Chapter 13
Competences and Professional Options of the Italian Graduates: Results from the Textual Analysis of the Degree Course Information Data

S. Balbi, C. Crocetta, M.F. Romano, S. Zaccarin, and E. Zavarrone

13.1 Introduction

The present chapter, developed within a research project promoted by the National Committee for the Evaluation of the University System (CNVSU, the complete report is available on www.cnvsu.it), analyses the Italian university offer, focusing on the communication adopted by the Universities to publicise their objectives and the results expected, also with respect to the potential employments. The analysed documents are the course information sheets of all the 3-year degree courses and some specialised (2-year) degree courses, contained in the OFF.F database of the Ministry of the University (MIUR) for the academic year 2005/2006. The research pursued the following aims:

1. reading the education offer focusing mainly on the competences foreseen for the graduates;
2. reading the foreseen job prospects for the graduates in a course;
3. analysing the consistency of the competences provided with the foreseen employment prospects;
4. analysing the consistency between the competences acquired by the 3-year graduates and the competences offered to whom decides for continuing with a 2-year degree course specialisation.

The analysis of points 1–3 is carried out through the use of typical *text – mining* procedures, while for point 4 the reference is to methods developed for the analysis of *multi-linguistic corpora.* In Sect. 13.2, a short overview on the adopted methods is presented, while in the following sections some general reflections on the topics are made, referring to the principal results for the degree in *Statistics* (class code 37), as an example.

S. Balbi (✉)

Dipartimento di Matematica e Statistica, Università di Napoli "Federico II", Napoli, Italy
e-mail: simona.balbi@unina.it

13.2 Textual Analysis of the University Education Offer

For the university system, as well as for any other aspect of our society, the topic of *communication* has gained – over the last few years – a growing importance. In addition to the obvious considerations connected with the "Internet" revolution, which makes new tools for interaction available, there are other stimuli which derive more specifically from the transformations that the Italian University system has undergone; transformations which are still going on.

On the one hand, the University belongs to the Public Administration, to whom *transparency* is required, in terms of the visibility of its work. This requisite is therefore required for accreditation of a university course. However, on the other hand, the communication regards also the autonomy which is granted to each single institution, in order to achieve a healthy competition. In this sense, therefore, it is a *communication* to be intended in a corporate sense, aimed at hitting one's own target market.

A further element to be considered derives from the indications elaborated within the Bologna Process with the purpose of creating a European common framework, whose practicability depends not only on having common objectives and tools, but on sharing their formulation.

There are, therefore, many forms of communications, starting from the advertising campaigns of the Universities, but also of the individual programs, in different *media*. Here, as mentioned, the attention is focused on those aspects which are more institutional, such as the presentations that each degree course sends to the Ministry, to be inserted in the database of the education offer OFF.F. These *schede* are the course information data sheets with a structure defined by the Ministry, which contain a variety of information. In particular, there are two sections which can provide the reader (the school-leaver or his/her family), a picture easily understandable by non specialists: these are two paragraphs written in free-form, one about the education objectives and the other about the employment prospects.

The subsequent analysis refers to the course information data sheets about the education offer of the Italian university system in the first phase of the fulfillment of the so-called 3 + 2 reform. Of the 2,339 3-year degree courses activated for the year 2005/2006 within the 42 classes (see Table 13.1) defined in the Ministerial Decree of the 4th August 2000, the portions of text relative to the *specific educational objectives (obiettivi formativi caratterizzanti)* and *job possibilities (ambiti occupazionali previsti)* were examined.

The compilation of the *schede* is, first of all, a requisite for the institution of a degree course. In this sense then, an administrative obligation exists. There is also a sort of framework, compiled by the Ministry, where, for each class of degree, the aims of the centrally specified qualifying program (*obiettivi formativi qualificanti*) and the areas of prospective employment are foreseen. The declaration that one course belongs to a specific class should, therefore, imply the adhesion to the education aims of that class.

It is, therefore, for the purposes of *competition* that each degree course is required to differentiate itself and give more space to the employment perspectives (in this sense the new reform in force from year 2008/2009 is more explicit on this point).

