

FACULTY OF ECONOMICS

Master of Science in Statistics (MScSTAT)

UNIVERSITY OF NEUCHÂTEL

MASTER THESIS

A statistical approach based on the correspondence
analysis of two latin corpuses:
Historia Augusta and *Corpus Tibullianum*.

Author:

Vittoria SACCO

Supervisor:

Prof. Yves TILLÉ

July 10, 2011

DEDICATION

With my utmost appreciation to Yves Tillé, professor of statistics at the University of Neuchâtel (yves.tille@unine.ch), who has always been available to answer questions and has given me essential advice. My gratitude goes also to Jean-Jacques Aubert, professor of classical philology and ancient history at the University of Neuchâtel (jean-jacques.aubert@unine.ch), for having revised the Latin part. Without him some fundamental pieces could not have been found. A special thanks to all my family and my friends, in particular to Lina Djazzar, Xavier Guillaume, Walter Maldonado, Natasha Appadoo and Elvire Beziz and to my classmates Nadège Houlmann, Hugo Mabillard and Lara Prati who have encouraged me to do this master and to trust myself. My acknowledgment goes to my aunt Carmela Lavia who has offered me a semester of Master. And finally, to my boyfriend, Grégoire Minisini, who has supported me emotionally, giving me serenity and the strength to overcome even the hardest days.

ABSTRACT

Two latin corpora titled "*Historia Augusta*" and "*Corpus Tibullianum*" are analysed by means of lexicographic methods. In the first corpus, some proofs are searched to confirm or refute that there is only one writer and if he has used or not some perspectives in each biography. The results show no distinct groups in the graphic visualisation. The hypothesis of one writer seems most probable. Viewing the factorial plan, a "chronological axis" separates earlier emperors from later ones. Two forms "de" and "item" have an important contribution to the first factorial axis and are more present in the biographies of the latest emperors. The second corpus includes the works of several authors of Latin elegiac poetry. The first correspondence analysis has led to finding two extreme points: the "Panegyric of Messalla" and elegy 3.20. One can conclude that both works are not by Tibullus and that they stand a part from the rest of "*Corpus Tibullianum*". For this reason, they have been removed from the initial corpus. After a second correspondence analysis, one can see the cluster for each book in the factorial plan view. The elegies with anonymous writers are in the middle of the factorial plan, so they cannot be attributed with any certainty to a writer at the exception of elegies 3.9 and 3.19 that are in the Sulpicia's sub cluster. Moreover, the First Book seems posterior to the second one because of the presence of a "chronological axis". In both cases, the hypotheses put forward by philologists seem to be confirmed by our statistical approach.

TABLE OF CONTENTS

DEDICATION	i
ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
CHAPTER	
INTRODUCTION	ix
1 Short reminding of Simple Correspondence Analysis	2
1.1 Coordinates	2
1.2 Chi-square distance	6
1.3 Criterion of adjustment	7
1.4 Inertia	7
1.5 Simultaneous representation	8
1.6 Interpretation of correspondence analysis	11
2 Data and Text Mining (DTM)	12
2.1 All about DTM	12
2.2 Digitizing	13
2.3 Computational steps	14
3 Historia Augusta	16
3.1 Issue	16
3.2 Data	18
3.2.1 Texts	18

	Page
3.2.2 Coding of the texts	18
3.3 First Analysis	18
3.3.1 Results	18
3.3.2 Contribution of the axes	19
3.4 To sum up	23
3.5 Second Analysis	24
3.5.1 Results	24
3.5.2 Contribution of the axes	25
3.5.3 Interpretation	32
3.6 Conclusion	39
4 Corpus Tibullianum	41
4.1 Issue	41
4.2 Data	45
4.2.1 Texts	45
4.2.2 Coding of the texts	46
4.3 First Analysis	46
4.3.1 Results	46
4.3.2 Contribution of the axes	48
4.3.3 Interpretation	54
4.4 To sum up	57
4.5 Second Analysis	57
4.5.1 Results	57
4.5.2 Contribution of the axes	60
4.5.3 Interpretation	65
4.6 Conclusion	68

LIST OF REFERENCES 71

LIST OF TABLES

Table		Page
1.1	Contingency table	3
1.2	Contingency table - Example	3
1.3	Table of frequencies	4
1.4	Table of frequencies - Example	5
1.5	Table of rows- <i>profiles</i>	5
1.6	Table of columns- <i>profiles</i>	5
3.1	Eigenvalues and Percentage of variance	20
3.2	Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order).	21
3.3	Principal parameters of correspondence analysis: contribution of words in building the second axis (decreasing order).	22
3.4	Words (frequency order)	25
3.5	Eigenvalues and Percentage of variance	26
3.6	Principal parameters of correspondence analysis: contribution of texts in building the first axis (decreasing order).	28
3.7	Principal parameters of correspondence analysis: contribution of texts in building the second axis (decreasing order).	29
3.8	Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order).	30
3.9	Principal parameters of correspondence analysis: contribution of words in building the second axis (decreasing order).	31
3.10	Contingency table	33
3.11	Lexical fields	38
4.1	Eigenvalues and Percentage of variance	47
4.2	Principal parameters of correspondence analysis: contribution of texts in building the first axis (decreasing order).	49

Table	Page
4.3	Principal parameters of correspondence analysis: contribution of texts in building the second axis (decreasing order). 50
4.4	Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order). 53
4.5	Vocabulary present only in elegy seven of the third book 55
4.6	Piece of the contingency table 58
4.7	Eigenvalues and Percentage of variance 59
4.8	Principal parameters of correspondence analysis: contribution of texts in building the first axis (decreasing order). 61
4.9	Principal parameters of correspondence analysis: contribution of texts in building the second axis (decreasing order). 62
4.10	Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order). 63
4.11	Principal parameters of correspondence analysis: contribution of words in building the second axis (decreasing order). 64
4.12	Piece of the contingency table 66

LIST OF FIGURES

Figure		Page
3.1	First factorial plan: frequencies classified by texts. Hypotheses view.	35
3.2	First factorial plan: frequencies classified by texts. Chronological order.	37
3.3	Correspondence analysis of the table texts-words.	39
4.1	First factorial plan: frequencies classified by texts.	56
4.2	First factorial plan: frequencies classified by texts. Hypotheses view.	67
4.3	Correspondence analysis of the table texts-words.	69

INTRODUCTION

After a discussion between Professor Tillé and Professor Aubert about statistical methods for determining the authorship, Professor Aubert has proposed to analyse two Latin corpora. He hopes to find something to support or to refute the current hypotheses and that would help the research considering the controversial nature of these manuscripts. Since I had expressed the desire to do something original as subject of my master thesis and which would require a method also used in economics, my starting field, my supervisor, Professor Tillé, proposed me these issues. At the beginning, I was a little bit scared of my recollection of Latin grammar, vocabulary and civilisation learned in high school. But, I have immediately accepted with enthusiasm the topic and was glad to work not just with one but two professors. Furthermore, if the results had been satisfactory I could have really helped Latinists who had been working for years on these manuscripts.

To begin with, there is a brief introduction about the tools that hide behind correspondence analysis, the chosen method. The next section explains which software is used to realize the analyses and its advantages and disadvantages. Finally, the analyses, the results and the conclusions about the two Latin's corpora are presented.

The first corpus, titled "*Historia Augusta*", will immerse the reader in the Imperial Rome, with its rules and its usurpers and background of political instability, lethal gossips and brutal conquests resulting into one of the greatest empire in the History. "*Historia Augusta*" includes thirty biographies of Roman emperors from Hadrian (117-138) to the advent of Diocletian (284). Using statistical methods, more precisely the correspondence analysis, the aim is to determine if these biographies have all the same writer or not and if the writer adopts various postures represented by different names (Aelius Spartianus, Julius Capitolinus, Vulcacius Gallicanus, Aelius Lampridius, Trebellius Pollio and Flavio Vopiscus). The first attempt at correspondence analysis failed as it was distorted by the overwhelming occurrences of the respective emperors first names. Each biography lists all members of the emperors family and their ancestors. Furthermore, the same praenomen usually recurs every other generation. This explains the high frequency of first names

and the connection of specific names with a single imperial dynasty. Consequently, the emperors first names have been removed from the corpus. After having made a second correspondence analysis with the corrected texts, some forms of words have been found that can help to drawing important conclusions. The use of “item” reflects alteration of the original text and suggests some kind of reworking or possibly reveals the addition in some biographies of a subjective element which does not reflect the historians/authors initial perspective. The use of “de” followed by an ablative instead of the archaic/classical genitive, contributes most significantly to axes building. The words “item” and “de” are associated in the factorial plan view with the latest emperors. This seems to emphasize a timeline in the writing of the biographies which opposes the earlier emperors to the later ones in the visualisation. Furthermore, the absence of distinct groups when viewing the factorial plan implies that the hypothesis of one writer made by historians cannot be rejected. The initial postures seem to be a stylistic device used by the same author.

The second corpus brings the reader even further back in time to meet one of the major eulogists. It includes the works of one or several major authors of Latin elegiac poetry. The “*Corpus Tibullianum*” is thought to include works not only by Tibullus, but also by some other poets of the so-called Messalla’s circle. Some poems may have been written by Sulpicia, making her the only Latin poetess whose work would have survived from the overall loss of classical literature. This corpus is divided into three books (or four, depending on which edition is used). The First and Second Books contain poems attributed solely to Tibullus. The Third (and, respectively, Fourth) includes elegies by one Lygdamus and by Sulpicia, in addition to the so-called “Panegyric of Messalla” and other poems of dubious origins. The statistical approach applied here aims at testing current hypotheses concerning the supposed authorship of each of the pieces contained in those three books. Accessorily, some attention will be paid to the specificity of those poems usually considered spurious. The first correspondence analysis highlighted two extreme points that blur the position of other elegies. Unsurprisingly, the “Panegyric of Messalla” has its own vocabulary and its topic is drastically different from the rest of the corpus. The other extreme point is Elegy 3.20, usually attributed to “*incerti autores*”,

which likewise displays a specific vocabulary. Words that are common in other elegies are totally absent from this one. One can conclude that both works are not by Tibullus and that they stand apart from the rest of the Corpus Tibullianum, which has then to be submitted to a second correspondence analysis. As a result, three clusters are visible in the plan view, corresponding to the above-mentioned three books. Within the Third one, sub clusters emerge and reflect the respective contributions by Sulpicia on the one hand, and Lygdamus on the other. Elegies by anonymous authors stand right in the middle of the factorial plan and therefore cannot be attributed with any certainty to any specific author. Exceptions are Elegies 3.9 and 3.19, within Sulpicia's sub cluster. Books One and Two seem to have been written in a chronological sequence: Book One appears to be earlier than Book Two as shown by the position of their respective elegies with regard to the axes. The choice of words and the apparent thematic development are consistent with those statistical data. In both cases, the hypotheses put forward by philologists seem to be confirmed by our statistical approach.

CHAPTER 1

Short reminding of Simple Correspondence Analysis

Correspondence analysis is an exploratory data analytic technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between rows and columns. Correspondence analysis has a high flexible data requirement. The only restriction is a rectangular data matrix with non-negative entries. In comparison with other methods yielding joint graphical displays, an advantage of correspondence analysis is that it produces two dual displays whose row and column geometries have similar interpretations, facilitating analysis and detection of relationships. Its primary goal is to transform a table of numerical information into a graphical display in which each row and each column is depicted as a point. Correspondence analysis shows the way variables are related and not just that a relationship exists.

1.1 Coordinates

Correspondence analysis is also a method for analysing qualitative variables. Throughout the presented analyses, the contingency table will be a particular case: the individual, giving rise to statistical counting for each cell of the table, will be the occurrence of a textual unit (form in our case). Since we have not chosen to do stem (see section “Computational steps”) each distinct and unmistakable word will be considered as a new form.

The contingency table has I rows and J columns. Each row i corresponds to a form and each column j corresponds to a text of the corpus. The n_{ij} denotes the number of times form i appears in text j . The observed data can be presented as in Table 1.1.

The marginal frequencies are denoted by $n_{i.}$ and $n_{.j}$ where:

$$n_{i.} = \sum_{j=1}^J n_{ij}$$

Table 1.1. Contingency table

Rows (words)	Columns				
	text 1	text 2	...	text J	Total
form 1	n_{11}	n_{12}	...	n_{1J}	$n_{1.}$
form 2	n_{21}	n_{22}	...	n_{2J}	$n_{2.}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
form I	n_{I1}	n_{I2}	...	n_{IJ}	$n_{I.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.J}$	n

is the total number of times that the form i appears in the corpus.

$$n_{.j} = \sum_{i=1}^I n_{ij}$$

is the total number of forms in a text j .

Moreover, let

$$n = \sum_{i=1}^I \sum_{j=1}^J n_{ij}$$

be the total number of forms in the corpus.

Taking into account what has been introduced before, a numerical example is presented. Imagine that you are interested in finding a relation between three biographies (Caracalla, Hadrianus and S.Severus) and the words “et” and “imperatores”. The Table 1.2 will be the new contingency table (the used numbers are fictitious).

Table 1.2. Contingency table - Example

Rows	Columns			
	Caracalla	Hadrianus	S.Severus	Total
et	25	35	15	75
imperatores	5	20	10	35
Total	30	55	25	110

The relative frequencies are noted as follows:

$$f_{ij} = \frac{n_{ij}}{n} \quad \left(\sum_{i=1}^I \sum_{j=1}^J f_{ij} = 1 \right)$$

Similarly,

$$f_{i.} = \sum_{j=1}^J f_{ij} = \frac{n_{i.}}{n}$$

can be also named the mass of the j th column.

$$f_{.j} = \sum_{i=1}^I f_{ij} = \frac{n_{.j}}{n}$$

can be also named the mass of the i th row.

Table 1.3 gives the contingency table of frequencies:

Table 1.3. Table of frequencies

Rows	Columns				Total
	text 1	text 2	...	text J	
form 1	f_{11}	f_{12}	...	f_{1J}	$f_{1.}$
form 2	f_{21}	f_{22}	...	f_{2J}	$f_{2.}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
form I	f_{I1}	f_{I2}	...	f_{IJ}	$f_{I.}$
Total	$f_{.1}$	$f_{.2}$...	$f_{.J}$	1

Table 1.4 contains the table of frequencies that correspond to the example given in Table 1.2. For instance, in order to get 0.23, 25 must be divided by 110.

The j th component of the i th vector has been taken as:

$$\frac{n_{ij}}{n_{i.}} \quad \text{for } i = 1, 2, \dots, I$$

This is called the *profile* of row i which is a vector of conditional densities.

The i th component of the j th vector are defined as:

$$\frac{n_{ij}}{n_{.j}} \quad \text{for } j = 1, 2, \dots, J$$

Table 1.4. Table of frequencies - Example

Rows	Columns			
	Caracalla	Hadrianus	S.Severus	Total
et	0.23	0.32	0.14	0.69
imperatores	0.04	0.18	0.09	0.31
Total	0.27	0.5	0.23	1

This is called the *profile* of column j which is a vector of conditional densities.

Table 1.5 corresponds to the table of rows-*profiles* and Table 1.6 corresponds to the table of columns-*profiles* applied to the above mentioned example. For example, in order to

Table 1.5. Table of rows-*profiles*

Rows	Columns			
	Caracalla	Hadrianus	S.Severus	Total
et	0.33	0.47	0.2	1
imperatores	0.13	0.58	0.29	1
Total	0.27	0.5	0.23	1

get 0.33, 0.23 must be divided by 0.69 (Table 1.5) and 0.85 is obtained by dividing 0.23 by 0.27 (Table 1.6).

Table 1.6. Table of columns-*profiles*

Rows	Columns			
	Caracalla	Hadrianus	S.Severus	Total
et	0.85	0.64	0.61	0.69
imperatores	0.15	0.36	0.39	0.31
Total	1	1	1	1

1.2 Chi-square distance

In correspondence analysis a distance is defined between profiles. A table of distances can thus be constructed separately for the row *profiles* and for the column *profiles*. In correspondence analysis, in place of the Euclidian distance, the Chi-square distance is used. The Chi-square distance between two row *profiles* i and i' is given by:

$$d^2(i, i') = \sum_{j=1}^J \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2$$

In symmetric fashion, the Chi-square distance between two column *profiles* j and j' is written as:

$$d^2(j, j') = \sum_{i=1}^I \frac{1}{f_{.i}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2$$

In fact, this distance only differs from the usual Euclidean distance in that each square is weighed by the inverse of the frequency corresponding to each term. Essentially, the reason for choosing the Chi-square distance is that it verifies the properties of distributional equivalency, expressed as follows:

1. If two columns having identical *profiles* are aggregated, then the distances between rows remain unchanged.
2. If two rows having identical distribution *profiles* are aggregated, then the distances between columns remain unchanged.

The property is important; it guarantees a satisfactory invariance of the results irrespective of how the variables were originally coded. If one applies these formula for the example in Table 1.4 one can find for the rows' distance:

$$\begin{aligned} d^2(et, imperatores) &= \frac{1}{0.27} \cdot \left(\frac{0.23}{0.69} - \frac{0.04}{0.31} \right)^2 + \frac{1}{0.5} \cdot \left(\frac{0.32}{0.69} - \frac{0.18}{0.31} \right)^2 + \frac{1}{0.23} \cdot \left(\frac{0.14}{0.69} - \frac{0.09}{0.31} \right)^2 \\ &\approx 0.461 \end{aligned}$$

In the same way one can find de columns' distance.

