

Proposition pour la section XVI:

Martin-D. Glessgen / Sabine Tittel

Bases de données textuelles et lexicographie historique: comment réunir les *Plus anciens documents linguistiques de la France* et le DEAF?

Le projet présenté par la suite réunit deux entreprises lexicologiques et philologiques de longue haleine: le *Dictionnaire Etymologique de l'Ancien français* (DEAF), rédigé depuis quatre décennies à Heidelberg, et les *Plus anciens documents linguistiques de la France* (DocLing), poursuivies également depuis les années 1970. Dans les deux cas, les années 2000 ont donné lieu à une informatisation systématique (cf. Tittel 2010, Glessgen var.) qui permet un rapprochement réciproque.

En synthèse, notre idée est d'intégrer les données lexicales des *DocLing* dans les nouveaux 'articles courts' du DEAF. Cela permettra de renforcer la part des sources documentaires dans le DEAF, mais aussi de créer une interface entre ce dictionnaire et une base textuelle de référence. En même temps, le vocabulaire des *DocLing* serait placé dans le cadre naturel de l'ancienne langue française par une harmonisation des lemmes.

L'idée est donc d'une grande simplicité, mais sa réalisation suppose une harmonisation très précise entre la structure lexicologique des deux entreprises et la création d'interfaces informatiques performantes. Une intégration réussie ouvrirait ensuite la voie vers d'autres dictionnaires et corpus textuels, comme l'AND ou le *Nouveau Corpus d'Amsterdam* (NCA).

Nous présenterons en premier lieu les structures lexicologique et informatique des deux bases de données en question. D'un point de vue lexicologique, il s'agira de préciser les interrogations soulevées par la lemmatisation, la structuration sémantique et les éléments syntagmatiques. D'un point de vue informatique, les deux systèmes sont basés sur un enregistrement à l'aide de MySQL et une représentation sous une forme XML, mais dans un cas dans une optique lexicographique, dans l'autre dans une optique avant tout textuelle.

Nous préciserons par la suite les interfaces concrètes entre les deux bases de données (lemme, sigle-DEAF, forme lexématique, datation et localisation, source, contexte, lien avec l'édition électronique des actes) et les interfaces pour les utilisateurs. En conclusion, nous réfléchirons sur la plus-value de cette alliance pour les deux entreprises en question et pour les études sur l'ancienne langue française ainsi que sur son impact sur la méthodologie en linguistique historique.

Très concrètement, le projet commun permettra entre autres de pallier à la faible présence des textes documentaires dans les études sur l'ancienne langue; d'un point de vue méthodologique, il pourra rompre le cloisonnement qui régit encore aujourd'hui les grandes entreprises de la recherche.

Bibliographie

- AND = Rothwell, William / Stone, L. W. / Reid, T. B. W., with the assistance of D. Evans, S. Gregory, David A. Trotter, P. Staniforth, *Anglo-Norman dictionary*, London: The Mod. Humanities Research Assoc., 1977–1992. Version AND2, Rothwell, William / Gregory, S. / Trotter, David A. et al., fasc. A-C et D-E, London (Maney/MHRA) 2005; version électronique ANDEI: <<http://www.anglo-norman.net>>.
- DEAF = Möhren, Frankwalt / Städtler, Thomas / Dörr, Stephen / Tittel, Sabine / Kiwitt, Marc (fondé par Baldinger, Kurt), *Dictionnaire étymologique de l'ancien français*, Québec, Les Presses de l'Univ. Laval, Tübingen, Niemeyer, 1971-; version électronique DEAFél: <<http://www.deaf-page.de>>.
- DocLing = *Les plus anciens documents linguistiques de la France. Édition électronique*, Collection fondée par Jacques Monfrin, poursuivie par M.-D. Glessgen, en collaboration aujourd'hui avec Frédéric Duval et Paul Videsott, <www.mediaevistik.uzh.ch/docling>, 2009 [nouvelle édition, 2012].
- Glessgen, Martin-D., « Le long chemin d'une charte vers l'ère multimédiale », in: Schweickard, Wolfgang / Overbeck, Anja / Völker, Harald (éds.): *Lexikon, Varietät, Philologie, Romanistische Studien. Günter Holtus zum 65. Geburtstag*. Berlin, de Gruyter, 2011, 649-664.
- / Alletsgruber, Julia, « Structure de la base de données lexicologique des *Plus anciens documents* », in: BEC 168 (2010) [2011], 83-94
- / Kihai, Dumitru / Videsott, Paul, « L'élaboration philologique et linguistique des 'Plus anciens documents linguistiques de la France, Édition électronique' », in: *Bibliothèque de l'École des Chartes* 168 (2010) [2011], 5-94.
- NCA = Stein, Achim / Kunstmann, Pierre / Gleßgen, Martin-D. (edd.): *Nouveau Corpus d'Amsterdam. Corpus informatique de textes littéraires d'ancien français (ca 1150-1350)*, établi par Anthonij Dees (Amsterdam 1987), remanié par A. S., P. K. et M.-D. G., Stuttgart: Institut für Linguistik/Romanistik, version 2-2, 2010 (<<http://www.uni-stuttgart.de/lingrom/stein/corpus/#nca>>).
- Tittel, Sabine, « Le « DEAF électronique » – un avenir pour la lexicographie », in: *Revue de Linguistique Romane* 74 (2010) 301-311.