Table 13.1 Three-year degree classes and number of courses. Year 2005/2006

Class codes and names	Number of courses activated
1 – Biotechnologies	53
2 – Legal Services	50
3 – Linguistic Mediation	41
4 – Architecture and Construction Engineering	62
5 – Literature	68
6 – Social Work	47
7 – Town, Regional and Environmental Planning	23
8 – Civil and Environmental Engineering	81
9 – Information Technology	141
10 – Industrial Engineering	168
11 – Modern Languages and Civilizations	66
12 – Biological Sciences	53
13 – Cultural Heritage Studies	74
14 – Communication Studies	73
15 – Political Science and International Relations	56
16 – Earth Sciences	31
17 – Business Economics	169
18 – Education Sciences and Teacher Education	66
19 – Public Administration	33
20 – Agriculture, Food Industry and Forestry	114
21 – Chemistry	61
22 – Aviation and Maritime Navigation	1
23 – Visual Arts, Music, Performing Arts and Fashion Studies	33
24 – Pharmacy and Industrial Pharmacy	62
25 – Physics	51
26 – Computer Science	56
27 – Environmental Sciences	59
28 – Economics	95
29 – Philosophy	46
30 – Geography	8
31 – Law	73
32 – Mathematics	48
33 – Physical Education and Sport	34
34 – Psychology	49
35 – Social Sciences for Co-Operation, Development and Peace	19
36 – Sociology	23
37 – Statistics	32
38 – History	29
39 – Tourism	23
40 – Animal Husbandry	27
41 – Technologies for the Conservation and Restoration of Cultural Assets	21
42 – Industrial Design	20
Total	2, 339

Source: Database OFF.F, year 2005/2006 http://offf.miur.it/index.html

However, although it is necessary to compile these forms, precise indications on what they should exactly say have not yet been issued.

On the theme of the Europeanisation of the system, the putting into effect of the *Bologna Process* pushes towards the arrangement of course information data sheets according to the so-called "Dublin descriptors", built on the following elements:

- knowledge and understanding
- applying knowledge and understanding
- making judgements
- communication skills
- learning skills

As far as the methodology adopted is concerned, the first descriptive analysis of the education offer of the whole university system was carried out from the point of view of *text mining*, given the mass of texts to be examined. The following analyses are based on the multivariate statistical techniques, opportunely adapted to the circumstance of having to work on textual data, instead of numeric data.

The first consequence is that our databases (made up of the *corpus* of the two sections of the course information data sheets) are unstructured and need to be organised in order to deal with the typical data structures used to perform textual analyses. Here we think of the so-called lexical tables, i.e. matrices cross-classifying courses and terms used for describing them. In building those matrices, we have to make different choice related to the minimal units to adopt, the frequency thresholds to consider them interesting for our aims. Moreover, we need to pre-treat the documents by lexicalisation for avoiding trivial cases of ambiguity. The documental base consists of about 2,400,000 occurrences, while concerning with documents length, it is worth noting that all the texts where written for being included in the OFF.F pre-defined window. Therefore, there are no interesting differences in their length.

The results obtained in our analysis have to be intended as stimuli for further investigation. The aim of exploration of the present work must therefore be borne in mind while reading the following.

The analysis suggested here is divided into three principal sections, characterized by different aims as well as by different dimensions of available data. The first aim is to understand the *specific education offer* and the *employment perspective expected* from the 2,339 degree courses (Sect. 13.3). Afterwards, the aspects about the consistency among these texts are analysed (Sect. 13.4), and then we deal with the themes connected to the 3 + 2 structure (Sect. 13.5). For reasons of space, only results for class 37 (*Statistics*) are shown.

13.3 Identification of the Competences Offered and the Job Possibilities

The tools chosen for this analysis, given the large dimensions of the *corpus* to be analysed, are those of the analysis of the textual data [6] and of *text mining*, using the *software* TALTAC. It was furthermore decided to treat separately the texts about

the competences and those referring to the employment possibilities, because of the different dimensions of the two bases of data, as the description of the job opportunities expected has received less interest at the local level. Thus, also different methodologies were used.

