

Cicero, Sigonio and Burrows: Investigating the Authenticity of the "Consolatio"

RICHARD S. FORSYTH
University of Luton, UK

DAVID I. HOLMES
The College of New Jersey, USA

EMILY K. TSE
University of California, Los Angeles, USA

[
Contact:

forsyth_rich@yahoo.co.uk

Cite as:

Forsyth, R.S., Holmes, D.I. & Tse, E.K. (1999). Cicero, Sigonio, and Burrows: investigating the authenticity of the "Consolatio". *Literary & Linguistic Computing*, 14(3), 1-26.

]

Abstract

When his daughter Tullia died in 45 BC, the Roman orator Marcus Tullius Cicero (106-43 BC) was assailed by grief which he attempted to assuage by writing a philosophical work now known as the *Consolatio*. Despite its high reputation in the classical world, only fragments of this text -- in the form of quotations by subsequent authors -- are known to have survived the fall of Rome.

However, in 1583 a book was printed in Venice purporting to be a rediscovery of Cicero's *Consolatio*. Its editor was a prominent humanist scholar and Ciceronian stylist called Carlo Sigonio. Some of Sigonio's contemporaries, notably Antonio Riccoboni, voiced doubts about the authenticity of this work; and since that time scholarly opinion has differed over the genuineness of the 1583 *Consolatio*.

The main aim of this study is to bring modern stylometric methods to bear on this question in order to see whether internal linguistic evidence supports the belief that the *Consolatio* of 1583 is a fake, very probably perpetrated by Sigonio himself. A secondary objective is to test the application of methods previously used almost exclusively on English texts to a language with a different structure, namely Latin.

Our findings show that language of the 1583 *Consolatio* is extremely uncharacteristic of Cicero, and indeed that the text is much more likely to have been written during the Renaissance than in classical times. The evidence that Sigonio himself was the author is also strong, though not conclusive.

Keywords: authorship attribution, Cicero, multivariate methods, neo-Latin, quantitative linguistics, text categorization, stylometry.

1. Introduction

This paper describes a study of the authorship of the *Consolatio Ciceronis*, a work known to be written by Cicero on the occasion of the death of his daughter Tullia in 45 BC. Like many works from so long ago, this text was eventually lost, and only fragments are known to have survived - seven fragments as quotations in the surviving works of Lactantius (c. AD 245-325) and one passage in a surviving work by Cicero himself, the *Tusculan Disputations*, Book I. More than sixteen centuries later, however, a book identifying itself as the *Consolatio* mysteriously reappeared, and was published in Venice and Bologna in 1583. It was edited by Carlo Sigonio (1522-1584), a prominent humanist scholar and skilled imitator of classical Latin.

However, no source manuscript of this text was ever made public, so, although its style appeared much like Cicero's, doubts about its authenticity were aroused almost at once. Its most outspoken critic was Antonio Riccoboni (1541-1599), who attacked it as a forgery by Sigonio in two successive publications *Iudicium* and *Accusator*. Another contemporary doubter was Latino Latini (1513-1593) who found in the text traces of what he took to be post-Christian thinking (e.g. a reference to the flight of the soul "ad futuram vitam") and post-classical usage (e.g. "homines" rather than "viri" used to denote specifically male human beings). Latini's letters were not published until many years afterwards, but they were widely circulated during the controversy (McCuaig, 1989) -- which was prematurely curtailed by Sigonio's death in 1584, without reaching a definite conclusion.

Although the *Consolatio* of 1583 has been included in numerous editions of Cicero's collected works, most modern scholars regard it as a forgery. However, very little textual analysis has been done on this question since Sage's book (Sage, 1910).

The main aim of the present work is to weigh the linguistic evidence for and against Cicero's, authorship, using modern stylometric techniques, unavailable in 1910. A secondary objective is to extend the applicability of methods previously used almost exclusively on English texts to an inflected language, namely Latin.

2. Background

As a background to the present study, it is necessary to be aware that the Latin language went through three major (and several minor) phases between the time of Cicero and that of Sigonio.

We use the term classical Latin to cover the period from about 100 BC to about 250 AD. This is often subdivided into the "Golden" and "Silver" Ages, with the former covering roughly the first century BC and the latter the next two and a half centuries. Cicero himself was the most notable prose author of Golden-Age Latin, and Virgil (70-19 BC) its pre-eminent poet. Prominent Silver-Age authors include Quintilian (c. 35-95 AD), Tacitus (55-120 AD) and the younger Pliny (62-114 AD).

But we know from inscriptions at Pompeii and other sources that classical Latin was already something of an artificial construct by the middle of the first century AD, and, after the fall of the Roman Empire early in the fifth century AD, it ceased to be a living language. However, it survived right through the Middle Ages as the language of diplomacy, law, scholarship and theology. In Western Europe, all who were educated wrote in Latin. We term the Latin of this period (over a

thousand years) medieval Latin. Because teaching was left almost entirely in the hands of churchmen, medieval Latin was predominantly ecclesiastical in nature.

The start of the third main phase in the development of Latin can conveniently be dated to the fourteenth century, with the revival of humanism and the renaissance of classical learning. In this phase the language escaped the confines of the cloister and became a vehicle for expressing secular concerns; indeed many prominent humanists of the time were suspected of pagan sympathies.

A pioneer of this movement was the poet and scholar Petrarch (Francesco Petrarca, 1304-1374), who devoted himself to reviving the literature of Classical Rome. He spent much time and effort searching for ancient manuscripts and was rewarded in 1345 by the discovery of one of the most celebrated finds of the period -- a batch of over 800 letters written by Cicero, which had been lost till that time.

The language of this third phase we call Neo-Latin. This represents an attempt to re-establish the Latin of the Golden Age, but it could never be a reproduction of that language: firstly, technology and society had changed too much in the interim, and secondly Neo-Latin was nobody's native tongue. In fact, many of its users were not proficient at speaking it, but only in writing it -- including Sigonio, according to McCuaig (1989).

After Petrarch, a succession of famous humanists strove to promote the pure style of classical Latin, with Cicero as their most respected model. Among these were: Gasparino Barzizza (1360-1430), Poggio Bracciolini (1380-1459), Pietro Bembo (1470-1547), Christophe de Longueil (1490-1522), Pietro Vettori (1499-1585), Jacopo Sadoletto (1477-1547) and Marc-Antoine Muret (Muretus) (1526-1585), as well as Carlo Sigonio himself. Although criticized by some, most notably Desiderius Erasmus (1466-1536), for slavish aping of Cicero (Scott, 1910), their views held sway at least until the end of the 16th century and influenced the syntax and vocabulary of Neo-Latin. As J.R. Hale puts it:

"the language of Cicero was imitated as part of a movement to restore the writing of Latin to the purity of its outstanding model. After the 1520s, at least in Italy, Ciceronianism was to become an orthodoxy" (Hale, 1971, p. 282).

This view is also attested by Nauert (1995, p. 49) who says: "students were expected to write or speak on an assigned topic in the approved 'Roman' way and in acceptably 'Ciceronian' Latin."

Paradoxically, according to Bodmer (1943), this desire to return to a prior state of perfection finished off Latin as a viable international medium. Thus the humanists killed the thing they loved.

"Pedantic attempts of the humanists of the fifteenth and sixteenth centuries to substitute the prolix pomposity of Cicero for the homely idiom of the monasteries hastened its demise. By reviving Latin, the humanists helped to kill it." (Bodmer, 1943, p. 313)

The relevance of this brief historical interlude to the current case is that if the 1583 *Consolatio* is genuine it was written by one of the foremost stylists of classical Latin, whereas if it is a forgery it will have been written by one of his imitators, in Neo-Latin. Consequently, if it does bear the hallmarks of Neo-Latin, it cannot be genuine -- whoever wrote it. At the heart of our problem, therefore, is the need to find a way of distinguishing between Cicero and Ciceronianism.

3. Materials

We have assembled a collection of writings by authors from classical Rome and Renaissance Italy, including generous selections from Cicero and Sigonio. These authors are listed in Table 1, in order of birth date, with our chief ‘suspects’ underlined.

Table 1 -- Authors Sampled.

Classical

Marcus Tullius Cicero (106-43 BC)
 Julius Caesar (c. 101-44 BC)
 Cornelius Nepos (c. 100-25 BC)
 Gaius Sallustius Crispus (86-34 BC) [= Sallust]
 Lucius Annaeus Seneca (c. 4 BC - 65 AD)
 Publius Cornelius Tacitus (c. 55-120 AD)

Neo-Latin

Pietro Vettori (1499-1585)
Carlo Sigonio (1522-1584)
 Marc-Antoine Muret (1526-1585) [= Muretus]
 Bernadino di Loredan (1533-??) [= Lauredanus]
 Antonio Riccoboni (1541-1599)

The concept of random sampling cannot truly apply in such a case, but we have endeavoured, particularly for our two major protagonists, Cicero and Sigonio, to achieve a breadth of coverage sufficient to permit estimation of the variability of the authors concerned.

In addition, we have also selected, for purposes of comparison, two other works which were long accepted as by Cicero, but which are now generally thought to be spurious. These are probably imitations of Cicero, but classical rather than Neo-Latin, namely, the *Epistula ad Octavianum* and the *Rhetorica ad Herennium*.

This gives us a total of more than 305,000 words of Latin, divided into 70 samples. The full list of the 70 text files used in this investigation may be found in Appendix I. It is ordered alphabetically by author (with the dubitanda preceding the works of known authorship).

Note that fragments of the *Consolatio* indisputably by Cicero, preserved as quotations in other works, were removed from the 1583 text prior to the analyses described below. These amounted to 368 words in total.

It should also be noted that mark-up codes (e.g. HTML) which were present in some of the samples that we obtained in electronic form, have been removed in all cases, leaving plain ASCII text in the

Roman alphabet without diacritics. Unless otherwise stated, all our analyses ignore the difference between upper and lower case; and so far we have made no use of punctuation marks.

4. Method of Approach

As Holmes (1994) has shown, a great variety of linguistic variables have been used in authorship studies. In this case, we decided to avoid excessive subjectivity by concentrating on variables which, in a sense, emerge from the texts under consideration.

