

Décrire le français au XX^e siècle

Marie-José Béguelin, Denis Apothéloz, Christophe Benzitoun,
Gilles Corminboeuf, José Deulofeu, Peter Lauwers et Dominique Willems

[Version non définitive d'un article publié dans :
Franck Neveu & Marco Fasciolo (éds), *Décrire une langue. Objectifs et méthodes*.
Paris : Classiques Garnier, 2025, 43-83.]

0. INTRODUCTION

Dans le sillon de la linguistique mondiale, la linguistique française a été marquée, à date récente, par une intensification de la collecte et de la mise à disposition de grands, voire de très grands corpus informatisés (CEFC, CRFC, FrWaC, OFROM, ESLO, Colaje, French Ten Ten Corpus, French Cow Corpus etc.). Cette évolution s'est accompagnée d'une transformation des savoir-faire, avec une importance grandissante accordée aux compétences liées à l'informatisation des données, à leur annotation, à leur fouille automatique, à leur traitement statistique. Les méthodes d'analyse outillées se sont affinées et de nombreux logiciels permettent désormais de mener des études descriptives exploitant non seulement le texte ou la transcription originale, mais aussi divers niveaux d'annotation¹, y compris prosodique (Praat, Elan, Winpitch...). Parallèlement à la linguistique de corpus s'est développée une linguistique dite 'expérimentale' qui, entre autres objectifs, vise à mettre au jour les possibles langagiers dans des situations 'de laboratoire', ce qui mobilise forces de recherche et moyens financiers. Les deux approches (orientée corpus, expérimentale) se sont d'abord ignorées, avant d'être adoptées conjointement dans certains travaux contemporains.

Ces évolutions représentent un atout majeur pour les études descriptives sur le français et pour la linguistique en général. Néanmoins, comme tout changement qui implique des innovations techniques, elles ne vont pas sans soulever des problèmes inédits. Identifier ces problèmes, esquisser des pistes pour en minimiser l'impact, tel est l'objectif que nous nous proposons dans ce chapitre.

Afin d'exploiter à bon escient les nouvelles opportunités, il est en effet nécessaire à nos yeux :

- (i) de recenser les obstacles auxquels se heurtent les méthodes outillées, afin de prévenir les risques de lacunes descriptives, voire de régressions qui pourraient découler de leur application généralisée (§ 1 : Impasses et défis actuels) ;

¹ Des plateformes telles que Nooj de Max Silberztein (<http://www.nooj-association.org/>), #LancsBox (Université de Lancaster), TXM (<http://textometrie.ens-lyon.fr/>) ou Sketch Engine (<https://www.sketchengine.eu/>) permettent le prétraitement et l'interrogation de corpus. V. § 1.2.

(ii) de garder à l'esprit les principes et « vérités fondamentales »² qui ont conféré à la linguistique sa spécificité et sa légitimité parmi les sciences humaines et exactes (§ 2 : Héritages méthodologiques) ;

(iii) de prendre, dans les domaines de l'enseignement et de la recherche, les mesures qui s'imposent afin de limiter les facteurs handicapants (§ 3 : Urgences).

Nous montrerons qu'il en va, ni plus ni moins, de la fiabilité et de la pérennité des recherches produites dans notre domaine d'études.

1. IMPASSES ET DÉFIS ACTUELS

Les auteurs de ce chapitre ont tous eu, à un moment ou à un autre, à s'interroger sur les données exploitables pour la description du français, les outillages et méthodes d'exploration appliqués à ces données, les modèles utilisés pour l'élaboration des analyses, mais aussi sur le contexte institutionnel de la recherche et de la formation en linguistique française. Dans la présente section, nous développerons plus particulièrement la question des données (§ 1.1) et des méthodes (§ 1.2) ; sur la question institutionnelle, nous nous limiterons à quelques observations générales (§ 1.3)³.

1.1. Les données

Les données mises à disposition des chercheurs, brutes et annotées, se sont multipliées au cours des dernières décennies. Dans le paysage francophone, Frantext fait figure de précurseur depuis son lancement en 1998 ; il comportait, fin 2023, 5658 références et 270 millions de mots. Le succès des études sur les SMS, les Tchats et les Tweets a permis d'y adjoindre des corpus représentatifs de types d'écrit non littéraires. De plus, le projet E-CALM vise à fournir un corpus longitudinal de copies d'élèves, en vue de l'analyse des processus d'apprentissage du français écrit. Du côté des corpus oraux, on dispose aujourd'hui de données variées, enregistrées dans différentes régions francophones de la planète, et des corpus longitudinaux sont accessibles en quantité non négligeable pour les études en acquisition (p. ex. CHILDES French⁴). Pendant plus d'une décennie, le French Treebank a été le principal corpus annoté en syntaxe exploité par les chercheurs. Mais on a vu se multiplier ces dernières années les corpus dotés de diverses couches d'annotations. Le projet *Universal Dependencies* a contribué à la structuration de la communauté autour de l'annotation syntaxique automatique⁵. Enfin, à date récente, la création de la plateforme

² Expression empruntée à Saussure, *De l'Essence double du langage (= ED)*, ÉLG p. 17.

³ Contrairement aux usages établis, nous ne renverrons pas à tous les travaux qui nous ont inspirés dans ce chapitre (rédigé au début de l'année 2021, actualisé en février 2022 puis en juillet 2024). Cela pour gagner de la place, et aussi parce que nos réflexions critiques visent des pratiques générales, auxquelles nous ne prétendons pas échapper nous-mêmes. Nous ne renvoyons pas systématiquement non plus aux références des bases de données et des outils d'analyse évoqués, qu'il est facile de retrouver en ligne.

⁴ <https://sla.talkbank.org/TBB/childes/French>

⁵ Voir les corpus « gold » dans match.grew.fr.

Ortolang, du consortium Corli ainsi que d'autres initiatives comme CorpusFinder⁶ ont témoigné d'un effort de centralisation et de recensement des données.

Tout n'est pas rose pour autant. Ainsi, beaucoup de très gros corpus écrits exploités aujourd'hui sont issus de la récupération automatique de données sur le Web ; ils sont peu fiables pour mener des descriptions linguistiques fines, car si la quantité est là, la nature et la provenance des documents réunis sont quasi invérifiables. Dans la plupart des corpus oraux, ce sont les entretiens qui dominent, si bien que la diversité des genres de parole reste mal représentée. De plus, il n'existe toujours pas, pour le français, de 'corpus de référence' comparable à ceux qui existent pour d'autres langues. Si le *Corpus d'Étude du Français Contemporain*⁷ comble en partie ce manque et rend des services appréciables, sa taille réduite (10 millions de mots dont 3,5 pour la partie orale⁸) ne permet pas une couverture suffisante. La récolte d'exemples 'à la volée', dans les interactions du quotidien, reste souvent encore le seul moyen d'obtenir certaines données cruciales pour la description et l'argumentation linguistiques. Quant aux corpus annotés automatiquement, ils sont sans doute utiles aux équipes qui ont conçu les protocoles d'annotation, mais il suffit d'une dissension sur les catégories retenues ou sur les principes d'étiquetage pour les rendre inexploitable par d'autres.

L'objectif visant à recueillir des données fiables pour l'étude descriptive du français a-t-il été atteint ? Dispose-t-on de corpus assez nombreux et diversifiés, et de qualité suffisante, pour réaliser un programme de 'classification linnéenne' des objets linguistiques, comme cela a été entrepris dans les sciences de la nature ? Poser ces questions revient à y répondre.

Le travail à partir de bases de données très vastes a d'autre part sur l'analyse linguistique deux effets pervers principaux, d'ailleurs reliés : la centration sur l'écrit (§ 1.1.1) et le calibrage normatif des données (§ 1.1.2).

1.1.1. Centration sur l'écrit (ou graphocentrisme)

Alors que les grammairiens d'autrefois procédaient le plus souvent par l'analyse d'exemples qu'ils inventaient ou empruntaient à des œuvres littéraires⁹, les promoteurs de la linguistique descriptive ont insisté sur la nécessité d'utiliser des données authentiques et plus exactement représentatives, provenant du parler d'un locuteur de la langue étudiée. Cet aspect méthodologique a été longuement développé dans *Structural Linguistics* de Z. S. Harris (1966 = 1951), ouvrage fondateur de la linguistique de terrain :

Investigation in descriptive linguistics consists in recording utterances in a single dialect and analyzing the recorded material. (Harris 1966 = 1951, p. 12)

⁶ <https://www.corpusfinder.ugent.be>.

⁷ www.ortolang.fr/market/corpora/cefc-orfeo.

⁸ Le British National Corpus (BNC) comporte 90% de textes écrits, pour 10% d'oral transcrit.

⁹ *L'Essai de grammaire française* de J. Damourette et E. Pichon (1911-1933) fait exception à cet égard. Cf. la réflexion de B. Laks (2013) qui oppose les linguistiques du *datum* vs de l'*exemplum*.

Dans l'introduction de son ouvrage, Harris met en place une distinction fondamentale entre *données primaires* et *données secondaires*. Les données primaires sont les paroles enregistrées. Les données secondaires sont les transcriptions écrites de ces paroles, élaborées par les chercheurs. Les données primaires elles-mêmes constituent déjà une sélection parmi les événements accompagnant les paroles dans la situation d'enregistrement. Ainsi les gestes en ont-ils d'abord été exclus : ils n'ont été pris en compte qu'au moment où il a été possible de faire facilement des enregistrements vidéo. Quant aux données secondaires, elles diffèrent selon les auteurs et Harris pose le problème de leur examen critique. Car les modalités de transcription dépendent des objectifs poursuivis par le linguiste : elles ne seront pas identiques s'il s'agit d'étudier la phonologie d'une langue ou sa syntaxe. Dans le second cas se pose la question de la transcription des morphèmes, qui peut être soit phonologique, soit orthographique. Pour les langues à tradition écrite, le choix de l'orthographe 'officielle' répond à un souci pratique de lisibilité, malgré son infidélité reconnue aux données primaires. Or, les linguistes qui ont entrepris de collecter du français parlé spontané ont amplement insisté sur les biais qu'introduit dans la transcription le recours à l'orthographe standard : C. Blanche-Benveniste et C. Jeanjean consacrent à les décrire 90 pages de leur livre de 1987 (p. 91-181). Dans un usage manuel des données, ces biais peuvent être surmontés grâce à l'expertise du linguiste ; mais ils sont bien plus difficiles à dominer quand on se fie, comme c'est souvent le cas aujourd'hui, à des logiciels d'analyse automatique. Les avertissements des précurseurs en reçoivent une actualité nouvelle.

Dans le cadre des méthodes outillées, les données secondaires occupent en effet, plus que jamais, le devant de la scène. Même l'étude à large échelle du français parlé s'y opère à partir d'un matériau transcrit et prétraité. Ce graphocentrisme de fait peut aller de pair avec une fétichisation des corpus édités et un oubli des enjeux épistémologiques du transcodage, pourtant fondamentaux :

En migrant du terrain à la bande puis de la bande à la page, l'oral se transforme [...] radicalement.
[...] de nombreuses interprétations de phénomènes se trouvent en effet incorporées dans la transcription *a priori*, alors qu'elles sont le but déclaré de l'analyse censée s'exercer *a posteriori* sur cette transcription. (Mondada 2000, §§ 4 et 5)

Dans le traitement de grandes masses de données, il est vite arrivé de sous-estimer l'effet configurant de la transcription, la décontextualisation qui va de pair, l'impact négatif de transcriptions imparfaites. Or dans un corpus quelconque, une banale omission du trait d'union de *peut-être* compromet d'avance le résultat de requêtes portant soit sur l'adverbe modalisateur, soit sur la 3^e personne du présent du verbe *pouvoir*. Et tout fait mineur de variation graphique (cf. *et cetera, etcetera, et caetera*, v. ex. *infra*) est susceptible, si l'on n'y prend garde, de nuire à la précision des mesures statistiques.