We started from the analysis of the specific educational objectives (competences), in a typical perspective of *mining*, using a two stage procedure of knowledge extraction [2]: the document is considered as a complex data, with a hyerachical structure within itself, the levels of which are made up of sentences, made up of phrases, made up of words (in their turn, made up of characters). To analyse the words (or groups of words), an initial selection of the interesting sentences was made, as they were relative to competences (*specific educational objectives*). In order to do this, we used, in a content analysis perspective, the identification "a priori" of *key-verbs* (e.g., acquire, apply, learn, have, know, demonstrate, carry out, be able to supply, teach, offer, operate, participate, possess, utilize, use, . . .) which it is assumed connect – within a sentence – words referring to the competences offered during the course of studies. These verbs were used to obtain a drastic reduction of the material to be analysed, as the *scheda* pertains also to organization and sometimes to considerations of a wider nature, independent from an analysis of the competences offered.

A field identifying a sentence as interesting (i.e. referred to competences) was built, through logical rules of the type: IF *acquire* = 1 THEN *interesting* = YES OR *apply* = 1 THEN *interesting* = YES OR . . . OR . . .). Therefore, all those sentences where *interesting* is different from YES were eliminated, and the textual analysis performed on the *sub-corpus* containing the sentences relative to competences (about 800,000 occurrences, with a 17,200 terms vocabulary).

Once the fragments of interest were identified, the construction of the repeated segments (defined as sequences of recurring words with a frequency higher than a fixed threshold (see [6]) was carried out as well as that of the nouns. For building the repeated segments, only the terms with a frequency higher than 5 were considered. Furthermore, the length of the segments was limited to a maximum of 4 terms.

In this way, the linguistic structures which characterised most of the courses belonging to the same class of degree were identified: e.g. short lists of terms used in a class of degree could suggest poor lexicon, somehow connected with a lesser (or more homogeneous?) education offer, from the point of view of the contents.

This analysis shows the necessity for more stringent indications from the Ministry about the contents to be assigned to these course information data sheets, so that they can work as a valid tool for an informed choice for those wishing to get useful information about the degree courses in which to enroll, and also for the transparency about the real education contents of a course, both in terms of competitiveness and autonomy.

The differences among the single classes of degree appear clearly: some classes of degree have an objective which is clearly defined and limited (e.g. *Industrial Design*), while others seem to want to cover large fields of human knowledge (e.g. *Agriculture, Food Industry and Forestry*). Furthermore, there is obviously the work carried out by each single University. Apart from the extreme case of class 22, in

which only one degree course was activated, more generally speaking, the different cultural matrices of the different courses expressed themselves in ways which are clearly typical of each area and definitely very variable.

The most concise vocabularies are those of the Classes 2 – *Legal Services*, 5 – *Literature*, 11 – *Modern Languages and Civilizations*, 13 – *Cultural Heritage Studies*, 18 – *Education Sciences and Teacher Education*, 23 – *Visual Arts, Music, Performing Arts and Fashion Studies*, 28 – *Economics*, 31 – *Law* and 35 – *Social Sciences for Co-operation, Development and Peace*. These vocabularies are substantially limited to a few words specific to each discipline, such as for example "legal" and "rules" for Class 2.

The Class 1 – *Biotechnologies* has a vocabulary which is well-constructed and not limited to terms which are specific to the discipline, but also relative to other disciplines ("chemistry", "physics", "mathematics", "informatics", "statistics"). Even larger vocabularies appear the Classes 4 – *Architecture and Construction Engineering*, 12 – *Biological Sciences*, 20 – *Agriculture, Food Industry and Forestry*, 29 – *Philosophy*, 40 – *Animal Husbandry*. These vocabularies are also characterized by a homogenous distribution of the occurrence of the words.