A number of studies have appeared recently (e.g. Burrows, 1989, 1992; Binongo, 1994; Burrows & Craig, 1994; Holmes & Forsyth, 1995; Forsyth & Holmes, 1996; Tweedie et al., 1998) in which the features used as indicators are not imposed by the prior judgement of the analyst but are found by straightforward procedures from the texts under scrutiny. Such textual features have been used by Burrows (1992) as well as Binongo (1994), among others, not only in authorship attribution but also to distinguish among genres. This approach involves finding the most frequently used words and treating the rate of usage of each such word in a given text as a feature. The exact number of common words used varies by author and application. Burrows and colleagues (Burrows, 1992; Burrows & Craig, 1994) discuss examples using anywhere from the 50 to 100 most common words. Binongo (1994) uses the commonest 36 words (after excluding pronouns). Greenwood (1995) uses the commonest 32 (in New Testament Greek). Most such words are function words, and thus this approach can be said to continue the tradition, pioneered by Mosteller & Wallace (1964 / 1984), of using frequent function words as markers.

In fact, these studies (and some others) can be lumped together as applications of what may be called the "Burrows Approach", which is outlined below.

1. Pick the N most common words in the corpus under investigation. N may be from 15 to 100. (Manual preprocessing is sometimes done, e.g. distinguishing "that"-demonstrative from "that"-conj.)
2. Compute the occurrence rate of these N words in each text or text-unit, thus converting each text into an N-dimensional vector of numbers.
3. Apply techniques of multivariate data analysis to reveal patterns, especially:
 - Principal Components Analysis;
 - Clustering;
 - Discriminant Analysis.
4. Interpret the results (with care!).

A striking success of this method is described by Burrows (1992) on prose works by the Bronte sisters. He took 4000-word samples of first-person fictional narrative from novels by the three sisters Anne, Charlotte and Emily, and was able to show that they fell into three distinct clusters. Given three such authors, linked by heredity and upbringing, writing in the same genre at around the same time, this was an impressive feat.

A number of studies have followed this approach, the majority of which have been on English-language texts. The central thrust of our investigation has been an application of this method to the *Consolatio*, along with our Latin control samples

4.1 Choice of Frequent Function Words

There is no definitive statement by Burrows (1992) or his successors on deciding exactly how many words to use. Generally about fifty are used, with the implication being that they should be among the most common in the language, and that content words should be avoided. In the absence of a precise specification, our procedure was as follows.

We picked twelve of our 70 texts, one sample from each author (treating "Anon" as a separate author for this purpose). The text sample chosen was the largest file of each author that did not exceed 10000 words in length. Although this did not give exactly equal coverage of all our authors, it gave a selection that was not dominated by any single author, time-period or topic. In so far as a bias exists in this selection, it is towards overselection of Cicero, who contributed 9749 words to a sample of 69235. On the basis of exactly equal contributions by each author he would have contributed 5770 words to this sample. We regard this as an acceptable departure from strict equality both because he is the central focus of our investigation and because of his position as a stylistic model.

This aggregation of 12 texts was then subjected to a word count, giving a word-frequency listing of which the top fifty words are shown in Table 2. This list shows that the top fifty words are mostly common content-free words, as required by the Burrows approach.

We have used orthographic words rather than lemmata (lexical entries) in the analyses that follow primarily for the sake of simplicity. Gurney & Gurney (1998) have reported that lemmatization helped them in tackling a Latin authorship problem (*Scriptores Historiae Augustae*), but the lemmatization of a large body of Latin text is no trivial matter. Software tools which partially automate this task do exist (e.g. <http://www.shef.ac.uk/uni/projects/hpp/stemmer.html>) but their usage requires quite a heavy investment in text pre-processing (Schinke et al., 1996). Moreover, lemmatization is somewhat contrary to the spirit of the Burrows method. (A follow-up study to assess the pros and cons of lemmatization in this and other Latin authorship problems would doubtless be valuable, but is beyond the scope of the present investigation.)

To determine the exact number of words to be used, we asked a Latinist (EKT) to scan down Table 2 until the first unequivocal content word. She decided that was number 47 ("rerum"), and so in all the analyses reported herein, we have used **46** variables, i.e. relative frequencies of the 46 words from "et" to "tamen" in the list below.

Table 2 -- Fifty Most Common Latin Words in Order of Frequency.

05/15/98 07:58:20 N = 69235.

Word	Frequency	Rank	% Freq.	Cumulative
et	1958	1	2.828	2.8280
in	1427	2	2.061	4.8891
est	858	3	1.2392	6.1284
non	805	4	1.1627	7.2911
ut	781	5	1.128	8.4191
cum	629	6	0.9085	9.3276
quod	570	7	0.8232	10.150
ad	551	8	0.7958	10.946
qui	539	9	0.7785	11.725
quae	506	10	0.7308	12.456
ac	505	11	0.7294	13.185
esse	490	12	0.7077	13.893
quam	457	13	0.66	14.553
atque	454	14	0.6557	15.209
ex	391	15	0.5647	15.773
a	387	16	0.5589	16.332
si	360	17	0.5199	16.852
sed	352	18	0.5084	17.361
aut	351	19	0.5069	17.868
se	319	20	0.4607	18.328
de	307	21	0.4434	18.772
enim	271	22	0.3914	19.163
etiam	268	23	0.387	19.550
neque	262	24	0.3784	19.929
autem	239	25	0.3452	20.274
ab	231	26	0.3336	20.608
nec	217	27	0.3134	20.921
sunt	206	28	0.2975	21.219
quo	199	29	0.2874	21.506
ita	196	30	0.283	21.789
ea	196	31	0.283	22.072
nihil	193	32	0.2787	22.351
quid	192	33	0.2773	22.628
sit	190	34	0.2744	22.903
hoc	190	35	0.2744	23.177
eo	183	36	0.2643	23.442
quidem	173	37	0.2498	23.691
vero	161	38	0.2325	23.924
vel	159	39	0.2296	24.154
tum	158	40	0.2282	24.382
quibus	153	41	0.2209	24.603
id	151	42	0.2181	24.821
eius	151	43	0.2181	25.039
per	147	44	0.2123	25.251
ne	144	45	0.2079	25.459
tamen	132	46	0.1906	25.650
rerum	128	47	0.1848	25.835
natura	128	48	0.1848	26.020
modo	124	49	0.1791	26.199
nam	116	50	0.1675	26.366

It should be noted that the first 46 words between them account for more than 25% of the tokens in this multi-author sample.

4.2 Syllable Counts

The standard Burrows procedure only uses word-frequency information, but it is our long-term intention to extend this by employing other sources of linguistic evidence. In the present study, a start was made towards this end by writing a syllable-counting program. This enabled us to compute not only the proportion of words of one, two, three, four syllables and so on, but also some information about syllabic transitions. Specifically, for each text the additional 22 variables described in Table 3 were computed.

Table 3 -- Syllabic Variables.

Variables	Meaning
S1, S2, S3, S4, S5, S6	Percentage of 1-6 syllable words in the text
[Syllable transitions:] ST11, ST12, ST13, ST14	Percentage of 1-syllable words that are immediately followed by 1-syllable, 2-syllable, 3-syllable and 4-syllable words (respectively)
ST21, ST22, ST23, ST24	Percentage of 2-syllable words immediately followed by words of 1, 2, 3, & 4 syllables
ST31, ST32, ST33, ST34	Percentage of 3-syllable words immediately followed by words of 1, 2, 3 & 4 syllables
ST41, ST42, ST43, ST44	Percentage of 4-syllable words immediately followed by words of 1, 2, 3 & 4 syllables

The rules of the procedure used for counting syllables are given in Appendix II.

In the analyses that follow, mention will be made of syllabic information when it is used. If no such mention is made, it can be presumed that the analysis is being performed using only the first 46 words of Table 2.

5. Results and Analyses

We present in this section a sequence of analyses, based on the Burrows approach, using the variables described in section 4, with the objective of shedding light on the question of who wrote the *Consolatio*.

5.1 Cicero

The initial investigation concerned only the 25 textual samples from the works of Cicero. A principal components analysis was carried out on the frequencies (rate per thousand) of the 46 most frequently-occurring words detailed above, the samples being labelled as 'orations' or 'non-orations' according to their genre. Figure 1 shows the textual samples plotted in the space of the first two principal components, which together account for 35.7% of the total variation in the data.

Figure 1 about here.

The genre effect is clearly visible along the direction of the first principal component, with the orations '0' falling generally to the left of the non-orations '1'. The accompanying scaled loadings plot, Figure 2, shows that non-orations are characterized by relatively high occurrences of the connectives "nec" and "enim", and words such as "est", "sit" and "sunt" which are forms of the verb "esse" (to be). Orations, by contrast, have high occurrences of "ac" and "atque". This discovery is no surprise since "enim" is an explanatory connective which would feature often in philosophical works or treatises which comprise the bulk of the non-orations whereas "ac" and "atque" are words with a more emphatic meaning typical of an oration.

Figure 2 about here.

5.2 Cicero and the Classical Dubitanda

The two classical dubitanda, *Epistula ad Octavianum* and *Rhetorica ad Herennium II*, were then added to the Cicero samples in 5.1 above. These are often included in the Ciceronian corpus but are generally accepted by classicists as not having been written by Cicero. Figure 3 shows the 27 textual samples plotted in the space of the first two principal components (PCs), which now account for 33.6% of the total variation in the data set. The dubitanda are labelled '12'. This shows that the *Epistula* fits in quite well as a Ciceronian oration, albeit only as a borderline. The *Rhetorica*, however, appears as an outlier along the direction of the second principal component and a look at the scaled loadings plot, Figure 4, shows that an exceptionally high usage of "ab", "ad", "id" and "aut" is associated with this placement. It is interesting to note that if we bring in the third principal component, not shown in Figure 3, then the *Epistula* also becomes an outlier, having a more negative value than any of the other texts. So we could sum up by saying (roughly) that the the first PC separates orations from non-orations, the second PC separates the *Rhetorica ad Herennium* from the rest and the third separates the *Epistula ad Octavianum* from the rest.

Figures 3 and 4 about here.

5.3 Cicero and the Classical Controls

Having looked at the genre effect within Cicero's works and at the classical dubitanda, we now turn our attention to the broader picture concerning the Ciceronian texts and all the classical control samples. A principal components analysis was conducted on the rates of occurrence of the 46 most frequently-occurring words for the samples from Caesar, Cicero, Nepos, Sallust, Seneca and Tacitus, and the resulting plot in the space of the first two principal components is shown in Figure 5, which accounts for 33.2% of the total variation.

Figure 5 about here.

This remarkable plot clearly shows how successful the word-frequency approach can be at discriminating between writers. The Cicero text samples, labelled '1', form a distinct group with the exception of *Pro Cluentio*. The Seneca samples, labelled '6', are on their own with the sole Tacitus sample, '7', appearing nearby. The Caesar samples, '3', and the Sallust samples, '5', are close together but these particular textual samples were concerned with military campaigns so perhaps this is not surprising. The Nepos samples, '4', form a tight grouping.