D'autres répercussions indésirables surviennent quand les transcrip-teurs appliquent certains automatismes au cours du transcodage, ou qu'une attention insuffisante est portée

aux sons correspondant aux extraits transcrits¹⁰. En ce cas, les ‘multi-lectures’ et autres ambiguïtés d’analyse peuvent rester inaperçues, alors qu’elles sont déterminantes pour identifier des phénomènes de réanalyse morphosyntaxique et de transcatégorisation en cours. Ainsi, l’accord du verbe au singulier est devenu courant en français parlé dans les structures de type [*un des X qui V*], y compris quand le relatif serait censé anaphoriser *les X*: p. ex. « *c’est aussi un des points négatifs qui a fait que j’ai voulu prendre cette colocation ici euh à Lausanne* », oral, OFROM¹¹. Le constat d’une variation bien installée sur l’accord concerné devrait donc inciter à diagnostiquer comme ambigus de nombreux exemples de même structure, où la différence entre accord au singulier et accord au pluriel est inaudible : « *ça c’est un des points qui révolte / révoltent beaucoup d’étudiants en sport justement* » (oral, OFROM). Or en pareille situation, le réflexe du transcripteur est le plus souvent d’accorder au pluriel, en fonction de la norme prescriptive en vigueur, ce qui aura des conséquences directes sur les comptages.

Dans une étude récente sur l’accord après des noms collectifs (*famille, monde, équipe*), C. Pusch a montré la nécessité, pour chaque occurrence orale rencontrée, d’élargir manuellement le contexte « afin de s’assurer qu’il y ait une forme verbale ou autre susceptible de se trouver dans une relation d’accord avec le nom repéré » (2020, § 15). La démarche qu’il préconise est exigeante, mais d’une indéniable salubrité méthodologique; à témoin son analyse de l’extrait suivant :

par exemple si tu prends euh toute une équipe là *qui est* en: % | _ | *qui embête* ou qui quoi *qui sont* euh un peu extrémistes dans leurs idées etcetera (OFROM, nos italiques, transcription fournie par la base¹²) [NB. Ce qui est transcrit par *en:* pourrait être une amorce de *embête*, et le *i* du *qui* précédant *embête* est éliminé par le locuteur.]

Si le linguiste s’en tient au segment fourni par le concordancier d’OFROM, qui s’interrompt après *en:*, il conclura à un accord au singulier (*est*) déclenché par le relatif *qui*. Or, comme le signale C. Pusch :

Ce n’est qu’en élargissant considérablement le contexte droit qu’on détecte – outre le verbe *embête* avec un accord scripturalement singulier mais phoniquement ambigu – le verbe *sont* qui par lien anaphorique se réfère toujours à *équipe* mais qui montre un accord non ambigu au pluriel. (*ibid.*)

Le choix qui consiste à transcrire, au plus près de la règle scolaire, *embête* au singulier, n’est donc pas anodin : il faut toute la perspicacité du linguiste pour analyser au plus juste cet accord verbal – en réalité ambigu, possiblement ‘associatif’.

¹⁰ Cela bien que « le fait de réécouter de façon indéfiniment répétée des fragments d’oralité exerce un effet de loupe agrandissant des phénomènes qui autrement échappent à l’ouïe, les faisant littéralement émerger comme tels (Auer, 1993) ; [...] » (Mondada, *ibid.*).

¹¹ Cf. : *c’est un des plus bel enregistrement de basson* (oral radio, France Musique, 1.5.2010). L’accord préconisé par la norme est en revanche respecté dans : *alors Neuchâtel ça a été je pense un des premiers cantons qui ont euh qui sont rentrés vraiment dans Harnos* (OFROM). Voir sur cette question Legros 2009.

¹² Les deux points notent un allongement, le signe % un segment inaudible [en fait *mmh*], et | _ | une pause.

Pour rappel, des choix interprétatifs de portée comparable sont engagés lors de la transcription de :

- [ja] *il y a / y a* (Blanche-Benveniste 2010) ;
- [pi] *puis / pis* (cf. Dostie 2004) ;
- [k] *que / qui* (élide devant voyelle) (v. extrait d’OFRON cité ci-dessus) ;
- [ki] *qui / qu’i(l)*, etc.¹³

Il y a bien d’autres situations où le transcripateur tend à « corriger » inconsciemment le texte oral (cf. § 1.1.2), p. ex. en rajoutant un morphème *quand* inexistant dans l’enregistrement (exemples repérés dans PFC), ou en réduisant un passé surcomposé en passé composé (dans OFRON). Si travailler sur des transcriptions peu vérifiées, faisant l’impasse sur les multi-lectures, nuit d’avance à la qualité des résultats espérés, le risque est évidemment démultiplié dès que sont exploitées en masse des données de seconde main.

L’inattention quasi généralisée aux problèmes d’ecdotique affecte aussi, cela va sans dire, les corpus d’écrit, en particulier pour ce qui touche à la ponctuation : ainsi, le théâtre de Marivaux a donné lieu, d’une édition à l’autre, à des découpages en ‘phrases graphiques’ fort différents, pour ne pas dire incommensurables (v. Goldzink 1991, p. 17-19). Or les options retenues en la matière, qui orientent la déclamation des acteurs, préconditionnent aussi l’analyse syntaxique (v. Béguelin 2002, p. 94-96), ce qui est insuffisamment thématisé dans les recherches.

1.1.2. La tentation normative

Elle s’exprime non seulement dans certaines interventions publiques consistant à préconiser, sur un point ou un autre, les ‘bonnes’ façons de dire, mais aussi à travers le tri plus ou moins conscient qui est fait des données, le prétraitement auquel elles sont soumises (§ 1.1.1), ainsi que les protocoles d’élucation mis en place pour les obtenir.

La tentation d’écarter de l’étude les données ‘non standard’, voire de discréditer le parler de certains locuteurs, est une tendance lourde en linguistique. L’astérisque a beaucoup servi à cette fin (v. notamment les critiques de Householder 1973, Berrendonner 1982), mais semble perdre un peu de terrain (v. à cet égard les options nuancées de la *Grande Grammaire du français* (= GGF) d’Abeillé *et al.*, 2021). La disponibilité des corpus, le succès des démarches ‘écologiques’, l’intérêt pour la variation, encouragent les chercheurs à appuyer leurs études sur des données d’observation. Pour reprendre un exemple qui défraya la chronique dans les années 1980, il serait difficile aujourd’hui de mettre un astérisque à « *il neige et elle tient* » (relevé dans une copie d’élève) – d’autant qu’ont été observées, dans l’intervalle, des productions comparables du genre « *il pleut et elle mouille* ».

¹³ Les erreurs de transcription affectent particulièrement les morphèmes ‘grammaticaux’ (articles, clitiques, auxiliaires) du fait de leur faible relief phonétique. Ainsi, dans Orféo, [saete] (= *ça a été*) a plus d’une fois été transcrit *ça été*.

Le normativisme peut néanmoins se manifester sous des formes plus subreptices : purge des scories et gommage des erreurs matérielles en vue du traitement automatique de grands corpus ; dérives des études qui, par commodité, étudient le français ‘spontané’ à partir de données radiophoniques ou journalistiques, ou à partir d’écrits issus des réseaux sociaux ; méconnaissance des faits minoritaires dans le cadre de l’approche statistique (§ 1.2.1) ; calibrage, parfois extrême, des protocoles de tests en psycholinguistique et recours à des artefacts normatifs en linguistique expérimentale ; études prosodiques faites à partir de phrases ou de textes lus en laboratoire... La GGF, évoquée *supra*, a été présentée lors de sa parution comme la « première grammaire parlante » de la langue française ; pourtant, une bonne partie des exemples oraux disponibles à l’écoute sur sa version en ligne sont de l’écrit lu (de manière fort peu naturelle), voire des transcriptions re-lues – ce qui renvoie à la difficulté que la linguistique rencontre à s’affranchir réellement de la représentation graphique, v. § 1.1.1.

Ainsi perdure ce déficit d’attention portée à l’oral que C. Blanche-Benveniste pointait en ces termes :

Lorsqu’il est question de langue française, de sa grammaire et de son lexique, c’est en général de langue écrite qu’il s’agit. La grammaire et le lexique de la langue parlée n’apparaissent dans les ouvrages de référence, la plupart du temps, que comme des curiosités marginales dignes d’un petit musée des horreurs de la langue. (Blanche-Benveniste 2003, p. 317)

La persistance d’une notion plus ou moins fantasmée de ‘français standard’ est un autre écueil fréquent. Elle tend à biaiser les analyses de chercheurs qui ont étudié le français à partir d’ouvrages prescriptifs et qui, sur cette base, sont portés à traiter comme des écarts les faits de langue parlée (p. ex. les redondances du genre *tout le monde il est gentil en ‘colloquial French’*). Dans les travaux en acquisition, la cécité face aux faits stigmatisés peut aller jusqu’à compromettre l’évaluation de l’interlangue des apprenants. Ainsi dans telle étude de FLE où la relative *le livre que je t’ai parlé*, frappée de l’astérisque, est décrite comme un transfert négatif à partir de la langue maternelle des apprenants, alors que la structure est courante dans la bouche des francophones. (La récente GGF, plus ouverte à la variation, devrait sur ce point exercer une influence positive.)

Se pose enfin la question des données qui ne figurent pas dans les corpus de référence et qui peuvent être indispensables à la description. Z. S. Harris relevait la difficulté à les obtenir de façon objective et proposait des pistes pour y parvenir :

If the linguist has in his corpus *ax*, *bx*, but not *cx* (where *a*, *b*, *c*, are elements with general distributional similarity), he may wish to check with the informant as to whether *cx* occurs at all. [...] Rather than constructing a form *cx* and asking the informant “Do you say *cx*” or the like, the linguist can in most cases ask questions which should lead the informant to use *cx* if the form occurs in the informant’s speech. At his most innocent, eliciting consists of devising situations in which the form in question is likely to occur in the informant speech. (1966 = 1951, p. 12, n. 2)

Près de soixante-quinze ans plus tard, c’est pourtant la solution de facilité qui prévaut, et les adeptes des méthodes dites de ‘linguistique expérimentale’ n’hésitent pas à interroger directement les sujets de leurs expériences sur l’acceptabilité des formes testées. Au risque

de sous-estimer, malgré les mises en garde de Harris (1966 = 1951), Labov (1975), Gadet (1989), Schütze (2016, 2020)..., les biais inhérents à tel ou tel design expérimental : impact des injonctions ; effets de la manière dont les questions sont formulées (non précisée dans certaines études !) ; propension des sujets à répondre ‘comme à l’école’ en fonction des attentes qu’ils attribuent aux testeurs, surtout quand ceux-ci sont leurs professeurs ; interférences de paramètres plus ou moins contingents¹⁴... Il arrive aussi que les approches formalistes fassent abstraction de l’identité du parleur ou de l’utilisateur des structures morphosyntaxiques à l’étude. Or il n’est pas indifférent, dans un corpus donné, que l’occurrence des relatives en *lequel-laquelle* ne soit observable que dans les interventions de l’intervieweur, linguiste de son état ; ou qu’un corpus de SMS de style informel, où l’accord ‘graphique’ sujet-V semble plutôt préservé, émane en fait de scripteurs solidement scolarisés, etc.