In general, it is appropriate to point out how the competences expressed in all the vocabularies are of a general character, though sometimes supported with terms which are specific to the discipline, such as "constructions", "building site", "technical and economic feasibility" (Class 4), "management of human resources" (Class 19), "commercialisation" and "technical assistance" and "quality control" (Class 20), "therapeutical preparations" (Class 24), "problems analysis" and "techniques of argumentation" (Class 29), "mathematical models" and "computational tools" (Class 32), "services to the individual" (Class 34), "properties of materials" (Class 41), "engineering", "planning" and "communication interfaces" (Class 42).

A noteworthy case is that of the degree courses in Engineering (8 – *Civil and Environmental Engineering*, 9 – *Information Technology*, 10 – *Industrial Engineering*): their vocabularies are overlapping (as well as the corresponding ministerial course information data sheets) and also the frequency of the occurrence often respects the same order, so that some doubts appear about the real need for this tri-partition.

The vocabulary of the competences for the class of degree 37 (*Statistics*) are quoted in Table 13.2. For this class, the vocabulary appears adequate, taking into account the presence of terms which describe competences which are specific to the discipline (*surveys, data sets, measurement, data analysis, data collections, statistical analysis, statistical, experimental, observational methods*) or general competences, but not generic ones (*measurement, logical-conceptual, method of research, technological-experimental*). There are also some terms referring to the areas of application or of employment (*economical, informatics, biomedical*).

As far as the employment perspectives are concerned, starting from those words which are considered to be of higher relevance on the basis of expert knowledge (*content bearing words*), we moved in a *data driven* manner, with higher attention to the linguistic aspects. The analysis carried out allows the identification of key words of the professions expected at the end of the course by the ministerial tables

Table 13.2 Vocabulary of the competences for Class 37 – Statistics (words with frequency > 1)

Words	Total occurrences	Words	Total occurrences
Statistics	102	Applications	8
Tools	45	Data analysis	8
Statistical	42	Methodological	8
Science	30	Data collection	8
Economical	28	Logical-conceptual	7
Technical	25	Statistical analysis	7
Social	24	Execution	7
Surveys	16	Method of research	4
Planning	15	Experimental	4
Applications	12	Biomedical	3
Statistical disciplines	11	Observational	3
Databases	10	Statistical methodology	3
Informatics	10	Culture of working	2
Measure	10	Technical-experimental	2

and taken up by the different courses of study, in the light of their own education offer, planned taking into account the potential demand, their internal resources and the characteristics of the geographical area.

The "repeated segments" describing the employment perspectives of the different classes were analysed, as they appear in the declarative statements of the individual degree courses. Their importance is measured by a lexicometric index IS [8], which measures the absorption of the single words in a segment.

The index is given by the rate between the frequency of the segment and the frequency of the single words which make it up, for the number of meaningful words of the segment (not articles, prepositions, conjunctions, adverbs). This tool allows us to distinguish the degree courses for the width of possible professional roles upon graduation (which is not necessarily a measure of the marketability of the degree in the world of work). Furthermore, it is necessary to distinguish the classes of degree with an exclusive professional role (doctors) and those where there are coexisting roles, such as, for example, those of Economics, which are defined generalist, as they offer a wide range of professional roles upon graduation.

From the textual analysis carried out on the employment possibilities, the picture appears to be quite varied. Given the high variability existing among the different course information data sheets proposed by MIUR, and among the latter and the degree programs activated at the different universities it is difficult to be able to gather any regularity.

The analysis carried out has highlighted the need to dedicate more attention to the description of the professional figures coming out from the different study paths proposed, also for the purpose of the verification of the congruity between the professional profiles expected and the education paths proposed (a need understood in the formulation of the *scheda* for the year 2008/2009). Furthermore, it would be opportune to add, on the *scheda*, together with the qualitative information, also some quantitative data about the effective possibilities of employment offered by the different paths and about the conformity of their contents to the working activities

carried out. In this way, useful elements could be provided to the potential student allowing him/her to make a well-informed choice.

Going back to the formulation of the course information data sheets, the duplicity of their objective emerges here even more clearly: providing *marketing* and *transparency*, in order to guarantee the potential student more information about the realistic employment perspectives, beyond simply catching labels. A higher degree of structuring (and standardization) of the course information data sheet of presentation of the employment perspectives would not affect the organizational autonomy of the Universities, however it would force them to carry out a more thought through evaluation about the future work placement of their own graduates.

