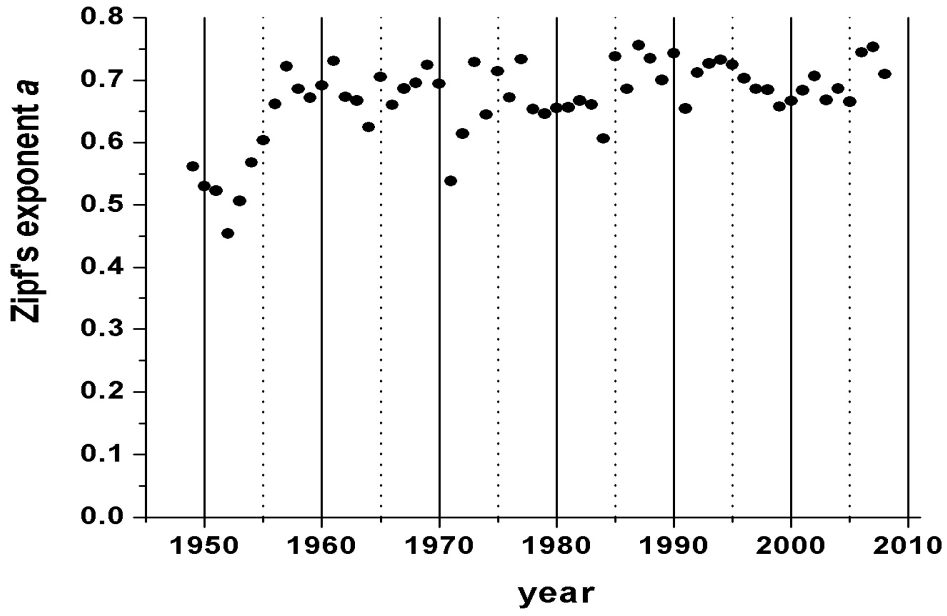

Figure 4.2.The convergence of parameter *a*

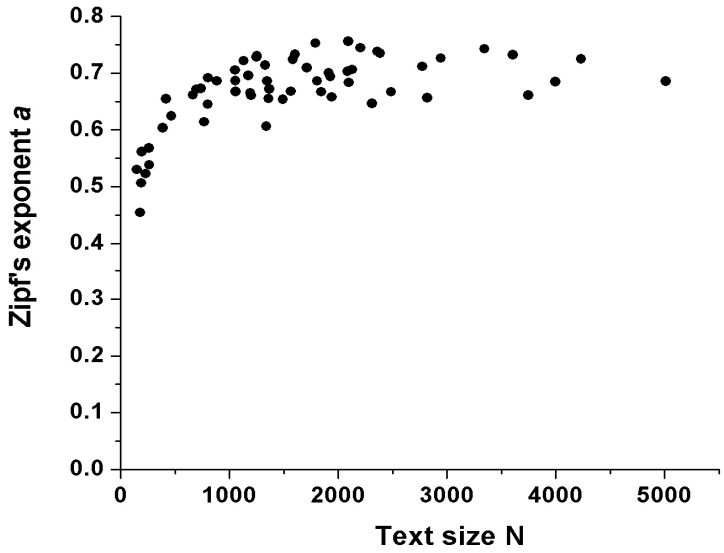

Figure 4.3. The relation of parameter *a* to text size *N*

As has been shown by Popescu, Altmann (2008) and Tuzzi, Popescu, Altmann (2009a), parameter *a* together with some other quantities can be used also for typological purposes.

4.2. Stratificational approach

Now, the power function can show whether a language tends to more analytism or synthetism but exactly this is the cause why it can be fitted to texts from different languages with different success. In highly synthetic languages the curve lies below the hapax legomena; in highly analytic languages it lies above the hapax legomena (Popescu, Mačutek, Altmann 2009:105ff). That means, it is not flexible enough to capture the deviations caused by morphology; nevertheless, it can signalize them.

The second non-realistic feature is the fact that it converges to zero which cannot be attained either by absolute or by relative frequencies. The smallest absolute frequency is 1; the smallest relative frequency is not fixed but it is not zero. The deviation is, of course, smaller with relative frequencies.

The third critical point is the fact that the power function does not tell us much about the text, the forces being effective during its creation, etc. This is why many researchers beginning with B. Mandelbrot tried to give the law a linguistic foundation. The enormous extent of this enterprise can be found on the Internet[1] documented by Wentian Li. As a matter of fact, no mathematical model

[1] http://www.nslij-genetics.org/wli/zipf

could capture the interplay of physical, biological, psychological, social and linguistic forces which are active during the text creation but partial models can furnish partial truth. In what follows we present the Popescu-Altmann-Köhler (2009) stratificational approach.

Any text, even if it is written only by one writer, is a construct consisting of practically infinite number of layers. The layers belong to different domains (physical, biological, etc.) but even in one domain there are as many strata as we are able to discern at the given state of research. Since we restrict ourselves to linguistics, one can find e.g. phonological, morphological, syntactic, semantic, lexical, idiolectal, sociolectal, dialectal, etc. strata, hence whatever we examine the result should be *eo ipso* a superposition of functions or distributions, representing a kind of mixed universe of discourse. And we realize with surprise that even if we penetrate in the depth of one stratum, we find the same picture in slight variations. Thus the idea of self-similarity will always appear when we venture to make a step in the depth in the same direction (cf. Hřebíček 2000).

Let us, at this place, consider only the words of a text. All of them can be classified in different classes according to the basis of classification which can be phonological, morphological, syntactic, semantic, discoursal, etc. The same word can belong in different classifications to different classes. But if a class is made up "correctly", then its elements abide by some rank-frequency law, i.e., their frequencies are different and "regularly" distributed. What happens, if we ignore a sub-classification (e.g. in parts-of-speech), join all classes and re-rank the frequencies? Then we automatically obtain a superposition of functions or a mixture of distributions differing only in parameters. The same event can happen if the text has been written by several authors. In order to capture the situation from this point of view, Popescu et al. (2009) proposed to use the superposition of exponentials in the form

$$f(r) = 1 + A_1 \exp(-r / a_1) + A_2 \exp(-r / a_2) + A_3 \exp(-r / a_3) + \ldots, \tag{4.2}$$

where 1 is added because there is no frequency smaller than 1. The formula has the advantage that for capturing the regularity of a rank-frequency sequence one exponential is usually sufficient in order to give results comparable to the Zipfian approach, but besides it shows the number of relevant strata: if two parameters a_i are equal, one of the terms may be skipped.

Though one always strives for reducing the number of parameters in the models, stratification is a phenomenon with which their reduction is possible only if they are redundant. Using this technique in the case of presidential addresses we used three terms of (4.2). In Table 4.3 the equal terms are highlighted and signalize a greater uniformity of the text. The goodness-of-fit is compared with that of Zipf´s power function. Because of more terms it is always better but even a one-term solution would be sufficient.

Table 4.3
Fitting (2) with three components to the individual texts
(from Tuzzi, Popescu, Altmann 2009a)

Text	A_1	a_1	A_2	a_2	A_3	a_3	R^2	Zipf R^2
1949Einaudi	7.6443	1.6698	2.8179	9.0029	2.6568	9.0029	0.9843	0.9381
1950Einaudi	5934.2065	0.1314	2.5977	3.2619	3.4366	11.0938	0.9680	0.9421
1951Einaudi	3.9695	2.7015	3.8897	2.7015	3.4125	13.9569	0.9693	0.9195
1952Einaudi	14.2438	0.4709	2.6599	7.1768	2.2826	7.1768	0.9701	0.8825
1953Einaudi	4.2712	2.6645	3.4545	2.6645	2.3721	14.0995	0.9489	0.8928
1954Einaudi	5.8840	2.6907	5.8840	2.6907	3.2691	17.6275	0.9800	0.9468
1955Gronchi	8.4966	4.1362	6.6965	4.1362	3.9743	23.0234	0.9848	0.9340
1956Gronchi	18.2388	3.6098	13.0374	3.6098	6.4362	31.4130	0.9750	0.9334
1957Gronchi	125.2218	0.8786	19.5045	9.6956	6.5511	54.2070	0.9966	0.9889
1958Gronchi	40.1958	1.9024	15.2168	9.4166	5.0742	48.2958	0.9777	0.9578
1959Gronchi	33.9712	2.2240	5.3585	9.3096	6.8189	30.9192	0.9872	0.9683
1960Gronchi	47.4827	1.8192	5.4082	10.9556	8.7466	29.4986	0.9945	0.9786
1961Gronchi	95.4670	1.3961	18.0649	13.1414	4.5681	71.7785	0.9872	0.9790
1962Segni	36.1129	2.2271	8.6352	11.4631	5.0045	41.5065	0.9829	0.9642
1963Segni	52.2356	1.4660	16.0616	11.8761	5.2227	59.0391	0.9883	0.9743
1964Saragat	11.4325	1.5246	10.5449	4.8139	5.8916	23.0845	0.9928	0.9668
1965Saragat	60.2752	2.0834	13.3610	17.1656	4.2634	54.3607	0.9847	0.9640
1966Saragat	34.3406	3.4197	15.6111	16.0344	4.4120	61.5389	0.9950	0.9430
1967Saragat	91.2582	0.6516	28.5931	7.6056	5.5109	57.5730	0.9927	0.9760
1968Saragat	61.4532	1.8357	16.2212	13.6566	5.6743	57.6195	0.9900	0.9734
1969Saragat	383809.9300	0.1069	48.9568	6.6664	9.9259	56.6177	0.9930	0.9759
1970Saragat	109.1318	0.9186	44.2931	11.8090	6.3860	91.0766	0.9966	0.9642
1971Leone	13.2030	1.2355	3.3502	16.6551	2.0066	16.6551	0.9816	0.9642
1972Leone	18222.6392	0.1368	15.3538	7.9291	5.4091	47.8131	0.9919	0.9692
1973Leone	58.7576	1.9100	32.1470	1.9100	13.1578	36.0208	0.9905	0.9844
1974Leone	19.2585	3.5130	13.8382	3.5130	7.4868	37.2214	0.9909	0.9644
1975Leone	67.5108	2.3533	17.3090	14.6730	4.6199	73.9528	0.9824	0.9598
1976Leone	48.2642	2.5209	17.5557	12.8206	6.1262	68.6035	0.9862	0.9557
1977Leone	87.4626	2.3195	20.5739	12.4245	5.7535	87.4776	0.9801	0.9628
1978Pertini	51.7380	3.5263	12.0036	20.4498	6.4901	78.6425	0.9878	0.9446
1979Pertini	65.1142	5.2342	17.2578	36.1223	5.1991	117.3082	0.9903	0.9132
1980Pertini	43.1443	4.4462	14.5189	21.1942	4.1853	88.2894	0.9877	0.9265

1981Pertini	76.4539	4.0594	26.2023	25.7265	7.5229	134.3922	0.9951	0.9444
1982Pertini	69.6833	3.7835	27.8234	19.6365	7.2609	124.0235	0.9931	0.9477
1983Pertini	80.0833	6.0098	38.0850	25.3670	7.9005	157.6009	0.9937	0.9063
1984Pertini	25.9186	3.1513	16.6236	13.1599	7.3192	77.4894	0.9951	0.9457
1985Cossiga	136.8596	2.2399	28.6480	16.4338	7.6528	108.3134	0.9873	0.9717
1986Cossiga	76.6676	1.6512	18.6207	17.4036	5.2327	76.1050	0.9949	0.9786
1987Cossiga	138.7106	1.8182	29.5079	14.3326	6.8221	89.7417	0.9812	0.9664
1988Cossiga	148.3059	1.8420	32.4753	18.0416	6.8291	111.1064	0.9954	0.9831
1989Cossiga	51.2245	2.1159	48.3362	2.1159	26.5383	33.1236	0.9827	0.9598
1990Cossiga	177.1306	2.1907	44.5506	16.6650	9.2062	120.5184	0.9826	0.9610
1991Cossiga	23.3737	1.9480	5.5257	7.7038	3.7792	28.7386	0.9617	0.9449
1992Scalfaro	45.4095	1.6037	93.3267	6.6096	14.8209	78.2459	0.9953	0.9513
1993Scalfaro	98.6053	2.2561	64.2523	8.9261	13.5482	84.1309	0.9877	0.9522
1994Scalfaro	190.1328	1.8193	52.2495	13.0644	13.6008	109.1463	0.9964	0.9833
1995Scalfaro	165.2531	3.3601	47.0084	17.0293	11.8799	137.3919	0.9942	0.9603
1996Scalfaro	75.4754	2.7234	31.1186	15.8561	5.0698	117.9454	0.9938	0.9639
1997Scalfaro	111.2014	3.0557	86.7475	16.3968	12.8793	151.8140	0.9944	0.9198
1998Scalfaro	131.1502	5.1954	28.0699	32.9316	8.7921	151.2300	0.9950	0.9240
1999Ciampi	38.4757	1.1304	45.4522	9.7402	8.2096	81.7525	0.9944	0.9420
2000Ciampi	77.3481	1.0461	34.2762	12.0572	7.9717	73.8609	0.9970	0.9593
2001Ciampi	9650.7466	0.1803	46.4997	10.8622	8.0033	86.6775	0.9967	0.9598
2002Ciampi	100.8667	1.4791	37.0863	12.2883	7.8884	87.8901	0.9950	0.9801
2003Ciampi	89.4191	0.7558	33.2284	9.4025	8.2983	63.4115	0.9951	0.9657
2004Ciampi	76.4494	1.9561	23.0445	16.6297	6.0687	87.2679	0.9952	0.9794
2005Ciampi	65.9313	1.3135	16.5153	14.1565	6.9560	55.0985	0.9955	0.9745
2006Napolitano	378.2224	0.6264	42.8270	12.7371	7.5295	88.7431	0.9940	0.9775
2007Napolitano	144.1288	1.3544	28.9434	14.7774	4.7310	101.0831	0.9968	0.9901
2008Napolitano	65.6537	2.4501	28.9426	12.5498	6.0662	78.7920	0.9948	0.9611

Comparing the results with those obtained from “monolithic” texts in different languages (cf. Popescu, Altmann, Köhler 2009) where two terms were always sufficient, we may conclude that the presidential speeches were mostly written by two authors; the President himself and his official consultant. Einaudi´s texts are either too short or were written by one person, and three speeches of Leone were written by one person only. A historical examination could be made easier if one used these results.

We may conclude that Zipf´s law holds for these texts but there are possibilities to make a step deeper into the genesis of the text. Needless to say, in this way one could join several views of the text, but one would be forced to use

more variables. The advancement of investigation will make such an enhancement of complexity necessary.

5. The parts of speech

In Chapter 3 we proposed the indicators *p* and *q* and showed that their sum tends to the golden section. Automatically the question arises whether this holds only for words or also for any other text entities and their different groupings (classes). One could count phonemes, letters, graphemes, syllables, morphemes, parts of speech, canonical forms, motifs, etc. It is neither our aim to check the complete palette of possibilities which is infinite because any linguistic entities are our conceptual constructs nor to extend the examination to other languages. Here we restrict ourselves to the classical parts of speech. In the presidential addresses we have distinguished the following classes

adjectives
adverbs
articles
conjunctions
interjections
nouns
numerals
prepositions
pronouns
proper nouns
verbs

i.e. 11 classes (I = 11) that were furnished by the tagging software. Not all of them are contained in all texts: interjections, proper nouns and numerals do not necessarily occur everywhere. The parts of speech have *eo ipso* a smaller, very restricted inventory (= number of classes); nevertheless, they can be ranked and the properties of the rank-frequency sequence can be studied. As is well known, each of the above classes can be subdivided in further classes (e.g. pronouns: personal, deictic, relational, interrogative,…) which may be studied separately. Hence POS is only one of the possible classifications. Here we shall scrutinize only Zipf´s law and the stability of the indicators *p* and *q*.

5.1. Zipf´s law

The frequencies of parts of speech in individual speeches are presented in Table 5.1. N is the text length, I the inventory which is maximally 11. Table 5.2. contains the fitting of Zipf´s power function together with the Zipf parameter c and a and the coefficients A_1, a_1 computed in Chapter 4. As can be seen, Zipf´s a is evidently neutral to increasing N as can be seen in Figure 5.1.

