Restaure, pour la diffusion des langues régionales

De gauche à droite une partie de l'équipe Restaure : Amalia Todirascu, Delphine Bernhard, Lucie Steiblé et Pascale Erhart.

L'occitan, le picard et l'alsacien. Quel point commun entre ces trois dialectes ? Pas leur origine. Respectivement langue d'Oc, langue d'Oïl - toutes deux romanes – et langue germanique, elles illustrent la diversité des langues régionales françaises. Parce que la constitution de lexiques ou dictionnaires informatisés est peu rentable, les locuteurs disposent de peu de ressources pour les accompagner au passage à l'écrit. C'est cette volonté de documenter la richesse des dialectes qui est à l'origine du projet Restaure.

Qui l'eut cru ? L'alsacien, proche de la langue de Molière

L'outil de reconnaissance optique de caractères développé par l'équipe du Lilpa pour numériser les textes en langue alsacienne a révélé qu'il existe autant de similitudes pour ce qui est des caractères utilisés à l'écrit entre l'alsacien et le français qu'entre l'alsacien et l'allemand.

Loin s'en faut, ce résultat ne remet pas en cause les racines germaniques du dialecte. Il s'explique notamment par le fait que les locuteurs sont de langue maternelle française et utilisent un clavier français pour rédiger. Bien expert, celui qui saura taper un eszett – ß !

Financé par l'Agence nationale de la recherche, le projet Restaure (Ressources informatisées et traitement automatique pour les langues régionales) réunit quatre équipes de recherche de Toulouse (occitan), d'Amiens (picard), d'Orsay (traitement automatique des langues) et de Strasbourg (alsacien). Le groupe strasbourgeois s'appuie sur les compétences croisées de quatre linguistes et deux spécialistes de l'automatisation de la langue*.

Des défis pour le traitement automatisé de la langue, les dialectes en présentent beaucoup. *« Les langues régionales ne sont pas normées, comme peuvent l'être le français ou l'anglais. Il n'y a pas d'équivalent alsacien de l'Académie française »*, explique Pascale Erhart, linguiste, membre de l'équipe du projet Restaure. Conséquence : il existe presque autant de variations à l'écrit que de rédacteurs. Cette diversité représente un casse-tête dès lors que l'on souhaite effectuer une recherche automatique de mot. Une seule lettre modifiée suffit pour qu'un mot ne soit plus reconnu par un programme informatique.

Au cœur de la réflexion des chercheurs : le passage *de* l'oralité *à* l'écriture

Au cœur de la réflexion des chercheurs : le passage de l'oralité à l'écriture, processus qui soulève de nombreux questionnements sur la structure même de la langue. Comment définit-on un mot ? Où se place l'apostrophe ? Quelle orthographe est la plus efficace ? Toutefois, le travail de l'équipe du Lilpa (Linguistique, langues, parole) ne se veut pas prescripteur de normes d'écriture. Et c'est également la volonté de l'équipe amiénoise spécialiste du picard. *« Ce parti pris, élégant, permet à la diversité d'une langue de s'exprimer. Car finalement, l'enjeu pour un locuteur est simple et unique, il souhaite être compris »*, conclut la linguiste Lucie Steiblé.

■ Fanny Cygan

*Respectivement Pascale Erhart, Lucie Steiblé, Christophe Gérard, Dominique Huck, Delphine Bernhard et Amalia Todirascu.