Once more, the result of the analysis carried out for the class 37 is presented. Among the possible employment outcomes for the graduates in statistical sciences, the ministerial decree proposes the professional activities in the field of learning and spreading of statistical knowledge, the management of qualitative and quantitative information on behalf of public or private companies and research institutions. The degree programs stress, instead the possibilities of employment in the fields of the planning and management of the *informative systems* and of the *professional activities* either as self-employment or as an employee of *private companies* or *research institutions*, not stressing work activities with specific contents referring to the professional competences of a statistician.

13.4 Consistency of the Education Paths with Employment Perspectives

For the evaluation of the consistency between competences offered and employment perspectives, we used a variant of Lexical Correspondence Analysis (LCA, [6]), one of the most common methods for the analysis of textual data. Aiming at analysing in depth the dependence of the description of the foreseen employment with the figure of the graduate described in the section "education objectives", the non symmetrical analysis of the co-occurrences [4] was adopted. This is a factorial technique aimed at the study of matrices having as generic element the frequency with which a term is used to describe the same object, in two different conditions, one logically, or temporally, antecedent to the other one. It is a variation of the non symmetrical lexical correspondences analysis (NSLCA), proposed by Balbi [1] for studying a relation of dependence of the vocabulary from a partition induced on the corpus.

NSLCA, instead of decomposing the association index χ^2, as usual in Correspondence Analysis, deals with the predictability index τ_b, proposed by Goodman and Kruskal (see also [5]). Let us consider an aggregated lexical table F (K,V), where K is the number of categories considered for documents and V is the number of terms in the vocabulary. F general element is f_{kv}, relative frequency of the v-th term in the k-th category. NSLCA studies the V conditional distribution $f_{kv}/f_{.v}$, with respect to the independence hypothesis, given by $f_{k.}$, where $f_{.v}$ and $f_{k.}$ are respectively the column- and row- marginal distribution. From a geometrical viewpoint, distances between the categories are measured in the usual Euclidean metric,

while distances between terms are measured in a weighted Euclidean metric. Compared with Lexical Correspondence Analysis, an interesting effect consists in giving a lower importance to infrequent words.

In this case, we analysed the 32 courses belonging to the class *Statistics*, in order to represent the internal consistency, by measuring how much the presence of one word in the section "specific education objectives" has influenced its use in the description of the "job possibilities". The two lexical tables cross-classifying *degree courses* (documents) and *objectives* (terms) and *degree courses* and *professional roles* (terms) were collapsed into a single matrix (terms × terms) *professional roles × objectives*. The generic element of the resulting matrix is given by the number of times the two terms are used jointly in the two corpora. The risk of obtaining a sparse matrix, lacking in information for the presence of too many 0's, was avoided by eliminating from the vocabulary all the so-called functional part-of-speech (i.e. conjunctions, articles, prepositions, adverbs), and limiting the analysis to the meaning words: nouns, adjectives, verbs.

In this analysis there are specific reading rules for the interpretation of the factorial planes:

- the scattering of words around the origin shows the strength of the dependence of the vocabulary from the information given on the documents, inducing their partition;
- two words are near, if they similarly depend from the partition;
- two categories of the partition variable are close if they similarly influence the use of words;
- a word is so far from the origin, the more it depends on the partition;
- a modality of the external variable is so far from the origin, the more it influences the use of the words.

The differences in metrics produces some important consequences. First of all, compared with usual LCA, the words with higher frequency are more "important" in the characterization of the documents, than those less frequent. The second characteristic is that the "joint plot" typical of LCA (the unique factorial plane where documents and terms are simultaneously represented) is difficult to be read, because of the different scales (as in principal components analysis), even though the origin (which represents the hypothesis of independence), the orientation of the axes and the percentage of the explained variability are common. The two factorial planes are therefore represented separately: Figure 13.1 represents competences, while Fig. 13.2 the employment perspectives.