Table 5.1
Frequencies of parts-of-speech diversification
in Italian End-of-Year Addresses

Text	**Parts-of-speech frequencies**	***N***	***I***
1949Einaudi	41,37,33,30,17,15,14,6,1	194	9
1950Einaudi	42,36,20,15,15,9,8,4,1	150	9
1951Einaudi	50,41,40,34,21,18,15,11	150	8
1952Einaudi	46,35,28,27,13,12,11,7	131	8
1953Einaudi	47,42,34,24,15,12,9,6,1	190	9
1954Einaudi	57,54,43,36,20,18,17,14,1	260	9
1955Gronchi	83,78,64,51,31,30,29,21,1	388	9
1956Gronchi	180,121,88,87,71,58,34,26	465	8
1957Gronchi	267,241,170,126,93,84,79,59,6,5	1090	10
1958Gronchi	201,162,131,127,82,74,63,42,3,1	886	10
1959Gronchi	181,135,92,80,72,71,36,29,1	697	9
1960Gronchi	196,161,112,106,78,63,45,38,3,2	794	10
1961Gronchi	304,244,184,162,111,105,75,65,2	1243	9
1962Segni	196,147,120,83,73,54,36,29	514	8
1963Segni	257,219,170,131,93,68,52,45,14,8	987	10
1964Saragat	102,85,84,64,42,40,28,17,3	447	9
1965Saragat	267,211,141,138,85,79,78,45,6,3	1033	10
1966Saragat	324,239,185,144,109,75,66,50,5,2	1189	10
1967Saragat	263,207,167,145,96,64,59,36,14,3,2	1045	11
1968Saragat	304,243,176,134,95,86,70,56,8,2	1164	10
1969Saragat	394,284,232,222,165,103,99,72,8,3,2	1573	11
1970Saragat	490,389,272,257,186,113,112,86,17,5,2	1918	11
1971Leone	70,51,37,35,30,17,11,6,3,2	252	10
1972Leone	182,149,134,111,69,45,45,24,5,3	747	10
1973Leone	298,232,205,174,103,97,76,63,1,1	1250	10
1974Leone	197,141,139,120,66,59,42,35,1,1	801	10
1975Leone	312,244,200,191,122,97,91,69,2	1319	9
1976Leone	321,239,211,196,113,112,97,73,3,1	1366	10
1977Leone	358,270,262,216,142,122,115,113,4,2	1594	10
1978Pertini	332,283,248,156,130,125,106,86,17,10	1403	10
1979Pertini	499,442,345,279,219,201,184,115,8,8,2	2291	11
1980Pertini	316,244,228,164,121,104,101,61,10,9,2	1349	11
1981Pertini	571,571,377,331,261,231,227,196,38,14,1	2818	11
1982Pertini	509,495,332,322,233,202,172,139,62,19,2	2476	11
1983Pertini	786,760,510,452,360,308,275,206,55,33,3	3726	11

1984Pertini	302,269,197,163,129,97,95,51,20,17	1180	10
1985Cossiga	612,427,404,289,207,192,120,93,10,3,2	2384	11
1986Cossiga	321,232,215,187,130,106,79,77,1,1	1349	10
1987Cossiga	501,414,349,248,184,163,107,103,11,11	1991	10
1988Cossiga	557,467,369,311,199,183,146,134,14,5	2345	10
1989Cossiga	441,399,302,231,154,145,102,101,31,6	1862	10
1990Cossiga	800,646,534,396,305,277,173,163,35,18	3177	10
1991Cossiga	95,71,64,57,48,29,26,22,4,2	408	10
1992Scalfaro	656,472,435,360,250,231,208,151,4,3,2	2761	11
1993Scalfaro	684,501,469,387,247,236,218,168,22,8,1	2941	11
1994Scalfaro	866,633,590,482,284,267,248,207,15,12,1	3605	11
1995Scalfaro	994,741,682,523,357,332,290,246,38,22,3	4206	11
1996Scalfaro	535,348,326,313,183,128,115,110,16,11	1985	10
1997Scalfaro	1113,1048,712,522,429,397,368,329,54,33,10	4916	11
1998Scalfaro	972,775,577,415,399,289,254,251,35,23,5	3951	11
1999Ciampi	504,347,291,278,206,110,89,82,24,9,1	1941	11
2000Ciampi	432,338,291,273,168,124,95,88,23,12	1734	11
2001Ciampi	549,395,338,262,224,109,96,89,18,15,2	2086	11
2002Ciampi	556,389,312,304,209,132,112,98,10,7	2069	10
2003Ciampi	408,297,231,214,142,112,79,75,4,2,1	1565	11
2004Ciampi	455,353,268,265,147,111,93,88,19,8	1737	10
2005Ciampi	290,235,181,166,114,89,55,40,12,10,1	1193	11
2006Napolitano	502,377,356,286,191,169,159,146,10,7,1	2204	11
2007Napolitano	419,352,274,242,144,123,115,104,13,5,3	1772	11
2008Napolitano	409,328,281,220,135,127,120,86,4,2,1	1713	11

Table 5.2
Fitting Zipf´s law and Popescu´s approach to parts of speech

	Zipf fitting: $y = c/x^a$			**Exp fitting: $y = 1 + A*\exp(-x/a)$**		
Text	**c**	**a**	**R^2**	**A**	**a**	**R^2**
1949Einaudi	47.07	0.5966	0.7315	55.4131	4.4061	0.9054
1950Einaudi	46.08	0.8131	0.8788	59.6584	2.9894	0.9664
1951Einaudi	44.26	0.7185	0.7681	56.1011	3.3544	0.9106
1952Einaudi	43.05	0.8300	0.8511	57.2212	2.8153	0.9477
1953Einaudi	53.39	0.7232	0.8121	66.9476	3.4773	0.9595
1954Einaudi	65.39	0.6288	0.7780	78.2885	4.1636	0.9330
1955Gronchi	94.85	0.6055	0.7765	112.0836	4.4270	0.9300
1956Gronchi	160.67	0.8950	0.9203	214.7349	2.7377	0.9597

1957Gronchi	297.24	0.7397	0.8308	364.2917	3.7208	0.9584
1958Gronchi	223.42	0.6749	0.8013	262.1934	4.2630	0.9408
1959Gronchi	192.28	0.7195	0.8828	228.9571	3.8226	0.9481
1960Gronchi	215.74	0.7381	0.8441	260.2692	3.7961	0.9645
1961Gronchi	329.68	0.6805	0.8636	393.3201	4.0293	0.9669
1962Segni	180.13	0.8950	0.8879	248.7756	2.6638	0.9816
1963Segni	281.31	0.7741	0.8393	352.5834	3.4962	0.9769
1964Saragat	112.73	0.6301	0.7636	134.6237	4.2649	0.9270
1965Saragat	289.04	0.7664	0.8653	350.8234	3.6439	0.9615
1966Saragat	350.29	0.8100	0.8867	436.2819	3.3792	0.9840
1967Saragat	292.45	0.7891	0.8406	358.7014	3.6043	0.9747
1968Saragat	332.60	0.7834	0.8779	410.9684	3.5014	0.9798
1969Saragat	429.42	0.7727	0.8389	512.3810	3.8062	0.9575
1970Saragat	541.52	0.7995	0.8561	662.1603	3.5796	0.9738
1971Leone	74.66	0.8131	0.8649	92.6163	3.3028	0.9644
1972Leone	203.65	0.7339	0.7917	250.6557	3.7467	0.9483
1973Leone	330.42	0.7133	0.8157	396.2084	3.9666	0.9510
1974Leone	216.28	0.7305	0.8063	260.6625	3.8525	0.9394
1975Leone	338.22	0.6500	0.8354	398.4730	4.2682	0.9506
1976Leone	350.97	0.6922	0.8137	412.1600	4.1778	0.9371
1977Leone	392.82	0.6576	0.7886	456.3595	4.4352	0.9205
1978Pertini	363.17	0.6967	0.8141	434.2952	4.0518	0.9481
1979Pertini	570.71	0.6963	0.7846	671.3088	4.2701	0.9402
1980Pertini	351.56	0.7338	0.8197	417.2399	4.0059	0.9573
1981Pertini	667.30	0.6584	0.7659	774.9393	4.5525	0.9208
1982Pertini	591.23	0.6646	0.7873	689.4811	4.4881	0.9476
1983Pertini	909.84	0.6817	0.7916	1068.0115	4.3566	0.9469
1984Pertini	323.95	0.7425	0.8219	398.3592	3.7149	0.9640
1985Cossiga	667.32	0.8046	0.8502	812.9602	3.5882	0.9670
1986Cossiga	349.87	0.6997	0.8180	412.0190	4.1275	0.9426
1987Cossiga	550.19	0.7489	0.8294	676.8828	3.6957	0.9666
1988Cossiga	618.77	0.7124	0.8265	744.3419	3.9614	0.9606
1989Cossiga	496.24	0.7193	0.8255	605.9822	3.8465	0.9673
1990Cossiga	869.28	0.7428	0.8439	1058.5020	3.7696	0.9722
1991Cossiga	103.02	0.6770	0.8091	120.0938	4.2254	0.9410
1992Scalfaro	717.94	0.7338	0.8168	840.8369	4.0915	0.9418
1993Scalfaro	753.09	0.7218	0.8171	880.3619	4.1586	0.9453
1994Scalfaro	953.87	0.7475	0.8247	1131.8520	3.9566	0.9498
1995Scalfaro	1093.88	0.7346	0.8327	1288.6011	4.0523	0.9575

1996Scalfaro	559.62	0.7727	0.8465	675.5472	3.6738	0.9426
1997Scalfaro	1265.46	0.7248	0.8116	1508.0736	4.0325	0.9437
1998Scalfaro	1067.40	0.7673	0.8594	1273.9222	3.8173	0.9653
1999Ciampi	546.17	0.7988	0.8544	659.4942	3.6409	0.9634
2000Ciampi	467.53	0.7286	0.8146	566.8884	3.8555	0.9542
2001Ciampi	599.44	0.8146	0.8610	736.5069	3.5044	0.9746
2002Ciampi	591.54	0.7840	0.8629	720.7311	3.5887	0.9622
2003Ciampi	445.42	0.8073	0.8603	541.8572	3.5698	0.9703
2004Ciampi	492.24	0.7735	0.8489	606.5151	3.5775	0.9653
2005Ciampi	325.46	0.7683	0.8320	395.4018	3.7380	0.9719
2006Napolitano	556.45	0.7099	0.8040	649.5857	4.2250	0.9381
2007Napolitano	471.12	0.7486	0.8191	567.3769	3.8705	0.9584
2008Napolitano	459.23	0.7550	0.8194	553.2141	3.8427	0.9571

Zipf´s approach yields a mean determination coefficient $R^2 = 0.8281$, a fact that shows that Zipf´s approach is rather restricted to word frequencies than to any other entities. On the contrary, the new approach from Chapter 4 yields very high determination coefficients with mean $R^2 = 0.9537$ and each $R^2 > 0.9$ which is a new corroboration of the greater adequacy of the stratification approach. It is very probable that the new approach can be applied to any kind of linguistic entities.

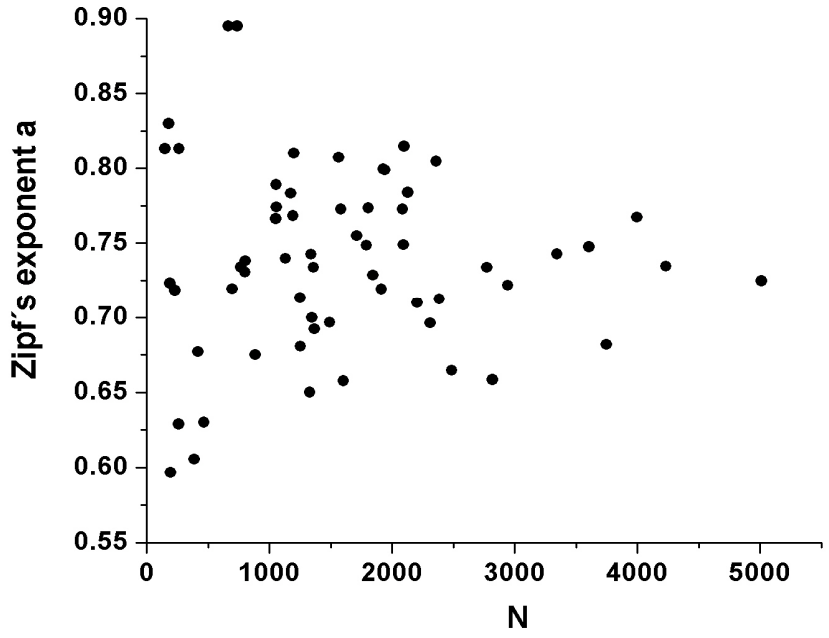

Figure 5.1. Increase of Zipf´s *a* with increasing *N*

Neither Zipf´s nor Popescu´s parameter *a* displays some developmental tendencies, a fact that confirms the (part-of-speech) homogeneity of the whole End-of-Year corpus.

Since the inventory of POS is small, it can be expected that in Popescu´s approach one exponential component will be sufficient to capture all data. This is shown in Table 5.2 but it can also be shown that if we apply more than one component, the parameters in the exponents are always equal. This is a sign of monostratality. In Table 5.3. we show the exponential fitting using 3 components to the texts of Einaudi.

Table 5.3.
Three-component exponential fitting to the texts of Einaudi

Text	A_1	a_1	A_2	a_2	A_3	a_3	R^2
1949Einaudi	18.4710	4.4061	18.4710	4.4061	18.4710	4.4061	0.9054
1950Einaudi	19.8861	2.9894	19.8861	2.9894	19.8861	2.9894	0.9664
1951Einaudi	18.7004	3.3544	18.7004	3.3544	18.7004	3.3544	0.9106
1952Einaudi	19.0737	2.8153	19.0737	2.8153	19.0737	2.8153	0.9477
1953Einaudi	22.3159	3.4773	22.3159	3.4773	22.3159	3.4773	0.9595
1954Einaudi	26.0962	4.1636	26.0962	4.1636	26.0962	4.1636	0.9330

The stratification approach has a further advantage allowing us to judge the "correctness" or adequacy of class formation. Though the mono-componential exponential is sufficient and displays good fitting, looking at Figures 5.2 and 5.3 we can see that in both cases the last three classes lie below the fitting function. Since this phenomenon is systematic, it would be necessary to analyze the status of these three classes. They are numerals, proper nouns and interjections. The interjections have a special status in different languages. Sometimes they contain phonemes used only in interjections; some texts do not contain them at all; they form a somewhat peripheral class. Proper nouns can simply be added to the nouns and the numerals can in some languages be subdivided to different other classes, e.g. nouns and adjectives. In some strongly synthetic languages they can be declined or admit different affixes. Proper nouns and interjections have no meaning. Thus the fitting could be improved if we pooled these somewhat amorphous classes before fitting. If we pool them before fitting, then both the power fitting and the exponential fitting improve slightly, $R^2 = 0.85$ and $R^2 = 0.98$ respectively in Ciampi 2005 and $R^2 = 0.90$ for Zipf and $R^2 = 0.97$ respectively in Scalfaro 1998.

Thus we see at least preliminarily some advantages of the exponential approach: it can be better substantiated than the power function, it yields better results and it better signalizes the so called "correctness" of data.

