

Intervention de A. Guerreau

Alain Guerreau est historien médiéviste, directeur de recherche CNRS au Centre de Recherches Historiques.

Il travaille sur un corpus de textes historiques. Les corpus historique se distinguent des autres corpus. Ils sont fixes : on ne peut pas en ajouter. Ils sont relativement courts et brefs, même complet, ce qui pose des problèmes par rapport aux statistique. Tous ces textes enfin ont été rédigés par des gens morts depuis une éternité : la loi et les traités internationaux sont comme ils sont : tout ce qui a été produit par les auteurs morts depuis plus de soixante-dix ans (avant 1940) est du domaine public, impossible à rebasculer dans le privé. Le philologue n'est pas propriétaire du texte, le texte est dans le domaine public. Donc pour résumer : les corpus historiques sont des corpus fixes, brefs, dans le domaine public.

Des problèmes techniques se posent pour textes relativement anciens ou même pour textes plus récents : on peut ne pas connaître la date de rédaction ou on la connaît mal, ce qui rend difficile toute analyse d'une évolution. On ne connaît pas non plus le lieu : il y a un latin dans chaque pays ! Enfin, beaucoup de textes non attribués ou mal attribués. En résumé : pas de date, pas de lieu, pas d'auteur !

Corpus disponibles en ligne et que signale A. Guerreau :

- Les Chartes sont un document émis par une institution souvent ecclésiastique entre le XXe et le XIIIe siècle. On peut en trouver beaucoup sur internet. Il y a une possibilité de se faire un corpus sur internet, mais on ne sait pas bien comment distinguer les textes les uns des autres. On pourrait espérer une géographie linguistique à partir de ce corpus : mais les premiers travaux à ce sujet sont encore extrêmement récents.
- Les *vitae sanctorum*, les vies de saints (bibliothèque hagiographique latine (IV^e-XVIII^e siècles) représentent un énorme corpus mais n'intéressent personne.
- Les fichiers de l'IGN sont désormais domaine public pour tous les chercheurs (micro-toponymie).

Les mots changent de sens, mais pas individuellement : on ne peut pas étudier les mots un par un, contrairement à ce qu'on a fait au XIX^e siècle. L'histoire de chaque mot est intéressante, mais elle n'ouvre pas vraiment : le vocabulaire n'est pas un tas de mots, c'est une structure, le plus intéressant, c'est la relation entre les mots. Il est rigoureusement impossible selon A. Guerreau de traduire une charte, car on ne désigne pas les mêmes choses. Pour la traduction de *campus*, « le champ » peut-il désigné à la fois le terrain cultivé avec une herse et le terrain cultivé avec une moissonneuse batteuse ? Le mot *campus* est strictement intraduisible. On touche là à un problème de la sémantique historique : il faut *essayer comme on peut de trouver des solutions, par approximations successives*. *Campus* va avec *terra*, avec *vinea*, *pratium*..., qui sont les principaux termes d'utilisation du sol.

Le texte de Saint-Augustin crée le système de représentation médiévale : lemmatisé, pressuré avec TXM, on peut à partir de lui comparer des fréquences. Les adjectifs *terremus* et *terrestris*, par exemple, veulent apparemment dire la même chose. Mais *terremus* est deux fois plus enfoui que *terrestris*. Il convient de comparer des listes de co-occurrences, d'observer les verbes qui vont avec les deux adjectifs respectifs, afin de voir si les hypothèses de départ tiennent debout. A. Guerreau travaille uniquement avec les lemmes, surtout qu'en latin, il n'y a pas de graphie stable.