The first principal component is separating the philosophical works, on the right, from the military and biographical works on the left, whilst the second principal component would seem to reflect temporal change. Those texts with positive (or almost positive) scores on this component are all texts written during the Roman Republic (BCE), and those texts with scores of -1 or less are written during the Roman Empire (CE). The associated scaled loadings plot, Figure 6, reveals that the CE texts are associated with high usage of "per", "ac", "et" and "nec", and the BCE texts with "cum", "quo" and "eo". The military and biographical texts, being of a narrative nature, are associated with high usage of "eius", a pronoun, whilst the philosophical texts are associated with high usage of "enim" and "quidem" which are explanatory and qualifying particles, which elaborate on a preceding clause.

Figure 6 about here.

An alternative analysis in this section may be provided by conducting a cluster analysis on the textual samples, using the 46 word rates as variables. Figure 7 shows the resultant dendrogram using average linkage as the clustering algorithm and Euclidean distance as the metric. We can see that works from the classical writers tend to cluster very well. There are some Cicero clusters which ultimately come together, a Seneca cluster (with the exception of *De Ira*, which groups with Tacitus forming a pair of outliers), a Nepos cluster and a Caesar/Sallust cluster as revealed above in the principal components analysis. The only change of any consequence is that the Cicero outlier is no longer *Pro Cluentio* but *Orator*! Our results with the two methods of analysis are mutually supportive.

Thus, even though we can interpret the first PC in terms of a genre effect and the second PC as a temporal factor, these works still tend to form clusters on the basis of authorship. In particular, the genre effect is not strong enough to disrupt the coherence of the Cicero cluster.

Figure 7 about here.

5.4 Sigonio and the Sixteenth Century Controls

Having successfully discriminated between classical writers using the frequencies of occurrence of the 46 common words, we now turn our attention to the sixteenth century textual samples, where we have writings from Lauredanus, Muret, Riccoboni, Sigonio and Vettori. Once again a principal components analysis was conducted on the data set and Figure 8 shows the texts plotted in the space of the first two principal components, which account for 31.9% of the total variation. The configuration obtained is less remarkable than that in the section above, but distinct groupings are nevertheless visible. The Sigonio texts, labelled '2', split quite dramatically into his two histories and remaining non-histories (mainly orations), whilst the Muret texts, labelled '9', split into three funeral orations and four scholarly orations. Clearly the genre effect is at play here as in the earlier analyses. Also visible are groupings for Lauredanus, labelled '8', Riccoboni, labelled '10', and Vettori, labelled '11'. The associated scaled loadings plot is shown in Figure 9.

Figures 8 and about 9 here.

An alternative analysis was again provided by using cluster analysis with average linkage as the algorithm and Euclidean distance as the metric. Figure 10 shows the resultant dendrogram. Whilst

the Sigonio histories are clearly clustered well away from his non-histories, the Muret samples exhibit a slightly different pattern to that revealed in the principal components plot, the funeral oration *Pro Antonia Rege Navarre ad Pium.* now appearing with his scholarly orations. Also, Vettori's *Oratio Petri Victorii in Max. II* now lies apart from the other Vettori texts. The clustering, although not as distinct as that shown in the principal components plot, is broadly supportive, and the word-frequency approach appears to have been successful as a sixteenth century authorial discriminator.

Figure 10 about here.

5.5 Cicero, Sigonio and the *Consolatio*

Having checked the efficacy of the set of frequencies of occurrence of the 46 common words as a discriminator for both batches of control texts, we can now concentrate on the main protagonists in the argument, namely Cicero, Sigonio and the textual samples from the *Consolatio*. Using the genuine Cicero and genuine Sigonio texts as defined groups, a stepwise discriminant analysis was run on the data. The four words chosen by the stepwise routine as being the best discriminators between Cicero and Sigonio were "ad", "ac", "neque" and "ab", and using these words the null hypothesis of no difference between the group means along the axis of the discriminant function was clearly rejected (Wilks' Lambda of 0.2420 and p-value 0.0000). When the discriminant function was employed to classify the texts from the known Cicero and Sigonio groups, it achieved an accuracy of 93.75% (without cross-validation), only the Ciceronian texts *Orator* and *De Imperio* being incorrectly classified in the Sigonio group.

The discriminant function score was then computed for the two text samples from the *Consolatio* (not used in developing this discriminant function). This assigned both samples to the Sigonio group. A graphic illustration of this result may be seen in Figure 11 which is a plot of the textual samples along the axis of the discriminant function. Here (as in figures 12-15) the horizontal axis is the score on the Canonical Discriminant function, which is a weighted sum of scores on each of the selected variables that maximizes the separation between the two categories. In Figure 11, four vertical symbols represent one text; Genuine Cicero texts are denoted by '1', genuine Sigonio by '2' and the *Consolatio* texts by #. We can identify the two misclassified Ciceronian texts and the assignation of the *Consolatio* to Sigonio.

Figure 11 about here.

5.6 Classical Latin, Neo-Latin and the *Consolatio*

Previous analyses in this study have shown that time or genre effects are often so marked that they can partly mask authorship. It seems entirely appropriate, therefore, to conduct a discriminant analysis on the two defined groups of Classical Latin texts and Neo-Latin texts, and then allocate the *Consolatio* to one of these two groups. A stepwise routine was employed, as in 5.5 above, on the 46 word occurrence rates. The words chosen by the routine as best discriminators between these time periods were "ac", "vel", "sed", "vero", "id", "ut", "ea", "neque", and "cum". The null hypothesis of no difference between the group means along the axis of the discriminant function was clearly rejected (Wilks' Lambda of 0.2178 for a p-value of 0.0000), and the function, when

applied to the original data without cross-validation, achieved a classification accuracy of 94.12% with Cicero's *Pro A. Licinio Archia Poeta Oratio* and *De Re Publica*, and the Tacitus sample (the most modern) being incorrectly assigned to the Neo-Latin group. The one Neo-Latin text incorrectly assigned to the Classical group was Riccoboni's *De Legum Laudibus Oratio*.

The two *Consolatio* textual samples were firmly ascribed to the Neo-Latin group by their discriminant function score. Figure 12 shows the plot of the text samples along the axis of the discriminant function. This time two vertical symbols represent one text; the symbol '1' denotes Classical Latin, the symbol '2' Neo-Latin and the # symbol the *Consolatio*. We can see quite clearly the two groups and how the *Consolatio* appears to be distinctly Neo-Latin in time. We have discovered that the *Consolatio* is neither Classical nor Ciceronian!

Figure 12 about here.

5.7 Syllabic Analysis

We have seen in section 4.2 that lengths of words by syllables for both single words and word-pairs were also counted. These particular counts have not yet been used. At this stage in the analysis it was decided to incorporate these counts of word-length by syllables into the data and re-run some of the previous analyses.

The addition of syllable counts into the analyses for sections 5.3 and 5.4, in which Cicero was compared with the Classical controls and Sigonio was compared with the Renaissance controls, confused the issue. In the plots of the first two PCs, the previously distinct groupings by author became blurred and overlapping in both the classical case and the sixteenth century case: it would seem that syllable counts play no positive role in authorship discrimination within time periods.

Returning to the analysis between time periods, a stepwise discriminant analysis was conducted on word and syllabic variables with the Classical Latin and Neo-Latin texts as the pre-defined groups. The variables chosen as being the best discriminators between the two time periods were the words "ac", "vel" and "vero" (all chosen in section 5.6 above), and the newly introduced syllabic variables ST13, ST22, ST43 and ST44, the first of these, for example, measuring the percentage of one-syllable words immediately followed by a three-syllable word.

The null hypothesis of no difference between the group means along the axis of the discriminant function was soundly rejected (Wilks' Lambda of 0.1738 for a p-value of 0.0000). When applied to the textual samples, without cross-validation, the discriminant function successfully classified 98.53% of the texts into their correct time periods with only the Tacitus text incorrectly classified as Neo-Latin. As commented previously, this text is the most modern of the classical texts.

The two *Consolatio* text samples were firmly ascribed to the Neo-Latin group, as in 5.6 above, by their discriminant score. Figure 13 shows the plot of the text samples along the axis of the discriminant function using exactly the same notation as in section 5.6. We seem to have discovered that while syllable counts are not very useful within time-periods they play an important role in discriminating between time-periods.

Figure 13 about here.

We finally return to the analysis comparing the Cicero texts with the Sigonio texts, and re-run this adding the syllabic variables to the word counts. A stepwise discriminant analysis with the genuine Cicero and genuine Sigonio texts as pre-defined groups revealed that the most effective discriminatory variables were the words "a", "ad", "enim", "est", "neque", "quibus", "quid", "sed", "vel" and "vero" (a substantially different listing from the analysis with words only), and the syllabic variable ST44. Once again, the discrimination between the groups was highly significant (Wilks' Lambda of 0.0276 for a p-value of 0.0000), and the discriminant function achieved an accuracy of 100%, without cross-validation, in assigning the text samples to their groups. When asked to assign the two samples from the *Consolatio*, the discriminant function gave us our first equivocal piece of evidence, the first sample being ascribed to Cicero and the second sample to Sigonio. Figure 14 shows the plot of the textual samples along the axis of the discriminant function, two vertical symbols representing one text, with the Ciceronian texts denoted by '1', the Sigonio texts by '2' and the *Consolatio* texts by the # symbol.

Figure 14 about here.

The introduction of syllable counts has moved the *Consolatio* to a borderline position between the two groups. If Sigonio is the author then it looks more Ciceronian than most of his output. The syllabic variable which has entered into the discriminant function at this point, ST44 -- the percentage of four-syllable words immediately followed by a four-syllable word, concerns words of above average length and may be indicative of a scholarly and technical second language rather than of a native language.

5.8 Chronometric Analysis

To shed further light on the date of the *Consolatio*, a stepwise regression was performed. This used 66 of our text samples (all apart from the dubitanda) with the century of composition used as the dependent variable. For these texts the century variable could take only three different values: 0 (first century BC), 1 (first century AD) or 16 (sixteenth century).