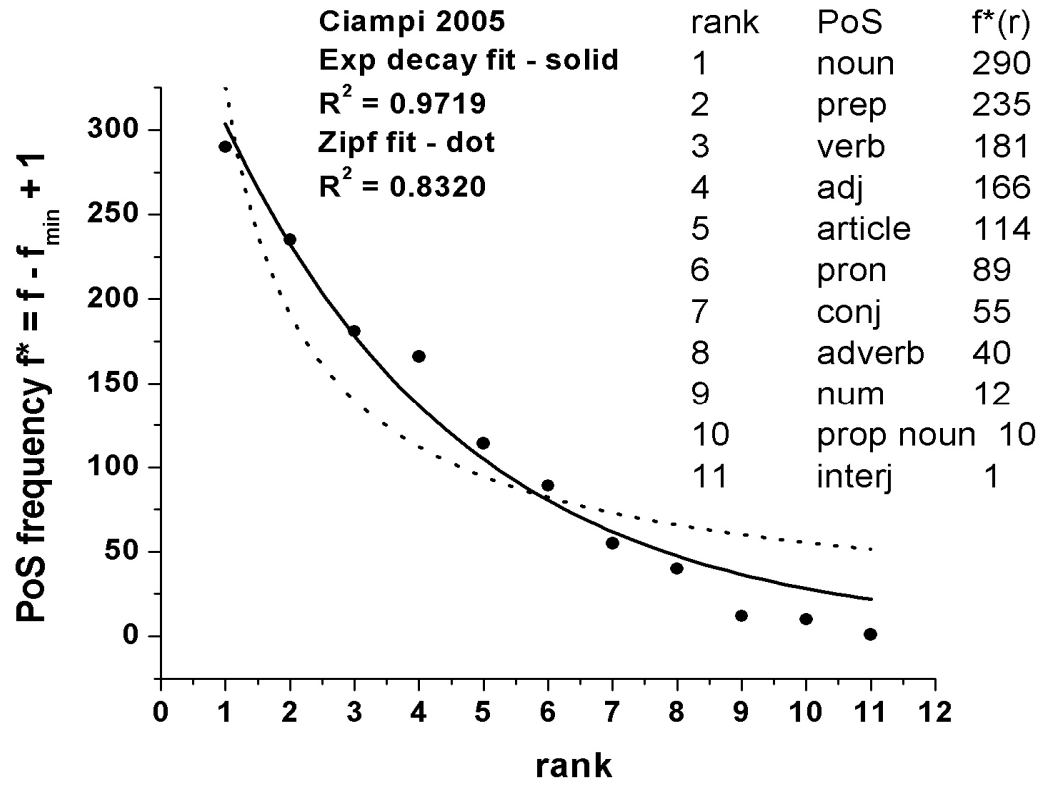

Figure 5.2. The fitting to the text Ciampi 2005

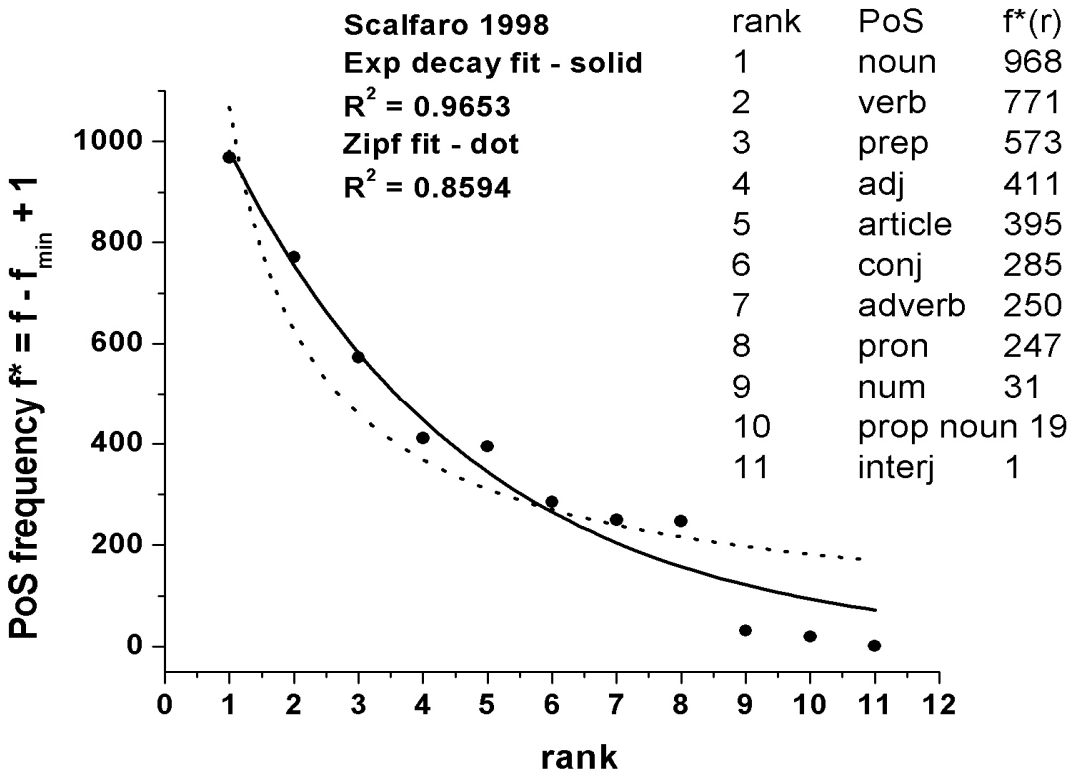

Figure 5.3. The fitting to the text Scalfaro 1998

5.2. The indicators *p* and *q*

In Chapter 3 it was shown that the sum of the indicators *p* and *q* converges to the golden section. Automatically the question arises whether restricted inventories like that of parts-of-speech do the same. While word-forms have a potentially infinite vocabulary (= inventory), parts-of-speech have a very restricted one. In our case we had maximally 11 classes but some texts contained only 8. Since the arc of POS is very short and the POS are not as conscious as words (one must learn them artificially in the school; they do not belong to the knowledge of the mother tongue), one can reasonably doubt that here one will find the golden section. But even if not, the next question arises: does only the use of a potentially infinite inventory lead to the golden section, and if so, what is the influence of inventory size on the sum of $p + q$? Evidently, here a very wide field of research arises because not only different inventories must be scrutinized but also a number of languages.

Here we shall concentrate only on parts-of-speech. As can be seen in Table 5.4., $p+q$ is only in two cases greater than 1.618…, and the mean of $p+q$ is 1.32

Table 5.4.
Indicators *p* and *q* of the Presidential speeches

Text	*N*	*R*	*f*(1)*	*h*	*L*	L_{max}	*p*	*q*	*p+q*
1949Einaudi	194	9	41	7.78	41.259	48	0.994	0.484	1.478
1950Einaudi	150	9	42	7.20	42.995	49	0.968	0.490	1.458
1951Einaudi	150	8	40	6.50	40.038	46	1.084	0.487	1.571
1952Einaudi	131	8	40	6.00	40.518	46	1.096	0.479	1.575
1953Einaudi	190	9	47	7.50	46.852	54	1.100	0.519	1.619
1954Einaudi	260	9	57	8.43	57.161	64	0.920	0.424	1.344
1955Gronchi	388	9	83	8.62	83.114	90	0.904	0.350	1.254
1956Gronchi	465	8	155	7.22	154.591	161	1.030	0.297	1.327
1957Gronchi	1090	10	263	8.87	262.656	271	1.060	0.253	1.313
1958Gronchi	886	10	201	8.85	200.544	209	1.077	0.284	1.361
1959Gronchi	697	9	181	8.72	180.644	188	0.953	0.279	1.232
1960Gronchi	794	10	195	8.80	194.686	203	1.066	0.295	1.361
1961Gronchi	1243	9	303	8.87	302.206	310	0.990	0.221	1.211
1962Segni	514	8	168	7.12	167.217	174	1.108	0.299	1.407
1963Segni	987	10	250	8.94	249.270	258	1.099	0.278	1.377
1964Saragat	447	9	100	8.47	99.850	107	0.957	0.338	1.295
1965Saragat	1033	10	265	8.87	264.875	273	1.032	0.253	1.285
1966Saragat	1189	10	323	8.89	322.316	331	1.101	0.252	1.353
1967Saragat	1045	11	262	9.33	261.673	271	1.120	0.289	1.409

1968Saragat	1164	10	303	8.74	302.256	311	1.130	0.256	1.386
1969Saragat	1573	11	393	8.97	392.744	402	1.161	0.233	1.394
1970Saragat	1918	11	489	9.54	488.701	498	1.089	0.212	1.301
1971Leone	252	10	69	7.50	69.194	77	1.201	0.492	1.693
1972Leone	747	10	180	8.70	180.389	188	0.988	0.278	1.266
1973Leone	1250	10	298	8.87	298.202	306	0.991	0.221	1.212
1974Leone	801	10	197	8.77	197.467	205	0.970	0.266	1.236
1975Leone	1319	9	311	8.88	310.214	318	0.988	0.214	1.202
1976Leone	1366	10	321	8.92	320.775	329	1.039	0.223	1.262
1977Leone	1594	10	357	8.94	356.658	365	1.051	0.209	1.260
1978Pertini	1403	10	323	8.91	322.278	331	1.103	0.233	1.336
1979Pertini	2291	11	498	8.84	498.182	507	1.125	0.184	1.309
1980Pertini	1349	11	315	9.00	314.757	324	1.155	0.252	1.407
1981Pertini	2818	11	571	10.29	571.239	580	0.943	0.165	1.108
1982Pertini	2476	11	508	10.44	507.190	517	1.039	0.197	1.236
1983Pertini	3726	11	784	10.68	783.110	793	1.022	0.162	1.184
1984Pertini	1180	10	286	8.84	285.493	294	1.085	0.248	1.333
1985Cossiga	2384	11	611	9.00	610.585	620	1.177	0.193	1.370
1986Cossiga	1349	10	321	8.90	321.344	329	0.969	0.208	1.177
1987Cossiga	1991	10	491	8.91	491.187	499	0.988	0.175	1.163
1988Cossiga	2345	10	553	9.10	552.170	561	1.090	0.182	1.272
1989Cossiga	1862	10	436	9.65	435.539	444	0.978	0.196	1.174
1990Cossiga	3177	10	783	9.50	782.123	791	1.044	0.157	1.201
1991Cossiga	408	10	94	8.68	93.794	102	1.069	0.406	1.475
1992Scalfaro	2761	11	655	8.96	654.916	664	1.141	0.173	1.314
1993Scalfaro	2941	11	684	9.86	683.221	693	1.104	0.180	1.284
1994Scalfaro	3605	11	866	10.17	865.299	875	1.058	0.162	1.220
1995Scalfaro	4206	11	992	10.50	991.120	1001	1.040	0.152	1.192
1996Scalfaro	1985	10	525	8.97	524.318	533	1.089	0.195	1.284
1997Scalfaro	4916	11	1104	10.58	1103.110	1113	1.032	0.141	1.173
1998Scalfaro	3951	11	968	10.47	967.292	977	1.025	0.154	1.179
1999Ciampi	1941	11	504	9.94	503.262	513	1.089	0.221	1.310
2000Ciampi	1734	11	421	9.25	420.201	430	1.188	0.235	1.423
2001Ciampi	2086	11	548	10.29	547.353	557	1.038	0.211	1.249
2002Ciampi	2069	10	550	8.94	549.312	558	1.094	0.191	1.285
2003Ciampi	1565	11	408	8.93	407.861	417	1.153	0.231	1.384
2004Ciampi	1737	10	448	9.25	447.371	456	1.046	0.207	1.253
2005Ciampi	1193	11	290	10.00	289.439	299	1.062	0.277	1.339
2006Napolitano	2204	11	502	9.25	501.400	511	1.164	0.204	1.368
2007Napolitano	1772	11	417	9.22	416.470	426	1.159	0.226	1.385

2008Napolitano	1713	11	409	8.94	408.835	418	1.154	0.221	1.375

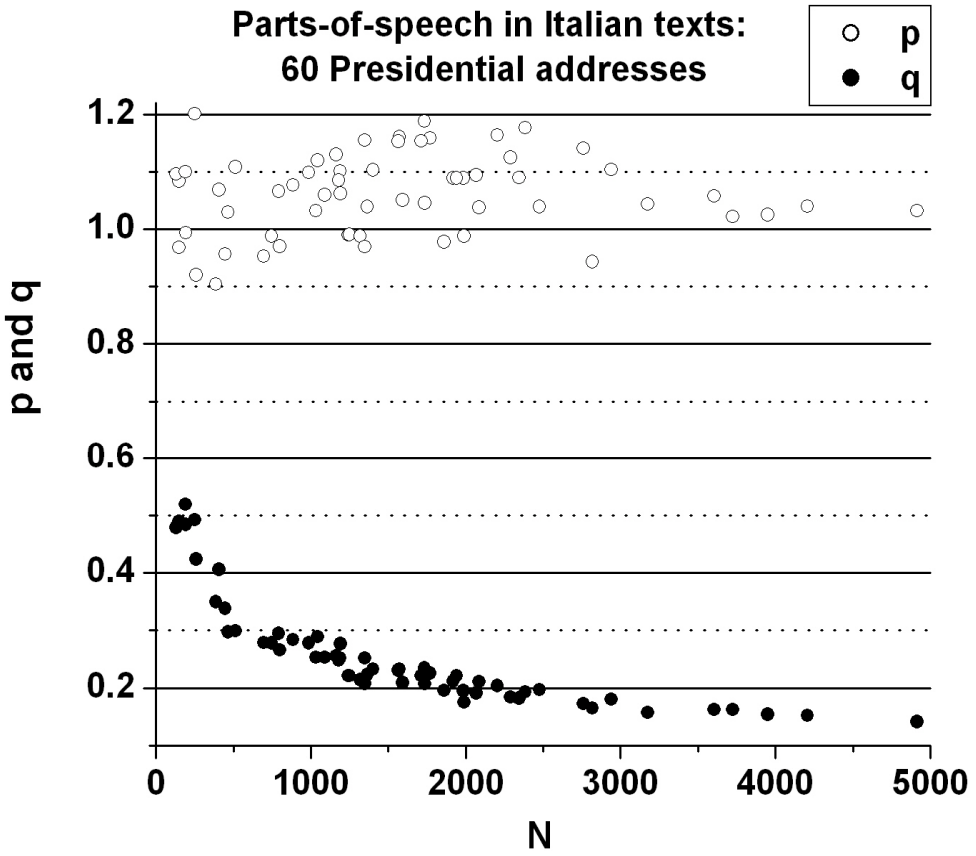

Figure 5.4. Indicators *p* and *q* for parts-of-speech in 60 Italian texts

In Figure 5.4, one can see that *p* is almost constant but *q* decreases with increasing text size. We can answer only the first question in very narrow circumstances: in the Italian Presidential speeches the rank-frequencies of parts-of-speech do not abide by the golden section, possibly because of the small inventory.

Since data from other languages are available (Schweers, Zhu 1991; Best 1994; Laufer, Nemcová 2009; Ziegler 1998, 2001; Sambor 1989), it is possible to compare them to the Italian data. In Table 5.5 where we present only the data and the *p, q* indicators (all computations can be found in Tuzzi, Popescu, Altmann 2009b) we see that the mean of $p+q$ for other languages is 1.1371. Since the standard deviation is here 0.0174, Italian or perhaps only the Presidential speeches lie outside of the 2-sigma interval. In any case, $p+q$ does not attain the golden section in the examined languages. Hence further analyses are necessary in order to state whether Italian has a special position.

The indicators *p* and *q* have their own regularities not depending on that of words or word-forms. *q* depends on the sample size but *p* could be considered a text characteristic within the given language.