It is worth noting how the reform has granted the opportunity to increase the offer of statistics training, motivating a wider differentiation of the degree courses in the different faculties. Mainly in those Universities without a Faculty of Statistics, the reform has meant planning something radically new. Unexpectedly, one of the roots which appears more frequently, on the graph describing competences, is *mercat-* (market), together with *marketing*. To these are added other words indicating further areas of application of statistical tools, such as *comunicazione* (communication), *finanza* (finance) and *assicurazioni* (insurance) at the bottom and *qualità* (quality)
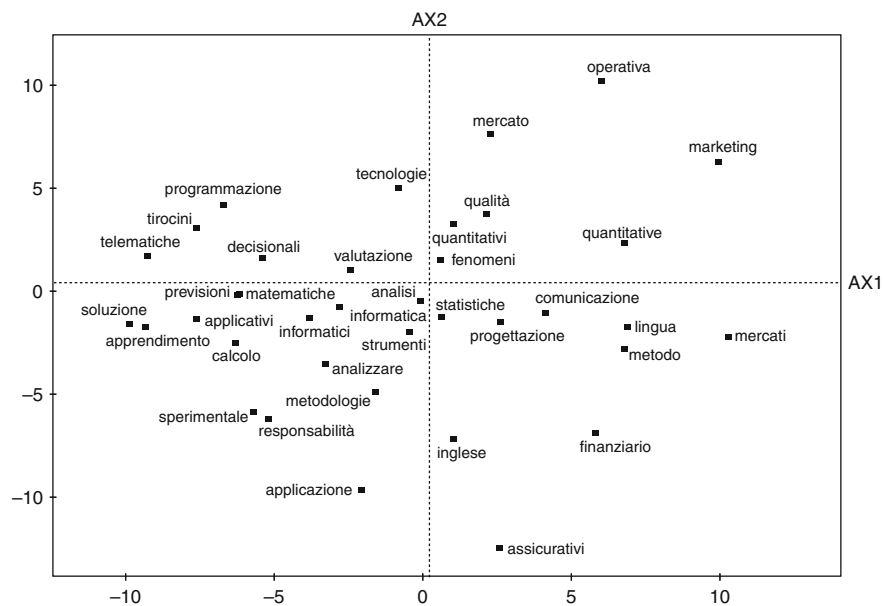
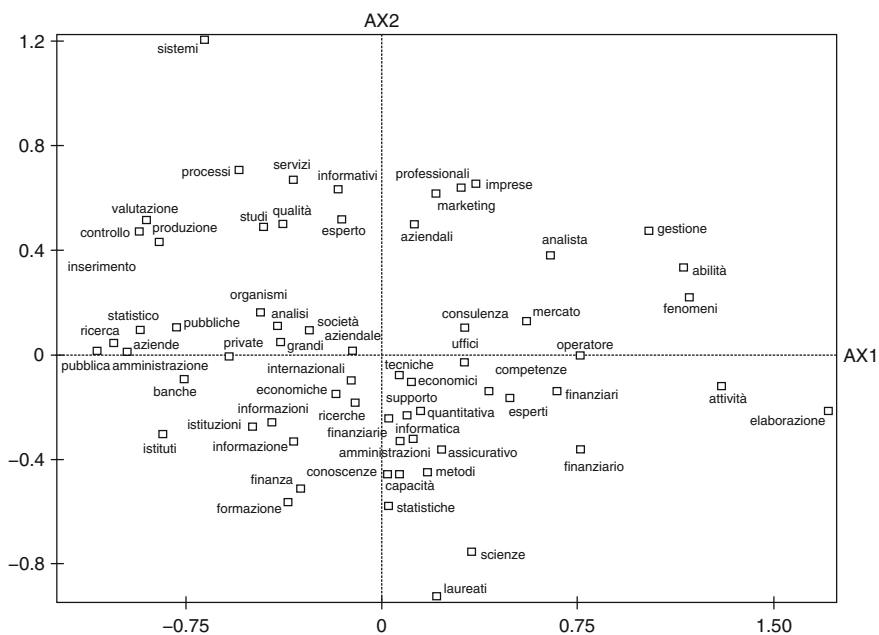**Fig. 13.1** The first factorial plane of the competences for Class 37 – Statistics

