
An Examination of the Authorship Attributions of Two Major Roman Authors

Lyman W. Gurney (lwgurney@telus.net)

Themis Research (ret.)

Penelope J. Gurney (pgurney@uottawa.ca)

University of Ottawa (ret.)

Introduction

This paper describes a stylometric analysis of 16 texts attributed by the manuscript tradition to two Roman authors of the first century BCE: Gaius Julius Caesar (100-44; all dates BCE, unless noted), and Gaius Sallustius Crispus, commonly referred to as Sallust (86-34). The 'control' consists of the *Lives of the Twelve Caesars* of Gaius Suetonius Tranquillus (77-121 CE at the earliest). All twelve works are attributed by the tradition to Suetonius, and our analysis, given in earlier publications, has corroborated this tradition of single authorship.

The texts analysed include the fourteen books attributed to Caesar, and two books by Sallust. The texts of Caesar comprise: the eight books on the war in which he conquered Gaul in the years from 58 to 50; the three books on the following Civil Wars (49-48), in which Caesar eliminated his great political rival Pompey and his Eastern army; and the three books on his final battles against the surviving generals: in Alexandria (48-47), Africa (47-46), and Spain (46-45). This study then completes a similar analysis of the works of Sallust: the war in Africa against the Numidian king Jugurtha (111-106); and the failed rebellion of the Roman noble Lucius Sergius Catilina, known as Catiline, in 63.

Of the eight books of the Gallic Wars, the question arises as to exactly when and by whom they were created. Some have argued that the first seven were written as a group, but that the eighth must be attributed to his general Hirtius, who claims in the work that he was responsible for both that and the later text on the war in Alexandria.

Of the further works, the three on the Civil Wars, from his crossing of the Rubicon in 49, to the battle of Pharsalus in Thessaly in 48, are generally accepted to have been authored by Caesar. There is considerable disagreement, however,

concerning the three wars in Egypt, Africa and Spain: the Alexandrian War, as noted, is claimed by Hirtius; the origin of the African War is uncertain; and the internal character of the text of the Spanish War clearly defines it to be the creation of an unknown person. There is little disagreement on this last score, since the text on the Spanish War comprises some of the worst Latin extant, and was probably intended to be the raw material for a more structured history.

Statistical Routines

The Stylometric Analysis of the 28 texts has been conducted by use of the SPSS routines Hierarchical Cluster Analysis, and Principal Component Analysis. Discriminant Analysis has then been used to test the group memberships suggested by the first two.

Data

The data for the statistical routines have been provided by a matrix of 9,000 by 58 real-valued elements that represent the normalized frequencies of occurrence of unique lemmas (dictionary head-words) in 58 texts. This matrix has been generated from the fully disambiguated texts of the 58 works of Caesar, Sallust, Suetonius, and the *Scriptores Historiae Augustae*, and involves a reduction from 329,000 to 305,000 numeric values to be handled after the removal of all proper nouns. This matrix has been sorted in decreasing order on the frequencies of lemmas, and lists also the number of texts in which each individual lemma is not found. It has therefore been easy to choose for analysis the most frequent function words, verbs, nouns & pronouns, and adjectives in the texts under consideration.

The main thrust of the research has been conducted on the data set of function words, but, because the most frequent verbs, nouns, and adjectives can be identified so accurately in a disambiguated and tagged text, it has been possible to compare the statistical results of these three parts of speech with those from function words, which themselves are considered in the literature to be standard as data in stylometric analysis. The first thing noticed, however, has been the necessity of comparing the results from a full set of lemmas, with a set from which the most frequent 2 or 3 have been removed. These few very frequent lemmas can apparently overwhelm the effects of the other lemmas, and skew the results slightly, but noticeably.

Analysis

In a test of 37 function lemmas (with the removal of the three most frequent: *et*, *in*, and the separable suffix *que*), the

twelve works of Suetonius, the 'control' works, demonstrate that the lives remain closely grouped, as found in our earlier research, with only the life of Titus being slightly removed from all others. When all 40 of the top function lemmas are involved, however, there are slight changes in the distances of most lives, and the life of Otho joins that of Titus at the further remove, although this removal does not involve any question of a change of authorship.

In an analysis of Sallust's *Catiline* and the *Jugurthan War*, the results from an analysis of the 37 most common function lemmas reinforce the manuscript tradition of authorship, with a very close association; with all 40 of the most frequent lemmas, however, the relative spacing of the two works approaches even a possible difference in authorship.

The analysis of 37 of the 40 most common function lemmas in the works of Caesar provides a greater complexity. There is, first of all, a very close association between *Gallic VIII* and the *Alexandrian War*, with both being clearly separated from most of the other works attributed: an apparently clear vindication of the claims by Hirtius to be the author of both. The *Spanish War*, that work of execrable Latin, is obviously of totally separate authorship. The most remarkable result, however, is the distance between Book I of the *Gallic Wars* (that book with the famous beginning: "Gallia est omnis divisa in partes tres" - "All Gaul is divided into three parts"), and the other works attributed to Caesar (other than the *Spanish War*).