Table 5.5
Characteristics of parts-of-speech diversification in other languages

Text	Parts-of-speech frequency	*N*	*I*	*p*	*q*	*p+q*
German 1	192, 161, 153, 112, 111, 104, 97, 70	448	8	0.8889	0.2835	1,1724
German 2	2032, 1939, 1532, 1338, 1179, 974, 914, 761	4589	8	1.0574	0.1033	1,1607
German SMS	2815, 2550, 2416,1606, 1459, 767, 541, 175	10937	8	1.0574	0.0669	1,1243
Portuguese	2586, 1607, 949, 819, 776, 680, 478, 440, 352	5528	9	1.0582	0.1076	1,1658
Brazilian Portuguese	2930, 2265, 1743, 1708, 1602, 1040, 936, 394	9474	8	1.0014	0.0719	1,0733
Latin	347, 173, 142, 98, 93, 59, 40, 39, 9	928	9	0.9044	0.2298	1,1342
Polish	144188, 79995, 71988, 56812, 33605, 31833, 21428, 18757, 8076, 650	460842	10	1.0000	0.0133	1,0133
Chinese	247, 228, 140, 133, 107, 81, 55, 27	810	8	1.0072	0.2460	1,2531

Since here we analyzed a special class of linguistic entities, a methodological remark is, perhaps, necessary. Whatever inventories, classes or entities of language we establish, they are conceptual constructs. They are neither true nor false and in many cases their establishment depends on the mother tongue and the erudition of the researcher. Linguistic entities are either prolific or non-prolific, depending on whether they can be used for descriptive (e.g. grammatical) or theoretical (e.g. use in hypotheses or laws) purposes. Their "existence" is an absolutely irrelevant problem. In the reality, there are no classes, there are only individuals. We unite them in classes in order to facilitate orientation, description, memorizing, etc. Natural language which represents "ethno-science" proceeds in the same way but there are no two languages that would subsume exactly the same meanings under exactly corresponding words. This is why a bilingual dictionary is not a list of one-to-one correspondences. Scientific language (meta-language) is a trial to establish classes for which laws could be stated, hence laws (or, at the beginning, empirical hypotheses) are the only criterion of the theoretical fruitfulness of linguistic classes. While for grammar that classification is prolific which allows us to form correct sentences in a high percentage of cases, i.e. enabling us to establish reliable rules, in theories we accept only those classifications that allow us to set up, test and systematize a hypothesis. Thus for language description a rule is the criterion of class fruitfulness, for language theory it is the existence of a law.

6. Vocabulary richness

Almost everybody concerned with text analysis tries to express the fuzzy concept of vocabulary richness (*VR*) and to capture this property of the texts. The number of operalizations is enormous (cf. Köhler 1995). Some of them are practicable, i.e. statistically treatable, other ones are simply some ratios which do not hold what they promise. A rather memento-like survey showing the hurdles of this type of indicators can be found in Wimmer, Altmann (1999).

Here we shall measure the vocabulary richness using a method proposed in Popescu, Altmann (2006) and used for 20 languages in Popescu et al. (2009: 29-34), in order to be able to compare the *VR* of Presidential speeches to both other Italian texts and texts from other languages. The greater will become the list of analyzed texts, the better we shall be able to tell something about this concept and estimate this property in texts. Needless to say, one could perform the examination differently but the comparability with 176 texts in 20 languages is very seducing.

For computing the indicator of *VR* we need three quantities: the text size *N*, the *h*-point as shown in Formula (3.1) of Chapter 3, and the cumulative relative frequencies of words up to the *h*-point, i.e. $F(h)$, the distribution function up to *h.* Then we set up the indicator

$$R_1 = 1 - \left(F(h) - \frac{(h-1)^2}{2N} \right). \quad (6.1)$$

For the given text *h* and *N* are constants, hence it is easy to derive the variance of (6.1) in form $Var(R_1) = F(h)[1 - F(h)]/N$ which can be used for setting up an asymptotic test of difference.

Our comparison will not be quite correct because in all other texts word-forms have been counted while in the Presidential addresses we reduced them to lemmas. But the difference will not be very aggravating because above the *h*-point mostly synsemantics occur having a simple form. Another problem is the fact that a word-form is not an unambiguous formation. In many languages there are controversies concerning their operational definitions. Even in Italian it is a decision problem whether "*l´università*" is one word-form or two words, since the beginning phonetic coalescence evokes a kind of fuzziness of boundaries. As emphasized on many places, language does not care for the crispness of classes, it is a work of grammarians.

The R_1 indicator for the Presidential speeches with all numbers necessary for its computation is presented in Table 6.1. The smallest R_1, 0.7173 can be found with Scalfaro 1997, the greatest one, 0.9134 with Einaudi 1952. Thus the range of values is rather small.

Table 6.1
Vocabulary Richness R_1 in 60 Presidential speeches

Text	N	V	h	F(h)	R_1
1949Einaudi	194	140	5.00	0.1753	0.8659
1950Einaudi	150	105	4.00	0.1533	0.8767
1951Einaudi	230	169	5.00	0.1478	0.8870
1952Einaudi	179	145	4.00	0.1117	0.9134
1953Einaudi	190	143	4.00	0.1368	0.8869
1954Einaudi	260	181	5.00	0.1677	0.8631
1955Gronchi	388	248	6.66	0.1932	0.8481
1956Gronchi	665	374	8.00	0.2060	0.8308
1957Gronchi	1130	549	12.00	0.2381	0.8154
1958Gronchi	886	460	11.00	0.2404	0.8160
1959Gronchi	697	388	9.00	0.2138	0.8321
1960Gronchi	804	434	10.00	0.2264	0.8240
1961Gronchi	1252	622	13.00	0.2428	0.8147
1962Segni	738	381	10.00	0.2358	0.8191
1963Segni	1057	527	11.66	0.2226	0.8312
1964Saragat	465	278	8.00	0.2086	0.8441
1965Saragat	1052	510	11.66	0.2511	0.8029
1966Saragat	1200	597	12.50	0.2425	0.8126
1967Saragat	1056	526	11.00	0.2367	0.8106
1968Saragat	1173	562	13.00	0.2515	0.8099
1969Saragat	1583	692	15.00	0.2798	0.7821
1970Saragat	1929	812	16.50	0.2809	0.7814
1971Leone	262	168	5.00	0.1450	0.8855
1972Leone	767	394	9.50	0.1909	0.8562
1973Leone	1250	616	12.00	0.2232	0.8252
1974Leone	801	426	9.00	0.1985	0.8415
1975Leone	1328	632	13.00	0.2538	0.8004
1976Leone	1366	649	13.00	0.2365	0.8162
1977Leone	1604	717	14.00	0.2618	0.7909
1978Pertini	1492	603	14.33	0.2495	0.8100
1979Pertini	2311	800	18.00	0.2787	0.7838
1980Pertini	1360	535	13.75	0.2747	0.7851
1981Pertini	2819	911	20.00	0.2778	0.7862
1982Pertini	2486	854	19.00	0.2836	0.7816
1983Pertini	3746	1149	23.66	0.3215	0.7470
1984Pertini	1340	514	13.66	0.2309	0.8289
1985Cossiga	2359	859	17.00	0.2866	0.7677
1986Cossiga	1348	561	14.00	0.2567	0.8060

1987Cossiga	2092	904	15.00	0.2715	0.7753
1988Cossiga	2384	875	19.00	0.3003	0.7677
1989Cossiga	1912	778	17.00	0.2819	0.7850
1990Cossiga	3345	1222	20.00	0.2996	0.7544
1991Cossiga	418	241	7.00	0.2081	0.8350
1992Scalfaro	2774	978	17.50	0.2987	0.7504
1993Scalfaro	2942	1074	18.60	0.3053	0.7473
1994Scalfaro	3606	1190	21.00	0.2967	0.7588
1995Scalfaro	4233	1341	22.66	0.3118	0.7436
1996Scalfaro	2085	866	16.00	0.2691	0.7849
1997Scalfaro	5012	1405	27.50	0.3528	0.7173
1998Scalfaro	3995	1175	23.50	0.3236	0.7398
1999Ciampi	1941	831	16.50	0.2612	0.8007
2000Ciampi	1844	822	16.00	0.2533	0.8077
2001Ciampi	2098	898	18.00	0.2755	0.7934
2002Ciampi	2129	909	17.00	0.2663	0.7938
2003Ciampi	1565	718	14.00	0.2409	0.8131
2004Ciampi	1807	812	15.00	0.2385	0.8157
2005Ciampi	1193	538	12.66	0.2354	0.8216
2006Napolitano	2204	929	16.50	0.2731	0.7814
2007Napolitano	1792	793	16.00	0.2751	0.7877
2008Napolitano	1713	775	15.00	0.2755	0.7817

It can easily be seen that *VR* is not a function of succession of Presidents. Taking averages of individual Presidents we obtain the results in Table 6.2.

Table 6.2
Italian presidential speeches ranked by decreasing mean VR

President	**mean VR**	**stdev of R_1**
Einaudi	0.8822	0.0183
Leone	0.8308	0.0330
Gronchi	0.8259	0.0122
Segni	0.8251	0.0085
Ciampi	0.8066	0.0110
Saragat	0.8062	0.0213
Pertini	0.7890	0.0255
Cossiga	0.7844	0.0276
Napolitano	0.7836	0.0035
Scalfaro	0.7489	0.0204

In order to test whether there is a difference between individual speeches, e.g. that between the greatest *VR* (Einaudi 1952) and the smallest *VR* (Scalfaro 1997) we set up the criterion

$$u = \frac{|R_{11} - R_{12}|}{\sqrt{Var(R_{11}) + Var(R_{12})}}, \quad (6.2)$$

insert the values from Table 6.1 and obtain

$$u = \frac{|0.7173 - 0.9134|}{\sqrt{\frac{0.1117(1-0.1117)}{179} + \frac{0.3528(1-0.3528)}{5012}}} = 8.01$$

a value which is highly significant. Hence the *VR* in the Speeches is not homogeneous, there are differences.

If we want to test the difference between individual Presidents then we use the standard deviations of the means, i.e. $s/\sqrt{n}$ where n is the number of texts. Testing Einaudi against Scalfaro we obtain $s_{\bar{R},Einaudi} = 0.0183/\sqrt{6} = 0.0075$ and $s_{\bar{R},Scalfaro} = 0.0204/\sqrt{7} = 0.0077$, hence our criterion is

$$u = \frac{|\bar{R}_{11} - \bar{R}_{12}|}{\sqrt{Var(\bar{R}_{11}) + Var(\bar{R}_{12})}} \quad (6.3)$$

and inserting the computed values we obtain

$$u = \frac{|0.8822 - 0.7489|}{\sqrt{0.0075^2 + 0.0077^2}} = 12.40,$$

that is, there is a significant difference between the *VR* of these two Presidents. The causes of such differences are not to be sought in the erudition or eloquence of the Presidents but rather in their striving for attaining some aims. The solving of such problems is rather a task for sociologists or political scientists.

In order to get a picture of the variability of this *VR* indicator we present the data from 20 languages taken from Popescu et al. (2009), however, in the source we computed the indicator with h^2, here we recomputed it with $(h\text{-}1)^2$ because the origin of the rank-frequency sequence is 1, not 0. The results are given in Table 6.3. The identification of individual texts can be found in the reference.

Table 6.3
VR for 176 texts in 20 languages (from Popescu et al. 2009: 31 ff.)

(B = Bulgarian, Cz = Czech, E = English, G = German, H = Hungarian, Hw = Hawaiian, I = Italian, In = Indonesian, Kn = Kannada, Lk = Lakota, Lt = Latin, M = Maori, Mq = Marquesan, Mr = Marathi, R = Romanian, Rt = Rarotongan, Ru = Russian, Sl = Slovenian, Sm = Samoan, T = Tagalog)

Text ID	**N**	**h**	**F(h)**	**R_1**	**Text ID**	**N**	**h**	**F(h)**	**R_1**
B 01	761	10	0.2286	0.8246	Lt 05	1354	8	0.1041	0.9140
B 02	352	8	0.2358	0.8338	Lt 06	829	7	0.0953	0.9264
B 03	515	9	0.2039	0.8582	M 01	2062	18	0.4617	0.6084
B 04	483	8	0.1988	0.8519	M 02	1175	14	0.4562	0.6157
B 05	406	7	0.2118	0.8325	M 03	1434	17	0.4888	0.6005
B 06	687	9	0.2344	0.8122	M 04	1289	15	0.4523	0.6237
B 07	557	8	0.1759	0.8681	M 05	3620	26	0.5221	0.5642
B 08	268	6	0.1716	0.8750	Mq 01	2330	22	0.5571	0.5375
B 09	550	9	0.2273	0.8309	Mq 02	457	10	0.4289	0.6597
B 10	556	7	0.2284	0.8040	Mq 03	1509	14	0.4884	0.5676
Cz 01	1044	9	0.1753	0.8554	Mr 001	2998	14	0.1301	0.8981
Cz 02	984	11	0.2124	0.8384	Mr 002	2922	18	0.1886	0.8609
Cz 03	2858	19	0.2988	0.7579	Mr 003	4140	20	0.1452	0.8984
Cz 04	522	7	0.1513	0.8832	Mr 004	6304	24	0.1950	0.8470
Cz 05	999	9	0.2262	0.8058	Mr 005	4957	19	0.1434	0.8893
Cz 06	1612	13	0.2593	0.7854	Mr 006	3735	19	0.1590	0.8844
Cz 07	2014	15	0.2895	0.7592	Mr 007	3162	16	0.1556	0.8800
Cz 08	677	8	0.1876	0.8486	Mr 008	5477	27	0.2657	0.7960
Cz 09	460	6	0.1957	0.8315	Mr 009	6206	26	0.1640	0.8864
Cz 10	1156	11	0.2137	0.8296	Mr 010	5394	27	0.2469	0.8158
E 01	2330	16	0.3004	0.7479	Mr 015	4693	21	0.1862	0.8564
E 02	2971	22	0.3618	0.7124	Mr 016	3642	18	0.1557	0.8840
E 03	3247	19	0.3745	0.6754	Mr 017	4170	19	0.1530	0.8858
E 04	4622	23	0.3674	0.6850	Mr 018	4062	20	0.1839	0.8605
E 05	4760	26	0.3683	0.6974	Mr 020	3943	19	0.1643	0.8768
E 06	4862	24	0.4093	0.6451	Mr 021	3846	20	0.1508	0.8961
E 07	5004	25	0.3675	0.6901	Mr 022	4099	21	0.2140	0.8348
E 08	5083	26	0.4552	0.6063	Mr 023	4142	20	0.1743	0.8693
E 09	5701	29	0.3898	0.6790	Mr 024	4255	20	0.1753	0.8671
E 10	6246	28	0.4092	0.6492	Mr 026	4146	19	0.1643	0.8748
E 11	8193	32	0.4129	0.6457	Mr 027	4128	21	0.1999	0.8485
E 12	9088	39	0.4547	0.6247	Mr 028	5191	23	0.1761	0.8705

E 13	11265	41	0.4663	0.6047	Mr 029	3424	17	0.1618	0.8756
G 01	1095	12	0.2749	0.7804	Mr 030	5504	20	0.1428	0.8900
G 02	845	9	0.2249	0.8130	Mr 031	5105	21	0.1589	0.8803
G 03	500	8	0.2400	0.8090	Mr 032	5195	23	0.1883	0.8583
G 04	545	8	0.2367	0.8083	Mr 033	4339	19	0.1348	0.9025
G 05	559	8	0.2272	0.8166	Mr 034	3489	17	0.1253	0.9114
G 06	545	8	0.1890	0.8560	Mr 035	1862	11	0.0956	0.9313
G 07	263	5	0.1711	0.8593	Mr 036	4205	19	0.1620	0.8765
G 08	965	11	0.2238	0.8280	Mr 038	4078	20	0.1721	0.8722
G 09	653	9	0.2129	0.8361	Mr 040	5218	21	0.1479	0.8904
G 10	480	7	0.1813	0.8562	Mr 043	3356	16	0.1159	0.9176
G 11	468	7	0.1624	0.8761	Mr 046	4186	20	0.1732	0.8699
G 12	251	6	0.1992	0.8506	Mr 052	3549	17	0.1677	0.8684
G 13	460	8	0.2000	0.8533	Mr 149	2946	12	0.0995	0.9210
G 14	184	5	0.1902	0.8533	Mr 150	3372	16	0.1260	0.9074
G 15	593	8	0.1568	0.8845	Mr 151	4843	23	0.2125	0.8375
G 16	518	8	0.1564	0.8909	Mr 154	3601	17	0.1402	0.8953
G 17	225	6	0.2133	0.8423	Mr 288	4060	17	0.1490	0.8825
H 01	2044	12	0.2495	0.7801	Mr 289	4831	19	0.1519	0.8816
H 02	1288	8	0.2073	0.8117	Mr 290	4025	17	0.1150	0.9168
H 03	403	4	0.1886	0.8226	Mr 291	3954	18	0.1629	0.8736
H 04	936	7	0.2041	0.8151	Mr 292	4765	19	0.1624	0.8716
H 05	413	6	0.2034	0.8269	Mr 293	3337	13	0.0869	0.9347
Hw 01	282	7	0.3298	0.7340	Mr 294	3825	17	0.1545	0.8790
Hw 02	1829	21	0.5489	0.5604	Mr 295	4895	20	0.1544	0.8825
Hw 03	3507	26	0.5441	0.5450	Mr 296	3836	18	0.1689	0.8688
Hw 04	7892	38	0.6405	0.4462	Mr 297	4605	18	0.1581	0.8733
Hw 05	7620	38	0.6382	0.4516	R 01	1738	14	0.2267	0.8219
Hw 06	12356	44	0.6475	0.4273	R 02	2279	16	0.2519	0.7975
I 01	11760	37	0.3423	0.7128	R 03	1264	12	0.2057	0.8422
I 02	6064	25	0.3130	0.7345	R 04	1284	10	0.1713	0.8602
I 03	854	10	0.2365	0.8109	R 05	1032	11	0.2141	0.8343
I 04	3258	21	0.3011	0.7603	R 06	695	10	0.2086	0.8497
I 05	1129	12	0.2524	0.8012	Rt 01	968	14	0.5072	0.5801
In 01	376	6	0.1729	0.8603	Rt 02	845	13	0.4769	0.6083
In 02	373	7	0.2038	0.8445	Rt 03	892	13	0.4294	0.6513
In 03	347	6	0.1643	0.8717	Rt 04	625	11	0.4224	0.6576
In 04	343	5	0.1137	0.9096	Rt 05	1059	15	0.4448	0.6477
In 05	414	8	0.2029	0.8563	Ru 01	753	8	0.1939	0.8386
Kn 003	3188	13	0.0994	0.9232	Ru 02	2595	16	0.2405	0.8029

