

# Des outils pour décrire certains corpus de poèmes et de chansons : les arbres additifs simultanés

Ludovic Lebart

CNRS (R) – ludovic@lebart.org

## Abstract

The statistical analysis of poetic texts and songs is a methodological challenge. These texts may include choruses, repetitions or versification constraints which question the statistical significance of lexical frequencies. We must then work on "bags of words" (presence or absence of words) which evidently constitute a tiny part of the artistic aspect of the texts, but nevertheless prove worthy of interest in terms of the study of an artwork. With this coding, the dimensionality and sphericity of the cloud do not allow good visualizations in a low-dimensional space. Correspondence Analysis (CA), despite its undeniable advantage of providing a simultaneous representation of words and texts, becomes insufficient. The calculation of additive trees is essential. In this contribution, we propose a new procedure for the simultaneous representation of texts and words for additive trees. Such procedure allows us to combine the advantages of CA and clustering. We briefly illustrate this new tool with terse examples of application to corpora of lyrics of three francophone iconic singers and poets: Brassens, Aznavour and Brel.

**Keywords:** Additive Trees, Correspondence Analysis, Data Visualization, Poetry, Lyrics.

## Résumé

L'analyse statistique des textes poétiques et des chansons est un défi méthodologique. Ceux-ci peuvent comporter des refrains, des répétitions ou des contraintes de versification qui mettent en question la signification statistique des fréquences lexicales. On doit alors travailler sur des « sac de mots » (présence ou absence de vocables) qui ne constituent qu'une partie infime de l'aspect artistique des textes, mais s'avèrent cependant dignes d'intérêt au niveau de l'étude d'une œuvre. Avec ce codage, la dimensionnalité et la sphéricité du nuage ne permettent pas de bonnes visualisations dans un espace de faible dimension. L'analyse des correspondances (AC) devient insuffisante, malgré sa précieuse propriété de représentation simultanée des mots et des textes. Le calcul d'arbres additifs s'impose. Nous proposons dans cette contribution une nouvelle procédure de représentation simultanée des textes et des mots pour les arbres additifs. Cette procédure permet de cumuler les avantages de l'AC et des classifications. On illustrera très schématiquement ce nouvel outil par des applications aux corpus de trois paroliers/chanteurs et poètes francophones Brassens, Aznavour et Brel.

**Mots clés :** Arbres additifs, Analyse des Correspondances, Visualisations, Poésies, Chansons.