Quant aux données obtenues par la méthode des sciences participatives, p. ex. dans certaines approches expérimentales, elles apportent certes des observations instructives sur la représentation que les sujets parlants se font de leurs pratiques langagières, qu’il est ensuite possible de confronter aux réalités observées sur le terrain. Dans les faits toutefois, ce sont souvent des étudiants en sciences du langage ou des universitaires, sollicités via des listes de diffusion ou via les réseaux sociaux, qui sont recrutés comme informateurs anonymes, sans que leur profil soit dûment documenté (si une bonne part des corpus oraux actuels offre un accès aux métadonnées, l’éllicitation telle que la pratiquent les sciences participatives ne permet pas d’accéder à ces informations pourtant essentielles). La sollicitation d’échantillons sociodémographiques plus diversifiés, auprès d’informateurs extra-universitaires moins rompus à l’exercice et moins familiers de l’écrit normé, apporterait à l’évidence d’autres résultats que ceux obtenus dans le biotope même du chercheur en sciences du langage, où sont surreprésentées les classes moyenne et supérieure.

1.1.3. Bilan partiel

Au total, on voit que l’établissement et l’exploitation des données linguistiques – qu’elles soient tirées de corpus existants ou obtenues expérimentalement – sont une tâche bien plus délicate et complexe qu’il n’y paraît. Ils requièrent une expertise multiple, plus nécessaire encore si les données secondaires comportent des annotations grammaticales (classes de mots, relations syntaxiques) en vue d’un recueil automatique de structures. On ne saurait par conséquent se passer, dans la recherche sur corpus, du point de vue de linguistes accomplis, dotés d’un solide esprit critique.

14 « [...] we should never underestimate subjects’ creativity in finding ways of looking at sentences that would not have occurred to us, or in being bothered by aspects of sentences that we find mundane. » (Schütze 2020 = 2019, p. 29)

1.2. Les méthodes

Les limitations inhérentes à la recherche outillée (§§ 1.2.1 et 1.2.2), auxquelles s'ajoute la propension à l'hégémonie de certains formalismes (§ 1.2.3), génèrent des obstacles qui viennent se superposer à ceux qui viennent d'être soulignés.

1.2.1. Approches outillées et question statistique

La recherche outillée sur la base de corpus informatisés, assortie de techniques statistiques de plus en plus raffinées, a permis d'élargir la couverture empirique de la recherche. Elle a favorisé l'exploration, dans les données, de 'patrons', de tendances, mais aussi le regroupement d'éléments au comportement distributionnel analogue (la *behavior profile analysis* de Gries 2010), et rendu possible l'identification de dimensions cachées sous une multitude de variables co-orientées (Van Wetteere 2018). Les techniques quantitatives donnent aux résultats obtenus une mesure de validité statistique et donc une certaine robustesse. À l'heure actuelle, les analyses sont souvent conduites sous R, qui est à la fois un langage de programmation et un environnement libre pour le traitement statistique de données. Autour de R s'est constituée une communauté d'utilisateurs, y compris de linguistes, qui publient des manuels mettant la linguistique quantitative à la portée de tous (Gries 2016 [2009], Levshina 2015, Desagulier 2017).

Si le traitement quantitatif donne à la recherche des allures d'objectivité, on ne saurait oublier pourtant que le choix des variables et la sélection des propriétés requièrent un travail interprétatif de la part du chercheur. Cela vaut pour les catégories sémantiques (telles que l'*animéité*), pragmatiques (statut topical ou non de tel constituant) et syntaxiques (nature verbale ou adjectivale des participes, etc.). La dimension lexicale n'y échappe pas : ainsi, les formes composées imposent des arbitrages qui ont un impact majeur sur les indices quantitatifs (Zeldes 2012, p. 98-106).

Tout traitement automatique soulève, en réalité, la question de la pertinence *linguistique* de l'investigation. F. de Saussure relevait à cet égard, dans ses notes des années 1890, que l'attention du locuteur est mobilisée en priorité par la fonction significative :

Un homme habitant le *Cher* peut passer sa vie sans se rendre compte que ce nom de son département ne diffère pas, en ses sons, du mot qu'il prononce dans *cher ami*. (ED, ÉLG, p. 49)

L'analyseur automatique s'appuie à l'inverse, de manière obtuse, sur la chaîne graphique – au risque de livrer *Vienne*, nom de ville, parmi les résultats d'une requête sur les formes conjuguées du V *venir* (expérience vécue dans OFROM). Les méthodes outillées achoppent ainsi sur ce fait crucial qu'en linguistique, l'objet (soit le Signe bifacial), n'est pas « donné d'avance » (v. § 2.1.2).

Face à cette difficulté, P. Lauwers (2014) adopte une stratégie d'évitement : dans son analyse quantitative de la détermination des pluriels lexicaux, il exclut a priori de l'échantillon retenu les items à polysémie/homonymie jugée gênante, tels *selles* 'fèces' / 'sièges placés sur le dos d'un animal'. Le traitement de la polysémie est également un des points faibles de la sémantique distributionnelle, dans la mesure où la position assignée à

un lemme dans un ‘espace sémantique’ reflète une sorte de moyenne des acceptions (Heylen et al. 2015). Toutefois, des avancées ont été réalisées depuis l’avènement de l’apprentissage profond, notamment grâce à l’apparition, en 2017, des modèles dits de « transformer », dont le mieux connu est sans doute BERT (Devlin *et al.* 2019), avec ses avatars français CamemBERT (Le *et al.* 2020) et FlauBERT (Martin *et al.* 2020). Il devient ainsi possible de désambiguïser des occurrences de termes polysémiques, voire de créer des représentations vectorielles statiques pour chacune des différentes acceptions du lemme concerné (Saha *et al.* 2023).

L’approche automatique et quantitative, toutefois, ne s’accommode guère de cas indécis. Et comme il est impossible de multiplier à l’infini le nombre de distinctions au sein d’une catégorie, le chercheur se voit parfois obligé d’en fondre certaines afin d’éviter des sous-catégories trop peu peuplées ou vides. La variable [animéité] illustre bien les différences de granularité possible dans les distinctions. Selon les cas, on peut prévoir une opposition à deux termes *animé/inanimé*, à trois termes *humain/animal/inanimé*, voire davantage si l’on entend prendre en compte les collectifs (*couple, équipe, marmaille*), les cas de métonymie (*Bruxelles / la banque ING a décidé que...*) et de métaphores (*les faucons au Parlement qui ont poussé à la guerre*), etc.¹⁵ Encouragées par les nécessités pratiques, les accommodations de ce type font obstacle à la comparabilité des études.

On a vu plus haut (§ 1.1.1) que le traitement quantitatif de masse entraîne une perte de contrôle de la part du chercheur, qui se trouve dans l’incapacité de connaître chacun des exemples dont il se sert. Il en va de même avec les outils mobilisés, qui peuvent être autant de ‘boîtes noires’ au sein desquelles ‘tokenisation’, lemmatisation, étiquetage morphosyntaxique sont invérifiables (v. p. ex. *Ngram Viewer*). Quant aux techniques statistiques, il existe de multiples algorithmes sous-tendant, p. ex., le regroupement, hiérarchique ou non, dans l’analyse de groupes (*cluster analysis*) – d’où la nécessité d’en confronter plusieurs, ne serait-ce que pour dégager ensuite les regroupements stables et qui ‘font sens’.

Enfin, le recours au *big data* peut paradoxalement induire un manque d’exposition aux faits, et conduire à méconnaître le rôle des données non intuitives ou sporadiques où se trouve parfois la ‘clé de voûte’ d’un sous-système grammatical, voire le point de départ d’un changement linguistique. En langue, le critère de fréquence et la tendance statistiquement significative ne devraient jamais servir d’argument pour se détourner du fait rare. Ce sont, p. ex., des créations éphémères dues à la coercition comme *C’est [...] sous le sifflet expert de Stéphane que le buzz final retentit* qui permettent de prouver l’existence d’une construction supra-lexicale [*sous N_{de contrôle} de N*] (Lauwers 2010). Et de longue date les romanistes ont averti du danger que l’on court à ignorer les exceptions :

Ce serait toutefois négliger un élément important de la réalité et oublier que ces exceptions sont le commencement d’une organisation nouvelle qui va transformer du tout au tout l’ancien système [...]. (Foulet, 1919 [1963], p. 356-357)

¹⁵ Sans compter les cas problématiques des *fantômes*, du *Saint Esprit*, des *anges*...

Plus une forme linguistique est isolée, plus elle est cognitivement saillante. Pour qu'elle prenne son essor, il suffit parfois qu'elle soit portée par une instance à haute visibilité médiatique, comme ce fut le cas du rarissime *abracadabrantisque*, remis au goût du jour en 2000 suite à une interview télévisée du président Jacques Chirac.

La sophistication croissante des démarches quantitatives ne saurait donc voiler les limites intrinsèques aux techniques exploitées. De ce qui précède, on retiendra que les outils de recueil d'exemples et les statistiques qui vont de pair ne livrent pas des analyses *linguistiques* à proprement parler : ils sont de simples moyens efficaces, mais imparfaits, aidant au rassemblement de données. Ni en amont (transcription, annotation, v. § 1.1, § 1.1.1), ni en aval (interprétation des résultats), ils ne peuvent remplacer une soigneuse observation qualitative. Sous peine de régressions, les approches quantitatives doivent donc s'appuyer sur l'analyse linguistique et demeurer à son service. Il faut à tout prix éviter de construire des forteresses statistiques sur des sables mouvants.

1.2.2. Priorité donnée au « visible »

L'ingénierie linguistique développée pour la fouille ou l'annotation automatique de données ne peut traiter que des phénomènes détectables par les instruments dont elle dispose ; l'exploration qu'elle permet est limitée, on l'a vu, aux faits immédiatement accessibles à un moteur de recherches. Maintes dimensions du système linguistique et de son exploitation lui échappent et ne restent pour l'heure accessibles que via un traitement 'manuel' des textes et des discours – de moins en moins prisé cependant.

Une tendance à sélectionner les objets de recherche en fonction des possibilités de l'instrument s'est exprimée à date récente dans l'inflation spectaculaire des travaux sur les unités lexicales et sur les marqueurs de discours, notamment dans le cadre de la théorie de la 'grammaticalisation' (v. les travaux référencés dans la base *Lexicales*, Dendale 2024). Le linguiste est ainsi conduit à privilégier l'explicite par rapport à l'implicite, les phénomènes marqués par rapport aux phénomènes non marqués. Et les recherches sur les 'relations de discours' se centrent sur celles qui sont exprimées à l'aide de connecteurs, de 'marqueurs discursifs', même s'il est démontré que ces relations sont en majeure partie implicites (Mann & Thompson 1988 ; Corminboeuf 2014). Cette option de facilité ruine d'avance toute ambition d'élaborer un modèle général des relations « rhétoriques » qui assurent la cohérence des discours.

La centration sur l'entité explicite a pour autre effet pernicieux d'assigner à tel segment – connecteur ou 'marqueur discursif' – tout le poids d'un processus interprétatif qui, en réalité, est global : une 'relation de discours' résulte de l'interprétation d'un faisceau d'indices disparates, qui mettent en jeu, en plus des connecteurs, les signifiés des verbes, les temps et modalités, les parallélismes structuraux, les mécanismes inférentiels, l'état des connaissances partagées, etc. Or si l'on veut être en mesure de calculer automatiquement les relations de discours, implicites comme explicites, il serait essentiel d'en fournir au préalable une modélisation non réductrice.

