

Setting bounds in a homogeneous corpus: a methodological study applied to medieval literature

Jean-Baptiste Camps*, Florian Cafiero**

*Laboratoire *Études et édition de textes médiévaux* (EA 4349)

Université Paris – Sorbonne

1, rue Victor Cousin

75005 Paris

Jean-Baptiste.Camps@paris-sorbonne.fr

<http://www.paris-sorbonne.fr/>

**École centrale de Paris

Grande voie des Vignes

92295 Châtenay-Malabry CEDEX

florian.cafiero@polytechnique.edu

<http://polytechnique.academia.edu/FlorianCafiero>

Abstract. The authors present here an exploratory and unspecific method that does not necessitate any *a priori* on the data – or any heavy transformation such as lemmatisation – that would have to be understood as a first step in the apprehension of a corpus. After a first phase of calibration, based on a control sample, the authors will introduce a method which heuristic value is to bring out, at different levels, internal divisions of different kinds (diachronic, diatopic, related to authorship or scribes, ...), that can then be analysed specifically. The authors illustrate this method by applying it to a corpus of Occitan medieval texts, the *vidas*. The corpus's authors and origins are in good part unknown.

1 A peculiar corpus: the *vidas*

1.1 Elements of context

Amongst the relatively small corpus of Occitan prose literature from the XIIIth century that has survived to this day, the *vidas*, along with the *razos*, constitute a corpus of roughly 225 texts known as the “biographies” of the troubadours since Raynouard’s edition in 1820. Several debates surround these texts’ origin: were they performed at first or were they created specifically to be integrated into manuscripts? Did they appear in Northern Italy or do they come from an older Languedocian tradition? Despite the lack of medieval unambiguous vocabulary, there is an apparent distinction between commentaries, explanation of a specific poem (a *razo*), and more general biographical narratives detached from a specific work (*vidas* as we call them).