Kn 004	1050	7	0.0819	0.9352	Ru 03	3853	21	0.2577	0.7942
Kn 005	4869	16	0.1327	0.8904	Ru 04	6025	25	0.2574	0.7904
Kn 006	5231	20	0.1357	0.8988	Ru 05	17205	41	0.2992	0.7473
Kn 011	4541	17	0.1132	0.9150	Sl 01	756	9	0.2037	0.8386
Kn 012	4141	19	0.1384	0.9007	Sl 02	1371	13	0.2392	0.8133
Kn 013	1302	10	0.1421	0.8890	Sl 03	1966	13	0.2477	0.7889
Kn 016	4735	18	0.1271	0.9034	Sl 04	3491	21	0.3861	0.6712
Kn 017	4316	18	0.1383	0.8952	Sl 05	5588	25	0.2770	0.7745
Lk 01	345	8	0.2870	0.7840	Sm 01	1487	17	0.5629	0.5232
Lk 02	1633	17	0.4489	0.6295	Sm 02	1171	15	0.5115	0.5722
Lk 03	809	12	0.4054	0.6694	Sm 03	617	13	0.5413	0.5754
Lk 04	219	6	0.2785	0.7786	Sm 04	736	12	0.5041	0.5781
Lt 01	3311	12	0.1130	0.9053	Sm 05	447	11	0.4832	0.6287
Lt 02	4010	18	0.1686	0.8674	T 01	1551	14	0.3972	0.6573
Lt 03	4931	19	0.1373	0.8956	T 02	1827	15	0.3969	0.6567
Lt 04	4285	20	0.1998	0.8423	T 03	2054	19	0.4747	0.6042

Taking means for individual languages as presented in Table 6.4 we see that vocabulary richness is not only matter of style or individuality but also of genre and language itself. The strongly analytic languages – even English – display a relatively low *VR* while strongly synthetic languages a rather high *VR*. This is caused by the plenty of word-forms proper to inflectional languages. Thus *VR* presented as above can be compared only within one language or lemmatization must be performed in both compared languages.

As can be seen, only Scalfaro has a lower *VR* than Italian literary texts (means of Pellico, Manzoni, Leopardi, Deledda, de Amicis), all other presidents have higher than average *VR*.

A comparison with Latin shows that its two daughter languages, Romanian and Italian, lost a great part of Latin synthetism.

More complicated would be the study of genres but we expect that the above approach, if developed from different views, could be helpful to solve some further problems of textology, too.

Table 6.4
Average VR in 20 languages

Language	**R_1**
Hawaiian	0.5274
Samoan	0.5755
Marquesan	0.5883
Maori	0.6025
Rarotongan	0.6290
Tagalog	0.6394
English	0.6664
Lakota	0.7154
Italian	0.7639
Slovenian	0.7773
Russian	0.7947
Hungarian	0.8113
Czech	0.8195
Romanian	0.8343
Bulgarian	0.8391
German	0.8420
Indonesian	0.8685
Latin	0.8891
Marathi	0.9264
Kannada	0.9057

7. Vector comparisons

The basic idea of this method is the direct comparison of the three conspicuous points controlling the shape of any rank-frequency distribution. These are the vocabulary V, the highest frequency f_1, and the h-point h. The most convenient way is to consider these real numbers as Cartesian components of a three-dimensional vector $\mathbf{T}(x, y, z)$ of a text and to compare any two texts by the angle τ_{12} between the corresponding vectors, as given by the well known formula

$$(7.1) \quad \cos\tau_{12} = \frac{\mathbf{T}_1 \cdot \mathbf{T}_2}{|\mathbf{T}_1||\mathbf{T}_2|} = \frac{(x_1x_2 + y_1y_2 + z_1z_2)}{\sqrt{x_1^2 + y_1^2 + z_1^2}\sqrt{x_2^2 + y_2^2 + z_2^2}}$$

Inasmuch as the considered coordinates (V, f_1, h) are always positive, the interval of τ is $(0, \pi/2)$. Now, it is quite natural to define as similar those rank-frequencies whose corresponding vectors are collinear ($\tau = 0$) or, in other words, whose vectors are obtained from each other by a re-scaling (scalar multiplication), say $\mathbf{T}_2 = c\mathbf{T}_1$ or $x_2 = cx_1$, $y_2 = cy_1$, $z_2 = cz_1$, where c is a real number. Geometrically, the shapes of these re-scaled distributions look fully similar and, for $c = 1$, they become identical. Obviously, if the τ-angle between the considered vectors increases, the cosine of τ (Figure 7.1) and the corresponding similarity degree of the distributions decrease. Therefore, in the following we shall use the τ-angle as a measure of *dissimilarity* of the compared rank-frequencies.

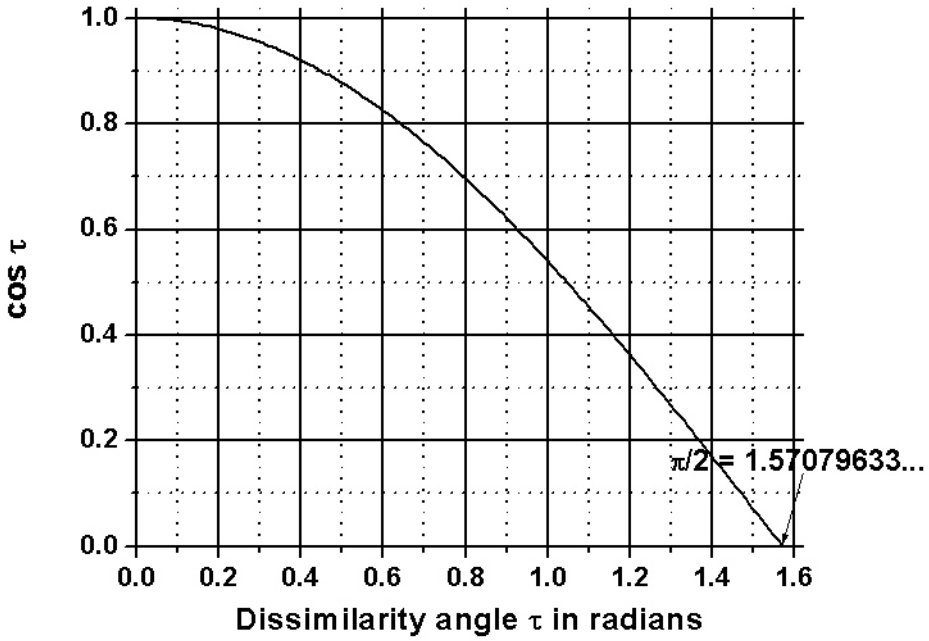

Figure 7.1. The cosine in the τ-angle interval $(0, \pi/2)$

Perhaps the most direct visualization of the τ-angle method is to consider a Cartesian coordinate system O*xyz* having *V* on O*x*, f_1 on O*y*, and *h* on O*z*. In this way, each text vector **T**(*x*, *y*, *z*) is represented by a point in this space. Let's now consider two such points, $\mathbf{T}_1(x_1, y_1, z_1)$ and $\mathbf{T}_2(x_2, y_2, z_2)$ and their corresponding straight lines, O$\mathbf{T}_1$ and O$\mathbf{T}_2$, connecting them to the origin O. The angle τ subtended by these lines is given by formula (7.1). Figure 7.2 simplifies the above presentation for the plane Oxy. The larger is the τ-angle, the smaller is the projection of the considered vectors on each other. However, it is clear that the upper limit of $\tau = \pi/2$ of complete dissimilarity ("orthogonal texts") is linguistically intangible because the coordinates (V, f_1, h) are always positive.

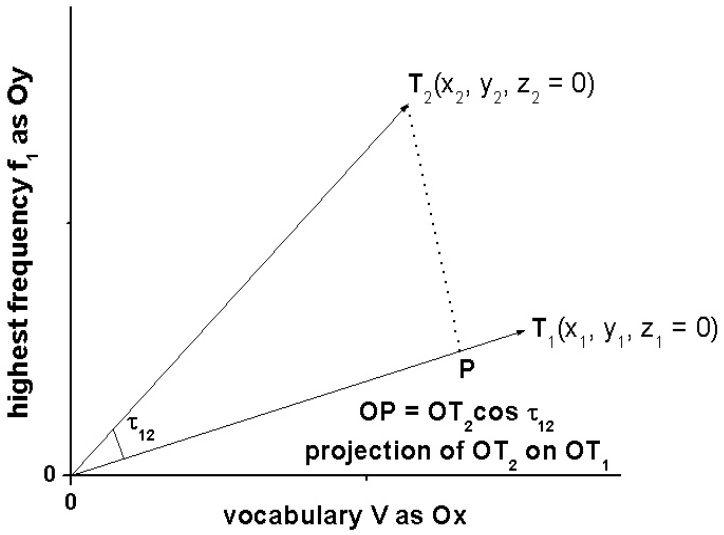

Figure 7.2. Vector comparison of rank-frequencies

For the sake of illustration consider the first two speeches by Einaudi in which we find:

	V	f_1	*h*
Einaudi 1949	140	10	5
Einaudi 1950	105	9	4

from which

$$\cos\tau = \frac{140(105)+10(9)+5(4)}{\sqrt{140^2+10^2+5^2}\sqrt{105^2+9^2+4^4}} = 0.9998968$$

$$\tau = \arccos(0.9998968) = 0.0143668\,.$$

Since $\tau \in <0,\pi/2>$, it would be possible to relativize or to normalize the dissimilarity in different ways in order to obtain greater numbers – the dissimilarities among the presidential speeches are in the interval <0.0003, 0.0950> – but this can be done in different ways and a testing procedure is not yet known, thus we leave it to future research. The normalized dissimilarity could be simply transformed in similarity by taking its complement, but this is, again, a further play with formulas.

Obviously, for the same purpose, we could consider the vector $\mathbf{Y}(V, f_1, L)$ as well, though the arc length L needs some more computation than the simple h-point. In principle, the advantage of L is that it follows strictly the shape of the rank-frequency distribution. However, and fortunately, the corresponding angle, υ, is practically proportional to τ, hence, the computation of one of them is sufficient. For convenience, in the following we shall use τ. It is to be emphasized that this kind of dissimilarity has nothing to do with content, semantics, etc. It is the dissimilarity of the shape of the rank-frequency function, of the organization of words in the texts, of their repetition structure.

An application of this simple method is given in Table 7.1 for texts of the same President and in Table 7.2 for texts of different Presidents.

Table 7.1
Vector dissimilarities within individual presidents

Einaudi	**1949**	**1950**	**1951**	**1952**	**1953**	**1954**	
1949	0	0.0144	0.0191	0.0244	0.0172	0.0095	
1950	0.0144	0	0.0334	0.0387	0.0313	0.0219	
1951	0.0191	0.0334	0	0.0053	0.0031	0.0131	
1952	0.0244	0.0387	0.0053	0	0.0077	0.0180	
1953	0.0172	0.0313	0.0031	0.0077	0	0.0103	
1954	0.0095	0.0219	0.0131	0.0180	0.0103	0	
Gronchi	**1955**	**1956**	**1957**	**1958**	**1959**	**1960**	**1961**
1955	0	0.0141	0.0536	0.0246	0.0207	0.0300	0.0433
1956	0.0141	0	0.0405	0.0118	0.0077	0.0169	0.0299
1957	0.0536	0.0405	0	0.0290	0.0330	0.0237	0.0106
1958	0.0246	0.0118	0.0290	0	0.0041	0.0054	0.0187
1959	0.0207	0.0077	0.0330	0.0041	0	0.0093	0.0226
1960	0.0300	0.0169	0.0237	0.0054	0.0093	0	0.0133
1961	0.0433	0.0299	0.0106	0.0187	0.0226	0.0133	0
Segni	**1962**	**1963**					
1962	0	0.0061					
1963	0.0061	0					

Saragat	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1964	0	0.0269	0.0080	0.0227	0.0246	0.0488	0.0301
1965	0.0269	0	0.0281	0.0053	0.0023	0.0221	0.0037
1966	0.0080	0.0281	0	0.0231	0.0258	0.0501	0.0307
1967	0.0227	0.0053	0.0231	0	0.0035	0.0270	0.0077
1968	0.0246	0.0023	0.0258	0.0035	0	0.0244	0.0057
1969	0.0488	0.0221	0.0501	0.0270	0.0244	0	0.0194
1970	0.0301	0.0037	0.0307	0.0077	0.0057	0.0194	0
Leone	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1971	0	0.0113	0.0384	0.0094	0.0295	0.0130	0.0411
1972	0.0113	0	0.0277	0.0067	0.0187	0.0042	0.0304
1973	0.0384	0.0277	0	0.0334	0.0091	0.0284	0.0028
1974	0.0094	0.0067	0.0334	0	0.0244	0.0051	0.0362
1975	0.0295	0.0187	0.0091	0.0244	0	0.0194	0.0118
1976	0.0130	0.0042	0.0284	0.0051	0.0194	0	0.0312
1977	0.0411	0.0304	0.0028	0.0362	0.0118	0.0312	0
Pertini	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1978	0	0.0013	0.0058	0.0174	0.0174	0.0150	0.0067
1979	0.0013	0	0.0067	0.0177	0.0177	0.0152	0.0070
1980	0.0058	0.0067	0	0.0124	0.0123	0.0105	0.0117
1981	0.0174	0.0177	0.0124	0	0.0003	0.0030	0.0239
1982	0.0174	0.0177	0.0123	0.0003	0	0.0031	0.0239
1983	0.0150	0.0152	0.0105	0.0030	0.0031	0	0.0217
1984	0.0067	0.0070	0.0117	0.0239	0.0239	0.0217	0
Cossiga	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1985	0	0.0218	0.0168	0.0037	0.0278	0.0109	0.0464
1986	0.0218	0	0.0095	0.0245	0.0072	0.0138	0.0247
1987	0.0168	0.0095	0	0.0203	0.0123	0.0062	0.0315
1988	0.0037	0.0245	0.0203	0	0.0308	0.0145	0.0492
1989	0.0278	0.0072	0.0123	0.0308	0	0.0182	0.0192
1990	0.0109	0.0138	0.0062	0.0145	0.0182	0	0.0373
1991	0.0464	0.0247	0.0315	0.0492	0.0192	0.0373	0
Scalfaro	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1992	0	0.0008	0.0226	0.0134	0.0188	0.0024	0.0045
1993	0.0008	0	0.0232	0.0139	0.0183	0.0026	0.0044
1994	0.0226	0.0232	0	0.0093	0.0415	0.0245	0.0268
1995	0.0134	0.0139	0.0093	0	0.0322	0.0154	0.0176
1996	0.0188	0.0183	0.0415	0.0322	0	0.0171	0.0149
1997	0.0024	0.0026	0.0245	0.0154	0.0171	0	0.0023

1998	0.0045	0.0044	0.0268	0.0176	0.0149	0.0023	0
Ciampi	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**
1999	0	0.0057	0.0195	0.0260	0.0083	0.0141	0.0211
2000	0.0057	0	0.0138	0.0203	0.0026	0.0084	0.0156
2001	0.0195	0.0138	0	0.0066	0.0113	0.0057	0.0037
2002	0.0260	0.0203	0.0066	0	0.0177	0.0119	0.0071
2003	0.0083	0.0026	0.0113	0.0177	0	0.0059	0.0131
2004	0.0141	0.0084	0.0057	0.0119	0.0059	0	0.0084
2005	0.0211	0.0156	0.0037	0.0071	0.0131	0.0084	0
Napolitano	**2006**	**2007**	**2008**				
2006	0	0.0075	0.0373				
2007	0.0075	0	0.0302				
2008	0.0373	0.0302	0				

The means of internal ("within") vector dissimilarities do not display a regular historical development. The extremes are represented by the so called "incomeplete" presidents, viz. Segni having 2 speeches and Napolitano having up to now 3 speeches. The mean τ radians are (historically) shown in Table 7.2.