Dans un domaine voisin qui est celui de l'anaphore, on a recouru de plus en plus à l'ingénierie informatique pour l'exploration et le traitement outillés de corpus importants. Dans le même temps, le champ des phénomènes étudiés a été drastiquement limité aux opérations de référence, *i. e.* d'identification et de désignation de référents, et aux expressions permettant d'accomplir ces opérations (principalement les syntagmes nominaux et leurs représentants pronominaux). Dans ce champ déjà restreint, les opérations de référence n'ont plus été considérées que sous l'angle de la coréférence, les faits 'associatifs' et d'anaphores 'floues' (Berrendonner 2014 ; Johnsen 2019) demeurant largement négligés. Illustrent ce mouvement les travaux, aujourd'hui nombreux, sur les « chaînes » de (co-)référence et sur la « résolution » des expressions anaphoriques. Il est difficile de ne pas voir dans cette double évolution, ingénierique et thématique, un lien de causalité, plus exactement un effet de l'outil sur le choix des objets de recherche. Cette évolution a au moins deux types de conséquences.

D'une part, elle induit une vision tronquée de ce que sont les mécanismes anaphoriques. En ne s'intéressant plus, sous l'appellation d'anaphore, qu'à des expressions nominales ou pronominales en usage coréférentiel, on en vient à oublier que l'anaphore est un mécanisme très général de conservation-prorogation de l'information dans le discours et dans l'interaction. À ce titre, elle inclut également l'ellipse (au sens morphosyntaxique du terme), tous les types de proformes, les connecteurs et conjonctions, certains temps verbaux, etc.

D'autre part, cette évolution a des effets, sans doute plus profonds et lourds de conséquences, sur la conception même de ce qu'est une anaphore. En privilégiant par nécessité technologique les *occurrences* d'expressions au détriment des *représentations* que ces expressions installent et des inférences qui en résultent, les travaux inscrits dans ce paradigme accèdent à une conception antécédentiste et substitutive des anaphoriques, plus ou moins hermétique aux processus mentaux qui président à leur production et à leur interprétation. Une notion comme celle de « chaîne de référence », ainsi que les métaphores spatiales auxquelles elle donne lieu (« maillon », « longueur », etc.), conduit assez naturellement à une vue de ce type, peu différente d'ailleurs de la perspective scolaire. Or, tout discours participe à l'élaboration de représentations dont la source n'est pas seulement d'origine verbale, mais aussi d'origine perceptuelle et cognitive ; et ces représentations conduisent elles-mêmes à toutes sortes d'inférences, qui déterminent les possibilités de référence ultérieures. La conception antécédentiste et localiste, *i. e.* « endophorique », des faits d'anaphore ne permet pas de rendre compte avec généralité de ces processus complexes. À cet égard, elle constitue une régression théorique.

S'agissant de la langue, '*cosa mentale*', une centration sur les faits apparents, tombant sous le sens, ne peut à notre avis qu'entraver le progrès des savoirs (§ 2.1.2). En plus des lacunes exposées ci-dessus, l'exclusivité donnée au 'visible' conduit à méconnaître les propriétés syntaxiques qui n'apparaissent qu'au prix d'une manipulation de la part du chercheur. Nous pensons ici à l'« expérience imaginaire » qui a longtemps fait l'essentiel de la méthode linguistique, via les tests de focalisation, d'antéposition, de pronominalisation, de mise en question, etc., qui donnent accès à la dimension paradigmatique de la langue (v. § 2.2).

Un exemple concret : si l'on s'intéresse à l'emploi de *en un mot* comme 'marqueur métadiscursif', il est tentant de s'appuyer sur une recherche quantitative dans un grand corpus, complétée par le commentaire d'exemples choisis. Ce faisant, on risque toutefois de manquer la différence entre l'emploi extraprédicatif de *en un mot* (« En un mot, *j'étais le contraire d'un soldat.* », J. Dutourd, Frantext), et son emploi régi intraprédicatif (« *Ma vie se résume en un mot : Art* », Web). Or cette différence de syntaxe externe a sa contrepartie au plan paradigmatique : dans l'emploi régi, *un* commute avec n'importe quel autre numéral (*cinquante-deux, cent, etc.*) et *mot* peut être remplacé librement par *phrase, paragraphe, article...*, alors que l'emploi 'adjoint' n'admet que la substitution de *un* par quelques numéraux de petite quantité (Béguelin 2021). C'est dans ce dernier cas seulement qu'*en un mot* mérite l'étiquette de locution (ou, si l'on y tient, de 'marqueur métadiscursif' ou de 'connecteur récapitulatif'), ce que la méthode purement quantitative risque de méconnaître ou de sous-estimer notablement.

Même dans une étude de pragmatique discursive, le tri syntaxique des données est donc une étape cruciale : au prix de l'investissement personnel du chercheur, elle permet d'aller au-delà des identités de pure forme, tout en garantissant la précision des descriptions et la solidité des mesures quantitatives. Ajoutons que la tendance à privilégier les phénomènes de surface guette même les Grammaires de Constructions et notamment leurs variantes *usage-based* : des propriétés aussi fondamentales que la nucléarité d'un complément sont ainsi devenues secondaires chez les tenants de la *Frame Semantics* (v. Boas 2003 sur les constructions résultatives).

1.2.3. 'Cadres théoriques', effets pervers du prestige de la formalisation

L'incitation forte, à partir des travaux de Chomsky, à présenter les analyses linguistiques dans un cadre théorique formalisé, a eu pour effet de mettre au second plan la nécessité d'une démarche descriptive préalable à la théorisation. Comme l'a relevé M. Gross (1979), la linguistique formelle a trop vite sacrifié l'étape taxonomique au profit de discussions (parfois byzantines) à propos de modèles élaborés sur des bases empiriques incomplètes. En bonne méthode pourtant, une formalisation devrait servir essentiellement à vérifier la cohérence et la non-contradiction d'une description préalable. Tel était le but (on l'a quelque peu oublié) du modèle proposé par Chomsky pour les structures syntaxiques, qui s'appuyait sur une description préexistante : l'analyse en constituants immédiats présentée au chapitre 16 de *Structural Linguistics* de Harris et illustrée par le cas de l'arabe marocain. Une reformulation sous forme de grammaire de réécriture a pour but de vérifier la consistance de l'hypothèse de départ et n'a pas d'autre intérêt que cela. Ce modèle de représentation/vérification a toutefois été adopté par l'école chomskyenne comme outil universel de description des structures syntaxiques, imposant certaines solutions d'analyse et en écartant d'autres *a priori*. Ainsi, les phénomènes de dépendance à distance à l'œuvre dans des séquences comme : « *à qui crois-tu qu'il faut que je m'adresse ?* » ont été traités systématiquement dans cette tradition comme résultant d'enchâssements de 'phrases'. Or un examen attentif des formes attestées dans les corpus (Bérard 2012) amène à préférer une description par constitution de prédicats complexes. Mais la notion de « complexe verbal », parce qu'elle est impossible à modéliser dans le cadre X-barre, apparaît aux yeux de

beaucoup d'auteurs comme trop peu rigoureuse et indigne d'être retenue. Toute théorie linguistique comporte une série d'axiomes de départ, de conditions préliminaires, qui limitent la classe des modèles possibles. Le problème survient quand on étend un modèle localement valide (ici, celui des syntagmes endocentriques) à la totalité de la langue, ce qui revient à en faire une condition sur la validité des grammaires.

Les orientations théorisantes ont relégué au second plan la question de la « description intermédiaire des données » (Mellet 2008, p. 15), préalable à la modélisation. Cette question est surtout posée par les spécialistes de typologie linguistique, dont certains appellent de leurs vœux un cadre descriptif « neutre », permettant de formuler les régularités morphosyntaxiques d'une langue et de les comparer à celles d'autres langues (Dixon 2010-2012). Mais l'idée même de cadre « neutre » a pu être ressentie comme une contradiction dans les termes :

Frameworks represent aprioristic assumptions that are likely to lead to a distorted description of a language. I argue against restrictive theoretical frameworks of the generative type, against frameworks of functional approaches such as Functional Grammar and Role and Reference Grammar, and against Basic Linguistic Theory. (Haspelmath 2008, p. 287)

Plutôt que de rechercher un ensemble de catégories susceptibles de tenir lieu de moyen terme descriptif, mieux vaudrait à l'évidence adopter comme cadre la pratique d'une méthodologie commune : la méthode structuraliste, rappelons-le, a justement été élaborée à cette fin (§ 2.2 *infra*). L'articulation d'une démarche classificatoire et de la construction de modèles mérite, quoi qu'il en soit, d'être remise au cœur de la recherche – d'autant plus que les approches outillées se bornent en général à reproduire les catégories et les relations de la grammaire traditionnelle (cf. § 1.2).

Un autre problème d'articulation entre données et modèle se présente en linguistique diachronique dans le cas des célèbres 'échelles' de grammaticalisation. Certaines régularités ont été observées dans les évolutions, et modélisées sous forme d'échelles ou de 'parcours' (de type '*lexical*' > '*grammatical*', '*concret*' > '*abstrait*', '*objectif*' > '*subjectif*' etc.). Mais à partir d'observations somme toute localisées, on a tiré abusivement une théorie générale, 'hors contexte', du changement linguistique, en conférant à ces parcours un caractère déterministe. Les contre-exemples pourtant abondent. Ainsi, dans le cas du supposé parcours '*concret*' > '*abstrait*', il s'avère que pour le verbe *entendre*, le sens 'physique' n'est devenu dominant qu'au XVII^e siècle : il est clairement second par rapport au sens d'intentionnalité de ce verbe polysémique (cf. lat. *intendere*). Et l'on a pu montrer aussi que *parce que*, censé illustrer le parcours préétabli *sens logique* > *sens pragmatique*, connaissait déjà les deux emplois en ancien français (Bertin 1987, Debaisieux 2013). En dépit d'un défaut de généralité dénoncé de toutes parts, les 'échelles' (et les parcours) de grammaticalisation continuent pourtant à servir d'hypothèses explicatives dans une part non négligeable des recherches sur le changement linguistique. Elles font des dégâts notamment chez les débutants, portés à les assumer avec candeur. Combinée à l'approche outillée de grands corpus, cette conception déterministe accréditée, dans les cas

les plus gênants, une vue chimérique du changement, ‘savante’ au mauvais sens du terme, déconnectée de la perspective des sujets qui en sont les acteurs (cf. § 2.1.2)¹⁶.

1.3. Les facteurs institutionnels

Pour compléter ce bilan, il convient d’évoquer le rôle de facteurs institutionnels dont la portée dépasse le domaine qui nous occupe, mais qui compromettent l’avancée des analyses descriptives sur le français.

1.3.1. Politique de la recherche et formation de la relève

L’institution, sans surprise, soutient l’emballement pour les approches outillées dont les effets pervers ont été décrits ci-dessus (§§ 1.2.1, 1.2.2). Les politiques de financement et de publication favorisent les travaux quantitatifs et statistiques, au détriment des démarches qualitatives plus classiques et de celles qui visent à diversifier les corpus ou à contrôler et perfectionner les transcriptions (il est conseillé, aujourd’hui, de minimiser la place de la collecte de données dans les demandes de financement...).