1.3 Criterion of adjustment

The goal is to obtain an approximate representation of the distances in dimension two. Geometrically all the row *profiles* can be represented in a space of dimension $N(I;J)$. The set of *profiles* is called the cloud of points. Thus, the proximities between the *profiles* and the average *profile* defined on the overall population can be represented. So, the cloud of points is considered as centered on its center of gravity. Choosing the *profiles* as coordinates gives the same importance to all modalities. However, the importance is returned through a “mass” assigned to each point (that is proportional to its frequency). It is enable to avoid any privilege to the class with low effectives and to respect the real repartition of the population. This “mass” intervenes in the calculus for estimating the coordinates of the center of gravity of the cloud and for the criterion of adjustment. For the calculus of adjustment, the quantity to maximize is the weighted sum of the square of the distances between the points and the center of gravity of the cloud using the Chi-square distance (i.e. the inertia of the maximum elongation of the line of the cloud).

1.4 Inertia

The inertia is an indicator of the dispersion of the cloud of points and measures the relation existing between two selected variables. There is a cloud of *profile* points with masses adding up to 1. These points have a centroid (i.e., the average *profile*) and a distance (Chi-square distance) between *profile* points. Each *profile* point contributes to the inertia of the whole cloud.

The total inertia of the contingency table is given by the following formula:

$$Total\ inertia = \sum_{i=1}^I f_i d^2(i, G) = \sum_{j=1}^J f_j d^2(j, G) = \sum_{i=1}^I \sum_{j=1}^J \left(\frac{f_{ij} - f_i f_j}{f_i f_j} \right)^2$$

where

$$d^2(i, G) = \sum_{j=1}^J \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - f_j \right)^2$$

and

$$d^2(j, G) = \sum_{i=1}^I \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - f_i \right)^2$$

and G represents the center of gravity. See for instance, Lebart et al. (2006).

1.5 Simultaneous representation

Now, the best simultaneous representation of the rows and columns of the initial data matrix is searched. This is equivalent to representing on the same axis the entire set of rows and columns in order to be close to the ideal situation explained as follows:

1. Each column point j is the barycenter of the row points i , these latter being affected by a mass proportional to their importance in the modality j . This mass is equal to $p_i = \frac{f_{ij}}{f_{.j}}$. For each column j , these masses are the *columns-profiles* of the initial data matrix with $\sum_{i=1}^I p_i = 1$.
2. Each row point i is the barycenter of the column points j , these latter being affected by a mass representing the slice of the modality j in the modality i . This mass is equal to $q_j = \frac{f_{ij}}{f_{i.}}$. For each row i , these masses are the *rows-profiles* of the initial data matrix with $\sum_{j=1}^J q_j = 1$.

So, relations that are strictly barycentric between the two sets are defined. The interpretation of the cross-proximity is not allowed between a row point and a column point because the two points are not in the same initial space. However, the position of a row point compared to all the points of the column (or vice-versa) is possible to be interpreted. The main reason of this simultaneous representation is given by a transition relation linking coordinates of a point in one space to those of all the points in the other space.

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1J} \\ f_{21} & f_{22} & \dots & f_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ f_{I1} & f_{I2} & \dots & f_{IJ} \end{pmatrix}$$

$$\mathbf{D}_i = \begin{pmatrix} f_{1.} & 0 & \dots & 0 \\ 0 & f_{2.} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & f_{I.} \end{pmatrix}$$

$$\mathbf{D}_j = \begin{pmatrix} f_{.1} & 0 & \dots & 0 \\ 0 & f_{.2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & f_{.J} \end{pmatrix}$$

$$\mathbf{S} = \mathbf{F}'\mathbf{D}_i^{-1}\mathbf{F}\mathbf{D}_j^{-1}$$

$$\begin{aligned} \mathbf{S} &= \begin{pmatrix} f_{11} & f_{21} & \dots & f_{I1} \\ f_{12} & f_{22} & \dots & f_{I2} \\ \vdots & \vdots & \ddots & \vdots \\ f_{1J} & f_{2J} & \dots & f_{IJ} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{f_{.1}} & 0 & \dots & 0 \\ 0 & \frac{1}{f_{.2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{f_{.J}} \end{pmatrix} \cdot \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1J} \\ f_{21} & f_{22} & \dots & f_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ f_{I1} & f_{I2} & \dots & f_{IJ} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{f_{.1}} & 0 & \dots & 0 \\ 0 & \frac{1}{f_{.2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{f_{.J}} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^I \frac{f_{i1}^2}{f_{i.}f_{.1}} & \dots & \sum_{i=1}^I \frac{f_{i1}f_{iJ}}{f_{i.}f_{.J}} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^I \frac{f_{iJ}f_{i1}}{f_{i.}f_{.1}} & \dots & \sum_{i=1}^I \frac{f_{iJ}^2}{f_{i.}f_{.J}} \end{pmatrix} \end{aligned}$$

and

$$\mathbf{T} = \mathbf{F}\mathbf{D}_j^{-1}\mathbf{F}'\mathbf{D}_i^{-1}$$

$$\begin{aligned} \mathbf{T} &= \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1J} \\ f_{21} & f_{22} & \dots & f_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ f_{I1} & f_{I2} & \dots & f_{IJ} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{f_{.1}} & 0 & \dots & 0 \\ 0 & \frac{1}{f_{.2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{f_{.J}} \end{pmatrix} \cdot \begin{pmatrix} f_{11} & f_{21} & \dots & f_{I1} \\ f_{12} & f_{22} & \dots & f_{I2} \\ \vdots & \vdots & \ddots & \vdots \\ f_{1J} & f_{2J} & \dots & f_{IJ} \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{f_{.1}} & 0 & \dots & 0 \\ 0 & \frac{1}{f_{.2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{f_{.J}} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^J \frac{f_{1j}^2}{f_{.j}f_{.1}} & \dots & \sum_{j=1}^J \frac{f_{1j}f_{Ij}}{f_{.j}f_{.I}} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^J \frac{f_{Ij}f_{1j}}{f_{.j}f_{.1}} & \dots & \sum_{j=1}^J \frac{f_{Ij}^2}{f_{.j}f_{.I}} \end{pmatrix} \end{aligned}$$

So, the principal axes are defined by:

$$\mathbf{S}\mathbf{u}_\alpha = \lambda_\alpha \mathbf{u}_\alpha$$

and

$$\mathbf{T}\mathbf{v}_\alpha = \lambda_\alpha \mathbf{v}_\alpha$$

For any α ($\alpha = 1, \dots, \min(I; J)$), u_α and v_α are the eigenvectors corresponding to the eigenvalues λ_α . \mathbf{F} represents the table of relative frequencies, \mathbf{D}_i is the diagonal matrix whose diagonal elements are f_i . and \mathbf{D}_j is the diagonal matrix whose diagonal elements are f_j . In addition, for example for \mathbf{S} , the matrix $\mathbf{F}'\mathbf{D}_i^{-1}$ contains the row *profiles*, whereas the matrix $\mathbf{F}\mathbf{D}_j^{-1}$ contains the column *profiles*. If $\phi_{i\alpha}$ is the coordinate of a row point i in the horizontal axis α and if $\psi_{j\alpha}$ is the coordinate of a column point j in the same axis α thus we have a system of symmetric relations:

$$\phi_{i\alpha} = \beta \sum_{j=1}^J \left(\frac{f_{ij}}{f_i} \right) \psi_{j\alpha} \quad (1)$$

$$\psi_{j\alpha} = \beta \sum_{i=1}^I \left(\frac{f_{ij}}{f_j} \right) \phi_{i\alpha} \quad (2)$$

Without the β coefficient, the *profile* of rows will be the barycentre of the *profile* of columns and vice versa. So, for relations (1) and (2) to be possible simultaneously, the coefficient β must be positive and greater than 1. Finally for any axis, we have the relation:

$$\beta = \frac{1}{\sqrt{\lambda_\alpha}}$$

where λ is the eigenvalue of the array when the weighted sum of the squares of the projections on the axes is maximized and the maximum axis of inertia of the cloud of the row-*profiles* (or column-*profiles*) is searched through the origin O. A demonstration of those properties is provided in Lebart (2010). Correspondence analysis can be presented as the research of the values of $\phi_{i\alpha}$ and $\psi_{j\alpha}$ corresponding to the smallest dilatator coefficient β . The formulas (1) and (2) are also true for the coordinates on the vertical but for different values of $\phi_{i\alpha}$, $\psi_{j\alpha}$ and β . Focusing only on the overall dispersion of the cloud of points is not enough but it is necessary to look at the existence of preferred direction of this cloud. To do this, the inertia of each axis (eigenvalues) and the corresponding percentage of inertia are consulted. This allows to deduce the shape of the cloud (spherical if there is a privilege direction or not spherical in the other cases).

1.6 Interpretation of correspondence analysis

The interpretation of the results of correspondence analysis comprises the interpretation of numerical results and factor graphics yielded by correspondence analysis. The former implies selection of significant axes and significant points (significant means “necessary to study in detail”; not in terms of statistical significance tests). How to interpret tables and visual representations of the factorial plan is explained throughout the next chapters.

CHAPTER 2

Data and Text Mining (DTM)

2.1 All about DTM

DTM is a statistical software used for analysing numeric and textual data. Its application concerns primarily the processing of responses to open ended questions in socio-economic sample surveys. The design and algorithms are created by Ludovic Lebart (Telecom-ParisTech, Paris, France) in collaboration with André Salem, (Univ. Paris 3, France), Monique Becue, (UPC, Barcelona, Spain) and Marie Piron, (IRD, Bondy, France). The software is free and can be downloaded from Lebart et al. (2010b).

To process simultaneous numerical and textual data, DtmVic offers different tools:

1. Complementary use of visualization techniques (Principal Component Analysis, Two-way and Multiple Correspondence Analysis) and clustering techniques (hybrid method using both hierarchical clustering and k -means technique; Self Organizing Maps: SOM, minimum spanning tree).
2. Assessments of visualization techniques: re-sampling techniques (bootstrap, partial bootstrap, total bootstrap, bootstrapping variables). Three kinds of total bootstrap are provided: total bootstrap type 1 (simple change of sign of the axes for the replicated analyses, when needed, total bootstrap type 2 (as type 1 with a correction for possible interventions of axes in addition) and total bootstrap type 3 (corrections of the replicated principal subspaces using procrustean analyses).
3. Contiguity analysis and related methods: Kohonen maps (SOM). These techniques, exemplified in example C3 of Tutorial C in a simple case, act as a missing link between principal axes methods and clustering.

In the recent version of the software, the set of bootstrap possibilities is extended. In the case of open ended questions, the statistician is dealing with two kinds of statistical units: on the one hand, the respondents and on the other, the words (or tokens). In the case of

correspondence analysis of a lexical table cross-tabulating, both units (respondents and words) can be bootstrapped and led to distinct statistical inferences.

The user can run the software on his data sets after changing a few parameters and complying with the input formats shown in the examples (data, dictionary and texts). Several importation procedures can make much easier the input of these data. For numerical data, the most frequent input is one (or several) Microsoft Excel, (r) file or Calc file (counterpart of Excel in the free package Open Office). The present version of this free software contains a series of typical examples of data processing and textual data processing.

Since June 2009 the user can consult:

- Tutorial A where five introductory commented application examples are available,
- Tutorial B which has four more advanced application examples relating to textual data,
- Tutorial C contains five advanced application examples relating to numerical data,
- and Tutorial D which is constituted by five examples of data importation.

These tutorials can be read directly from the main menu. They are also available as pdf files.

DTM presents some limitations:

- 22.500 as the maximum number of respondents (individuals, rows),
- 1.000 as the maximum number of variables (numerical or categorical),
- and 100.000 as the maximum number of characters for the response of an individual to a set of open-ended questions.

2.2 Digitizing

This method consists of ignoring the spelling form of word detected during the computational steps to retain only a number that will be associated with all the occurrences of this form. These numbers will be stored in a dictionary of forms specific to each

analysis. This allows reconstructing the graphic forms of the texts put in evidence by statistical calculus.

Another point concerns stemming: the process for reducing inflected words to their stem, base or root form. The stem is the form of a word after all affixes are removed. Verbs can be written in their infinitive form, substantives in singular, adjectives in masculine singular and some elided forms in no declined form. For example, a stemmer for English should identify the string “stemmer”, “stemming”, “stemmed” as based on “stem”. In all the presented analyses stemming is not used. The first reason is that we are not Latinists so for us it could be really hard and perhaps it would lead to mistakes or no sense. The second reason is that some researches seem to show that we lead to the same results even without using it Benzécri (1991).

2.3 Computational steps

Following the Tutorial A Lebart et al. (2010a), for each application one directory is created where in a second step the software will save different txt files containing the results. The text format type 1 is used because texts have free format. Since the texts are of different length, separators “*****” are used to distinguish between texts. The symbol “====” indicates the end of the file which is a text file (.txt): now DTM can be used.

After having opened the software, you can click on “data importation,..., exportation” and choose “importing dictionary, data and text” and then “textual data (free format)”. At this step, with the button “open text file”, the dataset is opened. The second step consists in doing the analysis. By coming back to the main menu you can go to “create a command file”. “VISUTEXT” is chosen since a simple correspondence analysis for texts format will be performed. The dataset of interesting must be specified with the key “open a text file”. The next step consists in selecting open questions and separators of words like: “.:;?!-()+=*”. Clicking on “Vocabulary and counts” the list of frequencies for each word in alphabetical order and in frequency order is obtained. The next step is maybe the most delicate one: a frequency threshold must be chosen. From

a statistical point of view, according to the reference book Lebart and Salem (1988a), using the correspondence analysis for the comparison of lexical profiles has sense only if the different forms of words appear with a certain frequency. So, at this point, the nonce words and also the rare forms would be taken off. Thus, the vocabulary considered will be really reduced. For all analyses the frequency threshold will be set at two (the words must appear in the corpus at least twice). The other steps are intuitive. You press on “continue”, then “create a first parameter file” and at the end “execute”. All the results are saved in the directory that was created at the beginning. To do the analysis all the tables that are in the directory are transferred in a Microsoft Excel document in order to store the results according to the criterions of interpretation.

Now you can look at the different graphs. After having returned to the main menu you click on “visualization, inference and classification” and choose the more convenient way to look the analyses. According to the cases to be presented, this will be “Plan view” and then “Active column” to see only the position of the texts and then “Active column + row” to see texts and words in two dimensional space. Thanks to these calculus the best representation of a multidimensional problem in a two dimensional space is obtained.

CHAPTER 3

Historia Augusta

3.1 Issue

The first corpus is a Latin handiwork which is a compilation of thirty biographies of Roman emperors, including aspiring emperors and usurpers, from Hadrian (177-138) to the advent of Diocletian (284). This work, known by the name of “*Historia Augusta*”, was allegedly written in the late third or early fourth century, although a later date is nowadays widely accepted. It covers the period from 117 to 285 AD, with a gap between the years 244-260 (lacking essentially the biographies of Philip the Arab and Decius). These texts relate emperors’ lives. They trace their origins, how they became emperors, how they ruled and the administered national and international relations. But they also portray these emperors’ character and privacy.

A complex and controversial problem lies behind the pages of this work. The historians have claimed that the writer had used six postures with six different identities and status (for instance, a soldier, a servant and a slave) which are:

- **Aelius Spartianus** would have written the lives of Hadrian, Aelius, Didius Julianus, Septimius Severus, Pescennius Niger, Caracalla and Geta;
- **Julius Capitolinus** would have reported the lives of Antoninus Pius, Marcus Aurelius, Verus, Pertinax, Clodius Albinus, Macrinus, the Maximini, the Gordians, Maxim and Balbin;
- **Vulcacius Gallicanus** would have only described the active life of Avidius Cassius;
- **Aelius Lampridius** is said to be the author of the lives of Commodus, Diadumenianus, Elagabalus and Severus Alexander;
- **Trebellius Pollio** would have composed the biographies of Valerian, of Gallieni, the Thirty Tyrants and Claudius the Goth;

- **Flavio Vopiscus** is said to have written the lives of Aurelian, Tacitus, Probus, the Quadriga of the Tyrants, Carus and his son.

All this has been accepted, albeit with reluctance, until H. Dessau's intervention in the late nineteenth century. In 1889 this pupil of Mommsen published an article that undermined everything that had been more or less admitted that far. For him, the six writers had never existed. A single person would have written these biographies not under Diocletian and Constantine, but in the late fourth century, at the time of Theodosius I. This thesis, which of course prompted reactions in different directions, is now widely accepted. The "*Historia Augusta*" is a hoax that deserves to be analyzed: first it is necessary to assess the historical value and then try to discover the goals pursued by the "forger".

Totally unaware of prejudices, the goal consists in determining by statistical methods which of these theses seems to be the more trustable. To do this, a first correspondence analysis is performed. There has been made no distinction between biographies in order to see which of them are grouped and which of them are remote in the factorial space. The first analysis also allows seeing if there are some extreme points or some distortions in the initial data set. The result thus obtained is not comprehensive because of the interaction of the emperors' first names (this is discussed later on in the section "First Analysis"). This implies the analysis to be carried out again after removing the emperors' first names. It is after the second analysis that a conclusion will be drawn. By the absence of distinct groups the hypothesis of several writers and even that of different points of view are rejected. It seems that the most recent hypothesis is the closest one to reality. But some words that separate the biographies in two groups seem to underline the presence of a "chronological axis" that suggests two different times of writing. Furthermore, the presence in some biographies of the adverb "item" supports the idea of a reworking or perhaps of different points of view used by the same writer but which would not correspond to those given by historians.

