

Le corpus et la vérité

Depuis une dizaine d'années, nous vivons le retour de l'empirisme en sciences (et plus généralement d'ailleurs). L'empirisme, ancré dans la culture anglo-saxonne (Hume, Locke, Russell) et s'opposant au rationalisme continental (Descartes, Kant, von Humboldt, Popper), propose que la vérité - toute la vérité - est contenue dans les données, et que **rien** ne peut contribuer à la découverte scientifique en dehors de leur inspection. La vérité émergera naturellement dès lors que nous accumulons des données de plus en plus détaillées et précises. Le rationalisme, au contraire, pense que les données resteront muettes si elles ne sont travaillées par le génie humain: c'est la rencontre des deux qui produit les connaissances. Si les données sont bien entendu indispensables, c'est une construction qui n'a pas de lien nécessaire avec elles et qui peut procéder de n'importe quelle lubie - une théorie - qui les fait parler. En linguistique, c'est la formule saussurienne qui incarne cette position: "c'est le point de vue qui crée l'objet" (et non pas: qui interprète l'objet).

De tous les temps et sous tous les cieux, les empiristes ont fait des rationalistes des affabulateurs qui spéculent sur un nuage et abusent les deniers du contribuable par leurs vues d'esprit: agir ainsi est ne pas satisfaire aux principes scientifiques de base. Or il va de soi que toutes les sciences sérieuses, adultes, couronnées d'applications matérielles (physique, chimie, biologie) sont rationalistes: la production des connaissances s'y fait par une dialectique entre données et théorie: les deux sont consubstantielles.

Après l'acquisition (rien de spécifiquement linguistique n'est inné, tout est fonction de fréquence et d'usage), l'architecture cognitive (le connexionisme) et la théorie phonologique elle-même (usage-based: la grammaire n'existe pas [sic]), le terrain d'exercice choisi par l'empirisme est le corpus: point de salut hors corpus, rien qui ne soit extrait du corpus ne vaut, tout raisonnement hors corpus est de la philosophie spéculative et non-scientifique.

En France et en phonologie, l'empirisme est représenté par Bernard Laks, qui illustre l'utilisation du corpus en tant qu'épée anti-rationaliste dans un article récent intitulé "pour une phonologie de corpus"¹.

Dans cette communication, mon objectif est :

- de montrer en quoi la phonologie est différente (ontologiquement parlant, par rapport à la morphologie, à la syntaxe et à la sémantique), et en quoi ceci requiert une attitude particulière envers les corpus.

- de montrer que le corpus est certes indispensable en phonologie, mais qu'il ne sera jamais qu'un outil qui restera stérile sans fécondation par le raisonnement théorique et indépendant. La question n'est pas d'être pour ou contre une phonologie de corpus - la réponse est évidente. La question est de savoir **quelle** phonologie de corpus on pratique. L'empirisme n'a pas le monopole du corpus.

- de démontrer, à partir d'un aspect particulier de la liaison en français, qu'il existe des informations phonologiquement pertinentes que les locuteurs savent (qui font partie de leur compétence), mais qui ne seront **jamais** contenues dans aucun corpus, fût-il de taille intergalactique. S'il n'y a pas d'alternative au travail sur corpus, il n'y a pas d'alternative au travail avec les informateurs non plus.

Et donc, partant, comme toute autre science, la phonologie est dualiste.

Références :

Scheer, Tobias 2004. En quoi la phonologie est vraiment différente. *Corpus* 3, 5-84.

Scheer, Tobias 2004. Le corpus heuristique: un outil qui montre mais ne démontre pas. *Corpus* 3, 153-192.

Articles téléchargeables sur (<http://www.unice.fr/dsl/tobias.htm>).

¹ Laks, Bernard. « Pour une phonologie de corpus ». *Journal of French Language Studies*.