The Minitab stepwise-regression procedure with standard defaults was allowed to choose from all 46 word variables as well as the 22 syllabic variables. From these it chose five, giving the regression equation below, which accounted for 82.36% of the variance.

$$\text{century} = -6.35 + 1.16 \text{ ST44} + 0.655 \text{ ac} + 1.35 \text{ vel} + 4.05 \text{ S6} - 0.579 \text{ ST13}$$

The variables in this linear equation appear in order of importance, with ST44 being the most important. The presence of three syllabic variables, including the most significant, compared with two word variables appears to confirm that syllabic information is useful for this kind of task. All coefficients are positive, indicating increased usage in the Renaissance texts, except for that of ST13 (percentage of 1-syllable words immediately followed by a 3-syllable word) which decreases in frequency from classical to Renaissance texts.

Summing up this formula qualitatively, it shows that the words "ac" and "vel" increase in frequency from classical to Neo-Latin times, as does the frequency of 6-syllable words. This latter attests to the more learned nature of Latin in the later period. The role of the syllable-transition variables is harder

to interpret, although an increase in the proportion of 4-syllable words followed by a 4-syllable word (ST44) would seem consistent with a move from native tongue to second language.

Having developed the above formula on securely dated texts, it was then used to estimate the date of the two halves of the *Consolatio*. Both pieces were placed firmly in the later period. The fitted value for the first half was 14.00 and for the second 15.56. For comparison, the mean value for our 40 classical texts was 1.28 and for our 26 Neo-Latin texts was 14.26. Thus both segments of the *Consolatio* were very close to the mean of our Neo-Latin sample, and over 13 centuries later than the mean of our classical sample; and while they fell well within the lower and upper quartile of our Neo-Latin sample, they were more extreme than the most extreme outlier of our classical sample (the *Agricola* of Tacitus, with a computed value of 13.00).

On this basis, the evidence of anachronism is extremely strong -- confirming Latini's suspicions at the time (see section 1). Taken together with the results of the discriminant analysis in section 5.6, this virtually eliminates the possibility of Ciceronian authorship.

How does this relate to the specific question of Sigonio's authorship? It allows us -- provided we accept that Ciceronian authorship is excluded -- to concentrate on finding the Renaissance author whose style matches most closely that of the *Consolatio*.

5.9 Discrimination among Neo-Latin Authors

For this analysis we excluded Lauredanus, of whom we have only 2 texts, as having too few samples for reliable estimation. This left Muretus, Riccoboni, Sigonio and Vettori under consideration.

A stepwise discriminant analysis procedure was then executed to find the variables which best distinguished Sigonio from the other three authors. The four most distinctive variables for this purpose were: ST13, "aut", "vero" and "hoc". Then the distance of the *Consolatio* was computed from each of the four authors on each of these four variables. This was standardized as a z-score, using the following formula

$$z_j = (x_c - m_j) / s_j$$

where z_j is the distance of the *Consolatio* from author j , x_c is the value of the variable in the *Consolatio*, m_j is the mean value for author j and s_j is the standard deviation of that variable in author j -- with j varying from 1 to 4. The results are summarized in Table 4. In this table an asterisk indicates a value outside the 95% confidence interval (by reference to the normal distribution) and a double asterisk indicates a value outside the 99% confidence interval.

Table 4 -- Distance of Consolatio from 4 Authors (Sigonio Markers).

Variable ↓	Muretus	Riccoboni	Sigonio	Vettori
ST13	1.99*	2.25*	-0.11	2.44*
aut	1.06	3.52**	-0.01	10.52**
vero	0.96	1.23	-0.11	5.86**
hoc	0.02	-0.18	-0.89	9.08**
ZSum =	4.03	7.18	1.12	27.9

All four z-scores fall within the 95% confidence interval for Sigonio. For Muretus one (variable ST13) falls outside, for Riccoboni two and for Vettori all four. The final row in this table gives the sum of the absolute (unsigned) z-scores, a convenient aggregate distance measure, corresponding to "City-Block" distance as used in nearest-neighbour classification (see: Beale & Jackson, 1990).

Using this measure, which is based on the variables that best differentiate Sigonio from his contemporaries, the *Consolatio* is clearly more like Sigonio than any of the other three. This evidence is compatible with the hypothesis that Sigonio wrote the *Consolatio*, and completely incompatible with the hypotheses that Riccoboni or Vettori wrote it (which were already highly implausible on other grounds).

On these figures alone it might just be possible to entertain the hypothesis that Muretus could have had a hand in the authorship of the *Consolatio* (although nobody has seriously made such a proposal). So, just for completeness, the same procedure was repeated, this time with the four most distinctive Muretus variables, which were: "ita", "ac", "ne", "quod". The results are summarized in Table 5.

Table 5 -- Distance of Consolatio from 4 Authors (Muretus Markers).

Variable ↓	Muretus	Riccoboni	Sigonio	Vettori
ita	-1.95	0.88	0.43	1.28
ac	-0.60	1.87	0.90	-0.74
ne	0.34	2.42*	1.03	2.60*
quod	1.80	0.34	1.08	1.20
ZSum =	4.69	5.51	3.44	5.82

In terms of aggregate distance, the order is the same as before: Sigonio is the closest match, followed by Muretus, Riccoboni and lastly Vettori; but the result is less clear-cut. Once again, all four matches fall within the 95% confidence limits for Sigonio, which is not true for Riccoboni or Vettori; but in this case so do the matches with Muretus (one marginally). Nevertheless, even using what might be termed the favourite markers of Muretus, the *Consolatio* appears more similar to works by Sigonio than by Muretus himself.

This impression was further confirmed by a stepwise linear discriminant analysis performed on Muretus and Sigonio only, which was 100% successful in assigning our 7 Muretus and 7 Sigonio samples to their correct sources, and gave both halves of the *Consolatio* to Sigonio. The variables chosen as discriminators by the procedure (using standard SPSS default settings) were the words "ita" and "ac", both of which are significantly more frequent in Muretus than in Sigonio. In fact, these are the first two markers in Table 5 above. A plot of both authors and the *Consolatio* segments on the discriminant axis is given as Figure 15. But as only two variables are used, the relationship between these two authors and the *Consolatio* can be better appreciated by a scatter diagram, using "ac" and "ita" as axes, which is shown as Figure 16. Visually it seems clear that, in respect of these two most discriminatory words, the *Consolatio* resembles Sigonio's writings more than those of Muretus.

Figures 15 and 16 about here.

Thus of the four Renaissance authors considered here, Sigonio's language is closest to that of the *Consolatio*.

6. Discussion

6.1 Substantive Conclusions

The findings from this analysis tend to support received opinion among Latin scholars that the *Consolatio* of 1583 is a work of Neo-Latin and not therefore the rediscovery of Cicero's long-lost text. Moreover it resembles Sigonio's style more than it resembles those of three other Neo-Latin control authors, namely Muretus, Riccoboni or Vettori.

In our view, the evidence presented here against Cicero's authorship of the 1583 *Consolatio* is compelling. The evidence that Sigonio himself was the author is also quite strong, although the effort required to reach that conclusion is tribute to his skill as a Ciceronian imitator.

6.2 Methodological Considerations

From the methodological viewpoint, we have demonstrated that the approach pioneered by Burrows (1992) works well enough to find differences between the language of Cicero and a number of his imitators; hence that it can be generalized to an inflected language, Latin. This agrees with the findings of Tweedie et al. (1998), who also worked on Latin.

We have also shown that syllabic information can be usefully added to the basic Burrows method in certain cases, thus extending the method somewhat. It is interesting, in this connection, to note that a number of studies by a group of researchers centred on Goettingen University, on many languages, including English, German and Latin, have found rather little variation between authors and genres in respect of word-length distribution (e.g. Wimmer et al., 1994; Best, 1996). Possible reasons for this apparent contradiction might be: (1) the Goettingen group have sought general models; (2) syllabic information is more useful for temporal discrimination than for authorship or genre as such; and (3) syllable-transitions give access to more useful information than plain syllable counts or the

distributions of such counts. We suspect that this is an area of quantitative linguistics that would repay further investigation.

To end on a cautionary note, we should add that we succumbed initially to the temptation (out of curiosity) to throw all our 70 samples from a dozen authors and two time periods sixteen centuries apart into a single large multivariate analysis. The results were confusing. Only when we split our problem into the series of steps recounted in section 5 above did some clarity begin to emerge.

Of course, picking an author from 11 possible candidates from different time periods where the distribution of genres between authors is inevitably unbalanced is asking rather a lot of any method. It is considerably harder than the classic stylometric problem solved by Mosteller & Wallace (1964 / 1984) of assigning 12 disputed political essays written around the same time to one of only two candidate authors -- Alexander Hamilton or James Madison. In that case, although the two authors' styles are remarkably similar, neither tried to mimic the other. Thus we should have known better than to expect enlightenment "in a single hit". Nevertheless, we are sure that the success of the Burrows method will tempt other workers, at least initially, to seek 1-stop insight, as we did. The fact that we were able to break our problem into 2 main stages -- first deciding that the suspect text belonged to the more recent time period, then finding the author from that time period whose style matched it most closely -- was essential to making our task feasible.

We would suggest that other researchers with similar multi-author or multi-genre problems should likewise seek ways of subdividing their task.

Acknowledgements

This research could not have been carried out without the kind help of Professor Jane Crawford (Loyola Marymount University) and Professor Bernard Frischer (UCLA). Thanks are also due to the British Academy (Grant Reference APN 3887) and to UWE, Bristol whose financial support have made this project possible. We also owe a debt of thanks to the various sources that we have drawn on in compiling our Latin text sample (Appendix I), especially the Lillard Classical Library.

References

- Beale, R. & Jackson, T. (1990). *Neural computing: an introduction*. Adam Hilger, Bristol.
- Best, K.-H. (1996). Results and perspectives of the Goettingen project on quantitative linguistics. *J. Quantitative Linguistics*, 5, 155-162.
- Binongo, J.N.G. (1994). Joaquin's Joaquesquerie, Joaquesquerie's Joaquin: A Statistical Expression of a Filipino Writer's Style. *Literary & Linguistic Computing*, 9(4), 267-279.
- Bodmer, F. (1943). *The Loom of Language*. G. Allen & Unwin, London.
- Burrows, J.F. (1989). 'An Ocean Where each Kind...': Statistical Analysis and Some Major Determinants of Literary Style. *Computers & the Humanities*, 23, 309-321.