3.2 Data

3.2.1 Texts

The “*Historia Augusta*” can be downloaded from the website The Latin Library - *Historia Augusta* (2008). It is divided into different biographies classified in chronological order. There is no translation and each phrase is numbered. The website Remacle et al. (2009) provides a translation of the corpus in French. The Hannick (2008) and *Historia Augusta* (2010) websites contain instead some basic information about the corpus. Some of the ideas for this analysis and interpretation have been inspired by Adams (1975) and Weil et al. (1976).

All the biographies are copy-pasted in a Microsoft Word document and all the notations that are meaningless are removed. The sample is composed of thirty texts which are of different length. The shortest one has only two pages, this is the biography of Septimus Severus, whereas the longest one has thirty pages and it concerns the emperor Divus Aurelianus.

3.2.2 Coding of the texts

The software used makes distinctions between capital and lowercase letter. For example, “Et” and “et” are two different words. To make the work easier and more relevant, the capital letters are removed from the original texts. In addition, some symbols used by Latinists that helped them to better understand the texts and that cut words into more pieces are taken out. After that the biographies are copy-pasted in a notepad document. Each biography bears the names of the emperors it refers to and is preceded by the symbol “****” as it is said in tutorial A Lebart et al. (2010a). Up to this phase, the hypotheses of authorships have not been taken into account. All the biographies are in the same document and the last one ends with the symbol “====” so that the software can read the document. From this point on, the analysis can start.

3.3 First Analysis

3.3.1 Results

Summary

The software has detected:

- 7711 as the total number of words,
- 3376 as the number of distinct words,
- and 43.8 as the percentage of distinct words.

Eigenvalues

The eigenvalues, bounding between 0 and 1 in correspondence analysis, presented in Table 3.1, are $\lambda_1 = 0.3013$ for the first axis and $\lambda_2 = 0.2848$ for the second axis. The percentage of variance or the percentage of inertia (ratio of each eigenvalue to its global sum) corresponding to these eigenvalues is respectively 4.98% for the first axis and 4.71% for the second axis. The percentage of variance measures the relative importance of each eigenvalue in the trace. In this case, the first factorial plan “explains” 9.69% of the total variance which is not great. The column named “Stars” represents the histogram of the eigenvalues. This column is the origin of the empiric procedure to judge the number of axis to retain. This histogram is studied in order to detect a changing on the slope. Each time that the histogram of eigenvalues presents a discontinuity one might suppose that something not random has interfered.

3.3.2 Contribution of the axes

This is where the reading and interpretation of Tables 3.2 and 3.3 are explained (Lebart and Salem (1988b)).

- The column “**Weight**” is about the margins of rows and columns.
- “**Disto**” treats the distance from the origin and contains the square distances from the axis’ origin, i.e. the distances of each *profile*, middle *profile* or margin. For example, in the first Table the word “Iulianus” has a really different *profile* comparing to the margin (71.07 as distortion). The same conclusion can be drawn for “dixisse” (42.84 as distortion) and “Maximus” (31.08 as distortion) in the second Table .

Table 3.1. Eigenvalues and Percentage of variance

Number	Eigenvalue	Percent.	Cumulat. Percent.	Stars
1	0.3013	4.98	4.98	80
2	0.2848	4.71	9.69	77
3	0.2748	4.54	14.23	74
4	0.2628	4.34	18.57	71
5	0.2563	4.24	22.81	70
6	0.2471	4.08	26.90	67
7	0.2424	4.01	30.90	66
8	0.2359	3.90	34.80	64
9	0.2307	3.81	38.61	63
10	0.2215	3.66	42.28	60
11	0.2199	3.63	45.91	60
12	0.2141	3.54	49.45	58
13	0.2117	3.50	52.95	58
14	0.2096	3.46	56.41	57
15	0.2052	3.39	59.80	56
16	0.2009	3.32	63.13	55
17	0.1933	3.20	66.32	53
18	0.1927	3.19	69.51	53
19	0.1867	3.09	72.59	51
20	0.1835	3.03	75.63	50
21	0.1808	2.99	78.62	50
22	0.1789	2.96	81.57	49
23	0.1752	2.90	84.47	48
24	0.1702	2.81	87.28	47
25	0.1661	2.74	90.03	46
26	0.1640	2.71	92.74	45
27	0.1544	2.55	95.29	42
28	0.1517	2.51	97.80	41
29	0.1333	2.20	100.00	36

Table 3.2. Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
Iuliano	0.001	25.24	2.35	-0.5	1.4	0.1	0.22	0.01
Severus	0.002	5.96	1.49	-0.37	1.4	0.1	0.37	0.02
item	0.002	9.35	-1.58	0.34	1.3	0.1	0.27	0.01
Albini	0	31.88	2.76	-0.31	1	0	0.24	0
Brittannia	0	31.88	2.76	-0.31	1	0	0.24	0
Iulianus	0.001	71.07	1.95	-0.7	1	0.1	0.05	0.01
Romam	0.002	4.86	1.34	-0.17	1	0	0.37	0.01
Septimum	0	31.88	2.76	-0.31	1	0	0.24	0
civile	0	31.88	2.76	-0.31	1	0	0.24	0
Claudium	0.001	21.97	-2.19	0.47	0.9	0	0.22	0.01
de	0.013	0.54	-0.46	-0.21	0.9	0.2	0.39	0.08
Cari	0.001	26.98	-1.73	-0.12	0.8	0	0.11	0
Claudio	0.001	8.74	-1.55	-0.13	0.8	0	0.27	0
in	0.031	0.19	0.27	-0.09	0.8	0.1	0.4	0.04
quasi	0.001	15.36	1.54	-0.16	0.8	0	0.16	0
Albinum	0.001	16.57	1.89	-0.23	0.7	0	0.22	0
Didio	0	25.24	2.35	-0.5	0.7	0	0.22	0.01
Gallieni	0.002	8.66	-1	-1.28	0.7	1.2	0.11	0.19
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 3.3. Principal parameters of correspondence analysis: contribution of words in building the second axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
Gordiani	0.002	12.15	-0.61	2	0.2	2.4	0.03	0.33
dixisse	0.001	42.84	0.21	3.07	0	1.9	0	0.22
Aelius	0.001	24.77	0.25	2.58	0	1.8	0	0.27
Gallieni	0.002	8.66	-1	-1.28	0.7	1.2	0.11	0.19
Maximus	0.001	31.08	-0.35	1.91	0	1.2	0	0.12
Gordiano	0.001	10.08	-0.57	1.99	0.1	1.1	0.03	0.39
Gordianus	0.001	10.04	-0.44	1.83	0.1	1.1	0.02	0.33
potens	0.001	9.23	-0.26	2.33	0	1.1	0.01	0.59
imperatores	0.001	7.18	-0.34	1.62	0	1	0.02	0.36
dona	0	13.99	-0.22	2.54	0	0.9	0	0.46
fuissent	0	13.99	-0.22	2.54	0	0.9	0	0.46
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

- The two columns of “**Coordinates**” are the coordinates of points in the two dimensional space. As an example, the word “Iuliano” has the coordinates (2.35; -0.5), which is positive in the first axis (abscissa) and negative in the second one (ordinate).
- The two columns of “**Absolute Contributions**” concern the importance of different elements (in this case the words and the texts) in building the axes. The sum of the column is equal to one hundred. For the first axis the most important words are two emperors’ first names “Iuliano” (1.4 as absolute contribution), “Severus” (1.4 as absolute contribution) and an adverb “item” (1.3 as absolute contribution) which is emphasized later on. On the other hand, the second axis has two other emperors’ first names “Gordiani” (2.4 as absolute contribution) and “Aelius” (1.8 as absolute contribution) and the verb “dixisse” (1.9 as absolute contribution).
- The last column called “**Squared cosine**”, which is the sum of the rows, is equal to one and shows the importance of the different axes in the explanation of each element. As for “Iuliano”, this is explained by 22% by the first axis and 1% by the second axis. Thus, it measures how well the display approximates the true position of the *profile*.

This data is definitely possible to have for all the other axes that are involved in the calculus. Here one deals with twenty dimensions since there are twenty texts. And the same goes for the representation in the spaces. Indeed, the software allows seeing the factorial plan with axes two and three, three and four and so on. Considering that the results are mostly based in the two dimensional space, it is not necessary to spend more time commenting these multiple views. The conclusion can be drawn on the two dimensional space thanks to the properties of simultaneous representation belonging to correspondence analysis.

3.4 To sum up

As the corpus is composed of biographies, the reader needs to know that it is not unusual to find at the beginning of each text an enumeration of all the members and

ancestors of the family of the concerned emperor. Furthermore, the same emperors' first name recurs every generation. This elucidates the high frequency of first names and why some of them characterize a line of emperors. As seen in the previous section, most of the words that contribute to the axes building are first names of emperors. For the reasons mentioned above, this analysis is not representative enough to draw a conclusion about the issue. Proceeding further would be meaningless as the graphs and other results are not relevant to this case. A new analysis should be conducted, leaving out the emperors' first names and focusing only on vocabulary words.

3.5 Second Analysis

After doing an exploratory analysis of the data set and having concluded that the emperors' first names could distort the results, the initial text file document is modified by removing all the emperors' first names in the biographies. The approach in the next analysis stays the same as the previous one. The results are always based on correspondence analysis. The interpretation of tables and graphs has not changed.

3.5.1 Results

Summary

The software has detected:

- 7823 as the total number of words,
- 3230 as the number of distinct words,
- and 41.3 as the percentage of distinct words.

Only for a descriptive point of view, Table 3.4 presents the frequency table. In this case, its main role is being the basis of the computational phase, but showing no interest for the interpretation of the results. By choosing to apply a threshold of two to "*Historia Augusta*" the most used words appears to be "et" that is mentioned 312 times and the less used word is "filias" that appears only twice (see subsection "Computational steps").

Table 3.4. Words (frequency order)

number according to alphabetic order	word	frequence
223	<i>et</i>	312
363	<i>in</i>	184
134	<i>cum</i>	129
826	<i>ut</i>	111
222	<i>est</i>	107
278	<i>fruit</i>	83
10	<i>ad</i>	82
⋮	⋮	⋮
3	<i>absumpto</i>	2
513	<i>occidendi</i>	2
256	<i>filiis</i>	2

Eigenvalues

In Table 3.5, $\lambda_1 = 0.2459$ for the first axis and $\lambda_2 = 0.231$ for the second axis. The percentage of variance or percentage of inertia (ratio of each eigenvalue to its global sum) corresponding to these eigenvalues is respectively 4.93% for the first axis and 4.63% for the second axis. The first factorial plan “explains” 9.55% of the total variance, which is not a great representation.

3.5.2 Contribution of the axes

In this subsection only the most significant results are presented and some emphasis is putting on the things more relevant for the conclusions.

Texts

Table 3.6 lists all the biographies. They are stored in decreasing order by the importance of building the first axis (see column Absolute Contribution f1). The two most characteristic texts are the biographies of Divus Cladius (16.3 as absolute contribution) and Septimus Severo (12.8 as absolute contribution) which really contribute to the po-

Table 3.5. Eigenvalues and Percentage of variance

Number	Eigenvalue	Percent.	Cumulat. Percent.	Stars
1	0.2459	4.93	4.93	80
2	0.231	4.63	9.55	77
3	0.227	4.55	14.1	75
4	0.2239	4.49	18.59	74
5	0.2082	4.17	22.76	69
6	0.2062	4.13	26.89	69
7	0.1975	3.96	30.84	66
8	0.1956	3.92	34.76	65
9	0.1914	3.83	38.59	64
10	0.1874	3.75	42.35	62
11	0.1847	3.7	46.05	62
12	0.182	3.65	49.69	61
13	0.1758	3.52	53.22	59
14	0.1711	3.43	56.64	57
15	0.1692	3.39	60.03	57
16	0.1671	3.35	63.38	56
17	0.1658	3.32	66.7	55
18	0.162	3.25	69.94	54
19	0.1568	3.14	73.08	53
20	0.1555	3.12	76.2	52
21	0.1503	3.01	79.21	50
22	0.1449	2.9	82.11	49
23	0.1393	2.79	84.9	47
24	0.1363	2.73	87.63	46
25	0.1348	2.7	90.33	45
26	0.1269	2.54	92.87	43
27	0.1241	2.49	95.36	42
28	0.1201	2.41	97.77	41
29	0.1116	2.23	100	38

sition of the first axis. In fact, these two texts are opposed in the factorial plan. If one looks at the column “Coordinates”, he will see that Claudius has negative abscissa (-0.95) and positive ordinate (0.68) whereas Severus has positive abscissa (1.02) and negative ordinate (-0.49).

Using the same criteria as the previous one, the second axis (ordinate) is characterized firstly by the biography of Macrino (25.8 as absolute contribution) and secondly by the one of Diadumenus (11.7 as absolute contribution)(see Table 3.7). One as to remember that the more two texts have the same words in common, the closer they are in the factorial plan which is exactly the case of these two biographies. If one looks at the coordinates of Macrino ($-0.5; -1.18$) and at that of Diadumenus ($-0.3; -1.1$), both have negative abscissa and ordinate.

Another remark is that in both discussed Tables the most characteristic texts are neither the longest nor the shortest ones. So this means that it is not the quantity of words that influence the analysis, but really the kind of words that are present in each text.

Words

Table 3.8 contains we have the words that have contributed to building the first axis. Certainly, the most important are “item” (1.8 as absolute contribution) and “de” (1.3 as absolute contribution) which hide fascinating things. The word “item” is really surprising as it means “likewise”, “the same”. According to historians this adverb is present when someone takes existing texts and adds in some words or phrases. This leads to modifying the original text. Indeed, the reworking of old texts was common in antiquity. In a sense it confirms the possible presence of different perspectives used by the same writer. In Latin, the preposition “de” followed by an ablative means “distance”, “remoteness”. In fact, in the evolution of Latin language the genitive has been replaced by “de+ablative”. Thus a chronology in the drafting of the texts can be marked. Moreover, these two words are in the same square in the factorial plan. Both have negative abscissa and positive ordinate (“item” has the coordinates ($-1.62; 0.81$) and “de” ($-0.49; 0.06$)). These can

Table 3.6. Principal parameters of correspondence analysis: contribution of texts in building the first axis (decreasing order).

Texts	Weight	Disto2	Coordinates		Absolute Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
D.Claudius	0.044	4.47	-0.95	0.68	16.3	8.9	0.2	0.1
S.Severo	0.03	6.02	1.02	-0.49	12.8	3.1	0.17	0.04
P.Niger	0.031	5.95	-0.84	-0.64	9	5.7	0.12	0.07
Carus, Carino and Numerian	0.035	5.4	-0.75	0.55	8.1	4.7	0.1	0.06
Caracalla	0.018	9.51	0.89	-0.49	5.8	1.9	0.08	0.03
Hadrianus	0.047	3.78	0.53	-0.24	5.4	1.2	0.08	0.02
C.Albinus	0.032	5.94	0.57	0.64	4.3	5.8	0.06	0.07
M.Aurelius	0.048	3.35	0.47	0	4.3	0	0.07	0
O.Macrinus	0.042	4.28	-0.5	-1.18	4.3	25.8	0.06	0.33
Gordiani Tres	0.062	2.53	-0.39	-0.09	3.9	0.2	0.06	0
Elagabalus	0.047	3.17	0.4	-0.1	3.1	0.2	0.05	0
Tyranni Triginta	0.064	2.66	-0.34	-0.41	3	4.6	0.04	0.06
Gallieni Duo	0.041	4.21	-0.39	0.25	2.5	1.1	0.04	0.02
L.Verus	0.016	8.14	0.6	-0.12	2.4	0.1	0.04	0
Aelius	0.02	8.85	-0.48	0.03	1.9	0	0.03	0
Firmus, Saturninus, Proculus and Bonosus	0.023	7.44	-0.45	0.66	1.9	4.3	0.03	0.06
Maximini Duo	0.057	3.15	0.28	0.04	1.8	0	0.02	0
D.Aurelianus	0.086	2.02	0.2	0.49	1.4	9.1	0.02	0.12
A.Pius	0.019	9.07	0.4	-0.06	1.2	0	0.02	0
Probus	0.037	4.39	0.29	0.42	1.2	2.8	0.02	0.04
Valeriani Duo	0.011	14.32	-0.51	-0.58	1.1	1.6	0.02	0.02
A.Cassius	0.025	5.52	-0.31	0.17	1	0.3	0.02	0.01
Pertinax	0.02	8.06	0.34	-0.26	0.9	0.6	0.01	0.01
Diadumenus	0.022	8.13	-0.3	-1.1	0.8	11.7	0.01	0.15
Commodus	0.028	5.9	0.18	0.02	0.4	0	0.01	0
D.Julianus	0.014	10.05	0.24	0.19	0.3	0.2	0.01	0
Geta	0.011	12.91	0.27	-0.56	0.3	1.5	0.01	0.02
Tacitus	0.032	4.86	0.14	0.38	0.2	2	0	0.03
S.Alexander	0.005	29.39	0.18	-0.44	0.1	0.4	0	0.01
Maximus and Balbinus	0.031	4.73	0.06	0.42	0	2.4	0	0.04

Table 3.7. Principal parameters of correspondence analysis: contribution of texts in building the second axis (decreasing order).

