

Philologie numérique de l'occitan médiéval, quelles perspectives?

Francesca Frontini

Paul Valéry Montpellier 3

CNRS, PRAXILING UMR 5267, France

Les humanités numériques et les technologies du langage offrent aujourd'hui des nouveaux outils pour l'étude des textes. En particulier, la philologie computationnelle est le domaine qui, au sein des HN, s'occupe de la création de ressources et infrastructures numériques pour l'étude des données textuelles et linguistiques dans les **langues historiques**¹ ; cette discipline, pourvue d'une forte composante technologique, est en plein développement et il n'est pas exagéré de dire qu'elle constitue un important moteur d'innovation scientifique.

Des nombreux projets visant à la construction de ressources et outils numériques ont été menés notamment pour les langues classiques (grec ancien et latin), mais aussi pour les langues pré-modernes, tels que par exemple l'ancien français, l'anglo-saxon, et bien évidemment l'ancien occitan, langue couvrant tous les usages sociolinguistiques à l'époque médiévale.

Ces efforts dans le domaine de la philologie computationnelle et des langues historiques sont le reflet du mouvement plus large pour la science ouverte, qui a investi les des sciences du texte. Les ressources et technologies doivent de plus en plus être *FAIR* (Findable = trouvables, Accessible = accessibles, Interoperable = Interopérables, Reusable = réutilisables, selon la définition de Wilkinson et al 2016).

Cela à son tour emmène à la mutualisation des ressources dans le cadre d'infrastructures de recherche, telles que les ERICS CLARIN² et DARIAH³ au niveau européen, et HUMA-NUM en France. De plus en plus les efforts des chercheurs portent non seulement à la numérisation des sources mais à la systématisation des connaissances par l'utilisation de formats et pratiques communes, ainsi qu'à la création de répertoires partagés de données et d'outils.

Le projet AcTo adhère à l'objectif de fond de ces infrastructures, qui vise à mettre à disposition des chercheurs et des étudiants un ensemble de ressources pour l'étude des textes et de la langue dans un "environnement numérisé virtuel" (ENV - VRE en anglais; Bellamy, 2012). En quoi est-ce que ces ENVs sont spécifiques pour les langues historiques? Pourquoi est-ce que les plateformes pour les langues modernes ne sont pas suffisantes? Il faut chercher la réponse à cette question au

¹ Les langues historiques ne sont pas uniquement les langues dont les parlants natifs ne sont plus disponibles (comme pour les langues classiques telles que Grec et Latin) mais aussi les variétés linguistiques où la compétence des parlants "natifs" n'est pas en soi suffisante.

² CLARIN, Infrastructure européenne pour les ressources linguistiques <<https://www.clarin.eu/>>

³ DARIAH, Infrastructure européenne pour les humanités numériques <<https://www.dariah.eu/>>

coeur de la philologie en tant que discipline, qui est au croisement entre études littéraires, historiques et linguistiques.

Dans mon poster je prends en compte les différentes dimensions de l'étude des sources anciennes - repérage, numérisation, édition, enrichissement, annotation linguistique et historique, analyse, reconstruction linguistique et historique - et je présenterais un ensemble de technologies et projets d'intérêt pour le développement d'un tel VRE pour l'occitan médiéval.

Bibliographie

- Bamman, David, and Gregory Crane. 2011. "The Ancient Greek and Latin Dependency Treebanks." In *Language Technology for Cultural Heritage*, edited by Caroline Sporleder, Antal van den Bosch, and Kalliopi Zervanou, 79–98. Theory and Applications of Natural Language Processing. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20227-8_5.
- Bellamy, Craig. 2012. "The Sound of Many Hands Clapping: Teaching the Digital Humanities through Virtual Research Environment (VREs)" 6 (2). <http://www.digitalhumanities.org/dhq/vol/6/2/000119/000119.html>.
- Caiti-Russo, Gilda Conférencier. 2015. "L'édition numérique du Petit Thalamus de Montpellier - Projet ANR / Gilda Caiti-Russo, professeure d'occitan médiéval, Université de Montpellier III Paul-Valéry." CIRDÒC-Mediatèca occitana. 2015. <http://www.occitanica.eu/omeka/items/show/11849>.
- Crane, Gregory. 2012. "The Perseus Project." In *Leadership in Science and Technology: A Reference Handbook*, 644–52. Thousand Oaks: SAGE Publications, Inc. <https://doi.org/10.4135/9781412994231>.
- Frontini, Francesca. 2018. "Reconstructing Urban Cartography Using Geo-Annotations." presented at the International Medieval Congress 2018 - Panel 1038 Collective Memory in Montpellier's Petit Thalamus: Digital Humanities, Language, and the Mapping of the Past, Leeds, UK, April 7. https://imc.leeds.ac.uk/dbsql02/AQueryServlet?*id=30&*formId=30&*context=IMC&chosenPaperId=NA&sessionId=7675&conference=2018&chosenPaperId=&*servletURI=https://imc.leeds.ac.uk/dbsql02/AQueryServlet.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (March): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Projets

Ancient Greek and Latin Treebank

<https://www.dh.uni-leipzig.de/wo/projects/ancient-greek-and-latin-dependency-treebank-2-0>

Mapping manuscripts migration

<http://mappingmanuscriptmigrations.org>

Perseus project

<http://perseus.tufts.edu>

Pelagios project

<http://commons.pelagios.org/>

REDEN

<http://reden.huma-num.fr/RedenOnline/site/input-tei.html>

Transcribus

<https://transkribus.eu/Transkribus/>