Fig. 13.2 The first factorial plane of the employment perspectives for Class 37 – Statistics

at the top. All these words are positioned on the right-hand-side of the diagram, while on the left there are words which are somehow more connected with the education of a statistician (*applicativi*-applicative, *previsioni*-forecasts, *decisionali*-decisional, *sperimentale*-experimental) next to generic terms (apprendimento-*learning*, soluzione-*solution*). It seems appropriate to point out how the root *statistic* – is very close to the origin of the axes, suggesting that it is being used generally in all degree courses.

In Fig. 13.2, describing the potential jobs for their graduates, it seems as if the statisticians are frightened of using the word which identifies their discipline and prefer turns of phrases instead (*quantitative*, *method*, *methodology*, *analysis*, *analyze*). There is a lot of mention of the disciplines ancillary to the study of statistics (*informatics*, *information technology*, *mathematics*, *calculation*), still on the left-hand-side of the graphic. In any case, the first axis opposes employees vs. consultants, but with poor relations with the opposition seen in Fig. 13.1. In conclusion, there doesn't seem to be a direct consequential link between the education offer and the employment offer, except for some specific professional roles (e.g. finance and insurance).

13.5 Coherence Between the Three Year Degree Courses and the Specialised Degree Programs

The objective to study the linguistic coherence between 3-year degrees and specialised 2-year degrees was pursued by adopting a method developed for the analysis of multilinguistic corpora, for comparing the latent semantic structures in translations, thanks to Procrustes rotations [3]. Perhaps this is a problem which will be less pressing with the new reform, as the structural connection between the 3 year degree course and the specialised degree appears to be loosening up, nevertheless considerations about the past could be profitable for the immediate future, for redefining of the degree courses.

The geometrical component of the method used is important. The documents are represented as points in a multi-dimensional system, spanned by the terms which make up the vocabulary of the analysed *corpus* (*vector space model*), or by their factorial transformation (*latent semantic indexing*). A measure of distance between two corpora can be obtained through Procrustes rotations, so called because they best adapt the scatterings which represent the documents of the two corpora, through centering, standardization, reflection and rotation of the factorial axes.

Briefly, it is necessary to reduce the number of terms used in the two linguistic *corpora* being compared, identifying the individual latent semantic structures (which can be assimilated to the factors of a principal components analysis) and then to compare the two semantic structures, measuring their distance, in terms of fitting of one to the other.

In addition to the quantitative result, represented by the goodness-of-fit index, it is also possible to represent graphically the scattering of the points which represent the declaratives of the individual degree courses.
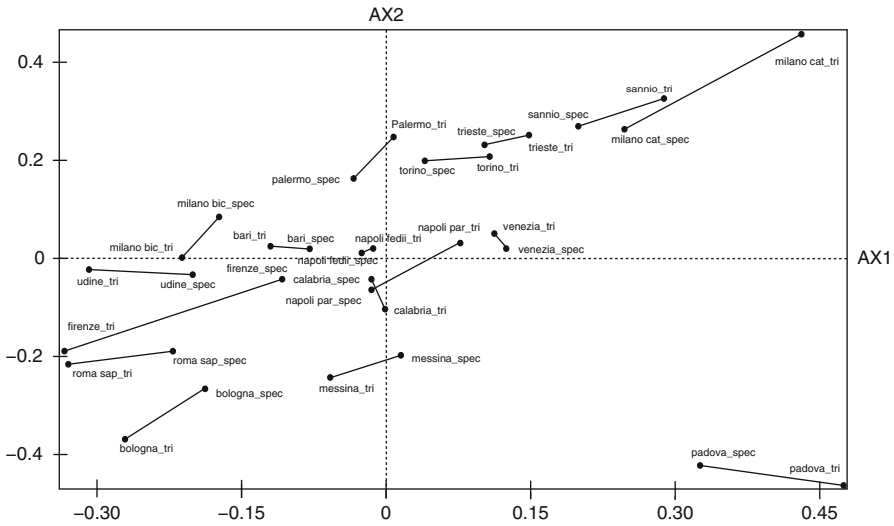
Fig. 13.3 The first factorial plane for Class 37 – Statistics and Classes 90S – Demographic and Social Statistics, 91S – Statistics for Experimental Research, 92S – Financial, Actuarial and Economic Statistics