Table 7.2
Mean internal ("within") dissimilarity

Einaudi	0.0178
Gronchi	0.0187
Segni	0.0061
Saragat	0.0186
Leone	0.0169
Pertini	0.0097
Cossiga	0.0163
Scalfaro	0.0138
Ciampi	0.0101
Napolitano	0.0250

However, if one omits the "incomplete" speech corpora (Segni, Napolitano) a clear decreasing trend appears, as can be seen in Figure 7.3, signalizing that the internal dissimilarity of the rank-frequency structure decreases; the texts of individual presidents become internally more uniform, a fact that can be interpreted only using the knowledge of the situation in the presidential bureaus. Nevertheless, the method yields a starting point for investigations of this kind.

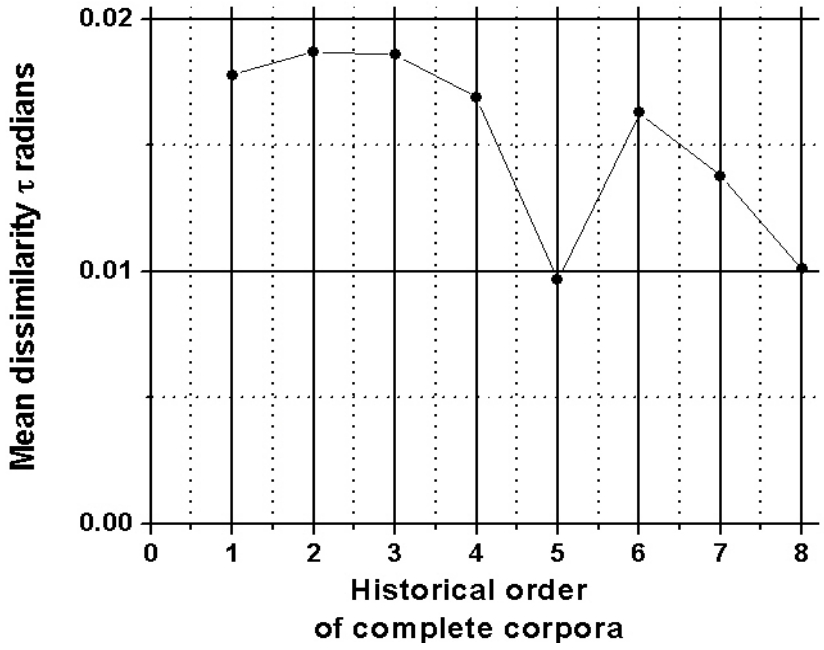

Figure 7.3. Mean internal ("within") vector dissimilarities of presidential speeches omitting the "incomplete" corpora (Segni, Napolitano)

In Table 7.3 we present the vector dissimilarities between individual presidents taking into account a comparison of "all with all".

Table 7.3
Vector dissimilarities of different presidents

Einaudi	**Gronchi**							**Segni**	
	1955	**1956**	**1957**	**1958**	**1959**	**1960**	**1961**	**1962**	**1963**
1949	0.0112	0.0155	0.0486	0.0212	0.0184	0.0261	0.0389	0.0224	0.0208
1950	0.0238	0.0185	0.0362	0.0145	0.0148	0.0173	0.0277	0.0133	0.0160
1951	0.0115	0.0255	0.0651	0.0361	0.0323	0.0415	0.0548	0.0385	0.0347
1952	0.0162	0.0298	0.0698	0.0408	0.0369	0.0462	0.0594	0.0434	0.0392
1953	0.0086	0.0225	0.0623	0.0333	0.0293	0.0386	0.0519	0.0358	0.0317
1954	0.0019	0.0128	0.0520	0.0230	0.0192	0.0284	0.0416	0.0254	0.0216

Einaudi	**Saragat**						
	1964	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1949	0.0080	0.0329	0.0149	0.0293	0.0307	0.0542	0.0364
1950	0.0137	0.0221	0.0208	0.0204	0.0203	0.0415	0.0258

1951	0.0222	0.0489	0.0221	0.0443	0.0466	0.0709	0.0519
1952	0.0272	0.0536	0.0262	0.0489	0.0513	0.0756	0.0565
1953	0.0195	0.0460	0.0190	0.0414	0.0437	0.0680	0.0490
1954	0.0093	0.0357	0.0099	0.0312	0.0334	0.0577	0.0388

Einaudi	**Leone**						
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1949	0.0059	0.0151	0.0404	0.0150	0.0319	0.0179	0.0430
1950	0.0164	0.0146	0.0294	0.0199	0.0223	0.0188	0.0316
1951	0.0181	0.0284	0.0561	0.0234	0.0470	0.0284	0.0588
1952	0.0232	0.033	0.0606	0.0275	0.0516	0.0326	0.0634
1953	0.0155	0.0254	0.0531	0.0203	0.0441	0.0253	0.0559
1954	0.0055	0.0152	0.0429	0.0109	0.0339	0.0157	0.0456

Einaudi	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1949	0.0202	0.0207	0.0241	0.0364	0.0363	0.0345	0.0137
1950	0.0144	0.0156	0.0145	0.0253	0.0251	0.0242	0.0121
1951	0.0349	0.0348	0.0402	0.0523	0.0523	0.0499	0.0285
1952	0.0396	0.0394	0.0450	0.0570	0.0570	0.0545	0.0333
1953	0.0321	0.0319	0.0374	0.0495	0.0494	0.0470	0.0257
1954	0.0218	0.0217	0.0270	0.0392	0.0392	0.0368	0.0154

Einaudi	**Cossiga**						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1949	0.0671	0.0453	0.0523	0.0698	0.0400	0.0582	0.0208
1950	0.0542	0.0326	0.0406	0.0566	0.0284	0.0461	0.0106
1951	0.0839	0.0623	0.0680	0.0868	0.0561	0.0741	0.0378
1952	0.0886	0.0671	0.0726	0.0916	0.0608	0.0787	0.0428
1953	0.0810	0.0595	0.0651	0.0840	0.0533	0.0712	0.0351
1954	0.0707	0.0492	0.0549	0.0737	0.0430	0.0610	0.0249

Einaudi	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1949	0.0519	0.0516	0.0737	0.0649	0.0346	0.0497	0.0474
1950	0.0400	0.0398	0.0607	0.0524	0.0251	0.0376	0.0355
1951	0.0679	0.0675	0.0903	0.0812	0.0493	0.0659	0.0636
1952	0.0725	0.0720	0.0950	0.0859	0.0538	0.0705	0.0682

1953	0.0650	0.0645	0.0874	0.0783	0.0464	0.0630	0.0607
1954	0.0547	0.0543	0.0772	0.0681	0.0362	0.0527	0.0504

Einaudi	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1949	0.0177	0.0212	0.0316	0.0379	0.0229	0.0279	0.0312	0.0650	0.0575	0.0300
1950	0.0192	0.0186	0.0224	0.0276	0.0186	0.0211	0.0205	0.0523	0.0449	0.0217
1951	0.0278	0.0333	0.0466	0.0531	0.0358	0.0416	0.0472	0.0814	0.0741	0.0445
1952	0.0320	0.0376	0.0511	0.0577	0.0401	0.0460	0.0519	0.0861	0.0788	0.0489
1953	0.0247	0.0303	0.0436	0.0502	0.0327	0.0386	0.0444	0.0785	0.0712	0.0415
1954	0.0152	0.0204	0.0335	0.0400	0.0228	0.0286	0.0341	0.0683	0.0609	0.0314

Gronchi	**Segni**		**Saragat**						
	1962	**1963**	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**
1955	0.0272	0.0231	0.0111	0.0374	0.0109	0.0328	0.0351	0.0594	0.0404
1956	0.0150	0.0097	0.0076	0.0243	0.0038	0.0193	0.0220	0.0462	0.0269
1957	0.0266	0.0308	0.0430	0.0163	0.0443	0.0212	0.0186	0.0058	0.0136
1958	0.0036	0.0025	0.0143	0.0128	0.0156	0.0083	0.0104	0.0348	0.0158
1959	0.0074	0.0025	0.0110	0.0168	0.0115	0.0120	0.0145	0.0388	0.0197
1960	0.0041	0.0072	0.0196	0.0074	0.0207	0.0033	0.0051	0.0295	0.0105
1961	0.0166	0.0203	0.0329	0.0060	0.0337	0.0106	0.0083	0.0164	0.0031

Gronchi	**Leone**						
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1955	0.0075	0.0168	0.0445	0.0120	0.0355	0.0170	0.0473
1956	0.0103	0.0045	0.0310	0.0024	0.0220	0.0029	0.0338
1957	0.0472	0.0369	0.0098	0.0429	0.0185	0.0379	0.0071
1958	0.0185	0.0079	0.0199	0.0142	0.0110	0.0097	0.0226
1959	0.0150	0.0039	0.0238	0.0101	0.0147	0.0058	0.0265
1960	0.0238	0.0132	0.0146	0.0193	0.0057	0.0145	0.0173
1961	0.0371	0.0265	0.0018	0.0323	0.0080	0.0274	0.0041

Gronchi	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1955	0.0234	0.0233	0.0288	0.0409	0.0408	0.0384	0.0171
1956	0.0105	0.0099	0.0164	0.0276	0.0276	0.0250	0.0066
1957	0.0302	0.0306	0.0250	0.0129	0.0129	0.0156	0.0366
1958	0.0012	0.0021	0.0046	0.0162	0.0162	0.0138	0.0078

1959	0.0029	0.0025	0.0087	0.0202	0.0202	0.0177	0.0047
1960	0.0066	0.0069	0.0028	0.0109	0.0108	0.0085	0.0131
1961	0.0198	0.0201	0.0149	0.0025	0.0027	0.0050	0.0264

Gronchi	**Cossiga**						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1955	0.0724	0.0509	0.0565	0.0754	0.0447	0.0626	0.0267
1956	0.0591	0.0381	0.0429	0.0623	0.0314	0.0490	0.0156
1957	0.0188	0.0040	0.0057	0.0218	0.0090	0.0100	0.0278
1958	0.0478	0.0265	0.0319	0.0508	0.0200	0.0380	0.0055
1959	0.0518	0.0305	0.0358	0.0548	0.0240	0.0419	0.0085
1960	0.0424	0.0212	0.0266	0.0455	0.0147	0.0327	0.0068
1961	0.0292	0.0090	0.0134	0.0324	0.0018	0.0194	0.0182

Gronchi	**Scalfaro**						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1955	0.0564	0.0559	0.0788	0.0697	0.0378	0.0544	0.0521
1956	0.0428	0.0423	0.0654	0.0562	0.0241	0.0410	0.0387
1957	0.0045	0.0048	0.0252	0.0163	0.0169	0.0023	0.0026
1958	0.0318	0.0313	0.0542	0.0451	0.0135	0.0297	0.0275
1959	0.0356	0.0352	0.0581	0.0490	0.0171	0.0337	0.0314
1960	0.0264	0.0260	0.0488	0.0397	0.0084	0.0244	0.0221
1961	0.0131	0.0127	0.0356	0.0264	0.0065	0.0111	0.0088

Gronchi	**Ciampi**							**Napolitano**		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1955	0.0164	0.0218	0.0350	0.0416	0.0242	0.0301	0.0358	0.0699	0.0626	0.0329
1956	0.0024	0.0078	0.0214	0.0280	0.0103	0.0162	0.0227	0.0565	0.0493	0.0192
1957	0.0386	0.0330	0.0191	0.0130	0.0304	0.0247	0.0179	0.0164	0.0090	0.0215
1958	0.0104	0.0059	0.0106	0.0171	0.0046	0.0070	0.0111	0.0453	0.0380	0.0088
1959	0.0065	0.0037	0.0143	0.0209	0.0046	0.0097	0.0152	0.0492	0.0419	0.0122
1960	0.0153	0.0099	0.0055	0.0119	0.0075	0.0046	0.0059	0.0399	0.0326	0.0043
1961	0.0281	0.0224	0.0086	0.0030	0.0198	0.0142	0.0077	0.0266	0.0194	0.0109

Segni								
	1964	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**	**Saragat**
1962	0.0164	0.0106	0.0188	0.0073	0.0083	0.0324	0.0140	
1963	0.0134	0.0146	0.0135	0.0097	0.0123	0.0366	0.0173	

	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**Leone**
1962	0.0206	0.0108	0.0180	0.0174	0.0096	0.0132	0.0206	
1963	0.0175	0.0063	0.0214	0.0121	0.0124	0.0074	0.0242	
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**Pertini**
1962	0.0046	0.0057	0.0017	0.0141	0.0140	0.0121	0.0101	
1963	0.0017	0.0004	0.0071	0.0179	0.0179	0.0153	0.0071	
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**Cossiga**
1962	0.0454	0.0238	0.0300	0.0483	0.0178	0.0359	0.0028	
1963	0.0495	0.0284	0.0334	0.0526	0.0218	0.0395	0.0079	
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**Scalfaro**
1962	0.0297	0.0293	0.0518	0.0429	0.0124	0.0275	0.0252	
1963	0.0333	0.0328	0.0558	0.0467	0.0147	0.0313	0.0291	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**Ciampi**
1962	0.0139	0.0095	0.0095	0.0156	0.0079	0.0079	0.0089	
1963	0.0081	0.0034	0.0119	0.0185	0.0027	0.0072	0.0130	
	2006	**2007**	**2008**					**Napolitano**
1962	0.0430	0.0356	0.0084					
1963	0.0469	0.0397	0.0098					

Saragat	**Leone**						
	1971	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**
1964	0.0042	0.0073	0.0342	0.0076	0.0253	0.0098	0.0369
1965	0.0311	0.0206	0.0075	0.0267	0.0032	0.0218	0.0101
1966	0.0091	0.0081	0.0348	0.0014	0.0258	0.0064	0.0376
1967	0.0268	0.0159	0.0118	0.0217	0.0027	0.0167	0.0145
1968	0.0288	0.0183	0.0097	0.0244	0.0025	0.0196	0.0123
1969	0.0530	0.0427	0.0154	0.0487	0.0243	0.0437	0.0127
1970	0.0343	0.0236	0.0041	0.0293	0.0050	0.0243	0.0069

Saragat	**Pertini**						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1964	0.0133	0.0134	0.0181	0.0304	0.0303	0.0282	0.0065
1965	0.0140	0.0143	0.0089	0.0035	0.0034	0.0024	0.0204
1966	0.0144	0.0138	0.0202	0.0314	0.0314	0.0288	0.0097
1967	0.0094	0.0095	0.0059	0.0084	0.0084	0.0057	0.0161
1968	0.0117	0.0121	0.0066	0.0058	0.0058	0.0039	0.0181
1969	0.0360	0.0364	0.0307	0.0187	0.0187	0.0213	0.0424
1970	0.0170	0.0172	0.0123	0.0018	0.0020	0.0020	0.0236

