

### *Projets de recherche*

– Entre 2004 à 2015, mes travaux de recherche constitués principalement par ma thèse de doctorat et ma thèse d'habilitation ont porté sur la première époque des langues romanes soit l'époque comprise entre le IX<sup>e</sup> et le XII<sup>e</sup> s. Ma thèse de doctorat a examiné l'émergence de l'ancien occitan à l'écrit à niveau régional et a permis de mettre au jour une élaboration graphématique complexe et insoupçonnée de la future langue des troubadours. Le rapport entre graphie et phonie a été l'une des préoccupations centrales de ce travail de même que les implications morphologiques de l'insertion du vernaculaire en contexte latin. Après avoir été retravaillé, cet ouvrage a été publié en 2011 (*Bibliothèque de Linguistique Romane*, 592 p).

– Ma réflexion sur l'émergence des langues (gallo)romanes à l'écrit m'a amenée, en 2010, à consacrer un article monographique aux choix lexicaux de Chrétien de Troyes. En prenant appui sur les premières attestations de mots du français, je me suis posé la question en quelle mesure l'œuvre du trouvère champenois semble refléter des usages linguistiques préexistants et en quelle mesure elle a contribué à l'innovation lexicale (article publié dans la revue *Romania* en 2013, 56 p.), sachant que le deuxième cas est beaucoup plus rare de ce que l'on a pu supposer jusqu'ici.

– À partir de 2010, j'ai exploité les résultats obtenus pour la grapho-phonétique de la période pré-textuelle pour étudier sur une base empirique les aspects lexicaux de la genèse des langues gallo-romanes, en élargissant ainsi très largement les espaces linguistiques à l'étude. J'ai pu prendre appui – grâce à l'achèvement concomitant du projet d'édition nancéien sur l'ensemble des documents originaux conservés en France antérieurs à 1120, corpus informatisé qui englobe de fait les domaines linguistiques français, occitan et francoprovençal. Je me suis intéressée pour cette recherche au lexique. La lecture des quelque 5000 actes latins originaux m'a permis de relever 500 lexèmes français, occitans et francoprovençaux philologiquement assurés, lexèmes que j'ai ensuite traités très en détail dans un *Trésor galloroman des origines*, complémentaire au *Französisches Etymologisches Wörterbuch* de Walther von Wartburg. Le relevé initial, de type manuel, a été élargi grâce aux fonctions de recherche dans les corpus électroniques des actes originaux latins et des *ca* 40 000 actes des *Chartae Galliae*. L'analyse interprétative de la trajectoire étymologique et géolinguistique de ces 500 lexèmes m'a permis de mieux cerner la nature complexe de la fragmentation lexicale de la Galloromania. Cette recherche de longue haleine, déposée comme thèse d'habilitation à l'Université de Zurich en octobre 2015 et suivie d'un oral devant la Faculté des Lettres en mai 2016 a depuis fait l'objet d'un important travail de réélaboration sur la base des rapports de la commission avant d'être publiée en novembre 2017 (*Travaux de Linguistique Romane*, dir. M. Aprile, Th. Städtler, A. Thibault, 768 pages).

– Après mon arrivée à l'Université de Neuchâtel en 2015, j'ai soumis avec ma collègue sur place Federica Diémoz un projet de recherche au *Fonds National Suisse* (FNS) qui prévoit une collaboration avec le centre du FEW à l'ATILF/Nancy. Ce projet (n°101212-162851), qui a pour objet « la scriptologie lexicale du domaine oïlique oriental », a été accepté et financé par le FNS au niveau demandé de 371 514.- CHF pour trois ans (prolongeable une 4<sup>e</sup> année). Le projet, qui a démarré en juillet 2016 et que je dirige seule depuis 2017, emploie actuellement deux jeunes chercheurs (contrats partiels de 70% chacun). Le noyau du projet consiste dans la rédaction d'un *Dictionnaire des régionalismes du français médiéval* (DRFM), dictionnaire qui est le point de départ pour des analyses et réflexions approfondies sur la régionalité lexicale au Moyen Âge, jusqu'alors sous-exploitée. L'identification d'une première nomenclature a pris appui sur le corpus électronique des *Documents linguistiques galloromans*, entièrement lemmatisé, et sera élargie grâce à un corpus de textes littéraires. Nous sommes en train d'étudier une voie d'exploitation de nos résultats dans le cadre de la lexicographie électronique qui engloberait, au moins par hyperliens, le FEW et le DMF et qui se baserait sur une version élargie du DEAF électronique. Nous poursuivons ainsi l'idée de contribuer à un réseau interactif de la lexicographie électronique médiéviste.

– En 2016, je suis devenue partenaire du projet des *Documents linguistiques galloromans* – projet centenaire de l'*Ecole des Chartes* coordonné depuis une quinzaine d'années par Martin Glessgen, à l'Université de Strasbourg d'abord (MISHA), à l'Université de Zurich ensuite. Grâce à l'obtention de plusieurs financements, j'ai pu intégrer une très grande partie des documents vernaculaires provenant du domaine francoprovençal et disponibles dans une édition philologiquement fiable. Le choix de quelque 1 000 documents des XIII<sup>e</sup>/XV<sup>e</sup> siècles (ca 700 000 mots) provenant de la région de Lyon, du Forez et de la Suisse romande, leur saisie sous format XML et leur intégration à cette base de données informatique a permis un élargissement considérable du corpus. C'est en prenant appui sur cette nouvelle ressource, et dans le prolongement du premier projet FNS, que mes recherches se sont davantage orientées vers la linguistique informatique. J'ai ainsi soumis en 2016 un deuxième projet auprès du FNS avec l'objectif d'étudier dans ce cadre « L'élaboration scripturale du francoprovençal au Moyen Âge » aux plans lexical, morphologique et grapho-phonétique. Cette nouvelle demande (n°100012-172826) a été également acceptée et financée au niveau demandé de 405 338.- CHF sur trois ans (prolongeable une 4<sup>e</sup> année). Ces deux projets FNS, très complémentaires, visent à décrire de manière philologiquement assurée et dans le cadre de la linguistique de corpus, l'élaboration linguistique des domaines oïlique oriental et francoprovençal qui se placent géographiquement et sociolinguistiquement dans une interface complexe qui offre de riches perspectives autant empiriques qu'interprétatives. Le financement obtenu a permis d'employer trois collaborateurs scientifiques (une doctorante, un post-doc et un informaticien) dont les contrats ont démarré le 1<sup>er</sup> février 2018.