- Burrows, J.F. (1992). Not unless you Ask Nicely: the Interpretive Nexus between Analysis and Information. *Literary & Linguistic Computing*, 7(2), 91-109.
- Burrows, J.F. & Craig, D.H. (1994). Lyrical Drama and the "Turbid Montebanks": Styles of Dialogue in Romantic and Renaissance Tragedy. *Computers & the Humanities*, 28, 63-86.
- Forsyth, R.S. & Holmes, D.I. (1996). Feature-Finding for Text Classification. *Literary & Linguistic Computing*, 11(4), 163-174.
- Greenwood, H.H. (1995). Common Word Frequencies and Authorship in Luke's Gospel and Acts. *Literary & Linguistic Computing*, 10(3), 183-187.
- Gurney, P.J. & Gurney, L.W. (1998). Authorship attribution of the *Scriptores Historiae Augustae*. *Literary & Linguistic Computing*, 13(3), 119-131.
- Hale, J.R. (1971). *Renaissance Europe 1480-1520*. Fontana / Collins, London.
- Holmes, D.I. (1994). Authorship Attribution. *Computers & the Humanities*, 28, 1-20.
- Holmes, D.I. & Forsyth, R.S. (1995). The "Federalist" Revisited: New Directions in Authorship Attribution. *Literary & Linguistic Computing*, 10(2), 111-127.
- McCuaig, W. (1989). *Carlo Sigonio: The Changing World of the Late Renaissance*. Princeton University Press: Princeton.
- Mosteller, F. & Wallace, D.L. (1984). *Applied Bayesian and Classical Inference: the Case of the Federalist papers*. Springer-Verlag, New York. [First edition: 1964.]
- Nauert, C.G. (1995). *Humanism and the Culture of Renaissance Europe*. Cambridge University Press.
- Sage, E.T. (1910). *The Pseudo-Ciceronian Consolatio*. University of Chicago Press, Chicago.
- Schinke, R., Greengrass, M., Robertson, A.M. & Willett, P. (1996). A stemming algorithm for Latin text databases. *J. of Documentation*, 52, 172-187.
- Scott, I. (1910). *Controversies over the Imitation of Cicero*. Teachers College, Columbia University: New York.
- Tweedie, F.J., Holmes, D.I. & Corns, T.N. (1998). The Provenance of *De Doctrina Christiana*, Attributed to John Milton: A Statistical Investigation. *Literary & Linguistic Computing*, 13(2), 77-87.
- Wimmer, G., Koehler, R., Grotjahn, R. & Altmann, G. (1994). Towards a theory of word-length distribution. *J. Quantitative Linguistics*, 1, 98-106.

Appendix I -- Latin Text Samples.

Sample ID	Words	Work	Source
CONS.A	8302	First half of 1583 Consolatio	UCLA
CONS.B	8301	Second half of 1583 Consolatio	UCLA
EPISTULA.OCT	1136	Epistula ad Octavianum	Loeb
RHET.HER	1763	Rhetorica ad Herennium	Loeb
CAESAR.BC2	6406	De Bello Civile, Liber II	Lillard
CAESAR.GAL	8184	Bellum Gallicum I	Lillard
CICERO.AMI	9276	Laelius de Amicitia	Lillard
CICERO.ARC	3112	Pro A. Licinio Archia Poeta Oratio	Lillard
CICERO.ATT	9443	Letters to Atticus I	Lillard
CICERO.B1	11586	Brutus, 1-169	Lillard
CICERO.B2	12989	Brutus, 170-333	Lillard
CICERO.CLU	3130	Pro Cluentio, 1-16	Bristol
CICERO.FIN	7179	De Finibus Bonorum et Malorum I	Lillard
CICERO.IC2	3118	In Catilinam II	OXTA
CICERO.IMP	6690	De Imperio Cn. Pompei (Pro Lege Manilia)	Lillard
CICERO.LEG	3195	De Legibus, 1-32	Loeb
CICERO.MAR	2797	Pro M. Marcello Oratio	Lillard
CICERO.MIL	1147	Pro Milone, 22-33	Loeb
CICERO.ND2	9749	De Natura Deorum II	Lillard
CICERO.OFF	9074	De Officiis I, 1-101	Lillard
CICERO.ORA	2717	Orator, 112-144	Loeb
CICERO.PH2	11420	Philippics II	Lillard
CICERO.PH7	2212	Philippics, VII	Lillard
CICERO.RE2	1657	De Re Publica II, 1-19	Loeb
CICERO.SEN	6835	Cato Maior De Senectute, 1-71	OXTA
CICERO.SEX	3959	Pro Sexto Roscio Amerino Oratio	Lillard
CICERO.SOM	2201	Somnium Scipionis (De Re Publica VI)	Lillard
CICERO.SUL	3633	Pro Sulla, 1-36	OXTA
CICERO.T1	7688	Tusculan Disputations I, 1-75	Loeb
CICERO.T2	5989	Tusculan Disputations II	Packard
CICERO.T4	7776	Tusculan Disputations IV	Packard
LAUREDAN.FRA	4329	In Funere Francisci Venerii ...	BL
LAUREDAN.MAT	4777	In Funere M. Antonii Trivisanii ...	BL
MURETUS.2	2549	De Laudibus	UCLA
MURETUS.4	3737	De Philosophiae et Eloquentia ...	UCLA
MURETUS.6	2030	Pro Antonia Rege Navarre ad Pium ...	BL
MURETUS.6C	1666	Ingressus Explanare Ciceronis Libros ...	BL
MURETUS.20	2256	In Funere Pii V Pont. Max.	BL
MURETUS.23	3652	De Utilitate Iucunditate ac Praestantia	UCLA
MURETUS.26	2652	In Funere Pauli Foxii	UCLA
NEPOS.ATT	3479	Atticus	Lillard

NEPOS.CAT	440	Cato	Lillard
NEPOS.DIO	1483	Dion	Lillard
NEPOS.HAN	2078	Hannibal	Lillard
NEPOS.MIL	1356	Miltiades	Lillard
RICCOBON.BEN	2906	In Obitu M. Mantuae Benavidii ...	VL
RICCOBON.NIC	2967	Ad Nicolaum Pontium Venetiarum ...	CUL
RICCOBON.LEG	4946	De Legum Laudibus Oratio	VL
RICCOBON.PAT	2769	Philosophorum in Patavino ...	BL
RICCOBON.RHO	1408	Civis Rhodigini et Patavini Oratio	BL
RICCOBON.STU	3743	Oratio pro Studiis Humanitatis	CUL
SALLUST.BC	3335	Bellum Catilinae, 1-22	Loeb
SALLUST.JUG	1508	De Bello Iugurthino, 104-114	Loeb
SENECA.BRE	5880	De Brevitate Vitae	Lillard
SENECA.CON	4881	De Constantia Sapientis	Lillard
SENECA.IRA	588	De Ira, 1-2	Loeb
SENECA.OTI	1775	De Otio	Lillard
SENECA.PRO	3725	De Providentia	Lillard
SIGONIO.1	2405	Pro Eloquentia I	UCLA
SIGONIO.2	2238	Pro Eloquentia II	UCLA
SIGONIO.5	3000	De Latinae Linguae Usu Retinendo	UCLA
SIGONIO.6	3075	De Laudibus Historiae	UCLA
SIGONIO.DD	5682	De Dialogo Liber, pp 435-448	BL
SIGONIO.H4A	5616	Historiarum de Regno Italiae IV, pp 89-98	BL
SIGONIO.H4B	4079	Historiarum de Regno Italiae IV, pp 98-104	BL
TACITUS.AGR	6740	Agricola	Packard
VETTORI.FUN	2545	Oratio Funebris de Laudibus Ioannis M.	UCLA
VETTORI.HAB	2214	Oratio Habita in Funere ad Iulium III	UCLA
VETTORI.LAU	4347	Liber de Laudibus Ioannae Austriacae	UCLA
VETTORI.PET	4240	Oratio Petri Victorii in Max. II	UCLA

Note on Sources:

BL	British Library
Bristol	Bristol Classical Press
CUL	Cambridge University Library
Lillard	http://patriot.net/~lillard/cp/latlib
Loeb	Loeb Classical Library, Harvard UP / Heinemann
OXTA	Oxford Text Archive
Packard	Packard Humanities Institute
UCLA	UCLA Research Library
VL	Vatican Library

Appendix II -- Procedure for Latin Syllable Counting.

The following procedure takes as input a word (w) which has been extracted as a string from the text being read and delivers an integer as result which is interpreted as the number of syllables in that word. The numbered steps are executed in order, as below.

It should be noted that upper-case letters in w will already have been converted into lower case before reaching this procedure, and that any changes made to w are local and have no effect on the text outside this procedure.

For counting purposes any of the characters "aeiouy@" is treated as a vowel: "@" is never present on input, but is a device to help deal with certain diphthongs. The characters "j" and "w" are used in this routine to replace "i" and "u" only in contexts where these letters should not be counted as full vowels.

Within string w:

1. 'qu' becomes 'qw'
2. 'gu' becomes 'gw' in front of a vowel
3. At the beginning of w:
 - 'i' becomes 'j' after 'ab', 'ad', 'con' or 'ob' in front of 'a', 'e', 'o' or 'u'
 - 'iniu' becomes 'inju'
 - 'interiac' becomes 'interjac'
 - 'iec' becomes 'jec' after 'in', 'inter' or 'sub'
 - 'i' becomes 'j' in front of a vowel
4. 'i' becomes 'j' between 2 vowels
5. 'oe' becomes '@' except between 'p' and 'm', 's' or 't'
6. 'ae' becomes '@'
7. 'au' becomes 'aw'
8. 'hui' becomes 'hwi'
9. At the end of w:
 - 'eu' becomes 'ew'
10. The number of vowels in w is counted and returned as the result.

Figures

- Figure 1 PCA Cicero: Genre Effect
- Figure 2 Scaled Loadings Plot Cicero: Genre Effect
- Figure 3 PCA Cicero and Dubitanda
- Figure 4 Scaled Loadings Plot Cicero and Dubitanda
- Figure 5 PCA Cicero vs. Classical Controls
- Figure 6 Scaled Loadings Plot Cicero vs. Classical Controls
- Figure 7 Cluster Analysis Cicero vs. Classical Controls
- Figure 8 PCA Sigonio vs. Neo-Latin Controls
- Figure 9 Scaled Loadings Plot Sigonio vs. Neo-Latin Controls
- Figure 10 Cluster Analysis Sigonio vs. Neo-Latin Controls
- Figure 11 Discriminant Analysis Cicero, Sigonio and the *Consolatio*
- Figure 12 Discriminant Analysis Classical Latin, Neo-Latin and the *Consolatio*
- Figure 13 Discriminant Analysis Words and Syllables, Classical and Neo-Latin
- Figure 14 Discriminant Analysis Words and Syllables, Cicero and Sigonio
- Figure 15 Discriminant Analysis Muretus, Sigonio and the *Consolatio*
- Figure 16 Muretus, Sigonio and the *Consolatio*.

