



Le Croissant : cartographie linguistique et traitement automatique du langage

Philippe Boula de Mareüil Directeur de recherche
au CNRS (LIMSI – UPR 3251)

Le Croissant, entre oïl et oc, est un domaine fascinant pour le linguiste qui (comme moi) y a enregistré une douzaine de locuteurs sur le terrain : fascinant parce qu'il représente une zone de turbulence linguistique, un espace où, à quelques kilomètres de distance, d'un village à l'autre, on observe d'importantes fluctuations. L'hybridation des formes que l'on constate permet de toucher du doigt, aujourd'hui, le fait qu'en France, jusqu'à l'expansion du français comme langue de grande diffusion, et dans tant de régions du monde, la variation devait être la règle générale dans le langage. On serait tenté de dire que le Croissant peut renouveler notre vision du langage.

Nous avons ainsi prévu un zoom sur le Croissant, dans l'Atlas sonore des langues régionales de France que nous constituons dans notre laboratoire, le LIMSI. Cet atlas prend la forme d'un site web (<https://atlas.limsi.fr>) présentant une carte de France interactive, avec plus de 200 points d'enquête sur lesquels on peut cliquer pour entendre (et lire) autant de versions, en langues régionales, d'une même histoire, la fable d'Ésope « La bise et le soleil », utilisée depuis un siècle par les phonéticiens pour décrire nombre de langues du monde. Le zoom permet de faire apparaître des communes qui sinon, à l'échelle de la France, seraient trop resserrées. À ce jour, une dizaine d'enregistrements du Croissant ont été cartographiés, des confins du Poitou méridional au Bourbonnais, en passant par la Marche. Ils ont pour la plupart été transcrits dans une graphie de type occitan, parfois dans une graphie francisée : objet de recherche en soi¹, la transcription orthographique des dialectes et langues minoritaires soulève d'autant plus de problèmes théoriques et pratiques pour le Croissant, à l'image de ce domaine de transition entre oïl et oc.

La conjugaison des verbes, dans des variétés romanes comme les parlers du Croissant, représente un autre enjeu linguistique. Sachant qu'il existe environ 40 types distincts de flexions verbales pour un parler donné, à multiplier par une soixantaine de formes pour les différents temps, modes et personnes, le traitement automatique du langage peut faciliter et accélérer l'analyse des données recueillies sur le terrain. Le LIMSI va ainsi développer des outils pour extraire les paradigmes verbaux directement à partir des enregistrements collectés, dans le cadre d'un projet de recherche centré sur le Croissant² qui sera l'occasion aussi d'enrichir l'atlas sonore.

¹ P. Boula de Mareüil, R. Sichel-Bazin, N. Quint, G. Adda (2017), « Norme et variation à l'âge des corpus informatisés pour les langues régionales de France », in C. Feuillard (éd.), *Usage, norme et codification : de la diversité des situations à l'utilisation du numérique*, EME Éditions, Bruxelles, pp. 217–222.

² Il s'agit du projet ANR '*Les parlers du Croissant*', décrit plus loin en détail dans ce numéro de Langues et Cité.