Saragat	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1964	0.0618	0.0401	0.0462	0.0646	0.0341	0.0523	0.0156
1965	0.0350	0.0139	0.0194	0.0381	0.0073	0.0254	0.0122
1966	0.0629	0.0420	0.0466	0.0661	0.0353	0.0528	0.0192
1967	0.0399	0.0191	0.0237	0.0430	0.0122	0.0299	0.0099
1968	0.0373	0.0161	0.0217	0.0404	0.0096	0.0277	0.0102
1969	0.0130	0.0089	0.0062	0.0160	0.0148	0.0058	0.0334
1970	0.0322	0.0120	0.0161	0.0354	0.0048	0.0222	0.0159

Saragat	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1964	0.0460	0.0456	0.0682	0.0592	0.0278	0.0439	0.0416
1965	0.0191	0.0188	0.0414	0.0324	0.0044	0.0170	0.0147
1966	0.0466	0.0461	0.0692	0.0600	0.0278	0.0448	0.0425
1967	0.0236	0.0232	0.0462	0.0370	0.0052	0.0217	0.0194
1968	0.0214	0.0210	0.0437	0.0347	0.0050	0.0193	0.0170
1969	0.0052	0.0060	0.0195	0.0109	0.0226	0.0057	0.0077
1970	0.0160	0.0155	0.0385	0.0293	0.0035	0.0140	0.0118

Saragat	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1964	0.0097	0.0133	0.0250	0.0315	0.0153	0.0207	0.0252	0.0594	0.0520	0.0231
1965	0.0225	0.0170	0.0040	0.0055	0.0145	0.0094	0.0017	0.0325	0.0252	0.0062
1966	0.0058	0.0115	0.0252	0.0317	0.0140	0.0199	0.0266	0.0603	0.0531	0.0230
1967	0.0174	0.0118	0.0023	0.0088	0.0092	0.0041	0.0043	0.0372	0.0300	0.0016
1968	0.0203	0.0148	0.0031	0.0074	0.0123	0.0076	0.0008	0.0348	0.0275	0.0047
1969	0.0444	0.0387	0.0249	0.0186	0.0362	0.0305	0.0237	0.0108	0.0034	0.0273
1970	0.0250	0.0194	0.0055	0.0019	0.0168	0.0111	0.0053	0.0296	0.0224	0.0079

Leone	Pertini						
	1978	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**
1971	0.0174	0.0175	0.0223	0.0346	0.0345	0.0324	0.0107
1972	0.0066	0.0064	0.0122	0.0240	0.0240	0.0216	0.0025
1973	0.0211	0.0213	0.0164	0.0042	0.0043	0.0061	0.0277
1974	0.0130	0.0124	0.0188	0.0300	0.0300	0.0274	0.0085
1975	0.0121	0.0122	0.0080	0.0058	0.0059	0.0030	0.0188
1976	0.0086	0.0077	0.0144	0.0251	0.0251	0.0224	0.0067
1977	0.0238	0.0240	0.0190	0.0066	0.0067	0.0088	0.0304

Leone	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1971	0.0660	0.0443	0.0504	0.0688	0.0383	0.0565	0.0197
1972	0.0556	0.0343	0.0397	0.0587	0.0279	0.0458	0.0111
1973	0.0282	0.0089	0.0120	0.0314	0.0024	0.0181	0.0198
1974	0.0615	0.0405	0.0452	0.0647	0.0338	0.0514	0.0179
1975	0.0372	0.0166	0.0210	0.0403	0.0095	0.0271	0.0119
1976	0.0566	0.0357	0.0402	0.0597	0.0289	0.0464	0.0142
1977	0.0254	0.0068	0.0093	0.0286	0.0033	0.0154	0.0222

Leone	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1971	0.0502	0.0498	0.0724	0.0634	0.0320	0.0481	0.0458
1972	0.0395	0.0391	0.0620	0.0529	0.0210	0.0375	0.0353
1973	0.0118	0.0114	0.0344	0.0252	0.0071	0.0100	0.0077
1974	0.0452	0.0447	0.0678	0.0586	0.0264	0.0434	0.0411
1975	0.0209	0.0204	0.0435	0.0343	0.0028	0.0190	0.0167
1976	0.0402	0.0397	0.0628	0.0536	0.0214	0.0383	0.0361
1977	0.0091	0.0087	0.0317	0.0225	0.0099	0.0072	0.0050

Leone	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1971	0.0127	0.0171	0.0291	0.0357	0.0192	0.0247	0.0294	0.0636	0.0562	0.0272
1972	0.0046	0.0061	0.0182	0.0248	0.0079	0.0135	0.0190	0.0531	0.0458	0.0161
1973	0.0291	0.0234	0.0096	0.0032	0.0208	0.0150	0.0092	0.0255	0.0183	0.0119
1974	0.0045	0.0101	0.0238	0.0303	0.0126	0.0185	0.0252	0.0589	0.0517	0.0216
1975	0.0201	0.0144	0.0008	0.0062	0.0119	0.0064	0.0030	0.0345	0.0273	0.0031
1976	0.0007	0.0050	0.0188	0.0253	0.0076	0.0135	0.0204	0.0538	0.0467	0.0165
1977	0.0319	0.0262	0.0123	0.0059	0.0236	0.0178	0.0118	0.0227	0.0156	0.0146

Pertini	Cossiga						
	1985	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**
1978	0.0490	0.0277	0.0331	0.0520	0.0212	0.0392	0.0062
1979	0.0493	0.0282	0.0332	0.0524	0.0216	0.0394	0.0075
1980	0.0437	0.0222	0.0283	0.0466	0.0161	0.0343	0.0040
1981	0.0316	0.0108	0.0159	0.0347	0.0038	0.0219	0.0156
1982	0.0316	0.0107	0.0160	0.0347	0.0038	0.0220	0.0155
1983	0.0342	0.0137	0.0181	0.0373	0.0066	0.0242	0.0141
1984	0.0554	0.0339	0.0397	0.0583	0.0277	0.0458	0.0098

Pertini	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1978	0.0329	0.0325	0.0554	0.0463	0.0146	0.0309	0.0287
1979	0.0331	0.0327	0.0557	0.0465	0.0146	0.0312	0.0289
1980	0.0280	0.0276	0.0502	0.0412	0.0108	0.0259	0.0236
1981	0.0156	0.0153	0.0380	0.0289	0.0051	0.0135	0.0113
1982	0.0157	0.0154	0.0380	0.0289	0.0053	0.0136	0.0113
1983	0.0179	0.0175	0.0405	0.0313	0.0024	0.0160	0.0137
1984	0.0395	0.0391	0.0618	0.0528	0.0213	0.0374	0.0352

Pertini	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1978	0.0093	0.0051	0.0117	0.0183	0.0042	0.0077	0.0124	0.0465	0.0392	0.0098
1979	0.0084	0.0038	0.0118	0.0183	0.0030	0.0073	0.0128	0.0467	0.0395	0.0097
1980	0.0151	0.0103	0.0079	0.0139	0.0084	0.0072	0.0072	0.0413	0.0339	0.0071
1981	0.0258	0.0202	0.0065	0.0032	0.0176	0.0122	0.0052	0.0291	0.0218	0.0089
1982	0.0258	0.0202	0.0066	0.0035	0.0177	0.0123	0.0051	0.0291	0.0218	0.0090
1983	0.0231	0.0174	0.0036	0.0034	0.0149	0.0093	0.0037	0.0315	0.0243	0.0060
1984	0.0071	0.0079	0.0184	0.0250	0.0093	0.0143	0.0188	0.0530	0.0456	0.0166

Cossiga	Scalfaro						
	1992	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**
1985	0.0165	0.0171	0.0066	0.0042	0.0353	0.0182	0.0204
1986	0.0085	0.0087	0.0283	0.0198	0.0155	0.0061	0.0050
1987	0.0013	0.0009	0.0227	0.0134	0.0188	0.0034	0.0052
1988	0.0199	0.0206	0.0051	0.0078	0.0385	0.0214	0.0236
1989	0.0119	0.0116	0.0342	0.0251	0.0083	0.0098	0.0075
1990	0.0063	0.0067	0.0166	0.0073	0.0250	0.0085	0.0107
1991	0.0311	0.0308	0.0529	0.0441	0.0147	0.0289	0.0266

Cossiga	Ciampi							Napolitano		
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**
1985	0.0572	0.0516	0.0377	0.0313	0.0490	0.0432	0.0367	0.0034	0.0098	0.0400
1986	0.0364	0.0309	0.0173	0.0119	0.0283	0.0229	0.0154	0.0197	0.0123	0.0197
1987	0.0409	0.0352	0.0215	0.0149	0.0326	0.0267	0.0211	0.0138	0.0076	0.0237
1988	0.0604	0.0547	0.0409	0.0346	0.0522	0.0464	0.0397	0.0071	0.0131	0.0432
1989	0.0296	0.0240	0.0102	0.0048	0.0214	0.0158	0.0089	0.0253	0.0179	0.0126
1990	0.0470	0.0413	0.0276	0.0211	0.0388	0.0329	0.0271	0.0077	0.0038	0.0298
1991	0.0149	0.0113	0.0119	0.0175	0.0101	0.0108	0.0106	0.0442	0.0367	0.0111

Scalfaro	Ciampi							Napolitano		
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
1992	0.0409	0.0352	0.0214	0.0149	0.0326	0.0268	0.0208	0.0137	0.0070	0.0236
1993	0.0404	0.0346	0.0209	0.0144	0.0321	0.0262	0.0205	0.0142	0.0077	0.0232
1994	0.0635	0.0578	0.0440	0.0375	0.0552	0.0494	0.0431	0.0090	0.0162	0.0463
1995	0.0543	0.0485	0.0348	0.0283	0.0460	0.0401	0.0340	0.0009	0.0075	0.0370
1996	0.0221	0.0163	0.0029	0.0040	0.0138	0.0079	0.0052	0.0325	0.0255	0.0049
1997	0.0390	0.0333	0.0195	0.0131	0.0308	0.0250	0.0187	0.0155	0.0084	0.0218
1998	0.0368	0.0311	0.0173	0.0109	0.0285	0.0228	0.0164	0.0178	0.0106	0.0196

Ciampi	Napolitano		
	2006	2007	2008
1999	0.0545	0.0474	0.0172
2000	0.0488	0.0417	0.0115
2001	0.0350	0.0279	0.0024
2002	0.0285	0.0215	0.0088
2003	0.0463	0.0392	0.0090
2004	0.0404	0.0334	0.0033
2005	0.0342	0.0269	0.0055

To make this large table more lucid, one can compute the mean vector dissimilarities between individual presidents. In this way one obtains the results in Table 7.4.

Table 7.4
Mean vector dissimilarities between presidents

Gronchi	Segni	Saragat	Leone	Pertini	Cossiga	Scalfaro	Ciampi	Napolitano	
0.0316	0.0286	0.0362	0.0306	0.0336	0.0576	0.0602	0.0333	0.0576	**Einaudi**
	0.0140	0.0201	0.0191	0.0161	0.0320	0.0325	0.0163	0.0370	**Gronchi**
		0.0161	0.0151	0.0093	0.0312	0.0330	0.0098	0.0306	**Segni**
			0.0197	0.0156	0.0279	0.0278	0.0158	0.0318	**Saragat**
				0.0164	0.0328	0.0032	0.0159	0.0380	**Leone**
					0.0274	0.0280	0.0115	0.0317	**Pertini**
						0.0170	0.0292	0.0224	**Cossiga**
							0.0293	0.0202	**Scalfaro**
								0.0324	**Ciampi**

It allows us to study the "leaning" of a president on his predecessors, the "influence" on his successors, and if we take the *mean of these means*, we obtain the position of the given president within his colleagues. Using Table 7.4 we obtain the following means of means in Table 7.5 showing that the first president was the most dissimilar; later on a tradition arose.

Table 7.5
The mean of mean vector dissimilarities

President	**Mean of means**
Einaudi	0.0410
Gronchi	0.0243
Segni	0.0209
Saragat	0.0234
Leone	0.0213
Pertini	0.0211
Cossiga	0.0308
Scalfaro	0.0279
Ciampi	0.0215
Napolitano	0.0336

If we take into account only the adjacent speeches, in order to discover some pattern, we obtain the results in Table 7.6 which, being somewhat narrower, obtains also the basic numbers from which we compute cos τ and the τ radians. The sequence of these numbers can be seen in Figure 7.4.

Table 7.6
Angles between the adjacent text vectors **T**(V, f_1, h)

Text				**adjacent vectors**	
ID	***V***	***f₁***	***h***	**cosτ**	**τ rad**
1949Einaudi	140	10	5		
1950Einaudi	105	9	4	0.9999	0.0144
1951Einaudi	169	9	5	0.9994	0.0334
1952Einaudi	145	7	4	1.0000	0.0053
1953Einaudi	143	8	4	1.0000	0.0077
1954Einaudi	181	12	5	0.9999	0.0103
1955Gronchi	248	16	7	1.0000	0.0019
1956Gronchi	374	29	8	0.9999	0.0141

1957Gronchi	549	65	12	0.9992	0.0405
1958Gronchi	460	41	11	0.9996	0.0290
1959Gronchi	388	33	9	1.0000	0.0041
1960Gronchi	434	41	10	1.0000	0.0093
1961Gronchi	622	67	13	0.9999	0.0133
1962Segni	381	35	10	0.9999	0.0166
1963Segni	527	46	12	1.0000	0.0061
1964Saragat	278	21	8	0.9999	0.0134
1965Saragat	510	52	12	0.9996	0.0269
1966Saragat	597	44	13	0.9996	0.0281
1967Saragat	526	51	11	0.9997	0.0231
1968Saragat	562	56	13	1.0000	0.0035
1969Saragat	692	86	15	0.9997	0.0244
1970Saragat	812	85	17	0.9998	0.0194
1971Leone	168	12	5	0.9994	0.0343
1972Leone	394	32	10	0.9999	0.0113
1973Leone	616	67	12	0.9996	0.0277
1974Leone	426	32	9	0.9994	0.0334
1975Leone	632	63	13	0.9997	0.0244
1976Leone	649	52	13	0.9998	0.0194
1977Leone	717	80	14	0.9995	0.0312
1978Pertini	603	53	14	0.9997	0.0238
1979Pertini	800	70	18	1.0000	0.0013
1980Pertini	535	50	14	1.0000	0.0067
1981Pertini	911	96	20	0.9999	0.0124
1982Pertini	854	90	19	1.0000	0.0003
1983Pertini	1149	118	24	1.0000	0.0031
1984Pertini	514	42	14	0.9998	0.0217
1985Cossiga	859	118	17	0.9985	0.0554
1986Cossiga	561	65	14	0.9998	0.0218
1987Cossiga	904	109	15	1.0000	0.0095
1988Cossiga	875	123	19	0.9998	0.0203
1989Cossiga	778	85	17	0.9995	0.0308
1990Cossiga	1222	155	20	0.9998	0.0182
1991Cossiga	241	22	7	0.9993	0.0373
1992Scalfaro	978	118	18	0.9995	0.0311
1993Scalfaro	1074	129	19	1.0000	0.0008
1994Scalfaro	1190	171	21	0.9997	0.0232

1995Scalfaro	1341	180	23	1.0000	0.0093
1996Scalfaro	866	88	16	0.9995	0.0322
1997Scalfaro	1405	167	28	0.9999	0.0171
1998Scalfaro	1175	137	24	1.0000	0.0023
1999Ciampi	831	66	17	0.9993	0.0368
2000Ciampi	822	70	16	1.0000	0.0057
2001Ciampi	898	89	18	0.9999	0.0138
2002Ciampi	909	96	17	1.0000	0.0066
2003Ciampi	718	63	14	0.9998	0.0177
2004Ciampi	812	76	15	1.0000	0.0059
2005Ciampi	538	54	13	1.0000	0.0084
2006Napolitano	929	125	17	0.9994	0.0342
2007Napolitano	793	101	16	1.0000	0.0075
2008Napolitano	775	75	15	0.9995	0.0302

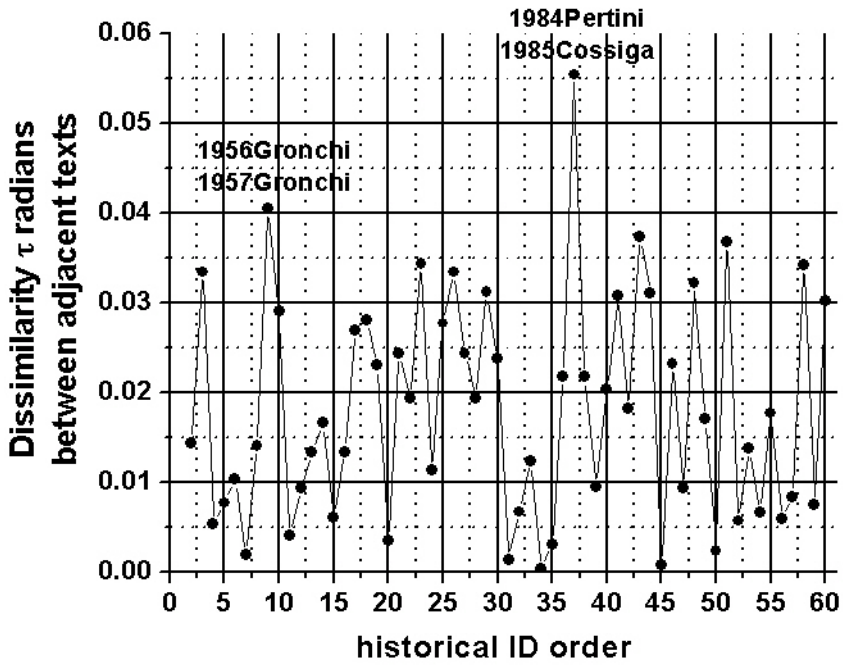

Figure 7.4. Angle τ radians between adjacent texts

The computation has been performed to eight decimal places but rounded to four. Though the subsequent texts may strongly differ, one cannot see a clear trend; the movement of dissimilarity is rather haphazard, as presented in Figure 7.4. The greatest dissimilarity is between Pertini´s last and Cossiga´s first speech.