Le volet lexicologique du projet entend développer ultérieurement la part lexicographique de mes recherches mais en se plaçant ici dans le cadre d'une réalisation électronique. Avec l'informaticien du projet, Pierre Menetrey, nous entreprendrons très concrètement un dédoublement du système de rédaction du *Dictionnaire de l'Ancien Gascon électronique* (DAGél, Heidelberg). Ce logiciel sur base MySQL, qui adapte lui-même celui du DEAF électronique, est idéal parce qu'il permet un retour constant aux textes sur lesquels il s'appuie, ce qui reste encore l'exception dans la lexicographie électronique romaniste. Il présente également l'avantage de s'inscrire dans une méthodologie informatique de pointe et commune à plusieurs projets européens.

C'est toutefois le volet scriptologique qui nous occupe le plus au plan de la programmation informatique. Toujours en dialogue avec l'informaticien du projet, nous réfléchissons à la manière la plus adéquate de développer la base de données grapho-phonétique et morphologique existante en la dotant d'outils de représentation dans l'espace de la variation à plusieurs niveaux: celle observée entre les domaines d'oïl et francoprovençal mais aussi à l'intérieur de chacun de ces domaines linguistiques. Nous venons d'achever une première étude de faisabilité dans le cadre d'un article monographique sur l'élaboration du francoprovençal médiéval (60 p., à paraître dans la *Zeitschrift für romanische Philologie*); la gestion de quelque 60 000 occurrences pour le paramètre à l'étude (les issues de lat. 'A [, précédé ou non de palatale) et la réflexion sur les différentes fonctions d'exportation et de relevés statistiques nous ont notamment servi pour mieux appréhender l'architecture globale des programmations nécessaires. Celle-ci prévoit différents outils d'exploitation statistique des interrogations du corpus, dans l'idée de mettre à la disposition de la communauté scientifique une base de données évolutive et commode à interroger, idéalement complémentaire avec les bases de données de référence déjà utilisées telles que la *Base du Français Médiéval* (BFM), le *Syntactic Reference Corpus of Medieval French* (SRCMF), le *Nouveau Corpus d'Amsterdam* (NCA) le Corpus MCVF ou encore le corpus SERMO.

– Dans un autre ordre d'idées, je me pose actuellement la question de savoir en quelle mesure les sources latines des IX<sup>e</sup>/XIII<sup>e</sup> siècles peuvent être exploitées pour la lexicologie romaniste dans les domaines ibéro- et, surtout, italomans. J'ai entrepris avec Seraina Montigel, assistante en linguistique à Zurich, une comparaison entre les lexèmes pré-textuels que j'ai pu identifier dans le cadre du *Trésor galloroman des origines* et leurs correspondants étymologiques dans les plus anciennes strates de l'italien et, également, de l'espagnol et du portugais. J'exploite en cela la documentation en latin médiéval, en ancien italien (et espagnol) et des dialectes modernes, notamment en réunissant les

informations riches et fiables, mais disparates du *Lessico Etimologico Italiano*. En collaboration avec Antonio Montinaro (Univ. del Molise), j'ai également commencé à appréhender les éléments vernaculaires dans les textes latins de l'Italie méridionale, notamment des Pouilles, et je suis en train d'évaluer les résultats des travaux dirigés par Rosanna Sornicola à Naples à partir des textes de la Campanie (cf. récemment *Sistemi, norme scritture. La lingua delle più antiche carte cavensi*, 2017).

– Parallèlement à ces entreprises médiévistes, je poursuis depuis 2015 une interrogation de nature sociolinguistique concernant l'époque révolutionnaire. Le projet prévoit l'édition critique des 37 réponses intégrales à l'enquête de l'abbé Grégoire à partir des manuscrits de la bibliothèque de Port-Royal et de la BN (ce dernier étant en grande partie inédit). Le volume entend proposer sur cette base nouvelle une interprétation systématique de la situation plurilingue de la France de la fin du XVIII<sup>e</sup> siècle en accordant ainsi toute l'attention due à ce qui fut la toute première enquête sociolinguistique au monde.

– Enfin, un manuel de consultation du FEW, écrit en collaboration avec Marguerite Dallas (assistante à l'Université de Zurich), Martin Glessgen (Zurich) et André Thibault (Sorbonne, Paris IV) est à paraître. Cet ouvrage à visée pédagogique, qui émane en grande partie de mon enseignement lexicologique, analyse et décrit les structures de ce dictionnaire complexe, dans le souci de rendre son utilisation plus accessible aux étudiants et chercheurs notamment francophones et, plus généralement, de fournir un outil d'apprentissage en lexicologie diachronique. Il est en cours de relecture finale et sera publié d'ici l'automne 2018.