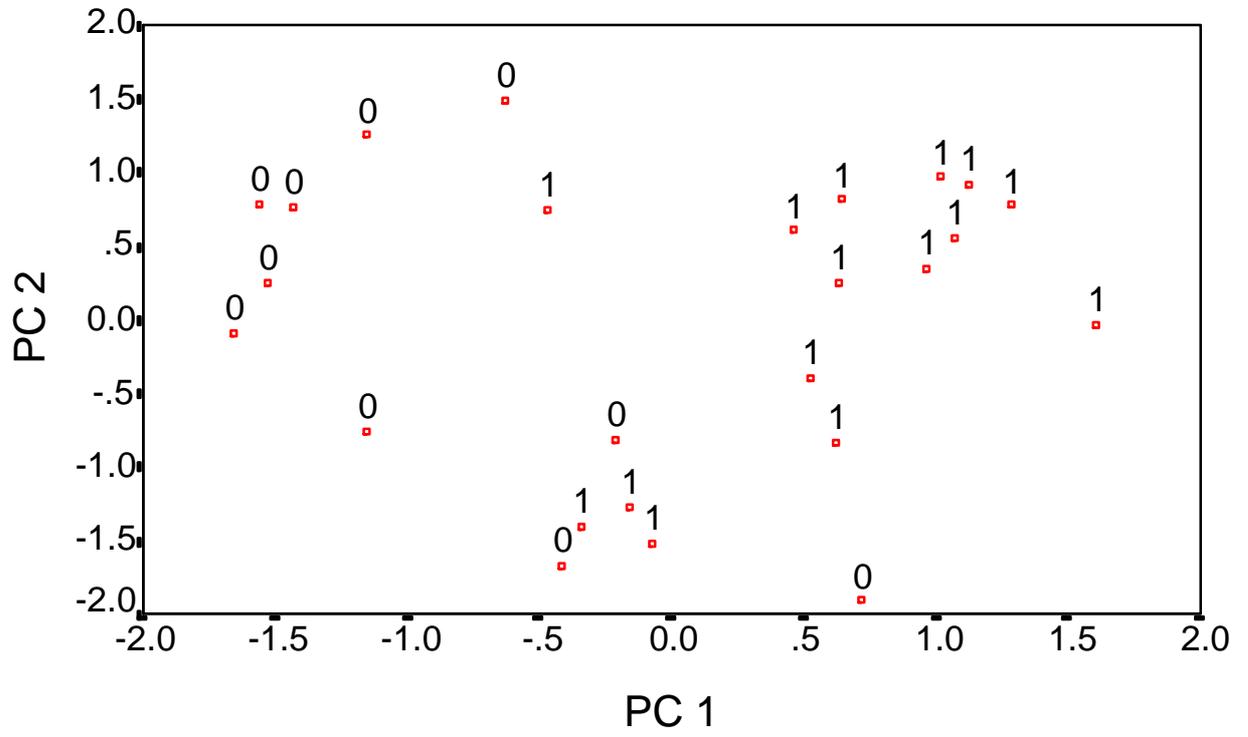


Figure 1: PCA Cicero: Genre Effect (0=oration)

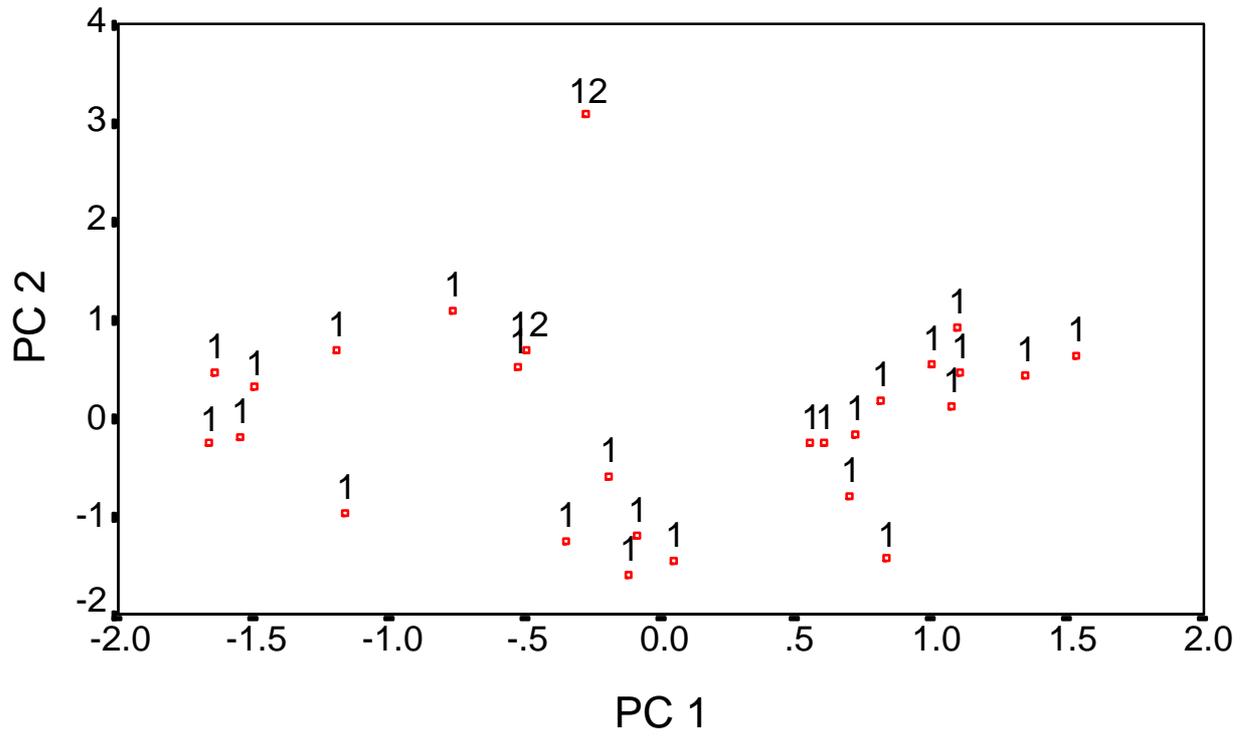


Figure 3: PCA Cicero and Dubitanda (12=Dubitanda)