Even if we take only the means of transition dissimilarities for individual presidents given in Table 7.7 and shown in Figure 7.5, the oscillation is very conspicuous.

Table 7.7
Transition dissimilarities within individual presidents

President	**Mean dissimilarity**
1. Einaudi	0.0142
2. Gronchi	0.0184
3. Segni	0.0061
4. Saragat	0.0209
5. Leone	0.0245
6. Pertini	0.0076
7. Cossiga	0.0230
8. Scalfaro	0.0141
9. Ciampi	0.0097
10. Napolitano	0.0302

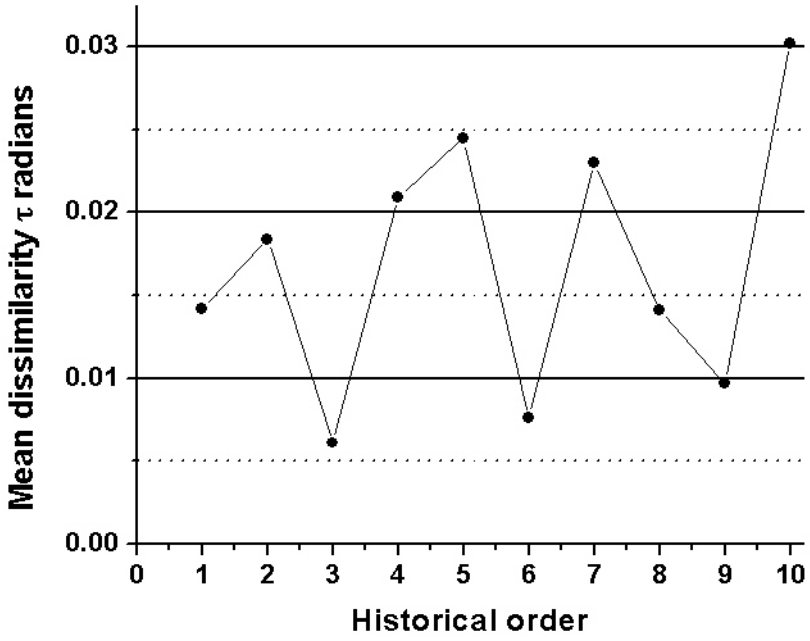

Figure 7.5. Subsequent mean transition dissimilarities

Another way of showing the sequence is to consider only the given transition dissimilarity from one president to the next one, as recapitulated in Table 7.8.

Table 7.8
Transitions between historical neighbours

"Neighbours"	**Mean dissimilarity**
1. Einaudi-Gronchi	0.0019
2. Gronchi-Segni	0.0166
3. Segni-Saragat	0.0134
4. Saragat-Leone	0.0343
5. Leone-Pertini	0.0013
6. Pertini-Cossiga	0.0557
7. Cossiga-Scalfaro	0.0311
8. Scalfaro-Ciampi	0.0368
9. Ciampi-Napolitano	0.0342

As can be seen in Figure 7.6 the transition dissimilarities increase slightly but irregularly. The "great jump" begins starting from Cossiga who strongly deviates from his predecessor, Pertini. Cossiga, from his first address, seems concerned with cutting loose from the very popular figure of his predecessor, and innovating in his choices of international and institutional themes. Afterwards, all presidents try to be dissimilar to their individual predecessor.

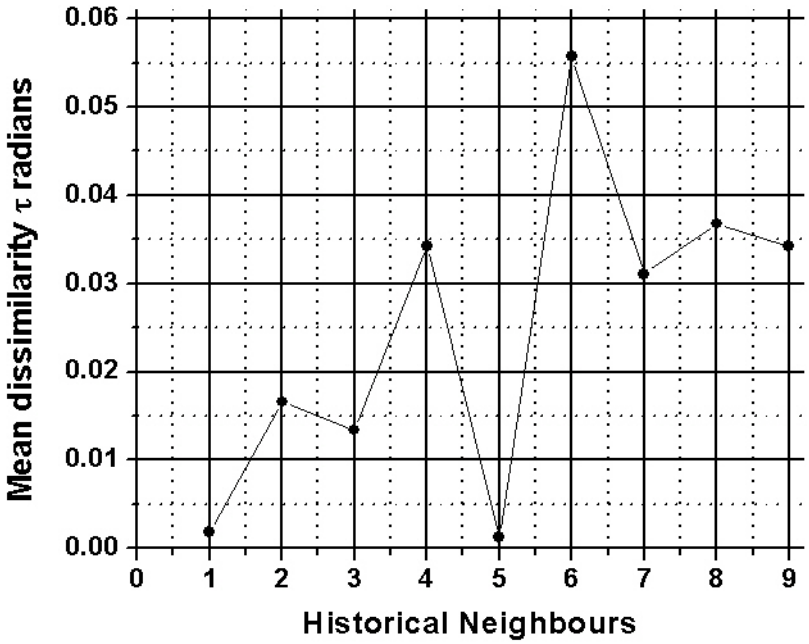

Figure 7.6. Transition dissimilarities between subsequent presidents

The present method is appropriate in all cases where the individuality of words or contents does not play any role. We used it for the sequence of presidential speeches but in the same way it can be used for discriminating poetry from other types of texts or even for the development of music or discriminating between musicians or musical styles. Its advantage is the simplicity of computation and the irrelevance of the types of textual units.

References

Bernardi, L., Tuzzi, A. (2007). Parole lette con misura (statistica). In: Cortelazzo, M.A., Tuzzi A. (eds.), *Messaggi dal Colle. I discorsi di fine anno dei presidenti della Repubblica: 109-134*. Venezia: Marsilio.

Best, K.-H. **(**1994). Word class frequencies in contemporary German short prose texts. *Journal of Quantitative Linguistics 1(2), 144-147.*

Brunet, E. (1988). Une mesure de la distance intertextuelle: la connexion lexicale. *Le nombre et le texte. Revue informatique et statistique dans les sciences humaines, Université de Liége*.

Buchanan, M. (2002). *Nexus. Small worlds and the groundbreaking science of networks.* New York-London: Norton.

Bunge, M. (1967). *Scientific Research Vol I.* Berlin-Heidelberg-New York: Springer

Carrol, J.B. (1969). Vectors of prose style. In: Doležel, L., Bailey, R.W. (Eds.), *Statistics and style: 147-155*. New York – London: Elsevier.

Cortelazzo, M.A., Tuzzi, A. (2006). Il discorso di insediamento del Presidente della Repubblica Giorgio Napolitano. Lessico e retorica. *LId´O Lingua italiana d´oggi III, 125-138.*

Cortelazzo, M.A., Tuzzi, A. (2007). *Messaggi dal Colle. I discorsi di fine anno die presidenti della Repubblica.* Venezia: Marsilio.

Dell´Anna, MV. (2005). Tra ufficialità e colloquialità. La lingua di Carlo Azeglio Ciampi. *LId´O Lingua italiana d´oggi II, 171-124.*

Doczi, G. (2005). *The Power of Limits: Proportional Harmonies in Nature, Art, and Architecture*. Boston: Shambhala Publications.

Dunlap, R.A. (1997). *The Golden Ratio and Fibonacci Numbers*, World Scientific Publishing, 1997

Hemenway, P. (2005). *Divine Proportion: Phi In Art, Nature, and Science*. New York: Sterling.

Hirsch, J.E. (2005). An index to quantify an individual´s scientific research output. http://arxiv.org/PS_cache/physics/pdf./0508/0508025.pdf,

Hřebíček, L. (2000). *Variation in sequences*. Prague: Oriental Institute.
http:/en.wikipedia.org/wiki/hirsch_number .

Huntley, H. E. (1970). *The Divine Proportion: A Study in Mathematical Proportion*. New York: Dover Publications.

Kelih, E. (2009). Slawisches Parallel-Textkorpus: Projektvorstellung von „Kak zakaljalas' stal' (KZS)". In: Kelih, E., Levickij, V., Altmann, G. (eds.), *Metody analizu teksta/Methods of Text Analysis:* 106-124. Černivci: ČNU.

Kelih, E. (2009a). Preliminary analysis of a Slavic parallel corpus. In: Levická, J., Garabík, R. (eds.), *NLP, Corpus Linguistics, Corpus Based Grammar*

Research. Fifth International Conference Smolenice, Slovakia, 25-27 November 2009. Proceedings: 175-183. Bratislava: Tribun.

Köhler, R. (1995). *Bibliography of Quantitative Linguistics*. Amsterdam-Philadelphia: Benjamins.

Labbé, C., Labbé D. (2001). Inter-Textual Distance and Authorship Attribution Corneille and Moliére. *Journal of Quantitative Linguistics 8, 213-231.*

Labbé, D. (2007). Experiments on authorship attribution by intertextual distance in English. *Journal of Quantitative Linguistics 14, 33-80.*

Laufer, J., Nemcová, E. (2008). Diversifikation deutscher morphologischer Klassen in SMS. *Glottometrics 18, 2009, 13-25*

Levin, J.I. (1964). Ob opisanii lingvističeskich ob'ektov, obladajuščich obščimi svojstvami. *Voprosy jazykoznania 13(4), 112 -119.*

Livio, M. (2002a). *The golden ratio: the story of phi, the extraordinary number of nature, art and beauty*. London: Headline.

Livio, M. (2002b). The golden ratio and aesthetics. *Plus Magazine, 22 (http://plus.maths.org/)*

Livio, M. (2002c). *The Golden Ratio: The Story of Phi, The World's Most Astonishing Number*. New York: Broadway Books.

Marquardt, S.R. (2002). Dr. Stephen R. Marquardt on the golden decagon and human facial beauty. *Journal of Clinical Orthodontics, 36(6), 339-47.*

Merriam, T. (2002). Intertextual distances between Shakespeare plays, with special reference to *Henry V* (verse). *Journal of Quantitative Linguistics 9(3), 261-273.*

Pacioli, L. (1509). *De divina proportione*, Luca Paganinem de Paganinus de Brescia (Antonio Capella) 1509, Venice.

Pauli, F., Tuzzi, A. (2009). The End of Year Addresses of the Presidents of the Italian Republic (1948-2006): discoursal similarities and differences. *Glottometrics 18, 40-51.*

Pereira, F., Tishby, N., Lee, L. (2008). Distributional clustering of English words. http://citeseer.ist.psu.edu/71393.html (03.28.2008)

Popescu, I-.I. et al. (2009). *Word frequency studies*. Berlin-New York: Mouton de Gruyter.

Popescu, I.-I., Altmann, G. (2006). Some aspects of word frequencies. *Glottometrics 13, 23-46.*

Popescu, I.-I., Altmann, G. (2007). Writer´s view of text generation. *Glottometrics 15, 71-81.*

Popescu, I.-I., Altmann, G. (2008). Hapax legomena and language typology. *Journal of Quantitative Linguistics 15(4), 370-378.*

Popescu, I.-I., Altmann, G. (2009). A modified text indicator. In: Kelih, E., Levickij, V., Altmann, G. (eds.), *Problems of quantitative text analysis*: *208-229.* Černivci: RUTA.

Popescu, I.-I., Altmann, G., Köhler, R. (2009). Zipf´s law – another view. *Quality and Quantity* (May 2009), DOI 10.1007/s11135-009-9234-y (2009)

http://www.springerlink.com/content/v268444738300073/?p=b28be84685b148ff8725b0cbc1671402&pi=0 (September 9,2009)

Popescu, I.-I., Mačutek, J., Altmann, G. (2009). *Aspects of word frequencies.* Lüdenscheid: RAM.

Sahlqvist, L. (2008). *Cardinal Alignments and the Golden Section: Principles of Ancient Cosmography and Design* (3rd Rev. ed.). Charleston, SC: BookSurge.

Sambor, J. (1989). Polnische Version des Projekts "Sprachliche Synergetik. Teil I. Quantitative Lexikologie. *Glottometrika 10, 171-197.*

Schweers, A., **Zhu, J.** (1991). Wortartenklassifizierung im Lateinischen, Deutschen und Chinesischen. In: Rothe (1991a): *157-165.*

Tuzzi, A., Popescu, I.-I., Altmann, G. (2009a). Zipf´s law in Italian texts. *Journal of Quantitative Linguistics 16, 4, 354-367.*

Tuzzi, A., Popescu, I.-I., Altmann, G. (2009b). Parts-of-speech diversification in Italian texts. *Glottometrics 19, 42-48.*

Tuzzi, A., Popescu, I.-I., Altmann, G. (2009c). The golden section in texts. *ETC – Empirical Text and Cluture Research 3, (submitted).*

Viprey, J.-M., Ledoux, C.N. (2006). About Labbe´s "Intertextual Distance". *Journal of Quantitative Linguistics 13(2-3), 265-283.*

Walser, H. (2001) *The Golden Section.* Washington, DC: The Mathematical Association of America.

Weiss, V., Weiss, H. (2003). The golden mean as clock cycle of brain waves. *Chaos, Solitons and Fractals 18: 643–652.*

West, D.B. (2001). *Introduction to graph theory.* Upper Saddle River, Nj: Prentice Hall.

Wimmer, G., Altmann, G. (1999). On vocabulary richness. *Journal of Quantitative Linguistics 6, 1-9.*

Ziegler, A. (1998). Word class frequencies in Brazilian-Portuguese press texts. *Journal of Quantitative Linguistics 5(3), 269-280.*

Ziegler, A. (2001). Word class frequencies in Portuguese press texts. In: Uhlířová, L. et al. (eds.) *Text as a linguistic paradigms: levels, constituents, constructs: Festschrift in honour of Luděk Hřebíček: 294-312.* Trier: Wissenschaftlicher Verlag.

Ziegler, A., Altmann, G. (2002). *Denotative Textanalyse.* Wien: Edition Praesens.

Zotti Minici, A. (2007). Dove guardano i presidenti a fine anno. In: Cortelazzo, Tuzzi (2007): *47-53.*

Author index

Subject index