



Publié sur *La Vie des Classiques* (<https://96.ip-213-32-20.eu>)

[Accueil](#) > Thematic features for intertextual analysis (Chris Forstall and Lavinia Galli Milić)

THEMATIC FEATURES FOR INTERTEXTUAL ANALYSIS (CHRIS FORSTALL AND LAVINIA GALLI MILIĆ)

The study of intertextuality in Classical poetry often presents itself as a specialized case of text-reuse detection: commentaries and other close readings of a work concern themselves with the identification and exegesis of phrases borrowed from earlier texts. Yet it has long been understood that larger-scale, structural parallelisms can also exist between texts (Genette 1997), and that these can provide the context necessary to establish an allusive or intertextual link between two phrases (Wills 1996). Automatic detection of intertextuality must take into account features at various scales: from individual phonemes to larger syntactic units and type scenes. This idea was proposed by, among others, Bamman and Crane (2008), and already forms one component of existing text-reuse search tools (for example, Büchler 2010). In this work in progress, we attempt to develop scene-level features to improve the automatic scoring of word-level text-reuse for its allusive significance. Considering small samples of 30 sequential verses in a bag-of-words model, we test three different featuresets: * tf-idf weights * tf-idf weights after an attempt to correct for authorship signal * Latent Dirichlet allocation with 50 topics We then perform unsupervised clustering on the samples using k-means. The results of this clustering are compared with a human classification both of the automatically-generated samples and of scenes as delineated by critics.
