# Comparison of latent semantic analysis and correspondence analysis as ordination methods in computational linguistics

Alvarez-Esteban Ramón[1], Bécue-Bertaut Mónica[2]

[1]Universidad de León – ramon.alvarez@unileon.es

[2]Universitat Politècnica de Catalunya – monica.becue@upc.edu

## Abstract

Analyzing large amounts of textual data in natural language requires data visualization tools to make sense of this information. For this purpose, latent semantic analysis is a widely used method in computational linguistics, while computational statistics favors correspondence analysis. Although these methods share the same goals and certain procedures, they have not been compared much. This is what we want to do here by investigating their ability to reveal and visualize any intrinsic patterns that the textual data may possess by projecting them onto reduced dimensional spaces.

**Keywords:** Correspondence analysis, Latent semantic analysis, Textual data visualization.

## Résumé

L'analyse de grandes quantités de données textuelles en langage naturel nécessite des outils de visualisation afin de donner un sens à ces informations. Dans ce but, l'analyse sémantique latente est une méthode largement utilisée en linguistique computationnelle, tandis que l'analyse des correspondances est plus choisie en statistique computationnelle. Bien que ces méthodes partagent les mêmes objectifs et certaines procédures, elles n'ont été que peu comparées. C'est ce que nous voulons faire ici en étudiant leur capacité à révéler et à visualiser les structures intrinsèques existantes dans les données textuelles à partir de leur projection sur des espaces de dimension réduite.

**Mots clés:** Analyse des correspondances, Analyse sémantique latente, Visualisation de données textuelles