Texts	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
O.Macrinus	0.042	4.28	-0.5	-1.18	4.3	25.8	0.06	0.33
Diadumenus	0.022	8.13	-0.3	-1.1	0.8	11.7	0.01	0.15
D.Aurelianus	0.086	2.02	0.2	0.49	1.4	9.1	0.02	0.12
D.Claudius	0.044	4.47	-0.95	0.68	16.3	8.9	0.2	0.1
Albinus	0.032	5.94	0.57	0.64	4.3	5.8	0.06	0.07
P.Niger	0.031	5.95	-0.84	-0.64	9	5.7	0.12	0.07
Carus, Carino and Numerian	0.035	5.4	-0.75	0.55	8.1	4.7	0.1	0.06
Tyranni Triginta	0.064	2.66	-0.34	-0.41	3	4.6	0.04	0.06
Firmus, Saturninus, Proculus and Bonosus	0.023	7.44	-0.45	0.66	1.9	4.3	0.03	0.06
S.Severo	0.03	6.02	1.02	-0.49	12.8	3.1	0.17	0.04
Probus	0.037	4.39	0.29	0.42	1.2	2.8	0.02	0.04
Maximus and Balbinus	0.031	4.73	0.06	0.42	0	2.4	0	0.04
Tacitus	0.032	4.86	0.14	0.38	0.2	2	0	0.03
Caracalla	0.018	9.51	0.89	-0.49	5.8	1.9	0.08	0.03
Valeriani	0.011	14.32	-0.51	-0.58	1.1	1.6	0.02	0.02
Geta	0.011	12.91	0.27	-0.56	0.3	1.5	0.01	0.02
Hadrianus	0.047	3.78	0.53	-0.24	5.4	1.2	0.08	0.02
Gallieni Duo	0.041	4.21	-0.39	0.25	2.5	1.1	0.04	0.02
Pertinax	0.02	8.06	0.34	-0.26	0.9	0.6	0.01	0.01
S.Alexander	0.005	29.39	0.18	-0.44	0.1	0.4	0	0.01
A.Cassius	0.025	5.52	-0.31	0.17	1	0.3	0.02	0.01
Gordiani Tres	0.062	2.53	-0.39	-0.09	3.9	0.2	0.06	0
Elagabalus	0.047	3.17	0.4	-0.1	3.1	0.2	0.05	0
D.Julianus	0.014	10.05	0.24	0.19	0.3	0.2	0.01	0
L.Verus	0.016	8.14	0.6	-0.12	2.4	0.1	0.04	0
M.Aurelius	0.048	3.35	0.47	0	4.3	0	0.07	0
Aelius	0.02	8.85	-0.48	0.03	1.9	0	0.03	0
Maximini Duo	0.057	3.15	0.28	0.04	1.8	0	0.02	0
A.Pius	0.019	9.07	0.4	-0.06	1.2	0	0.02	0
Commodus	0.028	5.9	0.18	0.02	0.4	0	0.01	0

be proofs that the biographies which are in this square are posthumous to the other ones because of the reworking of the corpus and the presence of a more recent language.

The last remark is about the word “epistola” which is found in some other texts written as “epistula”. This latter is in the same square as the two previous words. This could also be a mark of the evolution of Latin language. It seems that the form with an “o” is the Latin root whereas the one with an “u” comes from the Greek root. But it would be more reasonable to verify if this is not an error of coping or if it depends on the edition of the “*Historia Augusta*” that is used as support for the analysis. Nevertheless, another version has been checked Saur (2006) to see if the spelling changed but it seems correct. So, these two types of spelling could be hiding a different writer or more likely something chronological which could confirm the hypothesis introduced just above.

Table 3.8. Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
item	0.002	10.13	-1.62	0.81	1.8	0.5	0.26	0.07
de	0.014	0.47	-0.49	0.06	1.3	0	0.5	0.01
autem	0.007	1.26	0.53	0.1	0.8	0	0.22	0.01
dein	0.001	13.89	1.65	-0.89	0.8	0.3	0.2	0.06
epistola	0.002	7.52	-0.96	0.1	0.7	0	0.12	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

In Table 3.9, the column “Absolute Contribution f2” shows that the numbers in decreasing order do not have a big margin of difference or any big distinction, so this is difficult to interpret and say that the first five words really make a difference in building the second axis. Furthermore, they are common words so they do not help to draw any conclusion.

In order to help the historians find these “important” words which were discussed

Table 3.9. Principal parameters of correspondence analysis: contribution of words in building the second axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
graeco	0.001	10.16	-1.1	-2.03	0.3	1	0.12	0.41
en	0	22.55	-1	-2.46	0.1	1	0.04	0.27
impius	0.001	19.58	-0.48	-2.03	0.1	1	0.01	0.21
pius	0	22.55	-1	-2.46	0.1	1	0.04	0.27
vivos	0	22.55	-1	-2.46	0.1	1	0.04	0.27
versus	0.001	4.34	-0.96	-1.26	0.5	0.9	0.21	0.37
felix	0.001	11.19	-0.9	-1.93	0.2	0.9	0.07	0.33
au	0	16.17	-0.8	-2.38	0.1	0.9	0.04	0.35
cupit	0	16.17	-0.8	-2.38	0.1	0.9	0.04	0.35
erit	0	16.17	-0.8	-2.38	0.1	0.9	0.04	0.35
translati	0	16.17	-0.8	0.12	0.1	0.9	0.04	0.35
ad	0.015	0.37	-0.05	0.38	0	0.9	0.01	0.4
mille	0.002	10.62	0.4	1.03	0.1	0.8	0.01	0.1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

in the previous paragraphs, only a small part of the whole contingency table has been selected (Table 3.10) to look at the number of occurrences of these words and their positions in the texts (remember that the threshold is at least a frequency of 2 for each form of distinct words). An example of how to read this Table: as previously said, the word “item” (maybe one of the most interesting) is present six times in the biography of Divus Cladius and is completely absent in the biographies of O. Macrino, D. Aurelianus and Diadumenus. The consequence of this is that in the plan view (see the next section) Cladius will go far from Macrino and on the contrary some texts like Macrino and Diadumenus will be aggregated. If one follows the hypothesis concerning the word “item”, the texts that could have been altered or could present some additional parts then are the biographies of D. Claudius, P. Niger, A. Cassius and Gordiani Tres. These latter are in the left side of the plan view (see Figures 3.1 and 3.2). On the other hand, the preposition “de”, which is completely absent in the biographies of Diadumenus, S. Severo, Alexander, Iuliano and Aurelio, seems to characterize the right side of the plan view (see Figures 3.1 and 3.2). Whether there are present or not in the biographies, these two words appear to cut the factorial plan into two parts: on one side the recent emperors and on the other side the oldest ones. This suggests something chronological about the writing of the manuscript. For the different spelling of “epistola” doubts persist because some texts (D. Aurelianus and A. Cassius) have the two spellings. Therefore one may wonder if these are mistakes made by the author or made by the monks when they copied the corpus over the centuries.

Obviously, with this Table 3.10 it is also possible to see what kind of words are specific to one text (for example “vivos” only present in O.Macrino’s biography) or on the other hand to look at the most frequent ones like “et”, but this is not the real goal. That is why only certain rows were selected from the entire Table 3.10.

3.5.3 Interpretation

In figure 3.1 the initial hypotheses of authorship have been retraced. There is no certainty for the existence of a single writer of this corpus. To try and confirm this, the

Table 3.10. Contingency table

	de	epistola	epistolam	epistula	epistulae	epistulam	item
O.Macrinus	3	0	0	0	0	0	0
Diadumenus	0	3	0	0	0	0	0
D.Aurelianus	5	1	1	4	0	0	0
D.Claudius	7	4	0	0	0	0	6
C.Albinus	2	0	0	0	1	0	0
P.Niger	5	0	0	0	0	0	1
Carus, Carino and Numerian	5	0	0	0	0	1	0
Tyranni Triginta	10	0	0	0	0	0	0
Firmus, Saturninus, Proculus and Bonosus	3	0	0	0	0	0	0
S.Severus	0	0	0	0	0	0	0
Probus	2	0	0	1	0	0	0
Maximus and Balbinus	1	1	1	0	0	0	0
Tacitus	1	0	0	0	1	0	0
Caracalla	1	0	0	0	0	0	0
Valeriani Duo	2	0	1	0	0	0	0
Geta	1	0	0	0	0	0	0
Hadrianus	2	0	0	0	0	0	0
Gallieni Duo	4	0	0	0	0	0	0
Pertinax	2	0	0	0	0	0	0
S.Alexander	0	0	0	0	0	0	0
A.Cassius	4	1	0	1	0	1	1
Gordiani Tres	3	0	0	0	0	0	1
Elagabalus	3	0	0	0	0	0	0
D.Iulianus	0	0	0	0	0	0	0
L.Verus	0	0	0	0	0	0	0
M.Aurelius	1	0	0	0	0	0	0
Aelius	2	0	0	0	0	0	0
Maximini Duo	2	0	0	0	0	0	0
A.Pius	2	0	0	0	0	0	0
Commodus	1	0	0	0	0	0	0

emperors have been linked by chronological order for each hypothesis.

Here are some keys for interpreting the factorial maps:

- Points near the origin have undifferentiated *profile* distribution. As a consequence of the origin, they are placed at the centre of gravity of both clouds $N(I)$ and $N(J)$. This is the case for example of the biography of Commodus.
- The points which do not contribute essentially to the inertia of each axis are virtually identical to the average *profile* and are the ones near to the origin.
- Points of a cloud (or set) situated away from the origin but close to each other have similar *profiles* like the biographies of O. Macrino and Diadumenus. Geometrically, a particular row-*profile* would be attracted to a position in its subspace that corresponds to column variable categories proeminent in that row-*profile*.
- When correspondence analysis has more than two dimensions, proximity with one pair of axes may disappear when the other axes are (added) plotted.
- It is customary to summarize the row and column coordinates in a single plot. However, it is important to remember that in such plots, one cannot interpret the distances between row points and column points. The joint display of coordinates shows the relation between a point from one set and all the points of the other set but not between individual points between each set.
- A point makes a high contribution to the inertia of a principal axis in two ways: when it has a large distance from the barycentre, even if it has a small mass or when it has a large mass but a small distance. Considering all these points, it is necessary that the numerical results of correspondence analysis are all taken into account to interpret the results of correspondence analysis. That is exactly what has been done until now.

Looking at the visualization 3.1, since there are no distinct groups one can conclude that there are not several writers. It seems that this is a case of homogeneity of the

authors. The colours yellow (A. Spartianus), green (I. Capiolinus) and red (A. Lampridius) are really mixed and even the pink (F. Vopiscus) and the blue (T. Pollio), that are a little bit a part, are still near to the other colours. Finally, the biography of Cassius is near to the three different colours so it is very improbable that there is a unique writer for this biography, otherwise an extreme position in comparison to the other texts will be expected. The graph is messy which is really difficult to interpret.

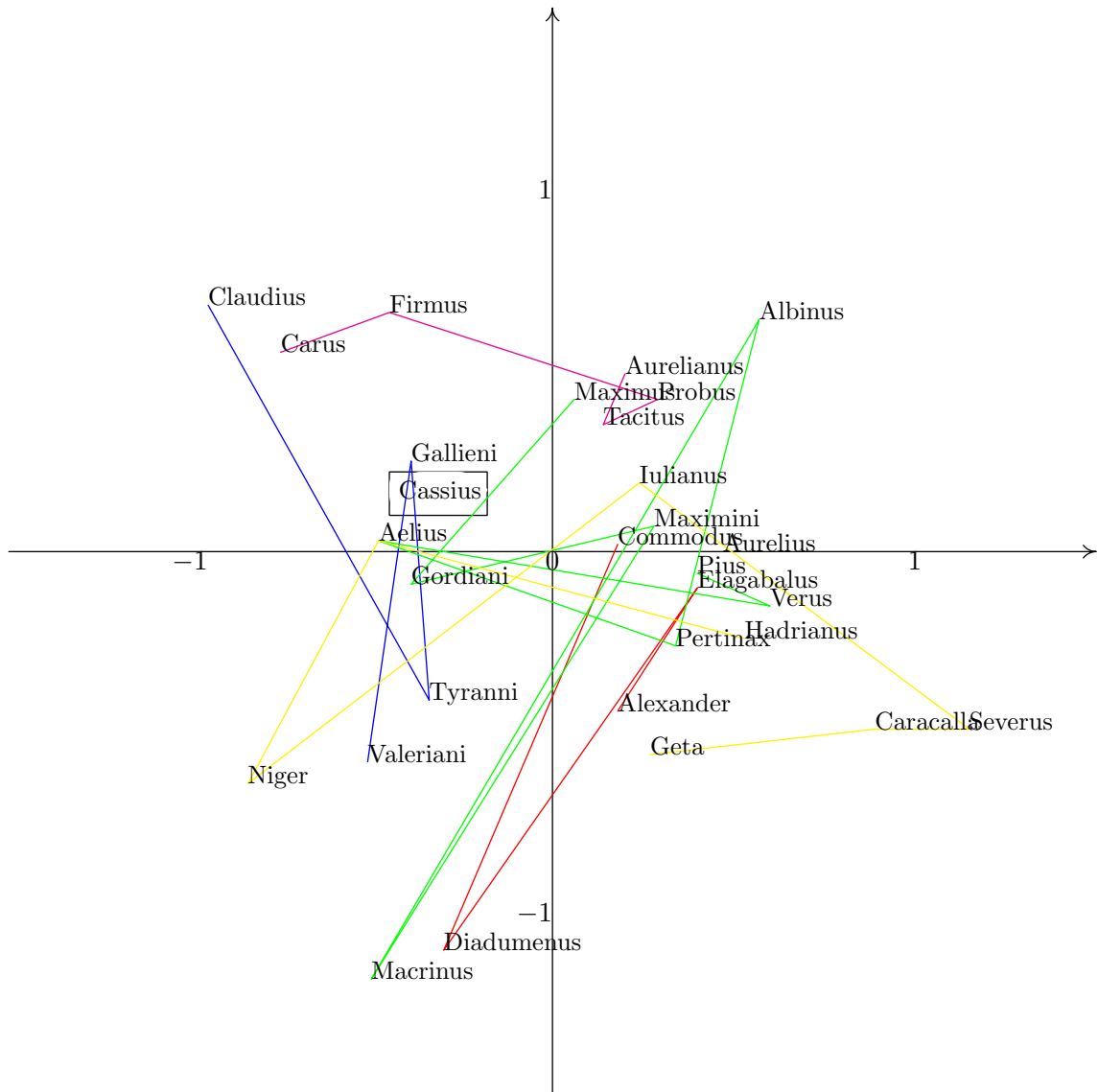


Figure 3.1. First factorial plan: frequencies classified by texts. Hypotheses view.

In Figure 3.2, the same graph as the previous one is shown but with another point of view. The emperors have been connected in chronological order, from Hadrianus to Carus, Carino and Numerian, without taking into account the initial hypotheses of authorships and leaving aside the usurpers Cassius and P. Niger and also Ablinus who only reigned in Britain and the potential emperor Aelius. This time something more interesting has been found. There is a clear opposition underlined by the pink line. To explain it better one must return to Table 3.8. In the top left square of the factorial plan there are the latest emperors who are associated with to the latest language. Remember the presence of “de” and “item” as characteristic words of the first axis and that their coordinates place them in the same square of later emperors. At the opposite side, there are the earlier emperors (as Hadrianus, A. Pius and L. Verus) with a more classical language like the spelling of “epistula” instead of “epistola” and the absence of “de”. This means that there is something chronological in the writing of the texts. It is hard to think that there is only one single writer but one still has to say that the style of a writer can change during his life time.