On a vu s’implanter d’autre part, dans le monde académique, une politique du chiffre qui pousse les chercheurs, jeunes ou moins jeunes, à morceler leurs travaux pour publier davantage, et si possible dans des supports jugés ‘prestigieux’. Les revues ont été cotées (sur des bases contestables, mais ceci est un autre sujet), et l’attention s’est concentrée sur des facteurs d’impact qui, pourtant, renseignent plus sur les thématiques à la mode que sur les qualités intrinsèques des parutions concernées. Soumis à la révision par les pairs (ce qui en soi est utile), les articles ont subi un formatage – y compris linguistique, via l’anglais académique¹⁷. Pour accéder à une revue anglo-saxonne, le francisant a été incité à inscrire son propos dans une logique privilégiant les concepts et les théories plutôt que la description de structures peu connues. D’où un style de rédaction *top down*, qui pose des hypothèses de recherche avant de passer à la démonstration. Appliqué de manière scolaire et peu imaginative, ce style tend à donner aux recherches un tour stéréotypé. Les auteurs se voient en outre contraints de traduire et de gloser leurs exemples, ce qui au final réduit le volume des faits traités : l’ambition descriptive cède le pas devant l’obligation de donner un vernis théorique au propos et de vulgariser à l’intention de non spécialistes.

Un peu partout, carrières et cursus doctoraux ont été calibrés dans le même esprit technocratique. En France, la durée de la thèse a été ramenée à trois ans, ce qui rend difficile la programmation de longue haleine, l’ambition interdisciplinaire, l’émergence de profils

¹⁶ Les ‘parcours de grammaticalisation’ posent un problème comparable à celui des ‘chaînes de référence’ évoquées au § 1.2.2. Alors que celles-ci empêchent, on l’a dit, une modélisation des processus cognitifs engagés dans la production et l’interprétation des anaphores, ceux-là font obstacle à une modélisation ‘en contexte’ des changements, ainsi que des facteurs qui en favorisent ou en inhibent la diffusion.

¹⁷ Il est surprenant que dans l’étude des langues modernes, les auteurs ne soient pas davantage incités à publier dans la langue cible, en espagnol sur l’espagnol, en allemand sur l’allemand, etc. Ce qui leur donnerait l’occasion de viser un lectorat d’experts et de démontrer, outre leur talent analytique, leur degré de maîtrise de la langue étudiée.

scientifiques inédits. F. Rastier (2013) a montré les effets délétères de l'idéologie managériale dont l'emprise s'est étendue sur le monde de l'enseignement et de la recherche. S'il est à l'évidence utile, comme on le fait aujourd'hui, de doter les doctorants en linguistique de compétences techniques (numériques, statistiques, etc.) et communicatives (participation à des colloques, mobilité scientifique, gestion de projets, collaborations), on peut néanmoins craindre que cette polyvalence mondaine ne finisse par nuire au 'cœur de métier' des jeunes chercheurs, qui devrait rester la langue. L'acquisition de telles compétences et leur application à des volumes de données de plus en plus imposants risquent de se faire au détriment de la culture linguistique générale, et en fin de compte, de la qualité des descriptions produites.

1.3.2. Absence de capitalisation du savoir

Dans certaines thèses récentes, on observe que la partie consacrée à l'état de l'art se réduit comme peau de chagrin. L'étude approfondie de la littérature n'est plus faite et les références citées se cantonnent au 'cadre théorique' où baigne l'auteur. Le clivage entre les modèles formalistes (notamment générativistes) et l'univers fonctionnaliste/cognitivistique prend à cet égard la forme d'un mur infranchissable. Il s'ensuit un dialogue moins informé et moins intense qu'autrefois avec une littérature par ailleurs de plus en plus pointue et pléthorique¹⁸. Ce repli sur soi favorise l'ésotérisme de certains courants et ne fait qu'attiser le problème du « faible taux de réinscription » dont pâtit la linguistique moderne (Auroux 1980), autrement dit de sa faible cumulativité.

L'oubli d'études anciennes et pourtant remarquablement documentées traduit la difficulté de la linguistique française à capitaliser ses acquis. Toutes sortes de travaux sont ainsi à (re)découvrir : outre ceux de F. de Saussure (écrits autographes et notes de ses étudiants, v. § 2.1), nous pensons à ceux qui demeurent précieux par les données qu'ils compilent et/ou par les analyses autorisées et parfois très fines qu'ils proposent de certains domaines (pour n'en citer que quelques-uns : Martinon 1927, Damourette & Pichon 1911-1933, Frei 1929, Blinkenberg 1928-1933, 1950, Gougenheim 1938, Bally²1944, Henry 1977 = 1956, Sandfeld 1970 et 1977, Eriksson 1993 ...). En linguistique historique, il conviendrait en outre de réintégrer le bilan, largement ignoré aujourd'hui, de 150 ans de recherches en grammaire comparée des langues indo-européennes : plus que jamais, la discipline a besoin d'entreprises collectives qui fassent la synthèse.

2. HÉRITAGES MÉTHODOLOGIQUES

Dans la première section de ce chapitre, nous sommes partis d'une réflexion sur l'accroissement des corpus, les effets filtrants de l'écrit et des normes, les limites des méthodes quantitatives et 'expérimentales' appliquées aux faits langagiers. Or, une telle réflexion renvoie à des thématiques qui ont été traitées aux sources mêmes de la linguistique structurale. À commencer par la question, chère à F. de Saussure, de l'objet de

¹⁸ « Comme la communauté publie beaucoup trop, chaque chercheur n'a plus le temps de lire vraiment tous les articles touchant à son domaine de recherche, il en vient à ne lire que les articles d'un 'réseau' ou les articles soutenus par un marketing efficace et ainsi, il est amené, en même temps, à se conformer à certaines 'idées' dominantes. » (Desclés 2020, p. 1)

la linguistique, du « point de vue » dont il émane, du moyen de valider les analyses produites par le savant (§ 2.1). Puis celle des principes d'une bonne description linguistique, abordée dans une perspective empiriste par L. Bloomfield (1933) et ses successeurs (§ 2.2). Indépendamment des théories au nom desquelles elles sont faites, les descriptions linguistiques qui s'appuient sur le socle méthodologique ainsi constitué demeurent à nos yeux les plus fructueuses.

2.1. L'objet de la linguistique

Saussure a posé les bases d'une science *sui generis*, dont l'objet est le « jeu des Signes » dans la vie sociale. On lui doit la reconnaissance du rapport de détermination réciproque entre *signifiant* et *signifié* – psychiques ou incorporels l'un comme l'autre – et celle de leur statut fondamentalement différentiel (§ 2.1.2). On lui doit aussi d'avoir montré que chaque entité linguistique entretient avec ses contemporaines des relations syntagmatiques en « parole effective » et des relations associatives en « parole potentielle » (c'est-à-dire virtuelle) :

Nous appelons *syntagme* la parole effective,
– ou la combinaison d'éléments contenus dans une tranche de parole réelle,
– ou le régime dans lequel les éléments se trouvent liés entre eux DANS LE DISCOURS par leur suite et précédence.

Par opposition à la *parallélie* ou parole potentielle, ou collectivité d'éléments conçus et associés par l'esprit, ou régime dans lequel un élément mène une existence abstraite au milieu d'autres éléments possibles. (Saussure, *ED*, ÉLG p. 61)¹⁹

À la linguistique, science des Signes, le linguiste genevois a fixé des prérequis auxquels font écho plusieurs des points évoqués ci-dessus²⁰.

2.1.1. Conditions d'une science de la langue

Indo-européaniste de formation, Saussure s'est trouvé, comme tous ses collègues, face à l'obligation d'interpréter les traces écrites d'états de langue révolus. Mais pour s'adonner à l'observation d'une langue vivante, le lituanien, il s'est aussi rendu en personne sur le terrain. Fort de cette expérience très complète, Saussure a vu dans l'écriture un écran qui, si l'on n'y veille, risque de voiler le « véritable objet »²¹ de la linguistique :

¹⁹ Nommés aujourd'hui, à la suite de L. Hjelmslev, *syntagmatiques* et *paradigmatiques*, ces deux types de rapports sont au fondement de la méthodologie structuraliste (§ 2.2).

²⁰ L'œuvre de F. de Saussure nous est parvenue dans des conditions éditoriales complexes (Rastier 2015). Le célèbre *Cours de linguistique générale* de 1916 s'appuie on le sait sur les notes d'étudiants qui avaient suivi les cours de linguistique générale donnés par Saussure à Genève en 1907, 1908-1909 et 1910-1911. Mais Ch. Bally et A. Sechehaye ont réorchestré le contenu en y mettant beaucoup du leur (v. CLG/DM, appareil critique). C'est pourquoi nous ne citons ici que les notes autographes de Saussure (ÉLG) et celles de ses étudiants (CLG/E), dont le contenu, sur certains points, s'écarte fortement de la vulgate de Bally & Sechehaye.

²¹ *Objet* est à prendre ici au sens de lat. *objectum* = gr. τέλος = « finalité d'une activité », v. T. De Mauro, CLG/DM n. 40, p. 415.

Forcément, [...], pour avoir des documents autant que possible sur toute époque, la linguistique devra s'occuper continuellement de la langue écrite, et souvent elle aura à emprunter ses lumières à la philologie pour se diriger mieux au milieu de ces textes écrits ; mais *elle fera toujours la différence entre le texte écrit et ce qu'il recouvre ; elle n'y verra que l'enveloppe, ou la façon extérieure de se faire connaître, de son véritable objet, qui est la langue parlée uniquement.* (É. Constantin, Notes du troisième cours de linguistique générale de F. de Saussure, 1910-1911, CLG/E 104, III C 6-7, nos italiques ; cf. *supra*, § 1.1.1)

En outre, il a mis en garde contre le biais normatif et prôné comme une évidence l'exhaustivité des descriptions :

Une femme ne dit plus *je me décollette* mais *je me décolte* ! Magnifique formation d'analogie ! Il est clair, non pas qu'il *faut* dire (**car il ne faut rien dire : car tout ce qu'on dit a sa raison d'être**) – mais il est clair qu'on a dit jusqu'à une époque récente [*je me décollette*, nous *nous décolletons*] comme *j'achète*, [*nous achetons*]. Très possible qu'on dise une fois *j'achte* (j'ai déjà entendu *je cachte*). – *récolter*. (Saussure, Deuxième conférence à l'Université de Genève, novembre 1891, ÉLG p. 162, nos caractères gras)

Toute la grammaire traditionnelle est une grammaire normative, c'est-à-dire dominée par la préoccupation de dresser des règles, de distinguer entre un certain langage dit correct et un autre dit incorrect, **ce qui exclut depuis le principe une vue supérieure sur ce qu'est le phénomène de la langue dans son ensemble.** (É. Constantin, *loc. cit.*, CLG/E 7, III C 1, nos caractères gras).

Sur ces questions sociologiquement sensibles (statut du parlé, attention portée aux écarts par rapport à la norme), les recommandations de Saussure n'ont toutefois guère été suivies d'effets, même chez ses disciples les plus proches²². La censure de l'oralité et des faits de langue 'réels' a perduré jusqu'à nos jours²³ et ne semble pas près de s'éteindre (§§ 1.1.1 et 1.1.2).

2.1.2. Propriétés du Signe et principe de différentialité

L'exploration automatisée de grands corpus, on l'a vu au § 1.2.2, privilégie ce qui est immédiatement détectable. Pourtant, Saussure a situé la linguistique « à l'extrême opposée

²² Cela ressort p. ex. des propos suivants :

[... ;] le sujet populaire qui s'embrouille dans *véneux* et *venimeux* ne fait pas du « français avancé » : il gâte, par ignorance, une distinction fine et utile. (Meillet 1929-1930, p. 148) V. dans le même esprit Bally 1963, § 44.