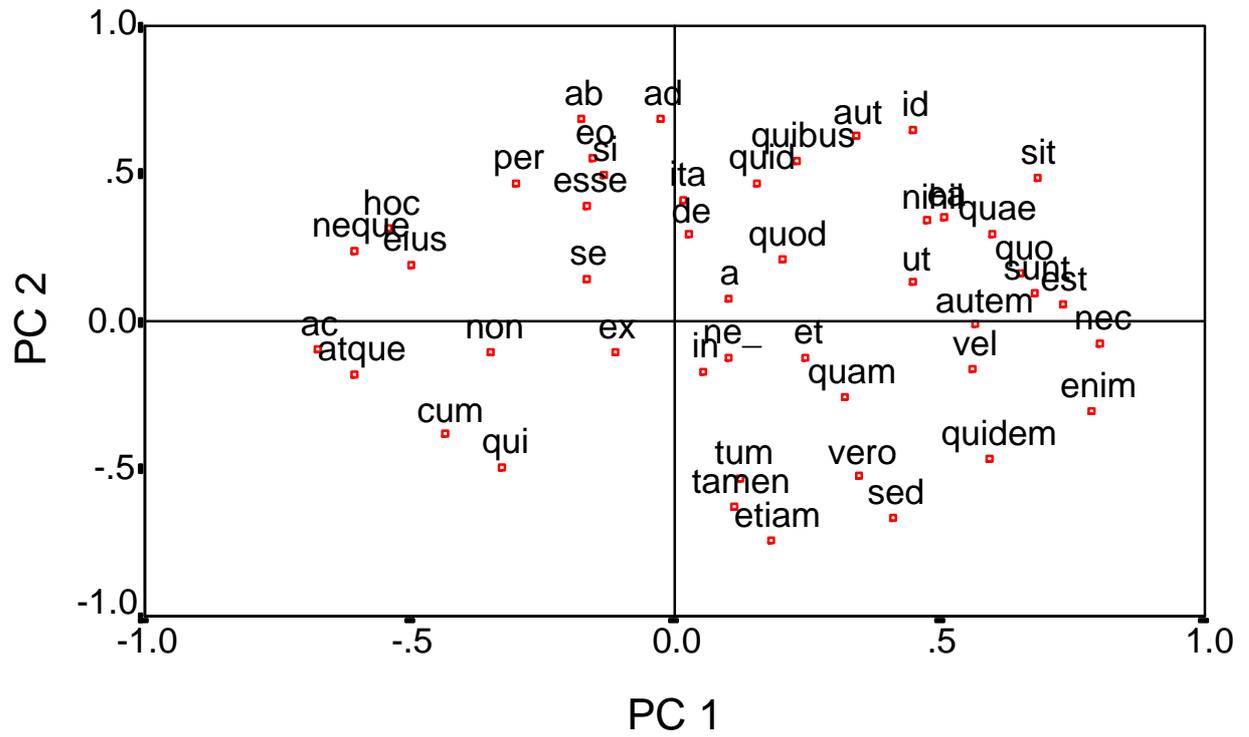


Figure 4: Scaled Loadings Plot Cicero and Dubitanda

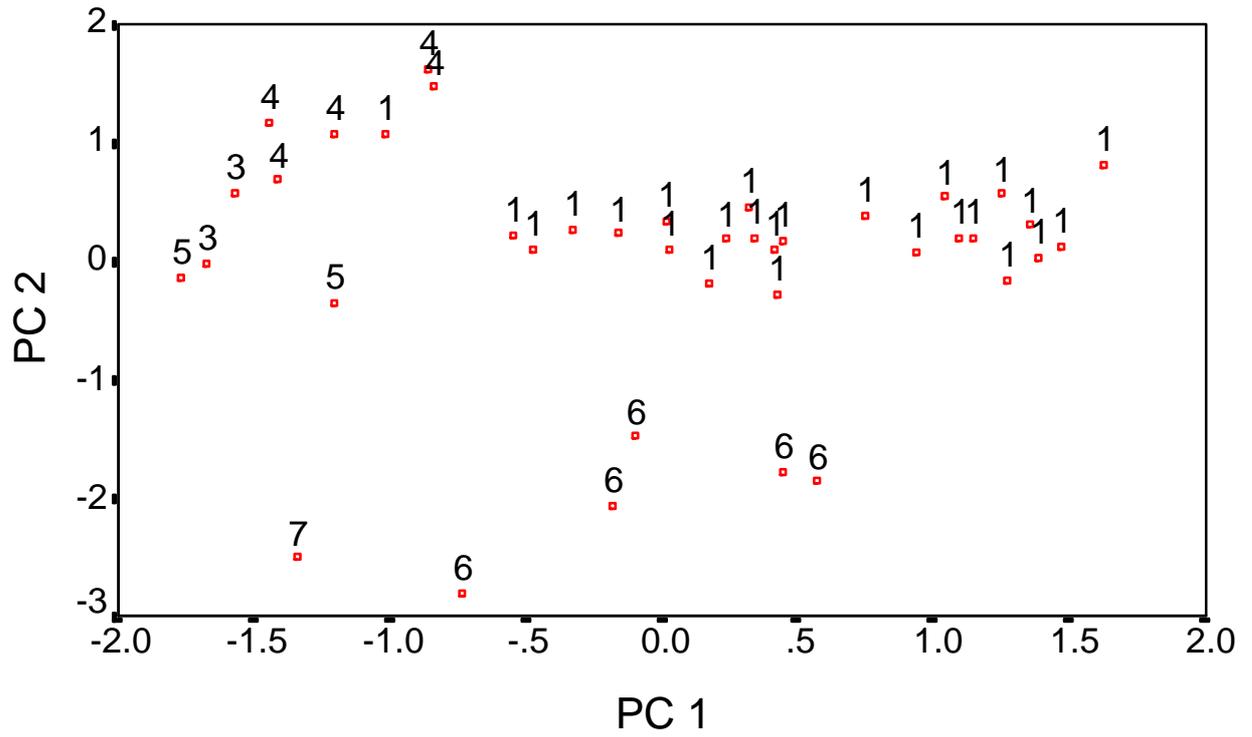


Figure 5: PCA Cicero vs. Classical Controls

[Key:

- 1 = Cicero
- 3 = Caesar
- 4 = Nepos
- 5 = Sallust
- 6 = Seneca
- 7 = Tacitus.]