Figure 3.3, represents the active columns and rows of the factorial plan. This time, in addition to the position of the different texts, the different words can also be detected. In order to get a better visualisation, only the words of which the sum of the contributions is higher than 0.4 have been reported in the factorial plan. Therefore, some lexical fields have been found. In purple one can see the words belonging to politics corresponding to some troubled periods in the Roman empire with the usurper Cassius or the tyrannies Frimus, Saturninus, Proculus and Bonosus. In green there are words referring to the behaviour or morality that are in the square of Valeriani Duo (the first was proclaimed emperor by merit and unanimous consent of all the Empire whereas the second one had received an excellent education; of humble spirit he was the opposite of his step brother Gallien). In the same square there can be also found P. Niger, loved by his people and in contraposition to Iulianus’s government and that of Tyranni Triginta which may indicate that the writer denounces this kind of insane government. In orange some words belonging to the lexical field of family can be seen. They are near to Hadrianus perhaps

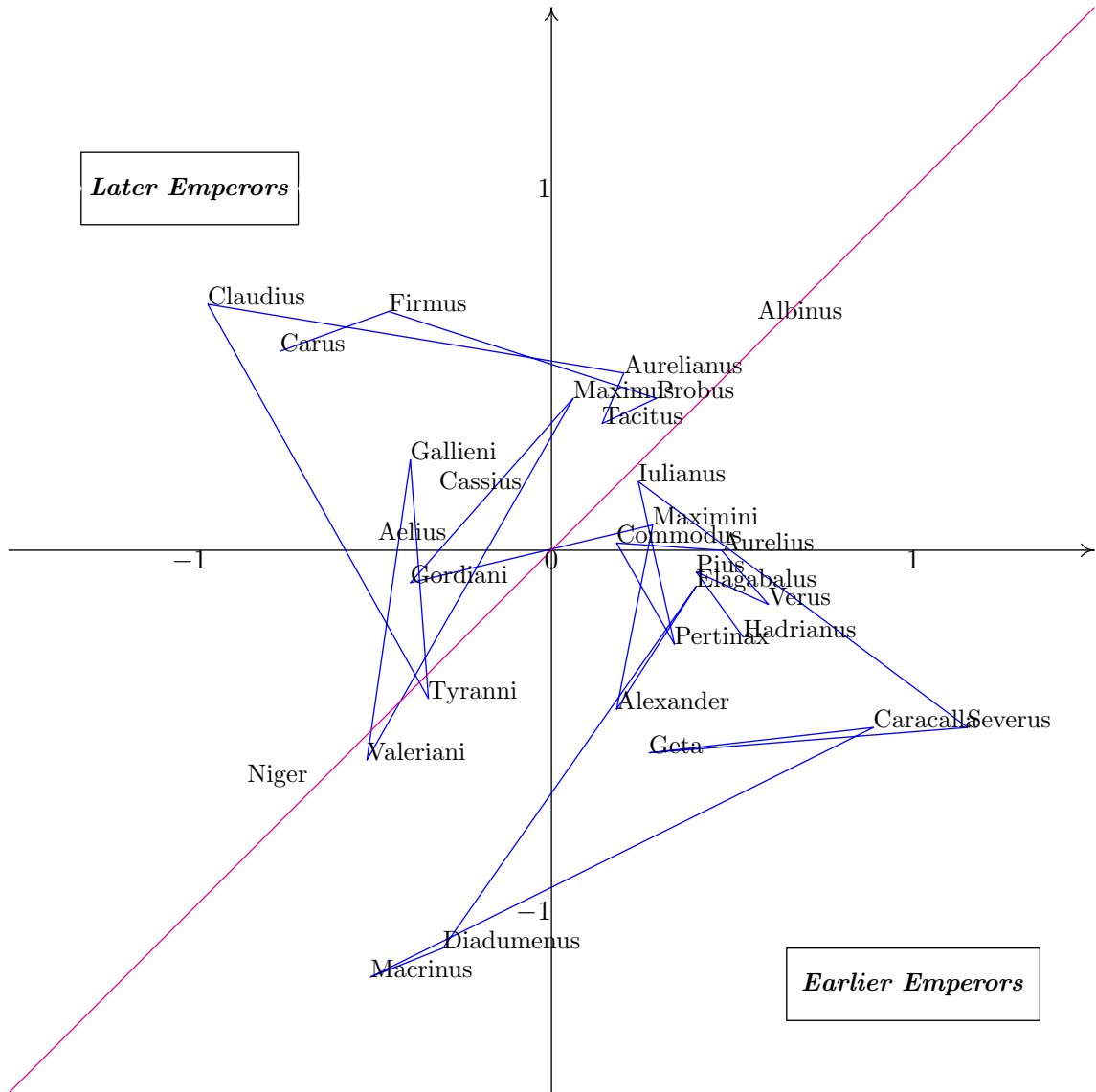


Figure 3.2. First factorial plan: frequencies classified by texts. Chronological order.

because he had adopted Aelius. Finally, in blue, words belonging to the field of war characterize the earlier emperors. To be sure of the results these fields have been verified in respect to their correspondence to the life of the emperors that are near to them. Moreover, it is not unusual to find some fragments of phrases written in the factorial plan. As an example, in the bottom left square it can be read “humani superi” or in the bottom right square “vitae privatae”. Table 3.11 has been created with the principal words of these different fields.

Table 3.11. Lexical fields

Politics	Morality	Family	War
<i>auctore</i>	<i>amavit</i>	<i>adoptaverat</i>	<i>amicos</i>
<i>epistolam</i>	<i>creditam</i>	<i>adulteri</i>	<i>britannia</i>
<i>epistulam</i>	<i>curia</i>	<i>filiis</i>	<i>castris</i>
<i>eruditus</i>	<i>digna</i>	<i>mulier</i>	<i>civiles</i>
<i>extant</i>	<i>felix</i>	<i>natus</i>	<i>civitates</i>
<i>loquitur</i>	<i>gloriae</i>	<i>pater</i>	<i>civitatis</i>
<i>oratione</i>	<i>graeco</i>	<i>privata</i>	<i>consulibus</i>
<i>ordo</i>	<i>homines</i>	<i>puer</i>	<i>exercitus</i>
<i>patria</i>	<i>humani</i>	<i>uxoris</i>	<i>fieri</i>
<i>populum</i>	<i>imperare</i>	<i>vitae</i>	<i>gallia</i>
<i>praefectum</i>	<i>imperatorem</i>	<i>vitam</i>	<i>galliam</i>
<i>principe</i>	<i>impius</i>		<i>imperatore</i>
<i>romana</i>	<i>intellectum</i>		<i>imperavit</i>
<i>scriptas</i>	<i>princeps</i>		<i>interfecto</i>
<i>senatum</i>	<i>populius</i>		<i>magno</i>
<i>sententiae</i>	<i>sperans</i>		<i>magnum</i>
<i>sermo</i>	<i>suspecto</i>		<i>memini</i>
<i>statimque</i>	<i>vates</i>		<i>militarem</i>
<i>successorem</i>	<i>versibus</i>		<i>mortis</i>
<i>sumpsit</i>	<i>versus</i>		<i>multos</i>
<i>tribunum</i>	<i>verum</i>		<i>occidendi</i>
<i>triginta</i>	<i>volens</i>		<i>occiso</i>
<i>tyrannos</i>			<i>persarum</i>
<i>urbem</i>			<i>praefecturam</i>
<i>urbis</i>			<i>procurator</i>
			<i>profectus</i>
			<i>provincia</i>
			<i>pugnavit</i>
			<i>romam</i>

view used by a same writer can be rejected because of the existence of a “chronological axis” and the intervention of the word “item”. However, the initial hypotheses of the six different points of view given by historians seem to be unfair. The fact that there are no distinct groups corresponding to these hypotheses goes to support the thesis of one single writer, therefore the most recent one given by H. Dessau.

Now it will be the job of the historians to look at the position of the words in the texts whose the importance has been underlined, to see the intrinsic meaning of their presence, their context and by what kind of words they are followed or preceded. Some suggestions, tools and emphasize on more determinant things for the analysis have been put forward rather than a really strict conclusion.

CHAPTER 4

Corpus Tibullianum

4.1 Issue

This time the corpus includes three Books (or four, it depends on the edition). The main information can be found in *Corpus Tibullianum* (2010). The subject of the First Book, assigned to Tibullus and counting 10 elegies, is represented by poems written between 30 BC and 26 BC. Five elegies arranged in no chronological order are dedicated to Tibullus' first love, Delia (perhaps the real name was Plania). Three elegies concern Tibullus' love for a boy: Marathus. Two others are for Messalla, his protector, and the last one is against the war.

First Book:

elegy 1 Tibullus refuses the war and the wealth that it brings. He supports the ideal of a simple life at the countryside with Delia. This elegy was written after the disease of Tibullus in Corfu.

elegy 2 This elegy is earlier than the previous one. Delia is married and cannot be engaged to Tibullus. Delia is the only love in the life of Tibullus and he is a hostage to her.

elegy 3 This elegy was written before the elegy 1.2. Tibullus is alone in Corfu. He commemorates the Death in loneliness and the "Champs Elysées" where Venus leads the lovers.

elegy 4 This elegy is dedicated to Marathus. Tibullus speaks in praise of Poetry. This is the most precious tribute to have been preserved to homosexual love.

elegy 5 The break with Delia is consumed: she has accepted offers from a rich admirer. In return, Tibullus mentions happy love at the countryside.

elegy 6 This elegy is the last one dedicated to Delia. Delia becomes a courtesan. Her

husband cannot control her and her excesses. Tibullus still believes in a shared love.

elegy 7 Tibullus addresses an eulogy to Messalla on the occasion of a birthday. He evokes Isis and Osiris, god of Egypt, of which Messalla was the governor.

elegy 8 Here Tibullus addresses an eulogy to Marathus but Marathus is not in love with him. His heart beats for a girl named Pholoe who loves someone else too.

elegy 9 Marathus was bribed by an old man who has money. Tibullus tells him that everything is over between them.

elegy 10 This elegy was written after 29 BC, year when Tibullus fell sick in Corfu. The author denounces the horrors of war, the fear of death and speaks in praise of Love.

According to historians, the Second Book (6 elegies) was published before the poet's death in 19 BC and is apparently incomplete. This time the female lover is called Nemesis (a fictitious name) and the elegies are arranged in chronological order.

Second Book:

elegy 1 This piece evokes the purification of fields: this is a hymn to the countryside, to deities and to Love.

elegy 2 This elegy is dedicated to a certain Cornutus. Tibullus addresses this elegy to him on the occasion of his birthday and wedding. Ironically, Tibullus tells him that his wife will be faithful.

elegy 3 This elegy is for Nemesis, the woman who breaks the heart of Tibullus. Tibullus dreams of being a slave to be by her side in Rome.

elegy 4 Tibullus expresses his rebellion and his despair over the greed and coldness of Nemesis. To seduce her, if necessary, he is ready to become a criminal and steal gold.

elegy 5 This is a national elegy addressed to the eldest son of Messalla on the occasion of his election as "quindecimuir" (officer who guards the Sacred Books).

elegy 6 This last elegy is also dedicated to Nemesis, courtesan who causes the despair of Tibullus.

The Third Book is more controversial. It comprises elegies by different authors in different styles, none of which can be assigned to Tibullus with any certainty. The natural conclusion is that a collection of scattered compositions, relating to Messalla and the members of his circle, was added as an appendix to the genuine relics of Tibullus. It cannot be precisely determined when this “Messalla collection” was created. But, it was definitely not until after the death of Tibullus (19 BC) and perhaps as late as the end of the 1st century AD. Besides the foregoing, two pieces in the collection called Priapea (one an epigram and the other a longer piece in iambics) have been attributed to Tibullus; but there is little external and no internal evidence of his authorship. Three main parts can be distinguished: for the first 6 elegies the writer calls himself Lygdamus and the love that he sings is for Neaera. Six other elegies might belong to Sulpicia, niece of Messalla, and sing her beloved. The elegy 3.7 is the “Panegyric to Messala”. The other seven elegies have “*incerti auctori*”.

Third Book:

elegy 1 This is the Kalends of March. The writer asks Perides if offering love poems to Nearea is enough to seduce her. He may well get an answer from her if his love is shared.

elegy 2 The author imagines the day of his funeral where Nearea mourns his death. He reads the inscription on his tomb: here appears for the first time the name Lygdamus.

elegy 3 Lygdamus cannot imagine a life without Nearea. He is ready to live in poverty so long as she is by his side.

elegy 4 Confused and unwilling to believe it, Lygdamus describes one of his dreams that he thinks is a warning. A spirit, a young daughter of the Gods, comes to see him and announces that Nearea is unfaithful and she prefers another man to him.

elegy 5 The writer invokes the Gods because he does not want to die young. He invokes the fear of Death.

elegy 6 Lygdamus mourns and despairs because Nearea has betrayed him. But he cannot stop loving her.

elegy 7 This is the “Panegyric of Messalla”. The writer sings the feat and the greatness of Messalla.

elegy 8 The author invokes the beauty of Sulpicia and asks Mars to come on Earth to admire her.

elegy 9 The writer complains that hunting keeps her lover away. She is ready to endure the perils of the forest to be alongside her lover.

elegy 10 The author invokes Phoebe so that it saves her beauty and brings back her health. Thus, he will make two people happy: the sick one and her lover.

elegy 11 Sulpicia speaks in praise of Cerinthe, her lover. She invokes the God of birthday so that their love may be eternal.

elegy 12 This is a tribute to Juno so that their love can be shared.

elegy 13 Finally comes Love. The writer expresses her joy.

elegy 14 Sulpicia expresses her desire to come back to the city: she finds the countryside boring.

elegy 15 The author is happy because she will be in Rome for the birthday of her beloved.

elegy 16 Sulpicia is delighted by the confidence shown by her lover.

elegy 17 Sulpicia is sick. She wants to heal but only if this is also the wish of her beloved.

elegy 18 In the desire to conceal her fever, she repents for having left her lover alone.

elegy 19 The author promises to be faithful and asks the same for her lover.

elegy 20 There are rumors about his mistress' infidelity. The writer would like to drive them out of his head because those cause him much trouble.

Here the goal is to verify that the elegies contained in the first two Books belong to Tibullus with a real distinction from the Third Book and maybe to make some hypotheses about the authors of the Third Book. Could it be possible to dissociate Lygdamus from Sulpicia and draw some conclusions concerning the "incerti autori"?

The steps to be followed are really similar to the ones of "*Historia Augusta*". The procedure is always a correspondence analysis. In order to confirm or to refute the hypotheses made by historians, all the elegies have been separated without making any distinction between Books in the hope to find distinct groups and to gather together elegies of the same Book.

The first analysis aims rather at exploring the data set and look at extreme points. Once the extreme points detected, this is important to try to understand what makes their singularity and then do the analysis again without them. At the end, the last visualisation could hopefully lead to distinct groups for each Book or even better for each hypothetical writer. In the best case, the elegies attributed to "incerti autori" would belong to a group which would make it possible to assign them an author.

4.2 Data

4.2.1 Texts

The texts can be downloaded from The Latin Library - Corpus Tibullianum (2008). In the Latin Library under Tibullus there is the "*Corpus Tibullianum*" divided into three Books. In the Third one each elegy is preceded by the name of the hypothetical writer. The same texts can also be found in Meurant (2002). They are also stored in three Books. The advantage of this website is that one can read the Latin text and its translation in French. In addition, it provides a list of the vocabulary present in the elegies and ordered in alphabetic order. For each form there is specified in which Book, elegy and verses it can be found and its frequency. To sustain the analysis some ideas

have been borrowed from Efron and Thisted (1987).

There are twenty-six elegies. The longest one has one hundred and twenty-two verses and is the elegy 2.5 attributed to Tibullus whereas the shortest ones, with only four verses, are elegy 3.15 from Sulpicia and elegy 3.20 from “incerti auctori”.

4.2.2 Coding of the texts

The elegies have been copy-pasted from the above quoted websites to a Microsoft Word document. As for the previous corpus the capital letters have been removed because the software used is always DTM. To be sure that the software can read the data set they have been converted into a notepad document. Each elegy is introduced by the symbol “****” and the last one is ended by “====”. All elegies are in the same notepad document because no distinction between Books has been made: it would be like starting with a neutral point of view. The analysis is based on a threshold of frequency two of each form. This choice has been made for exactly the same reasons explained in the last chapter. In general, the steps are the same as for the first corpus. If some modifications or changes are involved, they will be specified in the section where they appear. The interpretation of the tables, figures and results has the same reference as that of “*Historia Augusta*”.

4.3 First Analysis

4.3.1 Results

Summary

The software has detected:

- 12324 as the total number of words,
- 5329 as the number of distinct words,
- and 43.2 as the percentage of distinct words.

Eigenvalues

In Table 4.1, $\lambda_1 = 0.3348$ for the first axis and $\lambda_2 = 0.2969$ for the second axis. The percentage of variance or percentage of inertia corresponding to these eigenvalues

is respectively 4.62% for the first axis and 4.1% for the second axis. The first factorial plan “explains” 8.72% of the total variance which is not a great representation.

Table 4.1. Eigenvalues and Percentage of variance

Number	Eigenvalue	Percent.	Cumulat. Percent.	Stars
1	0.3348	4.62	4.62	81
2	0.2969	4.1	8.72	72
3	0.2824	3.9	12.61	69
4	0.2722	3.76	16.37	67
5	0.2626	3.62	19.99	64
6	0.2582	3.56	23.55	63
7	0.2479	3.42	26.98	61
8	0.2394	3.3	30.28	59
9	0.2341	3.23	33.51	57
10	0.2331	3.22	36.73	57
11	0.2272	3.13	39.86	56
12	0.2248	3.1	42.96	55
13	0.2187	3.02	45.98	54
14	0.2181	3.01	48.99	54
15	0.2138	2.95	51.94	53
16	0.2115	2.92	54.86	52
17	0.2107	2.91	57.77	52
18	0.2057	2.84	60.6	51
19	0.2032	2.8	63.41	50
20	0.1997	2.75	66.16	49
21	0.198	2.73	68.9	49
22	0.1906	2.63	71.53	47
23	0.1853	2.56	74.08	46
24	0.1825	2.52	76.6	45
25	0.1779	2.45	79.06	44
26	0.1765	2.44	81.49	44
27	0.1746	2.41	83.9	43
28	0.1643	2.27	86.17	41
29	0.1629	2.25	88.41	40
30	0.1598	2.2	90.62	40
31	0.1433	1.98	92.6	36
32	0.1407	1.94	94.54	35
33	0.1382	1.91	96.45	35
34	0.1325	1.83	98.27	33
35	0.1251	1.73	100	31

4.3.2 Contribution of the axes

The procedure will be the same as for the corpus “*Historia Augusta*”. The first step consists in analysing all the elegies, in looking at the results and in hoping to find interesting things. The main results are shown so as to draw the most evident and pertinent conclusions.

