

**Regarder (observer) la langue.
Considérations sur l'impact des corpus
sur l'évolution du domaine linguistique**

**Franck NEVEU
Professeur à Sorbonne Université
Faculté des Lettres**

Plan

Présentation

- Observer la langue

- Quelques mots sur le titre donné à cette présentation

Les observables chez Antoine Culioli

Émergence d'une linguistique de corpus

- Occurrence (token), forme attestée

- Contexte, entités effectives

- Attestation

- Astérisque

L'observable en TAL

- Extension du domaine d'emploi de l'*observable*

- Non-spécification de la nature matérielle de l'observable

L'apport des corpus à l'étude linguistique, l'exemple de la néologie par dérivation et l'exemple des amorces (G. Dal, G. Bergounioux)

Plan

Autres exemples d'apports déterminants de l'observation de données langagières en corpus pour l'étude linguistique

L'étude de l'auctorialité scientifique

Les références à la tradition grammaticale

Genèse textuelle des concepts scientifiques

Apport des données textuelles à la terminographie

Rôle de la terminologie dans l'approche des observables

Incidence des corpus sur le théorique

Applicabilité des « principes »

De l'*exemplum* au *datum*

Objectivisme naïf

Connexions interdisciplinaires

Un factuel sans définition, de la naïveté théorique à la naïveté empirique

Les effets d'illusion

Glissement épistémologique

Tournant quantitatif

Place ménagée à l'interprétation dans l'analyse

Préexistence d'un paradigme

Acte interprétatif comme réalité apriorique

Français Moderne 2016, n° 1

- Franck Neveu, *Observatoires et observables en linguistique française*
- Gabriel Bergounioux, Georgette Dal, *Les observables entre théorie et technologie. Deux exemples : la création lexicale et les amorces*
- Anne-Gaëlle Toutain, *La linguistique, 'science sans observatoire' ?*
- Salah Mejri, *Les observables dans le domaine phraséologique*
- Xavier-Laurent Salvador, *La synonymie textuelle est-elle un observable de la langue de spécialité ?*
- Fabrice Issac, *Outils et instruments pour la dictionnairique*

Antoine Culioli, 1968, « La formalisation en linguistique », *Cahiers pour l'analyse*, t. 9 : 106-109

« Observables et modèles: le problème de la relation entre un modèle, l'objet et l'observateur, n'est certes pas propre à la linguistique! Mais il a pour le linguiste une importance capitale, car, aux dangers habituels, viennent s'ajouter des difficultés supplémentaires.

a) La métalangue est la langue d'usage (dans le meilleur des cas, il faudra toujours utiliser la langue U pour parler sur la métalangue ou le système formel; en fait, le plus souvent, l'intrication entre terminologie et langue d'usage est telle que le linguiste se trouve pris au piège dont il voulait explorer le fonctionnement).

b) Le langage est une activité qui suppose, elle-même, une perpétuelle activité épilinguistique (défini comme 'activité métalinguistique non consciente'), ainsi qu'une relation entre un modèle (la *compétence*, c'est-à-dire l'appropriation et la maîtrise acquise d'un système de règles sur des unités) et sa réalisation (la *performance*) dont nous avons la trace phonique ou graphique, des *textes*.

c) L'activité langagière est *signifiante*: c'est parce qu'il y a, dans la communication, des opérations aux deux bouts que les énoncés prennent un sens (opérations complexes, car tout émetteur est en même temps, c'est-à-dire *au même moment*, récepteur, et réciproquement); mais on ne peut affirmer que les mots ont un sens sans être ramené à une conception outillère du langage, conçu comme un instrument dont la finalité explicite serait la communication entre des sujets universels qui, comme on le sait, se partagent le bon sens.[...]

d) On ne saurait ramener les problèmes de catégorisation à de simples généralisations fondées sur la fréquence. [...]

La conclusion de cette énumération est que l'on ne peut poser le problème des observables sans se donner une théorie de l'observation, en particulier, sans se demander où l'on poste les observateurs. [...] »

J.B. Watson, “Psychology as the Behaviorist Views it”, *Psychological Review*, 1913, 20, pp. 158-167

G. Bergounioux (2016), « [...] une telle conception, qui s'en tient aux données, convergerait avec des demandes issues d'une linguistique appliquée en émergence : l'enseignement des langues se mettait en quête de documents authentiques alors que le Traitement Automatique des Langues (TAL) requérait la constitution de ressources électroniques pour développer ses outils et expérimenter ses applications ».

G. Bergounioux (2016), « [...] pour nommer l'ustensile de table destiné à recevoir du poivre, on ait préféré *poivrier* à *poivrière*, mais que *salière* ait prévalu sur *salier* n'est pas prédictible, pas plus que ne le sont les transformations de valeur diachronique des unités qui ont fait passer *timbre* du registre de la musique à celui de la philatélie. C'est un problème d'*observables* qui s'esquisse, entre virtuel et effectif ».