²³ Ce qui est relevé en termes bien sentis par A. Culioli :

La linguistique a, de même, proclamé qu'elle entendait se dégager de la domination de l'écrit (et l'on connaît l'essor des études phoniques, du moins dans certains domaines). Mais en grammaire tout se passe comme si l'on continuait à se méfier de l'oral, avec ses contraintes spécifiques, et du parlé, je veux dire de la langue réelle dans laquelle nous sommes plongés. En d'autres termes, les linguistes continuent à fonder leurs analyses sur une langue artificielle : maigres énoncés pour jeux logiques, pseudo-oral qui n'est que de l'écrit manipulé, langue qui se veut à la page parce que Nixon, les CRS ou Brigitte Bardot entrent dans des exemples qui se piquent de désinvolture ; mais surtout, langue fortement censurée, derrière le masque conceptuel des usages, registres et niveaux, par des a priori théoriques et des interdits sociologiques. (1974, p. 6)

des sciences qui peuvent partir de la donnée des sens » (ED, ÉLG p 20). Car la linguistique a pour mission de classer des Signes, c'est-à-dire

des *accouplements d'objets hétérogènes* (signes-idées), nullement, comme on est porté à le supposer, de classer des objets simples et homogènes, ce qui serait le cas si on avait à classer des signes *ou* des idées. Il y a deux grammaires, dont l'une est partie de l'idée, et l'autre du signe ; elles sont fausses ou incomplètes toutes les deux. (Saussure, ED, ÉLG p. 20 ; dans cet extrait, *signe = signifiant*²⁴)

L'enjeu scientifique est d'autant plus délicat qu'en vertu du principe de différentialité, une entité linguistique « ne *signifie* pas, mais *vaut* : [...]. Elle vaut, par conséquent elle implique l'existence d'autres valeurs. » (ED, ÉLG p. 28, v. aussi p. 77). Aussi la mise en évidence des « faits opposables » est-elle un impératif liminaire pour l'étude linguistique, en même temps qu'une promesse de résultat :

Toutes les fois que, dans une branche quelconque de la linguistique, et en se réclamant d'un point de vue quelconque, un auteur s'est livré à une dissertation sur un objet de « phonétique », de « morphologie », de syntaxe déterminé – par exemple l'existence d'une distinction grammaticale de féminin en indo-européen, ou bien la présence d'un *ñ* cacuminal en sanscrit, – cela signifie qu'il a voulu étudier un certain secteur de faits *négatifs* et dépourvus en eux-mêmes de sens et d'existence, – son étude sera profitable *dans la mesure où il aura opposé les termes qu'il fallait opposer* ; pas autrement, et cela **dans un sens non banal** : à savoir que le fait dont il s'occupe n'existe littéralement pas ailleurs que dans la présence de faits opposables. (Saussure, ED, ÉLG p. 65, nos caractères gras)

Le principe de différentialité, au cœur de la doctrine, est envisagé par Saussure dans ses ultimes conséquences, ce qui le conduit à mettre en doute :

(i) *Le postulat de l'universalité des catégories.* – Ainsi, le datif du grec, langue qui n'a ni instrumental, ni ablatif, ne saurait correspondre, au plan fonctionnel, au datif indo-européen, bien que le même terme de *datif* soit appliqué aux deux états de langue (CLG/E 1562-1563, II R 288)²⁵.

(ii) *La différence entre sens propre et sens figuré.* – « Il n'y a pas de différence entre le sens propre et le sens figuré des mots (ou : les mots n'ont pas plus de sens figuré que de sens propre), parce que leur sens est éminemment négatif. » (Saussure, ED, ÉLG p. 80). Dans cette approche purement oppositive, un mot, à un moment donné, « peut contenir et enferme en germe, tout ce qui n'est pas hors de lui. » (*op. cit.* p. 75 ; comprendre : « tout ce qui n'est pas contenu dans les formes 'ambiantes' ou concurrentes »). Ce signifié sous-spécifié, dénué de substrat ontologique, peut être envisagé comme la contrepartie, au plan

²⁴ Dans les notes de l'*Essence double*, le mot *signe* désigne tantôt la dualité signifiant-signifié (sens distingué dans ce chapitre par une majuscule initiale), tantôt, quand il est opposé à *idée*, la simple *forme vocale*. L'introduction des termes *signifiant* et *signifié*, dans le cadre du Cours III, avait précisément pour but de lever cette ambiguïté (CLG/E 99, 1116).

²⁵ Sont dénoncées comme « anachroniques » les analyses d'inspiration étymologique, qui parlent d'une époque B à l'aide de notions (catégories) ou d'entités (mots) qui n'ont de validité que dans une époque A, et vice versa (Saussure, ED, ÉLG p. 23-24). V. Béguélin 2019.

de l'‘idée’, de la « fluctuation phonétique » à l'œuvre, par exemple, dans les diverses réalisations du *r* français (*op. cit.* p. 70-72).

(iii) *La possibilité d'une histoire des Signes et des entités grammaticales.* – Du fait qu'elles tirent leur identité des termes co-existant, les entités de langue ne sont saisissables qu'en synchronie²⁶. Le corollaire méthodologique est le suivant :

Il n'y a de « langue » et de science de la langue qu'à la condition initiale de faire abstraction de ce qui a précédé, de ce qui relie entre elles les époques. (Saussure, Notes pour un article sur Whitney, ÉLG p. 217).

Loin d'être soumis à des ‘parcours’ prédéfinis (§ 1.2.3), l'ordre diachronique échappe pour Saussure à toute forme de déterminisme (CLG/E 3280, N 10, p. 28a) : il est caractérisé par le « déplacement des valeurs d'où qu'il provienne » (A. Riedlinger, Notes du deuxième cours de linguistique générale de F. de Saussure, 1908-1909, CLG/E 2733, II R 59). Aussi la seule façon d'aborder scientifiquement la morphologie historique consiste-t-elle à confronter des « rapports de formes » dans des états de langues successifs (Saussure, Note Morphologie, ÉLG p. 185).

À plus d'un siècle de distance, le renouveau des études contrastives, typologiques et historiques, de même que les entreprises d'annotation à grande échelle, devrait inciter à se préoccuper davantage « de l'origine et du fondement des catégories avec lesquelles [la grammaire] opère »²⁷. Il ne serait pas inutile, d'autre part, d'étudier les retombées d'une conception purement oppositive du signifié, laissant libre cours, en tant que virtualités, aux effets de sens que nous qualifierions vraisemblablement aujourd'hui de pragmatiques, voire aux composantes dites « encyclopédiques » de la sémantique. Enfin, s'agissant de diachronie, il serait opportun de prêter l'oreille aux mises en garde méthodologiques du savant qui, dans son *Mémoire* de jeunesse, sut prédire l'existence des ‘coefficients sonantiques’ de l'indo-européen (= *laryngales*), près d'un demi-siècle avant que le déchiffrement du hittite ne vienne en confirmer la réalité (Saussure 1879 [1878]).

2.1.3. Validation des analyses

Un dernier point pour clore cette rubrique. En matière de langue, Saussure invite à distinguer les phénomènes *externes* tombant sous le sens, « directement saisissables », des phénomènes *internes* ou « de conscience » (*ED*, ÉLG, p. 17). Or la réalité linguistique réside exclusivement à ses yeux dans le fait présent dans la « conscience » ou le « sentiment des sujets parlants » :

[À propos de la réalité des unités morphologiques] <Grand principe :> Ce qui est réel dans un état donné du langage, c'est ce dont les sujets parlants *ont conscience à un degré quelconque ; tout ce dont ils ont conscience, et rien que ce dont ils peuvent avoir conscience.* (Saussure, Note Morphologie, ÉLG, p. 192, nos italiques)

²⁶ « [...] même par simple évidence *a priori*, ne vaut que ce qui est instantanément valable » (Saussure, CLG/E 1324 N = ÉLG p. 333). L'étymologie, dans la mesure où elle reflète un point de vue ‘savant’ sur les faits, étranger au sentiment des locuteurs, ne relève pas de la linguistique (ÉLG p. 43, p. 181).

²⁷ Culioli 1974, p. 6. Cf. la notice *Catégories* dans l'*Encyclopédie grammaticale du français*.

Rappelons-nous que tout ce qui est dans le sentiment des sujets parlants est phénomène réel. Nous n'avons pas à nous inquiéter de ce qui a pu provoquer ce sentiment. (*Loc. cit.* p. 185)

Le sentiment linguistique des sujets – source et contrepartie de leur savoir-faire langagier – s'exprime à tout moment dans l'activité de parole et notamment dans les néologismes :

En français, nous avons des séries comme :

pouvons : *peuvent*
voulons : *veulent*
nouveau : *neuve, neuf*
douloureux : *douleur*

Nous dégageons l'élément de différence et sa variation régulière plus ou moins facilement, avec plus ou moins de conscience, mais enfin, nous le dégageons ; **à preuve le mot *Babouvisme* créé de *Babeuf*!** (A. Riedlinger, Notes du premier cours de linguistique générale de F. de Saussure, 1907, CLG/E 2404, I R 1.96 ; *Babeuf* = révolutionnaire français qui a donné son nom à un courant de pensée)

Saussure voit dans ce critère de productivité le seul étalon permettant de confirmer les analyses du savant :

[À propos des préfixes tels que fr. *re-*]
Quelle est la preuve absolue, péremptoire que ces préfixes sont vivants ? **Ce ne sera que la création analogique**, c'est parce que je puis former *redémisionner*, *recontempler*, sans les avoir jamais entendues (cf. tous les *re-* que l'on place devant des mots qui d'après le dictionnaire ne l'admettent pas !). Et cela n'aura lieu qu'à propos de la parole sans que je réfléchisse, sans que je veuille dire *recontempler*, etc. ; donc ces préfixes sont bien vivants. (A. Riedlinger, *ibid.*, CLG/E 2590, I R 2.37-38, nos caractères gras ; cf. Saussure, Note Morphologie, ÉLG p. 184)

Ainsi, pour cerner le 'réel' langagier, le linguiste doit épouser le point de vue des locuteurs, quitte à mettre en veilleuse les connaissances qui font sa fierté de spécialiste : « [...] le linguiste doit, s'il veut comprendre un état de langue, se mettre lui-même dans cette perspective, et abandonner la perspective diachronique ou historique qui sera pour lui une gêne, un empêchement. »²⁸

Quel que soit son objet d'étude et quelle que soit sa méthode de prédilection, le linguiste aura toujours intérêt à se poser, avec la même humilité que Saussure, la question du statut et de la valeur des analyses auxquelles il aboutit²⁹. Relèvent-elles d'une perspective « objective », purement savante, sur les faits ? sont-elles au contraire ratifiées, directement ou indirectement, par l'analyse « subjective » du locuteur, qui « seule importe »³⁰, qui

²⁸ É. Constantin, *loc. cit.*, CLG/E 1496-1497, III C 371-372.

²⁹ « Moi, grammairien, je décompose *ek1wos* [= nominatif du mot i.-e. signifiant « cheval »] en *ek1wo+s* ou *canere* [= infinitif du verbe latin signifiant « chanter »] en *can+ere*. Mais dans la vie du langage, quelle est la contrepartie réelle et tangible, quelle est la sanction, quel est le phénomène positif qui donne une sanction à cette analyse ? » (Saussure, Note *Morphologie*, ÉLG p. 195)

³⁰ A. Riedlinger, Notes du premier cours de linguistique générale de F. de Saussure, 1907, CLG/E 2759-2761, I R 2.65-66.

constitue l'objet même de la science linguistique ? Cette question sera reformulée bien plus tard par K. L. Pike, avec sa distinction entre points de vue *étique* et *émique* sur les faits (Pike 1967, p. 38). À l'heure de la linguistique outillée, elle reprend toute son actualité et pousse à réfléchir sur la façon dont l'héritage saussurien pourrait être intégré dans les approches cognitives, également axées sur les conceptions/représentations linguistiques des locuteurs.