Texts

Table 4.2 lists all the elegies of the corpus. They are ordered in decreasing order by the importance of building the first axis (see column Absolute Contribution f1). The most characteristic text is the elegy 3.7 (“Panegyric to Messalla”) which has a high importance for building the first axis (70.7 as absolute contribution). The range of the absolute contribution of this elegy with the next one is very large. It jumps from 70.7 to 5.6 as absolute contribution! Later on, a visual representation will be drawn up. But, one can already say that elegy 3.7 is likely to be placed in the top left square: negative abscissa (-1.5) and positive ordinate (0.62). This elegy seems to be very singular. It could be interesting to see what determines this singularity.

This time, in Table 4.3, the elegies are stored in decreasing order for the importance of building the second axis. One can conclude that the three most important elegies are 2.1, 2.5 and elegy 3.7 with respectively 20.4, 14.1 and 13.5 as absolute contribution. This second axis may be characterized in the negative size by the Second Book (in column “Coordinates f2” one can read -1.17 and -0.84 as ordinates of elegies one and five), which till now has been attributed to Tibullus and in the positive size by elegy 3.7 that previously appears as a very detached piece of the corpus. Even here some emphasis is put on this elegy 3.7 that seems not to belong to Tibullus and to be farther from the rest of the corpus.

Words

Table 4.4 presents the words that have contributed to building the first axis. Before doing the vocabulary analysis, the reader has to be aware of some restrictions unique to Latin poetry. Written poetry in the Latin period is an art with complex roles. Firstly

Table 4.2. Principal parameters of correspondence analysis: contribution of texts in building the first axis (decreasing order).

Texts	Weight	Disto2	Coordinates		Absolute Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
L3.el.7	0.106	2.73	-1.5	0.62	70.7	13.5	0.82	0.14
L1.el.8	0.042	4.45	0.67	0.43	5.6	2.6	0.1	0.04
L1.e.12	0.055	3.51	0.57	0.5	5.4	4.7	0.09	0.07
L1.el.6	0.049	3.77	0.48	0.29	3.4	1.3	0.06	0.02
L1.el.9	0.046	4.37	0.43	0.62	2.5	5.9	0.04	0.09
L1.el.4	0.044	4.76	0.38	0.43	1.9	2.8	0.03	0.04
L1.el.5	0.04	4.53	0.4	0.27	1.9	1	0.04	0.02
L1.el.1	0.04	5.19	0.38	0.24	1.7	0.7	0.03	0.01
L2.el.5	0.06	3.74	-0.25	-0.84	1.1	14.1	0.02	0.19
L1.el.3	0.046	4.15	0.27	0.22	1	0.8	0.02	0.01
L3.el.20.incerti.autori	0.003	103.93	1.14	1.03	1	0.9	0.01	0.01
L3.el.1	0.014	18.86	-0.48	-1.52	0.9	10.5	0.01	0.12
L1.el.10	0.037	5.57	0.2	-0.04	0.4	0	0.01	0
L3.el.9.incerti.autori	0.013	16.89	0.32	0.17	0.4	0.1	0.01	0
L3.el.10.incerti.autori	0.015	12.59	0.25	-0.44	0.3	1	0	0.02
L3.el.17.Sulpicia	0.004	48.7	0.47	0.01	0.3	0	0	0
L2.el.2	0.011	20.8	-0.26	-1.44	0.2	7.5	0	0.1
L3.el.5	0.015	14.33	-0.19	0.01	0.2	0	0	0
L3.el.19.incerti.autori	0.014	10.59	0.22	-0.1	0.2	0	0	0
L1.el.7	0.029	7.53	-0.11	-0.12	0.1	0.1	0	0
L2.el.4	0.034	5.36	-0.07	-0.23	0.1	0.6	0	0.01
L2.el.6	0.03	5.99	0.1	-0.04	0.1	0	0	0
L3.el.11.incerti.autroi	0.013	13.72	-0.15	-0.31	0.1	0.4	0	0.01
L3.el.13.Sulpicia	0.005	38.8	-0.32	0.06	0.1	0	0	0
L3.el.15.Sulpicia	0.002	59.61	0.41	-0.23	0.1	0	0	0
L3.el.16.Sulpicia	0.003	44.97	0.28	0.22	0.1	0.1	0	0
L2.el.1	0.044	5.27	-0.04	-1.17	0	20.4	0	0.26
L2.el.3	0.041	4.68	0.03	-0.21	0	0.6	0	0.01
L3.el.2	0.015	14.29	0.05	-0.3	0	0.5	0	0.01
L3.el.3	0.019	11.48	0.05	0.14	0	0.1	0	0
L3.el.4	0.047	4.97	0.03	-0.12	0	0.2	0	0
L3.el.6	0.034	6.54	0.04	-0.11	0	0.1	0	0
L3.el.8.incerti.autori	0.011	20.33	-0.07	-1.31	0	6.6	0	0.08
L3.el.12.incerti.autroi	0.012	15.37	-0.1	-0.8	0	2.5	0	0.04
L3.el.14.Sulpicia	0.004	38.73	0.12	-0.13	0	0	0	0
L3.el.18.Sulpicia	0.003	54.04	-0.01	0.06	0	0	0	0

Table 4.3. Principal parameters of correspondence analysis: contribution of texts in building the second axis (decreasing order).

Texts	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
L2.el.1	0.044	5.27	-0.04	-1.17	0	20.4	0	0.26
L2.el.5	0.06	3.74	-0.25	-0.84	1.1	14.1	0.02	0.19
L3.el.7	0.106	2.73	-1.5	0.62	70.7	13.5	0.82	0.14
L3.el.1	0.014	18.86	-0.48	-1.52	0.9	10.5	0.01	0.12
L2.el.2	0.011	20.8	-0.26	-1.44	0.2	7.5	0	0.1
L3.el.8.incerti.autori	0.011	20.33	-0.07	-1.31	0	6.6	0	0.08
L1.el.9	0.046	4.37	0.43	0.62	2.5	5.9	0.04	0.09
L1.el.2	0.055	3.51	0.57	0.5	5.4	4.7	0.09	0.07
L1.el.4	0.044	4.76	0.38	0.43	1.9	2.8	0.03	0.04
L1.el.8	0.042	4.45	0.67	0.43	5.6	2.6	0.1	0.04
L3.el.12.incerti.autroi	0.012	15.37	-0.1	-0.8	0	2.5	0	0.04
L1.el.6	0.049	3.77	0.48	0.29	3.4	1.3	0.06	0.02
L1.el.5	0.04	4.53	0.4	0.27	1.9	1	0.04	0.02
L3.el.10.incerti.autori	0.015	12.59	0.25	-0.44	0.3	1	0	0.02
L3.el.20.incerti.autori	0.003	103.93	1.14	1.03	1	0.9	0.01	0.01
L1.el.3	0.046	4.15	0.27	0.22	1	0.8	0.02	0.01
L1.el.1	0.04	5.19	0.38	0.24	1.7	0.7	0.03	0.01
L2.el.4	0.034	5.36	-0.07	-0.23	0.1	0.6	0	0.01
L2.el.3	0.041	4.68	0.03	-0.21	0	0.6	0	0.01
L3.el.2	0.015	14.29	0.05	-0.3	0	0.5	0	0.01
L3.el.11.incerti.autroi	0.013	13.72	-0.15	-0.31	0.1	0.4	0	0.01
L3.el.4	0.047	4.97	0.03	-0.12	0	0.2	0	0
L3.el.9.incerti.autori	0.013	16.89	0.32	0.17	0.4	0.1	0.01	0
L1.el.7	0.029	7.53	-0.11	-0.12	0.1	0.1	0	0
L3.el.16.Sulpicia	0.003	44.97	0.28	0.22	0.1	0.1	0	0
L3.el.3	0.019	11.48	0.05	0.14	0	0.1	0	0
L3.el.6	0.034	6.54	0.04	-0.11	0	0.1	0	0
L1.el.10	0.037	5.57	0.2	-0.04	0.4	0	0.01	0
L3.el.17.Sulpicia	0.004	48.7	0.47	0.01	0.3	0	0	0
L3.el.5	0.015	14.33	-0.19	0.01	0.2	0	0	0
L3.el.19.incerti.autori	0.014	10.59	0.22	-0.1	0.2	0	0	0
L2.el.6	0.03	5.99	0.1	-0.04	0.1	0	0	0
L3.el.13.Sulpicia	0.005	38.8	-0.32	0.06	0.1	0	0	0
L3.el.15.Sulpicia	0.002	59.61	0.41	-0.23	0.1	0	0	0
L3.el.14.Sulpicia	0.004	38.73	0.12	-0.13	0	0	0	0
L3.el.18.Sulpicia	.003	54.04	-0.01	0.06	0	0	0	0

there is a metric. Generally a hexameter must be alternate with a pentameter. In addition, some hyphenations and figures of speech like assonance could be found. But, that is not all! The vocabulary plays a big role too. Not all the words can be used because of the themes treated but also because of the metric.

Here, the words “uel” and “seu” have the highest absolute contribution (2.1 for the first one and 1.4 for the second one). “uel” is an adverb which means: “or”, “if you want” (choose an expression among several). Whereas “seu” is a conjunction that can be written also “sive” and means: “or if”, “whether...whether”. In most cases the presence of adverbs determines the style of a writer. This could be a kind of a personal mark left by the author and that causes the remoteness of elegy 3.7 compared with the rest of the corpus in the correspondence analysis. So, this could be a good beginning to sustain the hypothesis of historians that do not attribute the “Panegyric of Messalla” neither to Tibullus nor to the “*Corpus Tibullianum*”. Until now, only the place on the axis of these two words can be known: both are in the top left square ((-1.96;0.74) as coordinates for “uel” and (-1.01;0.23) as coordinates for “seu”). There could be a connection between these words and elegy 3.7 since they are in the same square: this can be verified by looking at the contingency table.

“Uel” can be found:

- one time in elegy 2.4 (attributed to Tibullus),
- twelve times in elegy 3.7 (“Panegyric of Messalla”),
- two times in elegy 3.11 (attributed to “incerti auctori”),
- and one time in elegy 3.19 (attributed to “incerti auctori”).

This is quite clear that this word is used often in 3.7 and is almost absent in the other elegies. But, by looking at this direction a little bit closer, one can see that this word can also be found written as “vel” and that this form is only present in the First Book attributed to Tibullus (one time in elegies 1.4, 1.9 and 1.10 and two times in elegy 1.8). This is one of the explanations for the position of elegy 3.7 in the factorial plan and a

good proof to show its remotness, especially from the First Book.

“Seu” is present:

- two times in elegy 1.1 (attributed to Tibullus),
- four times in elegy 1.2 (attributed to Tibullus),
- one time in elegy 1.6 (attributed to Tibullus),
- two times in elegy 1.10 (attributed to Tibullus),
- three times in elegy 2.4 (attributed to Tibullus),
- two times in elegy 2.6 (attributed to Tibullus),
- two times in elegy 3.1 (attributed to Lygdamus),
- one time in elegy 3.5 (attributed to Lygdamus),
- eighteen times in elegy 3.7 (“Panegyric of Messalla”),
- four times in elegy 3.8 (attributed to “incerti auctori”),
- and two times in elegy 3.9 (attributed to “incerti auctori”).

This word seems commonly used in the elegies. This is not very surprising because it is a conjunction and like “et” it is very present throughout the poems.

If one compares it with “sive”, the latter can be found:

- two times in elegy 3.1 (attributed to Lygdamus),
- two times in elegy 3.3 (attributed to Lygdamus),
- two times in elegy 3.4 (attributed to Lygdamus),
- one time in elegy 3.5 (attributed to Lygdamus),
- and five times in elegy 3.7 (“Panegyric of Messalla”).

These two words are used a lot in the “Panegyric of Messalla” but they are also presented in the other elegies. Perhaps, in their frequency lies their importance. A last remark about “sive” is that it can only be found in the “Panegyric of Messalla” and in the elegies attributed to Lygdamus whereas “seu” is present also in the ones of Tibullus and “incerti autori”. The elegies attributed to Sulpicia do not seem to have these words. This is one of the consequences of the position of the texts on the factorial plan that one will see in Figure 4.1. The other tables have not been presented because at this step they are meaningless.

Table 4.4. Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
uel	0.002	5.89	-1.96	0.74	2.1	0.3	0.65	0.09
seu	0.005	2.6	-1.01	0.23	1.4	0.1	0.39	0.02
tellus	0.002	3.78	-1.53	0.74	1.1	0.3	0.62	0.14
orbem	0.001	6.35	-1.93	0.37	0.9	0	0.59	0.02
aduersis	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
aere	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
equum	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
fera	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
magnis	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
minus	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
qua	0.002	3.23	-1.06	-0.35	0.7	0.1	0.35	0.04
titan	0	8.48	-2.59	1.13	0.7	0.1	0.79	0.15
componere	0	17.52	-2.08	0.88	0.6	0.1	0.25	0.04
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

4.3.3 Interpretation

Figure 4.1 contains the visualisation of the above described results. Elegy 3.7 is really detached from the others so as to create another dimension to draw it. This has the consequence of creating an agglomeration of the other elegies at the centre without making any credible hypothesis of the authorship. Another extreme point, that has not been perhaps so clear in the numeric results, seems to be elegy 3.20 which is attributed to “*incerti auctori*”. This elegy is in the top right square and has 1 as absolute contribution for building the first axis (fifteen position out of thirty-six) and 0.9 as absolute contribution for the second axis (sixteen out of thirty-six). These two elegies (in red on the graph) stand out from the rest of the corpus by creating a non interpretable visualization. From now on, these two elegies will be analysed so as to show why they are so singular and appear so heterogeneous with all the “*Corpus Tibullianum*”.

Since the analysis is based on the occurrences of the forms of words the vocabulary of elegy 3.7 has been examined. That is why Table 4.5 reports the words that are only in elegy 3.7 with their related frequencies. The reader has to know that the first difference of this elegy is a stylistic one: this is the only one written only in hexameters. Moreover, this is dedicated to Messalla so it is less romantic and speaks more in praise of Messalla than the other elegies. Most of the words are first names that belong to mythology (Camenaes, Olympum, Phoebus and Titan). An epic lexical field has been found (actis, adversa, celerem, cita, cursus, equum, fera, hosti, impetus, magnis, peragit, pontus, subsistere and vincere), some unique adverbs (tantis and utrimque) and also, just for interest, here “uolucris” is used for “bird” but in the next analysis the word “avis” is used instead. Most of the words are also presented in Table 4.4 (adversis, equum, fera, magnis and Titan), but they are not the most important for building the first axis. The fact that the style of writing seems so specific in elegy 3.7 is some proof to support the thesis that this elegy belongs neither to Tibullus nor to the “*Corpus Tibullianum*” but was added later for some reason.

For elegy 3.10 another path has been followed. Table 4.6 contains a fragment of the contingency table. Some of the most striking words have been selected in order

Table 4.5. Vocabulary present only in elegy seven of the third book

Words	Frequency	Translation
actis	2	action
aduersa	2	adjective, which is opposite
aer	2	air
aetnaeae	2	Etna, volcano
altera	2	one of two
camenae	2	Camenes, nymph
celerem	2	go fast
celeremue	2	adjective, quick
chartis	2	in paper
cita	2	adverb, quickly
cognita	2	news
cursus	2	running
decus	2	ornament, everything look good
equum	3	horseman
fera	3	wild beast
hac	2	adjective demonstrative, this, that
hosti	2	enemy
impetus	2	moving forward, push forward
laudis	2	praise
magnis	3	great
memor	2	that has the memory of
minus	3	less
mundi	2	properly
obuia	2	on the road
olympum	2	Olympus
peragit	2	verb, push through
perlabitur	2	verb, equilibize
phoebo	2	Phebus
pontus	2	high seas
poterunt	2	verb, can
propior	2	which belongs to
signis	2	mark, sign
subsistere	2	verb, stop
tantis	2	adverb, however, as, of this quantity
terna	2	all three, each three, by three
titan	3	Titan
uincere	2	verb, defeat in war, be victorious
uolucris	2	bird
utrimque	2	adverb, on both side

