

# Une approche statistique basée sur l'analyse factorielle des correspondances de deux corpora latins: *l'Histoire Auguste* et le *Corpus Tibullianum*\*

Vittoria Sacco

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

Deux corpora latins ont été analysés à l'aide de méthodes lexicographiques afin de confirmer ou d'infirmer des hypothèses proposées par des historiens de la littérature au sujet de la singularité ou de la pluralité des auteurs des ensembles respectifs.

Le premier corpus, intitulé *Historia Augusta*, plonge le lecteur dans la Rome impériale, avec ses usurpateurs, son instabilité politique et ses intrigues mortelles, mais aussi ses conquêtes brutales qui l'ont amenée à former l'un des plus grands empires de l'histoire. *L'Historia Augusta* comprend trente biographies d'empereurs romains d'Hadrien à l'avènement de Dioclétien. Grâce à des méthodes statistiques, plus précisément à une analyse factorielle des correspondances (AFC), des indices permettent d'attester ou non que ces biographies ont toutes le même auteur et que certaines d'entre elles ont été écrites avec des postures variées sous des noms différents. La première analyse reposait excessivement sur les prénoms des empereurs. Comme le corpus est composé de biographies, le lecteur doit savoir qu'il n'est pas rare de trouver au début de chaque biographie une énumération de tous les membres et ancêtres de la famille impériale concernée. En outre, il était courant que le même prénom soit réutilisé en sautant une génération. Cela élucide la fréquence élevée des prénoms et explique pourquoi certains d'entre eux caractérisent une lignée d'empereurs. Pour ces raisons, l'analyse n'est pas suffisamment représentative pour pouvoir en tirer une conclusion. Ainsi, les prénoms des empereurs ont été retirés du fichier texte initial et une deuxième analyse a été effectuée. Cette fois, la visualisation graphique ne permet pas d'identifier des groupes distincts. L'hypothèse d'un seul auteur semblerait donc être la plus vraisemblable. Les mots « item », utilisé pour modifier le texte original, et « de » suivi d'un ablatif, substitut progressif du génitif archaïque et classique, contribuent le plus à la construction des axes. Dans le plan factoriel, ils sont associés avec les empereurs les plus récents, ce qui révèle une séquence chronologique dans l'écriture des biographies. Lorsqu'on visualise le plan factoriel, une ligne imaginaire semble opposer les empereurs anciens aux plus récents. La présence du mot « item » suggère un remaniement ou peut-

---

\* Le mémoire de maîtrise a été soutenu à l'Université de Neuchâtel en février 2011.

être un point de vue personnel dans certaines biographies. Mais celui-ci ne refléterait pas les postures initiales adoptées par l'auteur.

Avec le second corpus, on remonte encore plus haut dans le temps pour rencontrer un ou certains des plus importants poètes élégiaques latins. En effet, le *Corpus Tibullianum* regroupe les œuvres de Tibulle et d'autres poètes du cercle de Messalla, d'époque augustéenne (env. 40 av. J.-C à 14 apr. J.-C.). Certaines de ces élégies auraient été écrites par l'unique poétesse latine conservée, Sulpicia. Ce corpus est divisé en trois livres (ou quatre, selon les éditions). Le premier et le deuxième livre sont attribués à Tibulle. Le troisième livre comprend les élégies dites de Lygdamus, celles de Sulpicia, le « Panégyrique de Messalla » et d'autres d'origine douteuse. Dans ce cas, on cherche à vérifier les auteurs des trois livres et à identifier les écrivains des élégies anonymes. La première analyse des correspondances met en lumière deux points extrêmes qui engendrent une agglomération des autres élégies et l'impossibilité d'interprétation des résultats. Le premier point extrême n'est pas surprenant puisqu'il correspond au « Panégyrique de Messalla ». Ce dernier se caractérise par un vocabulaire qui lui est propre et sa thématique est bien différente de celle des autres élégies. Le second point extrême est l'élégie 3.20 qui est attribuée à « Incerti Autores ». Cette élégie présente un vocabulaire singulier. Des mots communs dans les autres élégies sont complètement absents de celle-ci. On peut conclure d'emblée que ces deux élégies ne sont pas des œuvres de Tibulle et constituent des cas à part au sein même du *Corpus Tibullianum*. Après avoir éliminé ces deux points extrêmes, on a procédé à une autre analyse des correspondances. Dans ce cas, trois groupes distincts apparaissent dans le plan factoriel. Ils correspondent aux trois livres. Pour le troisième on voit clairement le groupe des élégies de Sulpicia et celui des élégies de Lygdamus. Les élégies ayant des écrivains anonymes se trouvent dans le milieu du plan factoriel ce qui ne permet pas de les attribuer avec certitude à un des auteurs. Les seules exceptions sont les élégies 3.9 et 3.19, qui appartiennent visiblement au groupe de Sulpicia. De plus, selon la position des élégies dans le plan factoriel, le vocabulaire utilisé et une évolution des thématiques, le premier livre semble avoir été écrit avant le deuxième.

Pour conclure, les hypothèses initiales, formulées par les historiens de la littérature latine sur des arguments philologiques et littéraires, semblent trouver une confirmation dans l'approche statistique fondée sur une analyse factorielle des correspondances.