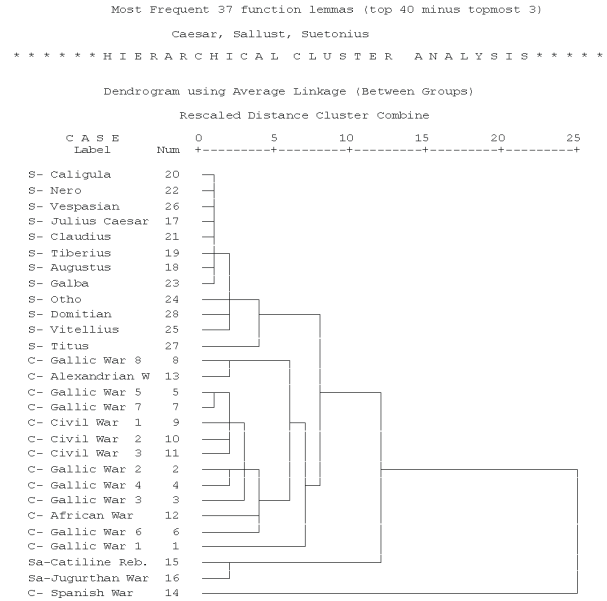
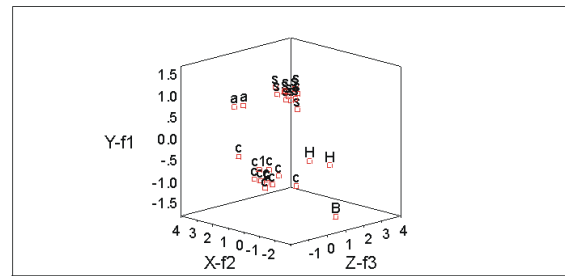


Figure 2

Principal Component Analysis

Caesar, Sallust, Suetonius

37 Function Lemmas



c-Caesar, a-Sallust, S-Suetonius, H-Hirtius

c1- Gallic War 1, B- Spanish War

Figure 3

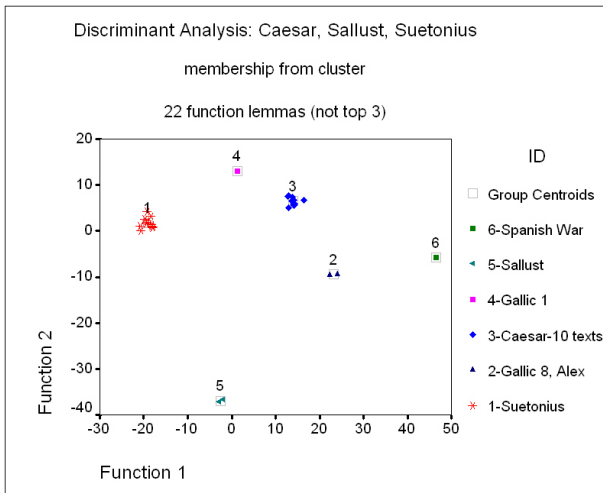


Figure 1

When all 40 top function lemmas are employed, most separations between the works attributed to Caesar increase. *Gallic I* and *Gallic IV* are now both at a further remove from the other works, and are hardly within reasonable attribution to Caesar. The *Alexandrian War*, however, is now very close to other texts, and far from *Gallic VIII*; and both the *African War* and *Gallic VIII* are now possibly beyond any reasonable attribution to Caesar, although the attribution of *Gallic VIII* to Hirtius remains possible. The *Spanish War* continues to be far distant from any other work in the 28 studied.

The analyses of verbs, nouns, and adjectives, all demonstrate considerable differences amongst the works. For example, *Gallic VIII* and the *Alexandrian War* become relatively distant in the analysis of nouns (less the top three), although they are still

closer to one another than to any other works. In the analysis of adjectives, however, they appear no longer to be related, and in addition, *Gallic VIII* becomes quite close to several other books of the *Gallic Wars* and *Civil Wars*.

Conclusions

It appears clear that the arguments in the stylometric literature, describing the necessity of using function lemmas, remain valid. Nonetheless, verbs, nouns, and adjectives must not be discarded as being inferior to function lemmas in the identification of authorship, since they provide valuable insights to the Latinist on the individual differences in word usage by the various authors.

The conclusion to be drawn from the differentiation between the use of all most common lemmas and those lemmas with the top 2 or 3 removed appears to be that a blind use of the most frequent lemmas can skew the results, and demonstrates that close cooperation between statistician and Latinist is required.

The overall conclusion on the authorship attributions to Caesar and Sallust are clear. The two works of Sallust appear to be correctly attributed. The attributions to Caesar remain complex, however: the text on the *Spanish War* is undeniably not that of any literate Roman; and *Gallic VIII* and the *Alexandrian War* appear definitely to be of separate authorship, and quite possibly that of Hirtius, as the tradition claims. The most striking fact, however, is that the first book on the *Gallic Wars* is of a high literary quality, yet is undeniably different from the other works of Caesar. Hence it now lies in the realm of the Latinist for a full analysis of the author's uses of all parts of speech, and the manner in which he apparently poured so much more effort into this first book that brought his conquests in a new land to the attention of the Roman People who would later be voting on his bitterly fought candidature for the Consulship.

Bibliography

Gurney, L.W., and P. Gurney. "The Scriptorum Historiae Augustae: History and Controversy." *Literary and Linguistic Computing* 13.3 (1998): 105-109.

Gurney, L.W., and P. Gurney. "Authorship Attribution of the Scriptorum Historiae Augustae." *Literary and Linguistic Computing* 13.3 (1998): 119-131.

Gurney, L.W., and P. Gurney. "Subsets and Homogeneity: Authorship Attribution in the Scriptorum Historiae Augustae." *Literary and Linguistic Computing* 13.3 (1998): 133-140.