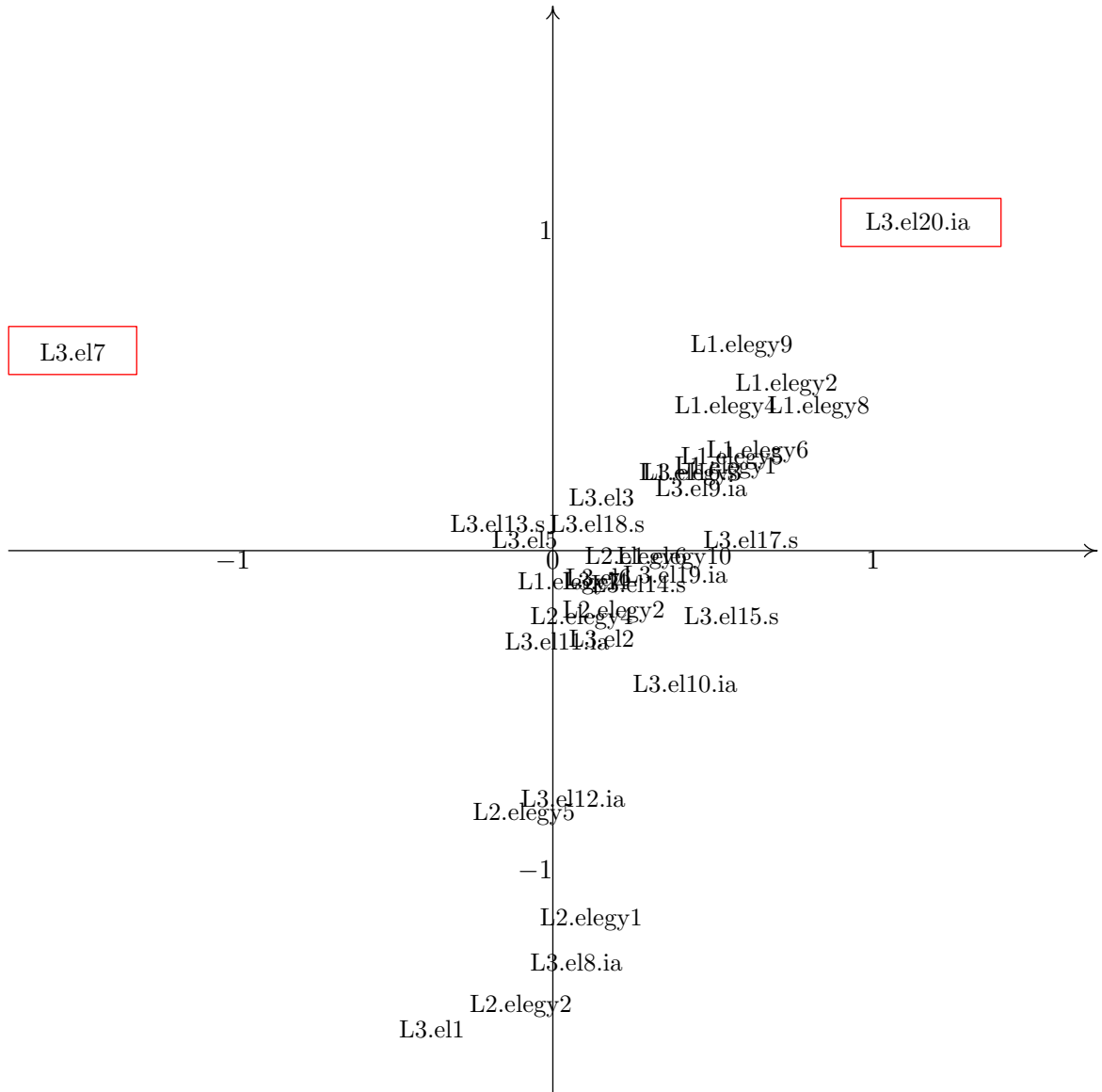


Figure 4.1. First factorial plan: frequencies classified by texts.

to understand why their position is in the extreme top right square on the graphic visualisation (see Figure 4.1). Since the threshold is two, all the words that are in the contingency table are present at least twice throughout the corpus. In fact, there are some words like “auribus”, “crebro” and “peccare” that can be found only in elegy 3.20 and once in the other elegies: in the First and Second Book hypothetical belonging to Tibullus. This explains the position of elegy 3.20 in the top right square of the previous

quoted graph. Another type of word that can be found in elegy 3.20 is a unique form specific for this elegy such as the word “rumor”. Or, on the other hand, those are totally absent in elegy 3.20 and commonly used in other elegies like “et”. It could be a good idea if historians continue in this way to look at the different words present in elegy 3.20 and compare them with the frequencies in the other elegies in order to reveal the particularity of the composition of this elegy and the choice of vocabulary made by the writer. It should be remembered that elegies 3.15 and 3.20 are the shortest ones (only four verses). This can also explain the least variety of vocabulary. But, unlike elegy 3.15 that is really in the middle of the factorial plan with coordinates (0.41; -0.23), elegy 3.20, with coordinates (1.14; 1.03), remains detached from the agglomeration perhaps created by elegy 3.7. This makes elegy 3.20, even if is not as radically different from the rest of the corpus as elegy 3.7, be enough singular so as to be far from the other elegies. This may perhaps indicate that it has a different writer.

4.4 To sum up

Because of the unique vocabulary and metric of elegy 3.7 and the singularity of elegy 3.20, the analysis has been performed once again but without these two extreme points. It can be said that these two elegies are not works by Tibullus, neither do they belong to the Corpus Tibullianum but have different writers.

4.5 Second Analysis

4.5.1 Results

Summary

The software has detected:

- 10892 as the total number of words,
- 4786 as the number of distinct words,
- and 43.9 as the percentage of distinct words.

Table 4.6. Piece of the contingency table

	auribus	crebro	et	peccare	rumor
L1.el.1	0	0	15	0	0
L1.el.2	0	0	20	0	0
L1.el.3	0	0	9	0	0
L1.el.4	0	0	7	0	0
L1.el.5	0	1	14	0	0
L1.el.6	0	0	10	0	0
L1.el.7	0	0	22	0	0
L1.el.8	0	0	15	0	0
L1.el.9	0	0	20	1	0
L1.el.10	0	0	15	0	0
L2.el.1	0	0	22	0	0
L2.el.2	0	0	2	0	0
L2.el.3	0	0	10	0	0
L2.el.4	0	0	15	0	0
L2.el.5	1	0	23	0	0
L2.el.6	0	0	8	0	0
L3.el.1	0	0	3	0	0
L3.el.2	0	0	7	0	0
L3.el.3	0	0	5	0	0
L3.el.4	0	0	15	0	0
L3.el.5	0	0	6	0	0
L3.el.6	0	0	10	0	0
L3.el.7	0	0	23	0	0
L3.el.8.incerti.adori	0	0	2	0	0
L3.el.9.incerti.adori	0	0	3	0	0
L3.el.10.incerti.adori	0	0	6	0	0
L3.el.11.incerti.adori	0	0	1	0	0
L3.el.12.incerti.adori	0	0	1	0	0
L3.el.13.Sulpica	0	0	0	0	0
L3.el.14.Sulpica	0	0	1	0	0
L3.el.15.Sulpica	0	0	0	0	0
L3.el.16.Sulpica	0	0	0	0	0
L3.el.17.Sulpica	0	0	0	0	0
L3.el.18.Sulpica	0	0	0	0	0
L3.el.19.incerti.adori	0	0	1	0	0
L3.el.20.incerti.adori	1	1	0	1	2

Eigenvalues

In Table 4.7, we can read that $\lambda_1 = 0.3101$ for the first axis and $\lambda_2 = 0.2866$ for the second axis. The percentage of variance or percentage of inertia corresponding to these eigenvalues is respectively 4.55% for the first axis and 4.2% for the second axis. The first factorial plan “explains” 8.75% of the total variance which is not a great representation.

Table 4.7. Eigenvalues and Percentage of variance

Number	Eigenvalue	Percent.	Cumulat. Percent.	Stars
1	0.3101	4.55	4.55	81
2	0.2866	4.2	8.75	75
3	0.272	3.99	12.74	72
4	0.2654	3.89	16.63	70
5	0.255	3.74	20.36	67
6	0.2476	3.63	23.99	65
7	0.2412	3.54	27.53	64
8	0.2395	3.51	31.04	63
9	0.2337	3.43	34.47	62
10	0.2304	3.38	37.85	61
11	0.2257	3.31	41.16	60
12	0.2226	3.26	44.42	59
13	0.2199	3.22	47.64	58
14	0.2184	3.2	50.84	58
15	0.2155	3.16	54	57
16	0.21	3.08	57.08	56
17	0.207	3.04	60.12	55
18	0.205	3.01	63.12	54
19	0.2026	2.97	66.09	54
20	0.1943	2.85	68.94	52
21	0.1889	2.77	71.71	50
22	0.1883	2.76	74.47	50
23	0.1819	2.67	77.14	48
24	0.1803	2.64	79.78	48
25	0.1777	2.61	82.39	47
26	0.1682	2.47	84.85	45
27	0.1643	2.41	87.26	44
28	0.1624	2.38	89.64	43
29	0.1504	2.21	91.85	40
30	0.1447	2.12	93.97	39
31	0.1431	2.1	96.07	38
32	0.1383	2.03	98.09	37
33	0.13	1.91	100	35

4.5.2 Contribution of the axes

Texts

In Table 4.8 the most characteristic texts for building the first axis are elegy 2.5 (16.3 as absolute contribution) and elegy 2.1 (12.6 as absolute contribution). Both have negative abscissa and positive ordinate so they are in the top left square of the graphic of the visual representation (elegy 2.5 has the coordinates $(-0.87; 0.58)$ and elegy 2.1 has the coordinates $(-0.9; 0.92)$). In relation to abscissa there seems to exist an opposition between the Third Book and the texts attributed to Tibullus. But, this will become clearer once with the interpretation of the visualisation.

In Table 4.9 the texts are classified in decreasing order by importance of building the second axis. The two elegies that have the highest contribution are elegy 3.4 (16.7 as absolute contribution) and elegy 2.1 (14.3 as absolute contribution). Until no more can be said. With a graphic representation and the most characteristic words interesting information can be obtained. As introduced before, it seems that the previous hypothesis is right. Elegy 3.4 has negative abscissa (-0.14) and negative ordinate (-0.96) whereas elegy 2.1 has negative abscissa (-0.9) but positive ordinate (0.92) which places them on opposite sides of the first axis.

Words

Table 4.10 shows the words that have contributed the most to the building of the first axis. Here, there are no eloquent words as in *“Historia Augusta”* but still certain emphasis can be put on some of them. In the first analysis, for the most characteristic texts (elegy 3.7) one of the unique words is “uolucris” meaning “bird”. On the contrary, “avis” (0.7 as absolute contribution) that also means “bird” appears in the second analysis. This difference in the vocabulary used to define the same thing can suggest the “pen” of two different writers. Another “strange” thing is that the top right square seems to be characterized by a higher presence of negation because there is “non” with 0.7 as absolute contribution. It would be interesting to see if “non” is present only once or more in the elegies of the top right square and if perhaps these elegies present other types of negation. This leads one to support the hypothesis of several writers.

Table 4.8. Principal parameters of correspondence analysis: contribution of texts in building the first axis (decreasing order).

Texts	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
L2.el.5	0.066	3.46	-0.87	0.58	16.3	7.9	0.22	0.1
L2.el.1	0.049	4.74	-0.9	0.92	12.6	14.3	0.17	0.18
L1.el.2	0.063	3.12	0.72	0.01	10.5	0	0.17	0
L3.el.1	0.015	17.04	-1.46	-1.48	10.1	11.4	0.12	0.13
L1.el.9	0.051	3.93	0.7	-0.03	8.1	0	0.12	0
L2.el.2	0.012	19.17	-1.45	0.06	8.1	0	0.11	0
L1.el.8	0.048	3.91	0.67	0.03	7	0	0.12	0
L1.el.4	0.05	4.27	0.55	0	4.8	0	0.07	0
L3.el.8.incerti.autori	0.013	18.45	-1.08	-0.42	4.8	0.8	0.06	0.01
L1.el.6	0.056	3.32	0.46	-0.1	3.8	0.2	0.06	0
L3.el.12.incerti.autori	0.013	14.03	-0.78	-0.28	2.7	0.4	0.04	0.01
L1.el.1	0.045	4.65	0.41	0.73	2.5	8.5	0.04	0.11
L1.el.5	0.045	4.07	0.4	0.05	2.3	0	0.04	0
L1.el.3	0.053	3.76	0.29	0.28	1.4	1.5	0.02	0.02
L2.el.4	0.038	4.86	-0.27	-0.09	0.9	0.1	0.02	0
L3.el.2	0.017	13.22	-0.4	-1.03	0.9	6.3	0.01	0.08
L3.el.11.incerti.autori	0.015	12.91	-0.39	-0.3	0.7	0.5	0.01	0.01
L1.el.7	0.033	7.06	-0.19	0.48	0.4	2.6	0.01	0.03
L2.el.3	0.045	4.24	-0.17	0.26	0.4	1.1	0.01	0.02
L3.el.6	0.038	6	-0.18	-0.93	0.4	11.3	0.01	0.14
L3.el.4	0.052	4.49	-0.14	-0.96	0.3	16.7	0	0.2
L3.el.5	0.016	13.4	-0.19	-0.27	0.2	0.4	0	0.01
L3.el.9.incerti.autori	0.015	15	0.18	-0.48	0.2	1.2	0	0.02
L2.el.6	0.033	5.49	-0.08	0.12	0.1	0.2	0	0
L3.el.10.incerti.autori	0.017	11.11	-0.16	0.04	0.1	0	0	0
L3.el.13.Sulpicia	0.005	36.31	-0.24	-0.91	0.1	1.4	0	0.02
L3.el.16.Sulpicia	0.004	41.93	0.34	-0.24	0.1	0.1	0	0
L3.el.17.Sulpicia	0.005	43.87	0.28	-0.78	0.1	1	0	0.01
L1.el.10	0.041	5.07	0.05	0.62	0	5.6	0	0.08
L3.el.3	0.021	10.37	0.08	-0.75	0	4.1	0	0.05
L3.el.14.Sulpicia	0.004	36.77	-0.14	-0.17	0	0	0	0
L3.el.15.Sulpicia	0.003	52.79	-0.04	-0.39	0	0.1	0	0
L3.el.18.Sulpicia	0.003	50.36	0.01	-1.28	0	2	0	0.03
L3.el.19.incerti.autrori	0.016	9.43	-0.04	-0.22	0	0.3	0	0

Table 4.9. Principal parameters of correspondence analysis: contribution of texts in building the second axis (decreasing order).

Texts	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
L3.el.4	0.052	4.49	-0.14	-0.96	0.3	16.7	0	0.2
L2.el.1	0.049	4.74	-0.9	0.92	12.6	14.3	0.17	0.18
L3.el.1	0.015	17.04	-1.46	-1.48	10.1	11.4	0.12	0.13
L3.el.6	0.038	6	-0.18	-0.93	0.4	11.3	0.01	0.14
L1.el.1	0.045	4.65	0.41	0.73	2.5	8.5	0.04	0.11
L2.el.5	0.066	3.46	-0.87	0.58	16.3	7.9	0.22	0.1
L3.el.2	0.017	13.22	-0.4	-1.03	0.9	6.3	0.01	0.08
L1.el.10	0.041	5.07	0.05	0.62	0	5.6	0	0.08
L3.el.3	0.021	10.37	0.08	-0.75	0	4.1	0	0.05
L1.el.7	0.033	7.06	-0.19	0.48	0.4	2.6	0.01	0.03
L3.el.18.Sulpicia	0.003	50.36	0.01	-1.28	0	2	0	0.03
L1.el.3	0.053	3.76	0.29	0.28	1.4	1.5	0.02	0.02
L3.el.13.Sulpicia	0.005	36.31	-0.24	-0.91	0.1	1.4	0	0.02
L3.el.9.incerti.autori	0.015	15	0.18	-0.48	0.2	1.2	0	0.02
L2.el.3	0.045	4.24	-0.17	0.26	0.4	1.1	0.01	0.02
L3.el.17.Sulpicia	0.005	43.87	0.28	-0.78	0.1	1	0	0.01
L3.el.8.incerti.autori	0.013	18.45	-1.08	-0.42	4.8	0.8	0.06	0.01
L3.el.11.incerti.autori	0.015	12.91	-0.39	-0.3	0.7	0.5	0.01	0.01
L3.el.12.incerti.autori	0.013	14.03	-0.78	-0.28	2.7	0.4	0.04	0.01
L3.el.5	0.016	13.4	-0.19	-0.27	0.2	0.4	0	0.01
L3.el.19.incerti.autrori	0.016	9.43	-0.04	-0.22	0	0.3	0	0
L1.el.6	0.056	3.32	0.46	-0.1	3.8	0.2	0.06	0
L2.el.6	0.033	5.49	-0.08	0.12	0.1	0.2	0	0
L2.el.4	0.038	4.86	-0.27	-0.09	0.9	0.1	0.02	0
L3.el.16.Sulpicia	0.004	41.93	0.34	-0.24	0.1	0.1	0	0
L3.el.15.Sulpicia	0.003	52.79	-0.04	-0.39	0	0.1	0	0
L1.el.2	0.063	3.12	0.72	0.01	10.5	0	0.17	0
L1.el.9	0.051	3.93	0.7	-0.03	8.1	0	0.12	0
L2.el.2	0.012	19.17	-1.45	0.06	8.1	0	0.11	0
L1.el.8	0.048	3.91	0.67	0.03	7	0	0.12	0
L1.el.4	0.05	4.27	0.55	0	4.8	0	0.07	0
L1.el.5	0.045	4.07	0.4	0.05	2.3	0	0.04	0
L3.el.10.incerti.autori	0.017	11.11	-0.16	0.04	0.1	0	0	0
L3.el.14.Sulpicia	0.004	36.77	-0.14	-0.17	0	0	0	0

Table 4.10. Principal parameters of correspondence analysis: contribution of words in building the first axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
uerba	0.002	3.19	-1.09	-0.2	0.8	0	0.37	0.01
auis	0.001	10.64	-2.1	0.04	0.7	0	0.41	0
non	0.013	0.5	0.39	0.01	0.7	0	0.31	0
ueni	0.001	5.71	-1.44	0.68	0.7	0.2	0.37	0.08
uos	0.001	5.22	-1.47	0.1	0.7	0	0.41	0
dicite	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
donate	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
huius	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
libellum	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
me	0.006	0.89	0.55	-0.15	0.6	0	0.34	0.03
qua	0.001	5.54	-1.17	0.08	0.6	0	0.25	0
quidquid	0.001	6.72	-1.17	0.08	0.6	0	0.2	0
uincula	0	82.1	-2.59	0.11	0.6	0	0.08	0
uir	0	36.65	-2.6	-1.33	0.6	0.2	0.19	0.05
urantur	0	82.1	-2.59	0.11	0.6	0	0.08	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

For the second axis the first most characteristic word is a female first name, maybe the main character to whom the elegies in the bottom left square (Third Book) are dedicated (see Table 4.11). It is not essential to spend more time on this word. The second one seems more interesting with 1 as absolute contribution. “Sive” which can also be written “seu” means “or”, “whether ... whether”. This conjunction is also present in the first analysis but with the spelling “seu” and is one of the most characteristic words for building the first axis and one which characterizes elegy 3.7 (see Table 4.4). This is another proof to sustain that elegy 3.7 is really far from the other elegies and could have been written by someone else.

