

Pascale Renders, *L'informatisation du Französisches Etymologisches Wörterbuch. Modélisation d'un discours étymologique*, Strasbourg, ELiPhi – Éditions de linguistique et de philologie, coll. « Travaux de linguistique romane – linguistique de corpus et philologie informatique », 2015, XIV + 326 p., ISBN 978-2-37276-003-4

Issue d'une thèse de doctorat soutenue en juin 2011 et réalisée en collaboration avec le laboratoire ATILF (CNRS), l'étude porte sur l'informatisation du Französisches Etymologisches Wörterbuch (25 volumes parus, entre 1922 et 2002, désormais abrégé FEW), outil majeur de la linguistique historique du français, dont la version imprimée présente cependant d'importantes difficultés de consultation, des incohérences structurelles et une grande complexité typographique. L'auteure se propose d'interroger la faisabilité et de dégager les enjeux d'une telle entreprise d'informatisation (ou « rétroconversion »), en proposant une modélisation du discours étymologique du FEW envisagé du point de vue de l'utilisateur.

Le projet est très concrètement interdisciplinaire, puisqu'il met en œuvre des compétences lexicographiques (en particulier : de lexicographie historique), lexicologiques et informatiques très poussées. Ces volets disciplinaires articulent également chacun des aspects théoriques et pratiques, qui convergent vers l'élaboration d'un logiciel complet et fonctionnel de rétroconversion du FEW.

L'auteure se livre d'abord à un travail d'objectivation rigoureuse de ce en quoi consiste exactement le FEW : sa richesse, mais surtout ses limites. En cela, elle offre un éclairage précieux sur la manière, éminemment tortueuse, dont s'est construit cet instrument unique pour l'étymologie française et romane, et dont ont été progressivement pensées sa rédaction puis son informatisation.

Cela dit, la grande originalité de la démarche de l'auteure, qui la distingue des travaux de lexicologie traditionnels, consiste surtout à vouloir prendre en considération trois types de contraintes de natures différentes : la structure de la version imprimée du FEW, les possibilités de traitement automatisé de ce corpus, et les besoins spécifiques de sa consultation. La question de l'accessibilité n'est donc pas seulement envisagée du point de vue du seul format éditorial (électronique *vs* papier), mais aussi plus largement du point de vue de la conception profonde du discours lexicographique, de sa compatibilité avec une interface informatisée, et de la variété des pratiques de consultation dont il peut faire l'objet. Pour cela, l'auteure s'est appuyée sur les réponses à un questionnaire soumis à des membres de la communauté internationale des chercheurs en linguistique française et romane. Elle livre ainsi une petite anthropologie des pratiques savantes de consultation et de lecture, ainsi qu'un éclairage très fin sur les besoins d'une communauté scientifique, sans antécédent à notre connaissance pour le champ de la lexicographie historique. Elle met en évidence une distinction et une complémentarité importantes entre la « dimension monographique » et la « dimension thesaurus » du FEW.

En prenant au sérieux la structure de la version imprimée, ainsi que les besoins scientifiques effectifs, l'auteure évite ainsi de présenter l'informatisation comme une solution magique qui résoudrait tous les problèmes actuels de consultation et multiplierait les modes de requête ; elle l'envisage plutôt comme une véritable opération de traduction susceptible de modifier le matériau de départ, et de biaiser

son interprétation.

L'auteure retrace ainsi très méthodiquement toutes les étapes de son entreprise d'informatisation, en accordant une large place à la modélisation des composantes des articles du FEW. Chaque question est affrontée en remontant à ses principes fondamentaux, de sorte que l'intérêt des discussions menées (par exemple, sur le traitement automatique de l'implicite, sur le choix du formalisme XML, sur les algorithmes de balisage sémantique) débordent largement le seul cadre du corpus envisagé par l'auteure, et touchent à toute entreprise d'informatisation d'un corpus textuel complexe, et à toute tentative d'automatisation de sa consultation. De même, les étapes de la rédaction de la quarantaine d'algorithmes sont détaillées méticuleusement, et, pour chaque algorithme, chacune des difficultés rencontrées est exposée, avec sa résolution, jusqu'au constat final de la faisabilité de la rétroconversion dans son ensemble. Ce constat est concrètement vérifié par la rétroconversion effective d'un corpus représentatif du FEW, qui donne lieu à une discussion critique des résultats produits par le travail des algorithmes.

Les perspectives ouvertes sont nombreuses, tant pour la lexicologie historique du français que pour les entreprises d'informatisation de corpus textuels. En effet, l'ouvrage fournit d'abord une base formelle qui pourront nourrir les apports successifs au FEW. Du même coup, il autorise la mise en relation du FEW avec d'autres bases documentaires informatisées. Ensuite, l'ouvrage fournit un modèle de protocole de rétroconversion appropriable par la communauté des linguistes, mais aussi transposable à d'autres initiatives du même genre. Enfin, l'ouvrage fournit les bases de travail pour l'élaboration d'un moteur et d'une interface de recherche performants sur ce corpus.

Il convient cependant de signaler que le volume se présente sous une forme assez aride, qui privilégie la grande systématisme, la haute technicité, et le souci d'exhaustivité, au détriment de la fluidité et de l'accessibilité aux non-spécialistes. C'est là une réserve tout à fait mineure, qui ne remet absolument pas en cause les grandes qualités de l'ouvrage. Porté par une très haute ambition interdisciplinaire, il adopte une démarche intellectuelle d'une grande originalité (la modélisation du discours étymologique, et la prise en compte des pratiques de consultation), qui débouche sur des apports scientifiques très substantiels (tant pour la compréhension du FEW que pour les modalités de son informatisation), et ouvre des perspectives de grande actualité pour la lexicologie historique romane, et pour la philologie numérique en général.

François Provenzano
Université de Liège