

Mise en ligne, mise à jour et mise en réseau
du *Französisches Etymologisches Wörterbuch*

Pascale Renders
Chargée de recherches FNRS
Université de Liège (Belgique)

Section 16 – Projets en cours ; ressources et outils nouveaux

La communication que nous proposons de présenter au XXVII^e Congrès International de linguistique et de philologie romanes poursuit deux objectifs. Le premier est de faire le point sur l'état du projet d'informatisation des 25 volumes du *Französisches Etymologisches Wörterbuch* (FEW). Le second est de présenter un projet fortement lié au précédent, qui vient d'obtenir un financement du Fonds National de la Recherche Scientifique en Belgique sous l'intitulé « Étude des modalités de mise à jour et mise en réseau du *Französisches Etymologisches Wörterbuch*, dans le cadre de son informatisation, avec d'autres ressources linguistiques notamment belgoromanes ».

La mise en ligne et la mise en réseau des ressources lexicographiques constituent un chantier actuellement très important en Europe (voir par exemple le CNRTL¹ à Nancy, le projet GiGANT² à Leiden ou le projet *Wörterbuchnetz* à Trèves). L'un des moyens mis en oeuvre consiste à « rétroconvertir » d'anciens ouvrages imprimés en y insérant, de façon semi-automatisée, des balises XML. Cette opération s'avère particulièrement délicate dans le cas d'ouvrages aux structures complexes, dont on veut rendre compte sans les dénaturer par une analyse a posteriori.

Parmi les ouvrages à informatiser de cette manière se trouve le FEW, œuvre incontournable dans le domaine de la linguistique française et romane et en même temps reconnue comme difficilement exploitable. Il est depuis longtemps acquis que l'informatisation du FEW, à condition de respecter les structures et la complexité de l'oeuvre, permettra de résoudre les problèmes d'accessibilité et ouvrira de nouveaux itinéraires de consultation. Par ailleurs, le FEW constitue un « dizionario-tetto » (Buchi/Renders à paraître) qui serait bénéfiquement mis en réseau avec des dictionnaires en ligne tels que le *Dictionnaire étymologique de l'ancien français* (DEAF) ou l'*Anglo-Norman Dictionary* (AND).

La faisabilité d'une informatisation du FEW est désormais prouvée (Renders 2011). Un logiciel dit « de rétroconversion » a été conçu, qui est capable d'analyser et de structurer tout article du FEW (codé en Unicode et muni d'un balisage typographique minimal) au moyen d'un balisage XML inséré de façon totalement automatisée. Le modèle de balisage a été défini de façon à répondre aux nombreuses attentes de la communauté scientifique. Les algorithmes qui forment le noyau du logiciel minimisent le nombre d'erreurs et les rendent facilement repérables.

Dès lors, une question se pose : pourquoi le FEW n'est-il pas encore sur Internet ? La réponse est simple. La saisie (préalable à la rétroconversion) des milliers d'articles du FEW requiert un investissement d'environ 70000 euros, non disponibles actuellement. La solution adoptée consiste donc à acquérir et à informatiser le texte de l'ouvrage en plusieurs tranches, par groupes d'articles. Cette solution est possible parce que le modèle de balisage du FEW – et, partant, le logiciel qui l'applique – a heureusement été conçu de façon à ce que le traitement de chaque article soit indépendant du reste de l'ouvrage. Nous expliquerons dans notre communication les démarches entreprises et les résultats concrets auxquels nous sommes parvenus.

En attendant l'acquisition et l'informatisation du texte complet du FEW, un nouveau projet de recherche a vu le jour, qui se penche sur deux questions : la mise à jour du FEW et sa mise en réseau avec d'autres ressources lexicographiques pour lesquelles il constitue une référence obligée. Ces deux problématiques sont liées, puisque la mise en réseau doit rendre disponible une partie des ajouts et corrections actuellement consignés dans des ouvrages annexes (cf. Greub à paraître) ou