S. Auroux, *La Raison, le langage et les normes*, PUF, 1998

« Il est évident que le langage préexiste à la grammaire et que quantité de phénomènes linguistiques existent et sont produits avant même que l'on en ait une quelconque représentation. Mais il peut sembler douteux que certains phénomènes très sophistiqués, comme les domaines de la quantification ou des anaphores, soient indépendants de toute représentation. La question n'est pas de savoir si on a des exemples avant la formulation de la règle (puisque l'exemple est l'équivalent d'une règle), mais si l'on a des attestations, avant que l'on discute des exemples. »

S. Auroux, *La Raison, le langage et les normes*, PUF, 1998

« Il s'agit [...] de formes non attestées, mais pensables au point que leur absence d'attestation est en quelque sorte le fait du hasard (formes perdues). Une forme astérisquée est une forme possible ; en ce sens, elle est tout à fait autre chose que l'anti-exemple tel qu'on le rencontre dans les traités normatifs. »

S. Auroux, *La Raison, le langage et les normes*, PUF, 1998

« L'astérisque, quelle que soit sa valeur, est un prédicat qui modalise l'existence. Pour le comparatisme, le fait se réduit à l'attesté, le non-attesté englobe aussi bien le possible que l'impossible. Pour la grammaire générative, la règle doit englober aussi bien l'attesté que le possible ; si elle explique la productivité, il n'y a pas de distinction théorique entre l'attesté et le possible, mais entre les deux et l'impossible. »

G. Dal (2016), « Là où, il y a une quinzaine d'années encore, mener des études en morphologie consistait, dans le meilleur des cas, à formuler des prédictions sur le lexique possible (lexique *attestable*, selon les distinctions opérées plus haut) à partir de l'observation du lexique construit attesté entendu le plus souvent comme le vocabulaire enregistré dans les dictionnaires de langue générale, la Toile, qu'on l'explore au moyen des moteurs de recherche commerciaux ou que, disposant des compétences idoines, on recourt à des outils informatiques conçus à cet effet, a ouvert un champ d'exploration dont la pleine mesure reste encore à prendre ».

Hathout N., Plénat M. & Tanguy L. (2003), « Enquête sur les dérivés en -able », *Cahiers de grammaire* 28 : 49-90.

Hathout N., Namer F., Plénat M. & Tanguy L. (2008), « La collecte et l'utilisation des données en morphologie », in Fradin B., Kerleroux F. & Plénat M. (éds), *Aperçus de morphologie du français*, Saint-Denis, Presses Universitaires de Vincennes : 267-287.

Dal G., Lignon S., Namer F. & Tanguy L. (2004), « Toile contre dictionnaires : analyse morphologique en corpus de noms déverbaux concurrents », communication au colloque *Les noms déverbaux*, Université Lille 3, Villeneuve d'Ascq, 23-25 septembre 2004.

Bergounioux (2016)

Exemple d'élément interrompu : « Vous êtes pas sup- fan d'un club »

Exemple d'élément réitéré : « ce qui faisait que euh ben n- nos enfants »

Exemple d'élément rectifié (aperture du /e/) : « é- m- elle m'a pas mis dehors »

Grossmann F., Rinck F. (2004), « La surénonciation comme norme du genre : l'exemple de l'article de recherche et du dictionnaire en linguistique », *Langages* 156 : 34-50.

Tutin A., Grossmann F. (2013), *L'Écrit scientifique : du lexique au discours, autour de scientext*, Rennes, PUR.

Neveu F. et Lauwers P. 2007, « La notion de 'tradition grammaticale' et son usage en linguistique française », *Langages*, 167 : 7-26.

Valette M., 2006 a, « La genèse textuelle des concepts scientifiques. Étude sémantique sur l'œuvre du linguiste Gustave Guillaume », *Cahiers de lexicologie*, 89 : 125-142.

Valette M., 2006 b, *Linguistiques énonciatives et cognitives françaises*, Paris, Champion.

Fuchs C. (2014), « Le tournant quantitatif en TAL et en linguistique : enjeux cognitifs », *L'Information grammaticale*, n° 142 : 8-13.

« [...] ce n'est pas le traitement des données par la machine (sans modélisation préalable) qui, sur le fond, fera surgir des connaissances inédites concernant le système de la langue : tout au plus permettra-t-il de mettre en évidence certaines régularités statistiques ayant trait à un certain type de pratique langagière dans des circonstances particulières. Là précisément peut résider le piège, pour le linguiste tourné quasi-exclusivement vers les ressources. Le travail sur corpus n'est évidemment pas incompatible avec la réflexion théorique, il peut même être bienvenu pour étayer un raisonnement de linguistique. Mais à la condition de respecter les contraintes d'une démarche scientifique et de savoir construire une véritable problématique théorique. Or, dans la phase actuelle, le linguiste semble bien souvent transformé en un travailleur de force qui n'en aurait plus ni le temps ni les moyens. A ce compte, la technicisation risque fort de renvoyer aux oubliettes le trésor de descriptions et de théories (non « outillées ») accumulées depuis des siècles (Lazard, 2013), en donnant l'illusion qu'un traitement de surface accompagné de quelques décomptes serait susceptible de révéler *proprio motu* les propriétés de la langue. Or les données langagières (même enrichies d'annotations diverses) ne se confondent pas avec le système de règles – aussi variable et labile soit-il – constitutif de la langue. »

Nicolai R. (2000), *La traversée de l'empirique. Essai d'épistémologie sur la construction des représentations de l'évolution des langues*, Paris, Ophrys (Bibliothèque de *Faits de langues*).

Nicolai R. (2007), *La vision des faits. De l'a posteriori à l'a priori dans la saisie des langues*, Paris, L'Harmattan.