For the class of degree under consideration (*Statistics*), thus two matrices were built, with the same number of rows, corresponding to the Universities where there are jointly activated degree courses in the 3 year degree and specialised degree courses in the connected classes and, in columns, the terms used in the course information data sheet about the education offer; in the first matrix about the 3 year degree and in the second one about the specialised degree programs. To the 3 year degree in statistics, many specialist degree courses can be connected, and only those where the word "statistics" appeared in the name were considered (Class 90S: *Demographic and Social Statistics*; Class 91S: *Statistics for Experimental Research*; Class 92S: *Financial, Actuarial and Economic Statistics*), involving a total of 17 Universities.

Once the latent structures were identified, the strength of the link for class of degree was measured, and the positions within the individual universities were shown (Fig. 13.3). In the figure, the final "tri"("spec") that follows the University name, indicates the 3 year (specialised) degree courses. The measure of fit is very low and amounts to 5.3 (the index is linked to the dimensions of the matrices analysed). The main variances are in the universities of Bologna, Milano Cattolica and Florence.

13.6 Final Remarks

From the whole of the analyses carried out, without simplifying the variety of the results obtained, some reflections have to be highlighted.

In predisposing the documentation, the universities show a substantial formal correspondence to the indications issued by the MIUR, even if sometimes we can note an excessive superficiality and, often, an excessive variability which is not always helpful in the choice. Furthermore, the areas where the model 3+2 was perceived more as extraneous show a greater difficulty in proposing more specific professional competences. In connection to the areas of employment expected, there doesn't seem to be a close consequence between the education offer and the employment possibilities indicated, except for some specific professional role. The analysis carried out has shown the necessity to dedicated more attention in the description of the professional roles at the end of the courses, trying to compare the competences offered to the students and what is effectively required in the job market, once they graduate. Finally, the connection between the 3 year degree courses and the specialist degree courses connected seems, in the cases examined, quite close; this positive judgement could, vice versa, turn itself into a contraindication in the light of the new reform, where this connection is, for some aspects, interrupted.

References

1. Balbi S (1995) Non symmetrical correspondence analysis of textual data and confidence regions for graphical forms. In: Bolasco S et al (eds) Actes des 3es JADT, Journées internationales d'Analyse statistique des Données Textuelles. CISU, Roma
2. Balbi S, Di Meglio E (2004) A text mining strategy based on local contexts of words. In: Purnelle G et al (eds) JADT04 Le poids des mots. Press Universitaire de Louvain, Louvain
3. Balbi S, Misuraca M (2005) Visualization Techniques in non symmetrical relationships. In: Sirmakessis S (ed) Knowledge mining (studies in fuzziness and soft Computing). Springer, Berlin
4. Grassia MG, Misuraca M, Scepi G (2004) Relazioni non Simmetriche tra Corpora. In: Purnelle G et al (eds) JADT04 Le poids des mots. Press Universitaire de Louvain, Louvain
5. Lauro NC, D'Ambra L (1984) L'analyse non symetrique des correspondances. In: Diday E et al (eds) Data analysis and informatics, vol III. North-Holland, Amsterdam
6. Lebart L, Salem A, Berry L (1998) Exploring textual data. Kluwer Academic Publishers, Dordrecht
7. Misuraca M (2005) La visualizzazione dell'informazione testuale. Contributi metodologici e applicativi, Tesi di Dottorato in Statistica, Dipartimento di Matematica e Statistica, Università di Napoli "Federico II"
8. Morrone A (1993) Alcuni criteri di valutazione della significatività dei segmenti ripetuti. In: Anastex SJ (ed) JADT93, Actes des secondes Journées Internationales d'Analyse Statistique de Données Textuelles, ENSTelecom, Paris