1 Pierrel/Buchi 2009; <http://www.cnrtl.fr/>

2 <http://www.inl.nl/onderzoek-a-onderwijs/projecten/gigant>

dans des ressources en ligne (TLF-Étym, DEAF ou AND, pour ne citer qu'eux). Les possibilités d'informatisation de ressources belgoromanes corrigeant le FEW, telles que l'*Atlas Linguistique de la Wallonie* (ALW), font partie de l'étude. Le XXVII^e CILPR sera l'occasion de présenter le projet et les premiers résultats. Dans le corpus d'articles testés à cette fin se trouvent notamment les articles du volume 19 (*Orientalia*), qui présentent l'intérêt d'avoir fait l'objet d'actualisations et de corrections systématiques (Arveiller 1999) et de pouvoir être mis en relation avec des notices du projet TLF-Étym consacrées aux arabismes du français (Baiwir 2010). Au-delà des possibilités et des choix informatiques, il s'agit d'étudier la façon dont la mise en relation du FEW avec des ressources en ligne susceptibles elles-même d'évoluer peut influencer les modalités de mise à jour du FEW lui-même.

Le FEW n'est donc pas encore disponible en version électronique, mais la mise en ligne effective d'une partie de l'ouvrage est néanmoins prévue à court terme, en attendant son informatisation complète. Elle permettra de fournir à la communauté scientifique un exemple, critiquable, de ce que peut être le FEW informatisé, mis à jour et mis en réseau.

Bibliographie

- ALW = REMACLE Louis/LEGROS Élisée *et al.*, *Atlas Linguistique de la Wallonie. Tableau géographique des parlers de la Belgique romane d'après l'enquête de Jean Haust et des enquêtes complémentaires* (10 volumes), Liège, 1953–.
- AND = ROTHWELL, W./GREGORY, S./TROTTER, D. A. (dir.), *Anglo-Norman Dictionary*, Londres, 2005–² [1977–1992¹]. <http://www.anglo-norman.net>
- ARVEILLER Raymond, *Addenda au FEW XIX (Orientalia)*, Tübingen, Niemeyer, 1999.
- BAIWIR Esther, Refonte des notices 'étymologie et histoire' des articles suivants du Trésor de la Langue Française informatisé (TLFi) : *achour* (2010), *alezan* (2010), *antari* (2010), *coufique* (2010), *fez* (2010), *ketmie* (2010), *amin* (2010), *sumac* (2010). <http://www.atilf.fr/tlf-etym>.
- BUCHI Eva/RENDERS Pascale, « 46. Gallo-romance I : Historical and etymological lexicography », in *Dictionaries. An International Encyclopedia of Lexicography. Supplementary volume : Recent developments with special focus on computational lexicography*, GROUWS Rufus *et al.* (éd.), De Gruyter, Berlin/New York (à paraître).
- DEAF = BALDINGER Kurt *et al.*, *Dictionnaire étymologique de l'ancien français*, Québec/Tübingen/Paris, 1971–. <http://www.deaf-page.de>
- FEW = Walther VON WARTBURG *et al.*, *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes* (25 vol.), Bonn/Heidelberg/Leipzig-Berlin/Bâle, Klopp/Winter/Teubner/Zbinden, 1922–2002.
- GREUB Yan, « Sur le FEW », *The Aberystwyth Colloquium. Present and future research in Anglo-Norman (21-22 juillet 2011)*, Aberystwyth, Prifysgol Aberystwyth University, à paraître.
- PIERREL Jean-Marie/BUCHI Éva, « Research and Resource Enhancement in French Lexicography : the ATILF Laboratory's Computerised Resources », in Silvia BRUTI, Roberta CELLA, Marina FOSCHI ALBERT (edd.), *Perspectives on Lexicography in Italy and Europe*, Newcastle upon Tyne, Cambridge Scholars Publishing, 79-117.
- RENDERS Pascale, *Modélisation d'un discours étymologique. Prolégomènes à l'informatisation du Französisches Etymologisches Wörterbuch*. Liège, Université de Liège, 2011 (thèse de doctorat).
- TLF-Étym = Projet TLF-Étym: mise à jour des notices étymologiques du Trésor de la langue française informatisé. Dossier de présentation, Collectif, Nancy, ATILF/CNRS/Université Nancy 2/UHP, 2005. <http://www.atilf.fr/tlf-etym> ; Voir aussi BUCHI Éva, « Le projet TLF-Étym (projet de révision sélective des notices étymologiques du Trésor de la langue française informatisé) », *Estudis romànics* 27, 2005, 569-571.