

A Statistical Approach based on Correspondence Analysis of Two Latin Corpora: the *Historia Augusta* and the *Corpus Tibullianum**

Vittoria Sacco

Institut de Statistique
Faculté des Sciences économiques
Université de Neuchâtel
Pierre à Mazel 7, 2000 Neuchâtel

Two Latin corpora are analysed by means of lexicographic methods to find support for current hypotheses made by philologists about their composition.

The first corpus, *Historia Augusta*, deals with Imperial Rome, its rulers and usurpers, and its background of political instability, lethal gossip and brutal conquests resulting in one of the greatest empires in History. The *Historia Augusta* includes thirty biographies of Roman emperors from Hadrian (117-138) to the advent of Diocletian (284). Using statistical methods based on correspondence analysis, this study sought to demonstrate that all thirty biographies were written by a single author adopting various postures represented by different names. 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 emperor's 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 were removed from the text and a second analysis was performed, the result of which shows no distinct groups in graphic visualisation. The hypothesis of a single writer seems most probable. The words "item", reflecting alteration of the original text, and "de" followed by an ablative instead of the archaic/classical genitive, contribute most significantly in axes building. On the factorial plan, these two words are associated with the most recent emperors, and thus emphasize a chronological sequence in the writing of the biographies. In view of the factorial plan, a chronological axis separates earlier emperors from later ones. The use of "item" suggests some kind of reworking or possibly reveals the addition in some biographies of a subjective element which does not reflect the historian/author's initial perspective.

The second corpus under examination brings us even further back in time. 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 circle. Some poems may have been written by Sulpicia, making her the

* Master thesis defended at the University of Neuchâtel in February 2011.

only Latin poetess whose work may 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) book includes elegies by one Lygdamus and by Sulpicia, in addition to the so-called “Panegyric of Messalla” and others 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. In addition, 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, subclusters 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 subcluster. 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.