Table 4.11. Principal parameters of correspondence analysis: contribution of words in building the second axis (decreasing order).

Words	Weight	Disto2	Coordinates		Abs. Contribut.		Squared Cosines	
			f1	f2	f1	f2	f1	f2
neaera	0.001	8.72	-0.81	-1.95	0.2	1.4	0.08	0.43
siue	0.001	11.21	-0.83	-1.77	0.2	1	0.06	0.28
dicite	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
donate	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
huius	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
libellum	0	66.52	-2.61	-2.77	0.6	0.7	0.1	0.12
magis	0	60.93	-1.01	-2.29	0.1	0.7	0.02	0.09
si	0.007	1.14	0.16	-0.54	0.1	0.7	0.02	0.25
cara	0.001	6.25	-0.4	-1.57	0	0.7	0.03	0.39
ad	0.004	1.08	-0.33	0.62	0.2	0.6	0.1	0.36
cura	0.002	6.97	0.08	-1.08	0	0.6	0	0.17
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 4.12 represents a piece of the contingency table. Only the words talked about earlier have been picked up. For example, one can clearly see that the first name “Neaera” is only in elegies 3.1, 3.2, 3.3, 3.4 and 3.6 which are associated to the writer called

Lygdamus. That is not a surprise since this is the name of his loved woman. In this Table to “non” have been added the other ways of creating the negation in Latin. Even if “non” is present in the three Books it is in majority in the First and Second one and the same goes for the other negations. Maybe, this can be explained by the topic of the elegies. In the First Book two impossible loves are related. Delia is already married and Marathus loves someone else. Nemesis, the loved woman of the Second Book, is austere with Tibullus and causes his despair. In brief, the First Book and the Second one seem to have a close thematic, with a high presence of negations, which is not the case of the Third Book even if preserves the same theme. The fact that the First Book and the Second one seem related by dramatic love whereas the Third Book seems to have happy ending love (this can also be proved by the vocabulary used) supports that the First Book and the Second one could have the same writer which would not be the case of the Third Book.

4.5.3 Interpretation

In Figure 4.2, the texts are represented in the factorial plan, which tried to plot the initial hypotheses and the heterogeneity of the different Books. Contrariwise to the “*Historia Augusta*” here there are some distinct groups that have been linked according to the initial hypotheses.

The pink colour represents the First Book, the red is the Second Book, the elegies that might belong to Lygdamus are in green, the ones of Sulpicia are in blue and finally those of “*incerti auctori*” are framed in black. There is an opposition between the Third book and the First and Second Book (first axis). Moreover, there is something chronological between the First Book and the Second one (positive second axis). The initial hypotheses seem to be checked. The First Book (top right square), the Second one (top left square) and also the Third one (bottom square) can be recognized. The last one is divided between the elegies belonging to Lygdamus and the ones of Sulpicia. The elegies belonging to “*incerti auctori*” are difficult to interpret because they are really in the middle of the factorial plan. What is sure is that they cannot be attributed to Tibullus.

Table 4.12. Piece of the contingency table

	avis	ne	neaera	nec	neque	ni	nihil	nil	nisi	non	siue	ueni	uerba	uos
L1.el.1	0	1	0	6	1	0	0	0	0	7	0	0	0	0
L1.el.2	0	1	0	9	1	0	2	0	0	9	0	0	0	0
L1.el.3	0	1	0	3	0	0	0	0	0	13	0	0	0	0
L1.el.4	0	4	0	4	0	1	0	0	1	6	0	0	0	0
L1.el.5	0	2	0	1	0	0	1	0	0	2	0	0	0	0
L1.el.6	0	3	0	5	0	0	1	0	0	10	0	0	0	0
L1.el.7	0	0	0	3	0	0	0	0	0	4	0	0	0	0
L1.el.8	0	1	0	8	0	1	1	0	0	7	0	0	0	0
L1.el.9	0	2	0	4	0	0	0	0	1	6	0	0	0	0
L1.el.10	0	0	0	3	0	0	1	0	0	4	0	0	0	0
L2.el.1	1	0	0	1	0	0	0	0	0	2	0	2	3	3
L2.el.2	1	0	0	2	0	0	0	0	0	0	0	0	1	0
L2.el.3	0	1	0	3	0	0	0	0	0	5	0	0	0	0
L2.el.4	0	2	0	8	0	0	1	1	0	2	0	0	0	1
L2.el.5	1	0	0	2	0	0	0	0	0	1	0	3	3	1
L2.el.6	0	3	0	1	0	0	0	0	0	2	0	0	1	0
L3.el.1	1	0	2	0	0	0	0	0	0	0	2	0	1	1
L3.el.2	0	0	1	1	0	0	0	0	0	2	0	0	0	0
L3.el.3	0	0	2	4	0	0	0	0	0	3	2	0	0	0
L3.el.4	0	2	2	18	0	0	0	0	0	7	2	0	1	0
L3.el.5	0	0	0	6	0	0	0	0	0	2	1	0	0	0
L3.el.6	0	0	1	5	0	0	0	0	0	1	0	0	4	1
L3.el.8.i.a	0	1	0	0	0	0	0	0	0	0	0	1	0	1
L3.el.9.i.a	0	1	0	0	0	0	0	1	0	0	0	0	0	0
L3.el.10.i.a	0	1	0	1	0	0	0	0	0	1	0	1	1	0
L3.el.11.i.a	0	0	0	1	0	0	0	0	0	0	0	0	1	0
L3.el.12.i.a	0	0	0	2	0	0	0	0	0	1	0	1	0	0
L3.el.13.s	0	1	0	0	0	0	0	0	0	2	0	0	0	0
L3.el.14.s	0	0	0	0	0	0	0	0	0	2	0	0	0	0
L3.el.15.s	0	0	0	1	0	0	0	0	0	0	0	0	0	0
L3.el.16.s	0	2	0	0	0	0	0	0	0	0	0	0	0	0
L3.el.17.s	0	0	0	0	0	0	0	0	0	1	0	0	0	0
L3.el.18.s	0	1	0	0	0	0	0	0	0	0	0	0	0	0
L3.el.19.i.a	0	0	0	2	0	0	0	1	0	0	0	0	0	0

The only elegies belonging about which some supposition can be made are elegy 3.9 and elegy 3.19 of the Third Book that are in the cluster belonging to Sulpicia. It has been verified that elegy 3.19 is well represented and that there is no other dimension hidden behind it. Indeed, if one looks at the absolute contribution of this elegy for the other axes, one realizes that its weight is not significant (axis 3 =1.3, axis 4 =0, axis 5 =0), which means that it is well represented in the graph. So, it seems that it is very close to the writing of elegies belonging to Sulpicia.

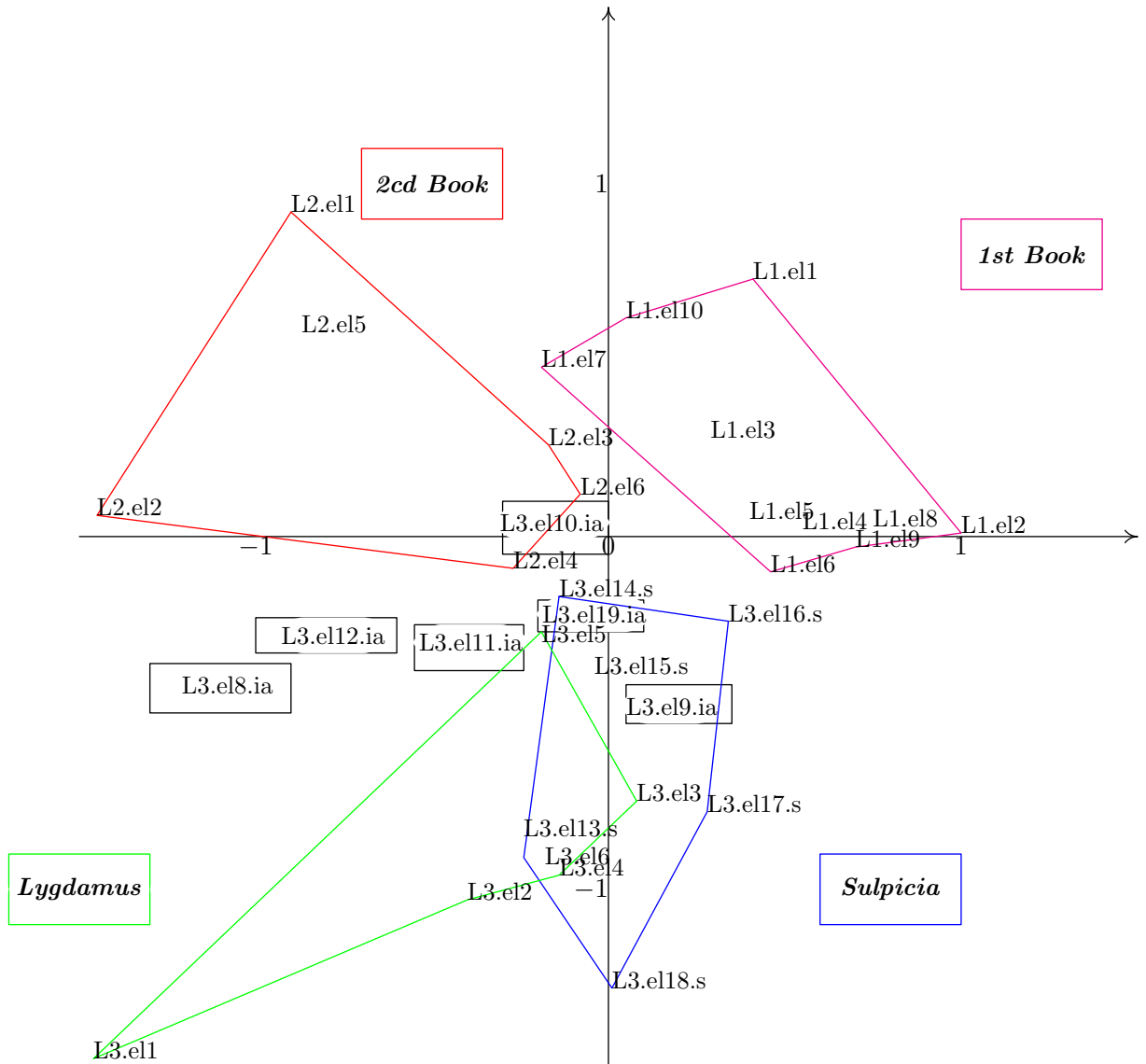


Figure 4.2. First factorial plan: frequencies classified by texts. Hypotheses view.

In Figure 4.3, there is the same representation as the previous one but with the words added in. In order to get a better visualisation, only the words of which the sum of the contributions is higher than 0.4 have been reported in the factorial plan. This time it is not like in *“Historia Augusta”* where distinct lexical fields have been found. But, by looking at the vocabulary, one can see that the “chronological axis” formed by the First Book and the Second one is more understandable. It seems that there is an evolution in the thematic. On the right, there are the elegies dedicated to Marathus (elegy 1.4, 1.8 and 1.9) which continue with the ones dedicated to Delia (elegies 1.1, 1.2, 1.3, 1.5 and 1.6) and go on with elegy seven that is still in the other square and is the one in Messalla’s honour. In the left square, but still in the centre, there are the elegies to Nemesis (elegy 2.3, 2.4 and 2.6) another woman loved by Tibullus, while the farthest have a more bucolic theme with a refined style (*festas, agricola, Saturnus, Celeres, sancte, honores ...* See Figure 4.3 words underlined in blue). That appears logical from a chronological point of view. The Second Book is a more mature Tibullus with an evolution of the written subject (see other examples in Labbé (1983) and Labbé (1990)). So, it seems that the hypothesis made by historians of the First Book and the Second one are right: they have the same author (Tibullus) and the First Book seems older than the Second one.

4.6 Conclusion

The results seem to confirm the hypotheses of historians. Three distinct groups correspond to the three different Books. For the “*incerti auctori*”, it cannot be said that they belong to Tibullus since they are close to the middle of the factorial plan. Perhaps one can suggest that elegy 3.9 and elegy 3.19 belong to Sulpicia because, as seen in the visualisation, they are in the cluster of Sulpicia. The presence of a “chronological axis” sustains the hypothesis that the First Book is written before the Second one. Finally, for elegy 3.7 and elegy 3.20, which present very different vocabulary and are treated as extreme points, one can conclude that they are far from the three Books and perhaps have been added to them later.

To conclude, the same method is applied to two different corpora with very different results. The first one seems to have the same writer for all its texts whereas the second one is likely to have several writers due to the presence of distinct groups. Surely, this is only one of the methods for analysing this kind of texts. It can be completed by some other ones such as classification or maybe validated by techniques of bootstrap. Furthermore, no absolute conclusion can be drawn. It is only possible to give historians some tools to go on in their work and make it easier, put some emphasis on things that calculus have put forward and which have a high importance for aggregated or distinguished texts.

Bibliography

- Adams, L. (1975). A statistical analysis of the book of Isaiah in relation to the Isaiah problem. *Provo*. Brigham Young University.
- Benzécri, A. (1991). Typologies des textes latins d'après les occurrences des formes des mots-outil. *Les Cahiers de l'Analyse des Données*, XVI(4):439–465.
- Corpus Tibullianum (2010). Corpus Tibullianum. [Online]. Available: <http://en.wikipedia.org/wiki/Tibullus>. Consulted on 24/09/2010.
- Efron, B. and Thisted, R. (1987). Did Shakespeare write a newly discovered poem? *Biometrika*, 74:445–455.
- Hannick, J.-M. (2008). Historiographie gréco-romaine. University of Louvain, [Online]. Available: http://bcs.fltr.ucl.ac.be/encyc-1/historia_augusta.htm. Consulted on 24/09/2010.
- Historia Augusta (2010). Historia Augusta. [Online]. Available: http://en.wikipedia.org/wiki/Augustan_History. Consulted on 24/09/2010.
- Labbé, D. (1983). François Mitterrand - Essai sur le discours. *La pensée sauvage*. Grenoble.
- Labbé, D. (1990). Le vocabulaire de François Mitterrand. *Presse de la Fond. Nat. des Sciences Politiques*. Paris.
- Lebart, L. (2009-2010). Multivariate Analysis - Course Material. University of Neuchâtel - Master of Science in Statistics.
- Lebart, L., Morineau, A., and Piron, M. (2006). *Statistique Exploratoire Multidimensionnelle*, volume 4, pages 138–153. Dunod, 4th edition.

- Lebart, L. and Salem, A. (1988a). *Statistical Analysis of Textual Data. Open Questions and Lexicometry (French)*, volume 5, page 139. Dunod-Gauthier-Villars [Centrale des Revues Dunod-Gauthier-Villars].
- Lebart, L. and Salem, A. (1988b). *Statistical Analysis of Textual Data. Open Questions and Lexicometry (French)*, volume 3, pages 89–97. Dunod-Gauthier-Villars [Centrale des Revues Dunod-Gauthier-Villars].
- Lebart, L., Salem, A., Becue, M., and Piron, M. (2010a). Dtm-vic: Data and Text Mining - Visualization, Inference, Classification - Tutorial A - Example 4. version 5.0.
- Lebart, L., Salem, A., Becue, M., and Piron, M. (2010b). Dtm-vic: Data and Text Mining - Visualization, Inference, Classification - version 5.0. [Online]. Available: http://www.dtmvic.com/05_SoftwareF.html. Consulted on 01/07/2010.
- Meurant, A. R. A. (2002). Bibliotheca Classica Selecta (BCS) - Lupa Capitolina Electronica (LCE). [Online]. Available: <http://agoraclass.fltr.ucl.ac.be/concordances/intro.htm>. Consulted on 24/09/2010.
- Remacle, P., Renault, P., Fournier, F.-D., Murcia, J., Vebr, T., and Carrat, C. (2009). L'antiquité grecque et latine du moyen Âge. [Online]. Available: <http://remacle.org/bloodwolf/historiens/histaug>. Consulted on 24/09/2010.
- Saur, K. (2006). *Bibliotheca Teubneriana Latina*, volume 4. Brepols.
- The Latin Library - Corpus Tibullianum (2008). The Latin Library - Corpus Tibullianum. [Online]. Available: <http://www.thelatinlibrary.com/classics.html>. Consulted on 01/07/2010.
- The Latin Library - Historia Augusta (2008). The Latin Library. [Online]. Available: <http://www.thelatinlibrary.com/sha.html>. Consulted on 01/07/2010.
- Weil, G., Salem, A., and Serfaty, M. (1976). Le livre d'Isaïe et l'analyse critique des sources textuelles. *Revue (R.E.L.O) LASLA*, 2. Liège.