2.2. Acquis méthodologiques

Pour décrire scientifiquement les langues, nous sommes redevables également à l'école structuraliste, notamment américaine, qui a mis en place des méthodes d'investigation éprouvées, applicables universellement. L'apport de ces méthodes, sous-estimé, voire dénigré de nos jours, se révèle pourtant considérable.

2.2.1. La démarche structuraliste

Au départ, il y a le souci de fonder la démarche scientifique sur l'observation des faits : « Toute construction théorique a toujours été précédée d'un long travail d'accumulation systématique des données » (Gross 1975, p. 9 ; v. § 1.2.3). Ces données, essentiellement empiriques, s'incarnent dans un corpus, échantillon de l'usage des locuteurs, augmenté le cas échéant de jugements de leur part.

Dans le traitement de ces données, les principes d'immanence et de différentialité (développés notamment par L. Hjelmslev, 1943) assurent l'objectivité de la démarche :

Puisque le linguistique réside dans l'association indissoluble entre signifiant et signifié, les invariants linguistiques sont des relations qui ne mettent pas en jeu seulement des formes ou seulement des régions de l'espace sémantique, mais la corrélation entre signifiants et signifiés. (Lazard 2006, p. 258).

À partir de cette définition du Signe comme objet d'étude spécifique de la linguistique, assumée jusque dans les approches formelles les plus récentes (Boas & Sag 2012), et d'une approche différentielle du sens, les structuralistes ont mis au point des méthodes rigoureuses de traitement des données :

(i) Distinction et hiérarchisation des niveaux d'analyse, avec prise en compte de la double articulation distinguant unités distinctives et unités significatives. On ne peut sous-estimer l'impact de la description du phonème comme unité abstraite, *émique*, purement distinctive, sur les analyses proposées aux autres niveaux (cf. Goldsmith 2005 ; Pike 1967, p. 38).

(ii) Distinction des deux axes (v. § 2.1), permettant d'organiser l'étude de la structure en fonction des opérations principales à l'œuvre dans la construction des énoncés : sélection d'une entité sur l'axe paradigmatique, dans une série *in absentia*, et combinaison de cette entité avec d'autres, *in praesentia*, sur l'axe syntagmatique (Jakobson 1956, p. 48).

Cette distinction³¹ a notamment été exploitée en syntaxe de l'oral, où les phénomènes de 'bribes', longtemps exclus des descriptions, peuvent être envisagés comme les traces, dans l'axe syntagmatique, de l'effort de sélection paradigmatique auquel est soumis le locuteur (Blanche-Benveniste *et al.* 1997, p. 18-22).

(iii) Procédures de découverte des structures linguistiques par l'application systématique des tests de *substitution* et de *permutation* (cf. l'« expérience imaginaire » déjà évoquée au § 1.2.2). L'examen méthodique de la distribution des unités langagières (qu'il s'agisse des phonèmes, des morphèmes ou d'unités plus larges) aboutit à une analyse en constituants immédiats et à une classification de ces constituants (Harris 1966 = 1951, 1970). Une couverture empirique aussi large que possible et le caractère contrôlable des opérations de substitution et de permutation garantissent la robustesse de la description. Ces démarches inductives constituent un préalable à la quête des propriétés ou contraintes susceptibles d'expliquer la non-équiprobabilité des combinaisons. Comme l'écrit D. Leeman :

Une structure est définissable comme un réseau d'unités définies les unes relativement aux autres par des propriétés et des relations formant système, découvertes par l'accumulation des observations et des tests ainsi que par la confirmation ou infirmation d'hypothèses partielles qui surgissent au cours du travail scientifique [...]. (2010, p. 15)

Sur ce fondement, des hypothèses précises peuvent être formulées, que ce soit au sein d'une langue donnée, dans le domaine de l'acquisition ou dans celui de la comparaison inter-langues.

La mise en rapport de la forme et du sens peut, par la suite, être approfondie et raffinée : le sens (linguistique) ne se laisse en effet saisir qu'à travers les combinaisons d'unités et les restrictions observables dans les constructions. C'est là un des objectifs de la grammaire constructionnelle, la construction étant définie comme l'appariement d'une forme et de propriétés sémantiques (v. Goldberg 1995 et, pour une synthèse, Legallois 2016). La description des contraintes propres à chaque construction se prolonge par l'établissement d'un réseau, montrant comment les constructions héritent des propriétés de constructions plus générales et les transmettent à d'autres.

2.2.2. Illustrations

L'application du principe de différentialité et des méthodes d'analyse qui en découlent demeure garante non seulement de la qualité, mais aussi du caractère innovateur des analyses qui sont proposées dans la description des langues. Quelques exemples le rappelleront pour achever notre section 2.

³¹ Qui recoupe les notions saussuriennes de *syntagme* et de *parallélie*, cf. § 2.1. Jakobson se distingue cependant du linguiste genevois par sa conception convenue du signe, défini chez lui via l'antique formule *aliquid stat pro aliquo* = « quelque chose tient lieu de quelque chose ». Dans son commentaire conclusif à Jakobson 2020, N. La Fauci montre à quel point cette conception du signe comme « signe de... », fondée sur une notion de « renvoi » et tout droit issue du sens commun, est étrangère au concept saussurien de Signe, où se manifeste une relation, interne à la langue, entre un signifié et un signifiant dont l'un est la pure contrepartie de l'autre, et vice versa.

– « *Une chose* » – L’analyse d’un des emplois du SN *une chose* proposée par C. Blanche-Benveniste (1986) illustre l’efficacité de la démarche structuraliste. Par une analyse distributionnelle en plusieurs étapes, portant sur des exemples oraux de type *il y a une chose que je me demande, c’est s’il est sincère*, l’auteure montre que *une chose* peut faire partie d’une construction remarquable, proche de la pseudo-clivée (*ce que je me demande, c’est s’il est sincère*).

L’application des critères de substitution et d’équivalence révèle tout d’abord qu’il ne s’agit pas d’un groupe nominal ordinaire, car la substitution par un autre lexème de sens voisin est impossible³², tout comme celle par un pronom *en...une* :

- * il y a *un objet* que je me demande c’est s’il est sincère
- * je m’*en* demande *une* c’est s’il est sincère

Le SN *une chose* n’est cependant pas ici un simple équivalent de *ce* dans *ce que je me demande c’est si...*, puisqu’à la différence de ce pronom, il peut recevoir des déterminations adjectivales :

la *seule* chose que je me demande c’est s’il est sincère

D’autre part, à la différence des constructions voisines à SN définis contenant des nominalisations (*la constatation, la contrainte*), l’élément verbal introduit par *c’est* ne peut être considéré comme un ‘complément’ de *chose* :

- la seule constatation à faire c’est qu’il était devenu dominateur
- la constatation qu’il était devenu dominateur (s’impose)
- il y a une chose embêtante c’est qu’ils sont venus
- ? la chose embêtante qu’ils sont venus (est un obstacle)

Il est enfin possible de montrer que *une chose* n’entretient pas de sélections lexicales avec le V introduit par *c’est*. Cette compatibilité extrême en fait un outil particulièrement utile comme relais d’une forme verbale et confirme la tendance du français à exprimer deux fois la valence d’un verbe, par une proforme d’abord, suivie d’une réalisation lexicale (comme dans *je la connais, ton histoire*). La démarche appliquée dans cette étude, essentiellement descriptive et inductive, n’a rien de mécaniste : elle fait émerger de réelles découvertes, dans un aller-retour constant entre données et hypothèses.

– *De + infinitif*. – C’est également la méthode structuraliste qui a permis d’identifier, dans les structures du type *il craint d’échouer*, un *de* démarcatif de proposition (= Comp), homonyme de la préposition casuelle ablative présente dans *il s’occupe de son canari / il s’en occupe*. En effet, *il craint d’échouer* commute avec *il le craint / il craint ça / il craint l’échec*. V. Huot 1981.

– *La prétendue différence de sens entre je et moi*. – Une différence sémantique de ce genre est activement recherchée par certains auteurs. Or la question est sans objet : *je* et *moi*, dont l’un est un pronom clitique, l’autre un SN tonique, sont en distribution

³² À moins qu’il ne s’agisse d’un autre nom ‘postiche’ (Kleiber 1987, p. 109) comme *truc*.

complémentaire et par conséquent n'entrent jamais en opposition l'un avec l'autre (Groupe de Fribourg 2024).

– *L'évolution des marques de genre dans les langues romanes.* – La « disparition du neutre » est un lieu commun de la linguistique diachronique latino-romane. N. La Fauci (2016) a ré-envisagé cette question dans une perspective saussurienne, assumant le principe de l'identité oppositive des entités morphologiques. Or, relève La Fauci, en latin, le neutre est le genre de l'indifférence fonctionnelle entre sujet et objet. Et cette indifférence fonctionnelle est précisément ce qui s'est répandu dans les langues romanes, au rebours de la description qui est donnée habituellement du phénomène. Une foi naïve dans les catégories héritées de la tradition grammaticale est ainsi de nature à empêcher l'appréhension rationnelle du processus évolutif. (Laquelle permettrait pourtant de s'épargner bien des errances sur la question du genre).

3. URGENCES

Nous avons bon espoir que les nouvelles générations de linguistes parviendront à relever les défis qui se posent aujourd'hui à la linguistique française et à sa version 'outillée'. Trois mesures simples seraient, nous semble-t-il, de nature à préparer le terrain.

La première consisterait à faire un état des lieux de la discipline. La masse des travaux descriptifs et des parutions (n. 18) rend actuellement peu visibles aussi bien les analyses consensuelles que celles sujettes à controverses, et fait obstacle à l'identification des domaines de la grammaire pour lesquels la documentation et les études font défaut. Dans ces conditions, toute entreprise de mutualisation et de synthèse des résultats de la recherche est bienvenue. Le projet *Encyclopédie Grammaticale du Français* (EGF), en libre accès sur le site *encyclogram.fr*, auxquels prennent part les auteurs de ce chapitre, se présente comme une modeste mise en œuvre en ce sens. Chaque notice de l'EGF vise à présenter une synthèse critique des connaissances accumulées sur l'objet dont elle traite. Les rédacteurs sont invités à rendre compte des conceptualisations théoriques en présence, à relever les points de convergence et de divergence entre écoles, à faire le point sur la couverture empirique des différents modèles, à (re)mettre à l'honneur les études pérennes et indémodables, à identifier les secteurs encore vierges ou en jachère, etc.

La deuxième mesure consiste à raviver le débat épistémologique. Par ses capacités de cumul et de conservation, la technologie nous fait entrer dans l'ère de la philologie numérique. On peut prévoir que les données ne seront plus statiques, mais qu'elles évolueront au gré des exploitations, entraînant une mutation des objectifs de la linguistique de corpus. Il importe, à cet égard, de penser mûrement chaque choix qui interviendra dans l'établissement des données, et de mener sur les finalités de l'ingénierie linguistique et de la linguistique computationnelle une réflexion épistémologique approfondie – d'autant plus nécessaire que la majeure partie de l'activité langagière « se déroule de manière silencieuse, inobservée et, il ne faut jamais l'oublier, inobservable » (La Fauci 2005, p. 8). C'est ici que la fréquentation des ouvrages fondateurs de la discipline s'avère fructueuse, voire indispensable, raison pour laquelle nous leur avons accordé une place importante dans ce chapitre.

La métalangue du linguiste, qui est un de ses instruments de travail, mérite, elle aussi, d'être soumise à l'examen critique. Le piège lié à toute métalangue est l'effet de naturalisation qu'elle produit. Cela se traduit, chez beaucoup d'auteurs, dans l'assomption du caractère atemporel, universel, des catégories qu'elle inclut. En diachronie, cette assomption conduit tout simplement à des anachronismes³³ ; en typologie, elle peut faire croire à des équivalences en réalité factices.