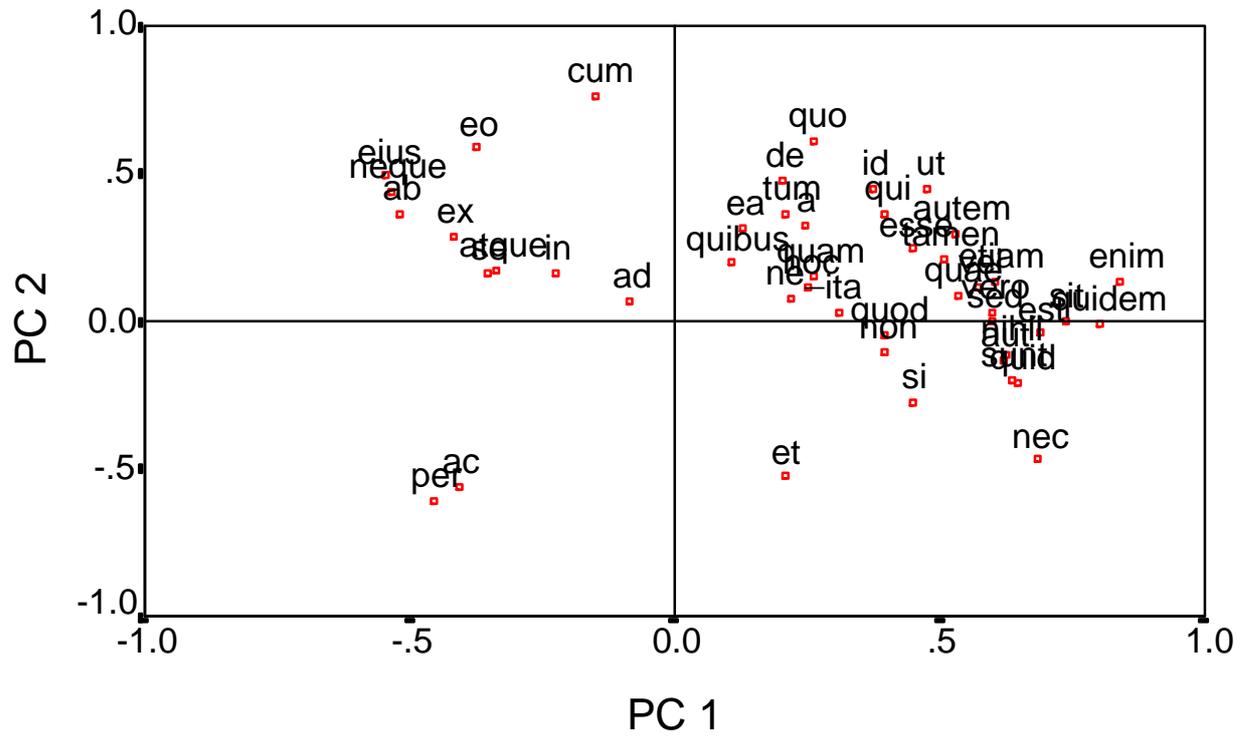


Figure 6: Scaled Loadings Plot Cicero vs. Classical Controls

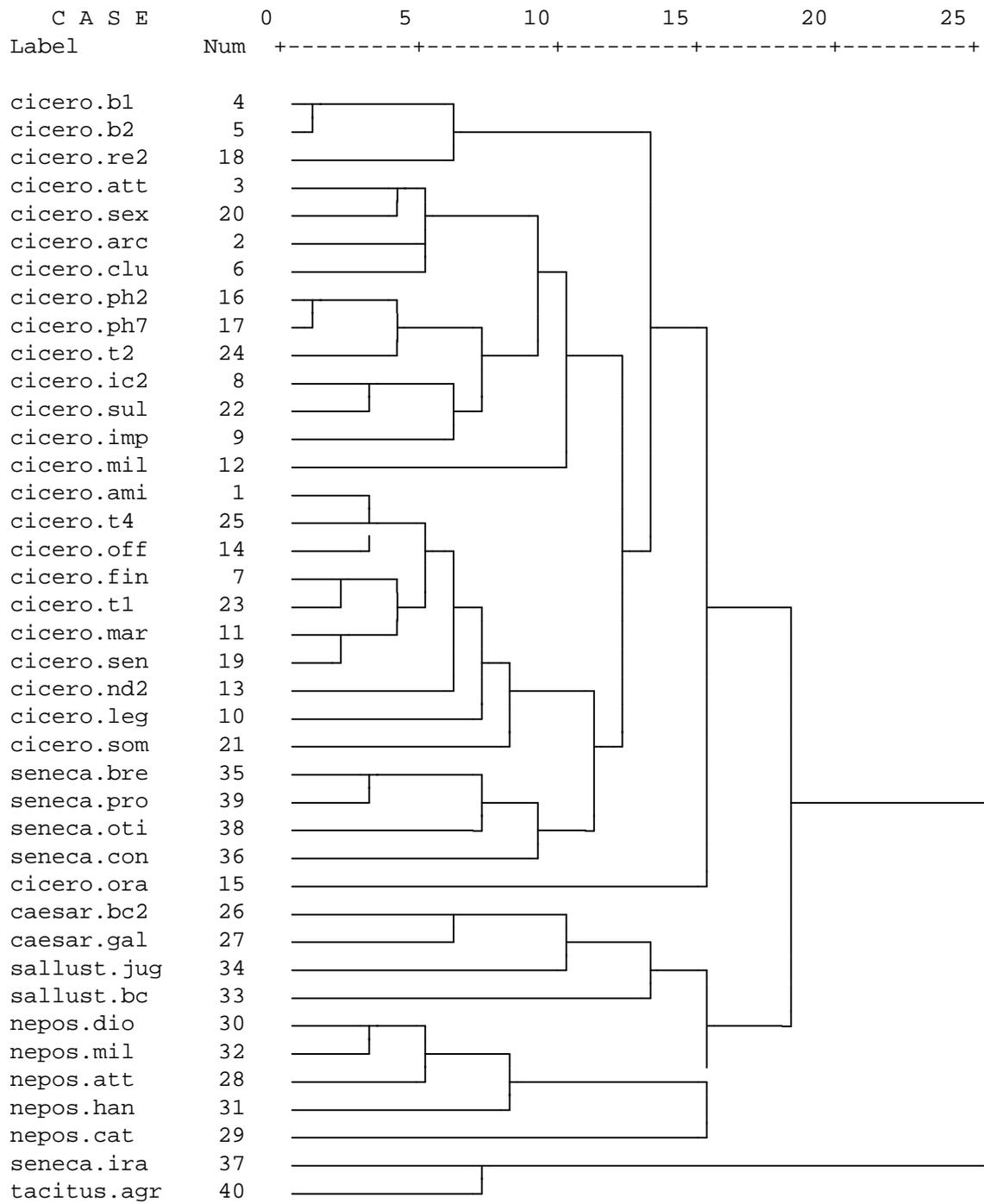


Figure 7: Cluster Analysis Cicero vs. Classical Controls

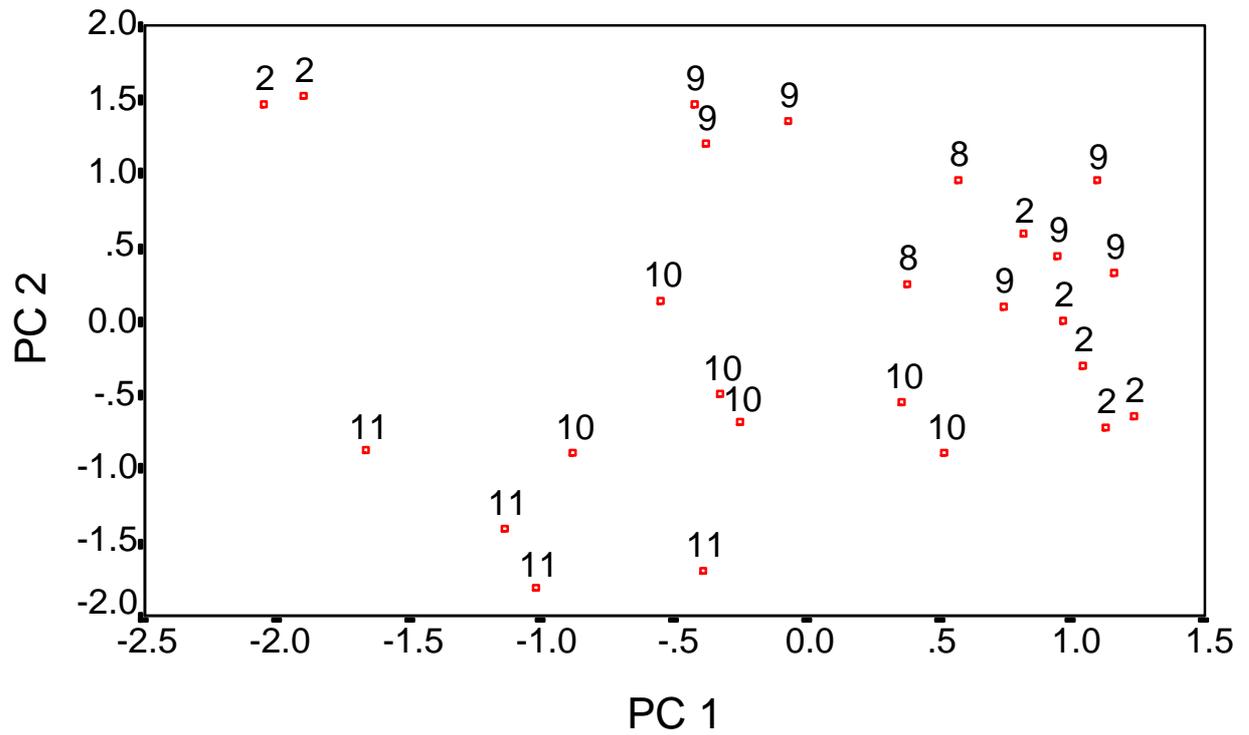


Figure 8: PCA Sigonio vs. Neo-Latin Controls

[Key:
 2 = Sigonio
 8 = Lauredanus
 9 = Muretus
 10 = Riccoboni
 11 = Vettori.]

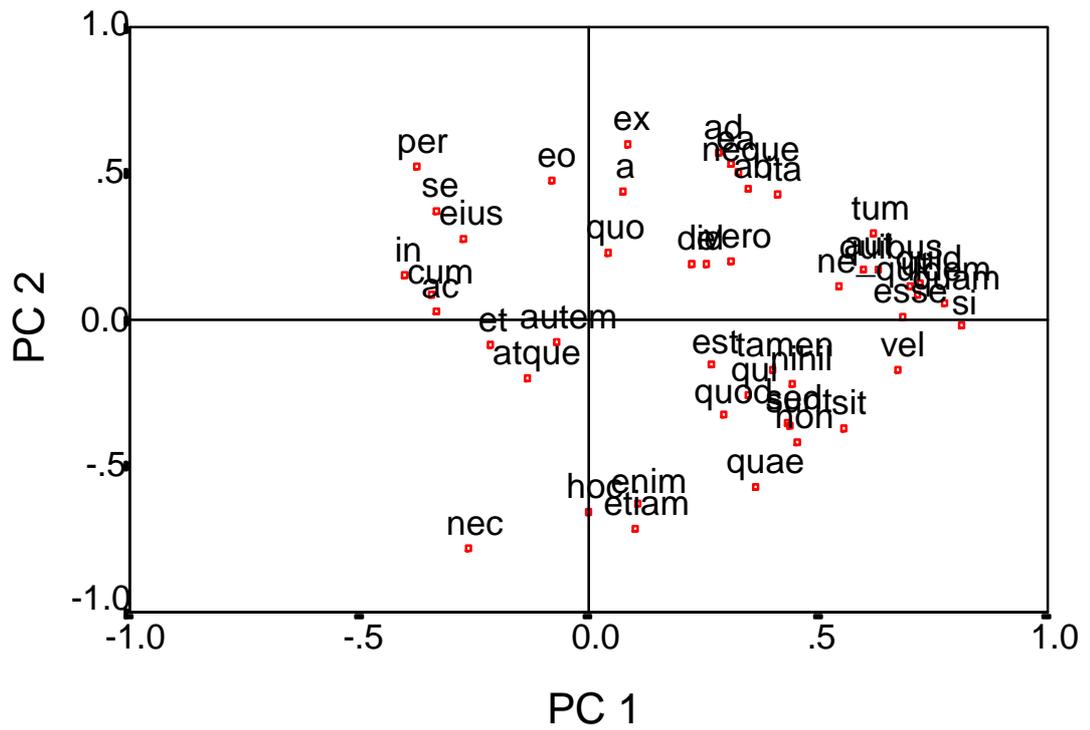


Figure 9: Scaled Loadings Plot Sigonio vs. Neo-Latin Controls

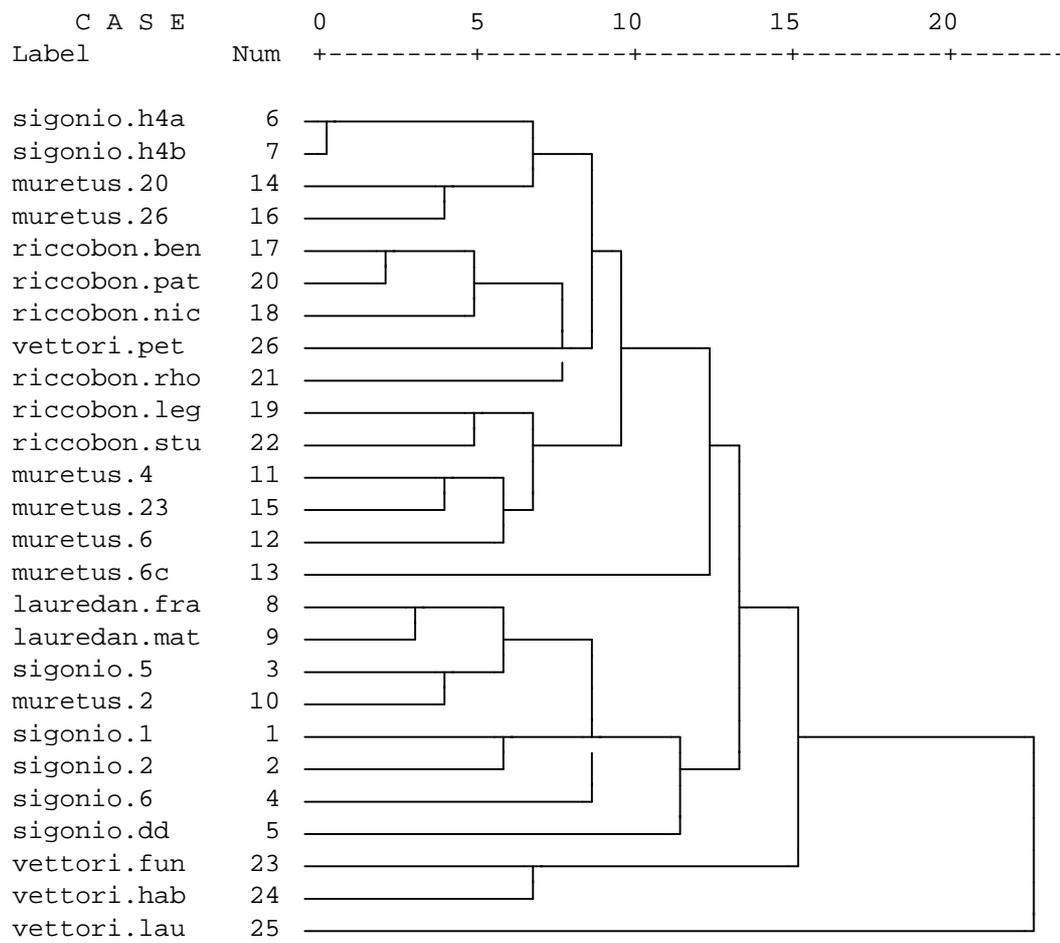


Figure 10: Cluster Analysis Sigonio vs. Neo-Latin Controls

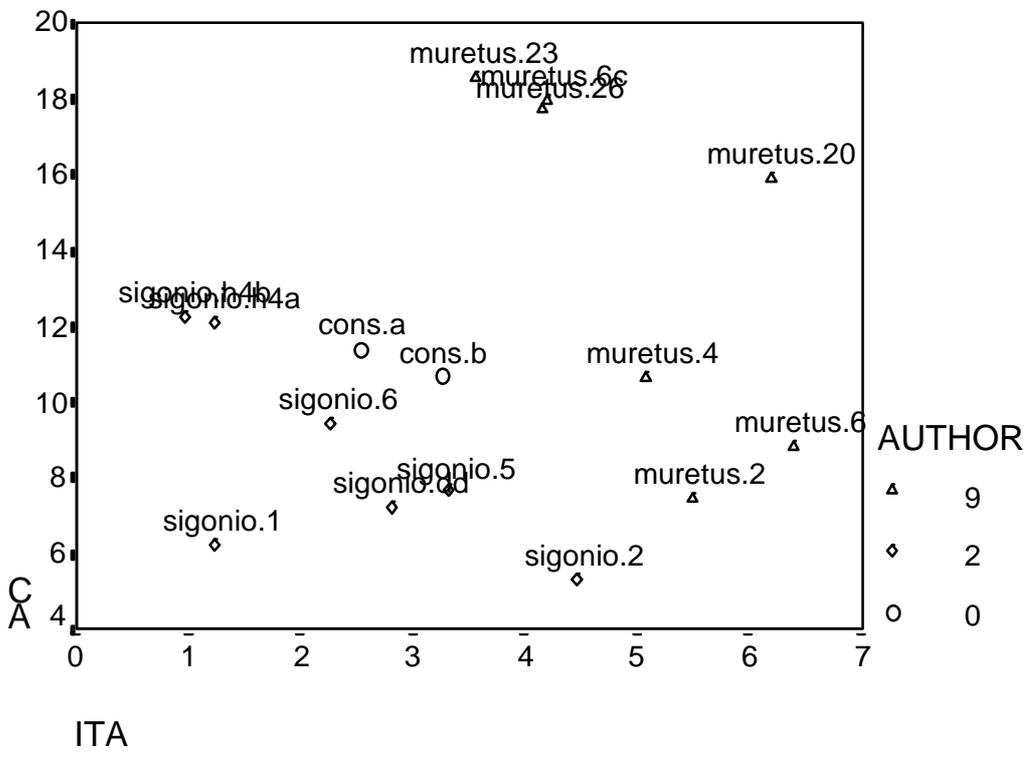


Figure 16: Muretus, Sigonio and the *Consolatio* .