

Autour du Corpus de la Renaissance littéraire gasconne (CoRLiG)

Gilles Guilhem Couffignal*, Lucence Ing†

3 avril 2017

Le projet CoRLiG (*Corpus de la Renaissance littéraire gasconne*) a pour but de favoriser l'étude d'œuvres littéraires multilingues (occitan, français, langues anciennes et étrangères) produites dans la seconde moitié du XVI^e siècle dans le Sud-Ouest de la France. Pour ce faire, une chaîne de traitement numérique complète est mise en place, depuis l'acquisition des données par reconnaissance des caractères (OCR) jusqu'à l'exploitation textométrique et la publication, web et papier. Le principe retenu est de procéder par constitution d'un corpus électronique à petite échelle, permettant la mise au point progressive d'outils réutilisables (modèles d'OCR, DTD, protocoles d'exportation des données pour analyse textométrique ou publication) et favorisant la prise en compte d'œuvres singulières plurilingues, tout en ouvrant la voie à une base de données de large empan pouvant concerner l'ensemble des productions en occitan prémoderne.

De nombreux projets de corpus ont déjà vu le jour dans les études médiévales. Alors que la *Concordance de l'occitan médiéval* (COMM I et II) est devenue un usuel incontournable, des publications en cours telles que celle du Petit Thalamus de Montpellier ou du *Corpus électronique gascon* ont montré leur intérêt comme projets fédérateurs d'équipes interdisciplinaires ou comme vecteurs de nouvelles analyses linguistiques. Ces grands chantiers, essentiels aujourd'hui à la recherche en domaine occitan, ont demandé, et demandent encore, l'investissement prolongé de nombreux moyens, humains, techniques et financiers.

Les récentes avancées dans certains domaines tels que la reconnaissance de caractères, la standardisation de techniques d'édition électroniques dédiées au texte ancien et, enfin, le développement de logiciels facilitant la mise en exploitation de corpus de toutes tailles nous permettent d'envisager d'autres types de projets. Nous voudrions mettre en avant différentes méthodes qui peuvent être mises en œuvre ponctuellement par une équipe restreinte, de deux à trois personnes.

*Maître de conférences en histoire de la langue, Université Paris-Sorbonne, EA STIH.

†École nationale des Chartes.

La pratique de la constitution d'un corpus électronique est pour nous un moyen supplémentaire d'interroger la place de l'occitan dans la littérature du xvi^e siècle. C'est pourquoi nous nous penchons notamment sur le cas des œuvres plurilingues, que l'occitan soit précédé ou suivi de pièces en d'autres langues, ou qu'il soit directement mis en regard avec le français. La mise en corpus est ainsi l'occasion non seulement de mieux outiller la description de l'occitan prémoderne, mais encore d'articuler cette description avec d'autres projets seiziémistes francophones.