

Tobias Scheer
Guylaine Brun-Trigaud
Université Côte d'Azur, CNRS 7320

L'atlas Dees électronique

La communication présente l'atlas Dees électronique 2022 (ADE22) récemment mis en ligne (<http://atlasdees.unice.fr/wordpress>). Elle aborde la constitution du corpus qui l'alimente (2,2 Mio de mots, ramenés à 98.230 formes uniques) ainsi que, à l'aide d'un exemple choisi, son fonctionnement et le type de questions auxquelles il peut répondre.

L'atlas Dees 1987 (AD87) est basé sur 200 textes littéraires datés des 13^e et 14^e siècles et son originalité réside dans le fait qu'A. Dees les a localisés dans l'espace à l'aide de 268 critères linguistiques dégagés par l'étude de chartes (dans son atlas des chartes de 1980). Il a ainsi attribué les textes à des segments d'un maillage géographique de 87 zones, ce qui lui a permis ensuite de produire des cartes montrant la distribution diatopique pour des ensembles de données choisis.

Le propos de l'ADE22 est de rendre vivant et librement accessible l'AD87, i.e. de reconstruire, autant que faire se peut, l'instrument que l'équipe Dees avait mis au point dans les années 80, qui a permis la confection des deux atlas mais n'a jamais été rendu public ni même décrit, et qui est aujourd'hui perdu. Dees et son équipe ont choisi 517 cartes publiées dans l'AD87 selon ce qu'ils jugeaient important. L'AD87 était ainsi figé à tout jamais en fonction des choix et des intérêts de ses concepteurs. Le propos de l'ADE22 est de rendre vivant le travail de Dees et de son équipe, de façon à qu'il soit accessible à tous et permette l'affichage diatopique de données librement choisies. A cette fin, l'utilisateur peut formuler ses propres requêtes dans le corpus des 200 textes, puis afficher la distribution diatopique des résultats.

Le moteur de recherche exploitant le corpus a été fait sur mesure pour le projet. Il permet des recherches par forme et lemme afr., ainsi que par lemme latin. L'input à la requête sont ou bien des chaînes de caractères fixes, ou bien des expressions régulières. La communication explique les choix qui ont été faits lors de la constitution du corpus et de l'alignement progressif du matériau de base fourni par les fichiers informatique Dees d'origine dont nous avons disposé grâce à Piet van Reenen et du NCA : la lemmatisation a été opérée à l'aide du lexique Frolex 3.0, mais demeure largement imparfaite (Frolex ne connaît que 58% des formes uniques). Sur cette base des lemmes reconnus, Frolex permet l'alignement avec les formes latines du FEW (qui ne notent pas la quantité vocalique), qui ensuite ont été à leur tour alignées avec celles du Gaffiot électronique (qui fournit la quantité). Le travail sur l'alignement des formes latines n'est pas achevé : la recherche par forme latine fonctionne mais actuellement présente des défauts majeurs.

Enfin, comme l'AD87, l'ADE22 est un atlas lexical : les parties du discours (avec une grille très fine) sont identifiées par un code à trois chiffres que Dees a associé à chaque mot, mais pour l'instant on ne peut faire des requêtes dans ce code. Par ailleurs, le corpus n'est pas annoté pour d'autres propriétés, notamment morpho-syntaxiques. Toutes ces questions sont autant de pistes pour l'avenir : meilleure lemmatisation, recherche par code Dees, alignement avec le latin, annotation morpho-syntaxique.

Afin de montrer à quel type de questions l'ADE22 peut répondre, une idée de G. Straka est mise à l'épreuve. Lat. -C(i)ca produit une variation concernant le voisement du réflexe de la vélaire : pour °grānica par exemple on trouve *grange* autant que *granche*, et cette variation est systématique en afr. L'ayant examinée dans les toponymes, les articles du FEW et une carte de l'ALF (*gallica*), Straka (1979 [1970] : 359) conclut que la syncope s'est répandue du Nord-Est vers le Sud (importée par les Francs et suivant leur sillon). Le voisement intervocalique en revanche aurait pris le chemin inverse : venant du Sud, il a gagné du terrain en allant vers le

Nord. Ainsi au Sud et à l'Ouest la voisée $dʒ$ a été doublement favorisée : le k d'origine a été atteint par le voisement plus tôt qu'ailleurs, et il a dû patienter plus longtemps avant d'être figé par la syncope. Il en va de même, à l'inverse, pour f dans le Nord-Est : ici la syncope s'est appliquée de bonne heure, et le voisement s'est fait attendre. La base empirique de cette analyse était limitée par les moyens de l'époque, et l'ADE22 est de nature à confirmer ou réfuter la distribution diatopique conçue par Straka.