La troisième mesure consiste à repenser la répartition des tâches et à rééquilibrer les contenus de formation. Rien ne remplace, dans l'approche d'une langue, une connaissance intime, immersive, des données, indispensable au repérage dans les corpus des faits dignes d'intérêt. C'est pourquoi il est temps de contrebalancer la formation en termes d'outils (statistiques notamment) par une formation solide en linguistique descriptive, faisant place à la dimension historiographique. Tout linguiste ne peut (ni ne doit nécessairement) devenir statisticien ou informaticien, de même que tout informaticien ou statisticien ne peut s'improviser linguiste. En revanche on peut attendre beaucoup de la collaboration entre spécialistes des deux domaines. Le rééquilibrage souhaité en faveur des méthodes descriptives peut passer par l'organisation régulière d'ateliers et par la mise à disposition d'ouvrages introductifs, détachés, autant que faire se peut, des « cadres théoriques » dont l'étanchéité fait obstacle au brassage des idées.

Une dernière remarque pour conclure. Il arrive fréquemment, en linguistique, qu'un problème qui se pose à un niveau de la structure trouve sa solution à un autre niveau, ou par recours à des domaines d'interprétation extra-grammaticaux. Ainsi, la découverte des « coefficients sonantiques » de l'indo-européen, déjà évoquée plus haut (fin du § 2.1.2), est une conséquence directe de la description que Saussure a d'abord donnée des *alternances significatives* dans les états de langue qu'il étudiait : la phonétique, en l'occurrence, doit tout à la morphologie. Il existe de même des phénomènes syntagmatiques qui ne peuvent être éclairés que par recours à la prosodie, à la pragmatique inférentielle (v. ce qui est dit des relations de discours et de l'anaphore au § 1.2.2), à l'analyse conversationnelle. C'est pourquoi une formation généraliste, couvrant les différentes dimensions de l'activité langagière, nous semble si importante pour assurer la qualité et la fécondité des descriptions à venir.

RÉFÉRENCES

Abeillé Anne, Godard Danièle (sous la dir. de), avec la collab. de Delaveau Annie et Gautier Antoine, 2021, *La Grande Grammaire du français* (GGF), 2 tomes, Paris Actes Sud.

Auer Peter, 1993, "Über". *Zeitschrift für Literaturwissenschaft und Linguistik*, 90/91, p. 104-138.

Auroux Sylvain, 1980, « L'histoire de la linguistique », *Langue française*, 48, p. 7-15

³³ Les auteurs de la *Grande Grammaire historique du français* (Marchello-Nizia *et al.*, 2020) se sont trouvés confrontés de plein fouet à ce problème (v. Combettes 2012 ; Béguelin 2022).

- Bally Charles, 1963 (¹1909), *Traité de stylistique française*, Genève Georg et Cie S.A.
- Bally Charles, ²1944, *Linguistique générale et linguistique française*, Berne A. Francke.
- Béguelin Marie-José, 2002, « Clause, période, ou autre ? La phrase graphique et la question des niveaux d'analyse », dans : Charolles Michel, Le Goffic Pierre & Morel Mary-Annick (éds), *Y a-t-il une syntaxe au-delà de la phrase ? Verbum XXIV*, n° 1-2, p. 85-107.
- Béguelin Marie-José, 2019, « Synchronie vs diachronie, ou les vicissitudes d'une dualité incomprise », dans : Arabyan Marc, Bronckart Jean-Paul & Escudé Pierre (éds), *Les Langues dans la vie. Hommage à Tullio De Mauro*, Limoges Lambert-Lucas, p. 163-192.
- Béguelin Marie-José, 2021, « Syntaxe externe et fonctionnement sémantique de *en un mot* », dans : Berrendonner Alain (éd.), *Les Périphériques. Syntaxe et sémantique, Verbum XLIII*, n° 2, p. 245-272.
- Béguelin Marie-José, 2022, Compte rendu de : Christiane Marchello-Nizia, Bernard Combettes, Sophie Prévost & Tobias Scheer (éds), *Grande Grammaire historique du français*, 2 volumes, Berlin/Boston, De Gruyter, 2020, 2185 pages, dans : *Scolia 36*, p. 172-192. En ligne : <https://doi.org/10.4000/scolia.1958>.
- Bérard Lolita, 2012, *Dépendances à distance en français contemporain. Étude sur corpus. "C'est ce qu'on pense qui devrait être fait"*, thèse de l'Université de Lorraine.
- Berrendonner Alain, 1982, *L'Éternel grammairien*, Berne Peter Lang.
- Berrendonner Alain, 2014, « Anaphores louches et dualités », dans : Fossard Marion & Béguelin Marie-José (éds), *Nouvelles perspectives sur l'anaphore. Points de vue linguistique, psycholinguistique et acquisitionnel*, Berne Peter Lang, p. 169-186.
- Bertin Annie, 1987, *L'Expression de la cause en ancien français*, Genève Droz.
- Blanche-Benveniste Claire, 1986, « 'Une chose' dans la syntaxe verbale », *Recherches sur le français parlé*, 7, p. 141-168.
- Blanche-Benveniste Claire, 2003, « La langue parlée », dans : Yaguello Marina (éd.), *Le Grand Livre de la Langue française*, Paris Seuil, p. 317-344.
- Blanche-Benveniste Claire, 2010, « Où est le *il* de *il y a*? », *Travaux de Linguistique*, 61, p. 137-153.
- Blanche-Benveniste Claire & Jeanjean Colette, 1987, *Le Français parlé. Transcription et édition*, Paris Didier Érudition, Institut national de la Langue française.
- Blanche-Benveniste Claire, Bilger Mireille, Rouget Christine, Eynde Karel van den, 1997, *Le Français parlé. Études grammaticales*, Paris Éditions du CNRS.

Blinkenberg Andreas, 1928-1933, *L'Ordre des mots en français moderne*, Copenhague Levin & Munksgaard (2 vol.).

Blinkenberg Andreas, 1950, *Le Problème de l'accord en français moderne : essai d'une typologie*, Copenhague Munksgaard.

Bloomfield Leonard, 1933, *Language*, New York Holt, Rinehart and Winston. Traduction française, 1970, Paris Payot.

Boas Hans, 2003, *A Constructional Approach to Resultatives*, Stanford CSLI.

Boas Hans & Sag Ivan (eds), 2012, *Sign-Based Construction Grammar*, Stanford CSLI.

Combettes Bernard, 2012, « De quelques problèmes spécifiques à l'élaboration d'une grammaire historique », *Langue française*, 176, p. 69-83.

Corminboeuf Gilles, 2014, « L'identification des relations de discours implicites : le cas de l'adversation », Congrès Mondial de Linguistique française (CMLF'14), p. 2367-2382, publication électronique : www.shs-conferences.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01049.pdf

Culioli Antoine, 1974, « À propos des énoncés exclamatifs », *Langue française*, 22, p. 6-15.

Damourette Jacques & Pichon Édouard, 1911-1933, *Des mots à la pensée. Essai de grammaire de la langue française*, Paris Éditions d'Artrey.

Debaisieux Jeanne-Marie, 2013, « Autour de parce que et de puisque », dans : Debaisieux, Jeanne-Marie (dir), *Analyses linguistiques sur corpus. Subordination et insubordination en français*, Paris Hermès, p. 185-245.

Dendale Patrick, 2024, *Lexicales* (édition 22.2). Bibliographie en ligne d'études linguistiques portant sur des unités lexicales et grammaticales du français, (6800+ références), publication électronique : <https://www.uantwerpen.be/lexicales>.

Desagulier Guillaume, 2017, *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*, New York Springer.

Desclés Jean-Pierre, 2020, « Proposition d'évaluations par *disputatio* », manuscrit.

Devlin Jacob, Chang Ming-Wei, Lee Kenton & Toutanova Kristina, 2019, « BERT: Pre-training of deep bidirectional transformers for language understanding », in: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1, p. 4171-4186, Minneapolis, Minnesota, Association for Computational Linguistics.

Dixon Robert M. W., 2010-2012, *Basic linguistic theory*, 3 vol., Oxford Oxford University Press.

Dostie Gaétane, 2004, « Considérations sur la forme et le sens. *Pis* en français québécois. Une simple variante de *puis* ? Un simple remplaçant de *et* ? », *Journal of French Language Studies*, 14(2), p. 113-128. doi:10.1017/S0959269504001607

Dryer Matthew, 2001, “What is a basic linguistic theory?”, online: <https://www.acsu.buffalo.edu/~dryer/blt>

EGF = *Encyclopédie grammaticale du français*, en ligne : encyclogram.fr

Eriksson Olof, 1993, *La Phrase française. Essai d'un inventaire de ses constituants syntaxiques*, Göteborg Acta Universitatis Gothoburgensis.

Foulet Lucien, 1919 [1963], *Petite syntaxe de l'ancien français*, 3^e édition revue, Paris Champion.

Frei Henri, 1991 = 1929, *La Grammaire des fautes*, Genève Slatkine.

Gadet Françoise, 1989, *Le Français ordinaire*, Paris Armand Colin.

Goldberg Adele, 1995, *Constructions. A Construction Grammar approach to argument structure*, Chicago Chicago University Press.

Goldsmith John, 2005, “Review: The Legacy of Zellig Harris. Language and Information into the 21st Century”, *Language* 81 (3), p. 719-736.

Goldzink Jean, 1991, « Note sur la présente édition avec une petite digression sur la ponctuation », dans : Marivaux, *Les Acteurs de bonne foi, La Dispute, L'Épreuve*, Paris Garnier Flammarion, p. 17-19.

Gougenheim Georges, 1938, *Système grammatical de la langue française. Manuel de morphologie et de syntaxe descriptives*, Paris d'Artrey.

Gries Stefan Th. 2010, “Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics”, *The Mental Lexicon* 5 (3), p. 323-346.

Gries Stefan Th., [2009] 2016, *Quantitative Corpus Linguistics with R*, New York Routledge.

Gross Maurice, 1975, *Méthodes en syntaxe*, Paris Hermann.

Gross Maurice, 1979, “On the failure of generative grammar”, *Language* 55 (4), p. 859-885.

Groupe de Fribourg, 2024, *L'Inversion du sujet clitique et ses fonctions en français contemporain*, Berne Peter Lang.

- Harris Zellig S., 1951, (*Methods in*) *Structural Linguistics*, Chicago University of Chicago Press (First Phoenix edition 1960, Seventh Impression 1966).
- Harris Zellig S., 1970, *Papers in structural and transformational linguistics*, Dordrecht D. Reidel Publishing Company.
- Haspelmath Martin, 2008, 'Framework-free grammatical theory', in: Heine Bernd & Narrog Heiko (eds), *The Oxford Handbook of Linguistic Analysis*, chap. 13, Oxford Oxford University Press, p. 287-310.
- Henry Albert, 1977 = 1956, *Études de syntaxe expressive : ancien français et français moderne*, Bruxelles Éditions de l'Université de Bruxelles.
- Heylen Kris, Wielfaert Thomas, Geeraerts Dirk & Speelman Dirk, 2015, "Monitoring Polysemy: Word Space Models as a Tool for Large-Scale Lexical Semantic Analysis", *Lingua: International Review of General Linguistics*, 157, p. 153-172.
- Hjelmslev Louis, 1943 (trad. 1968), *Prolégomènes à une théorie du langage*, Paris Éditions de Minuit.
- Householder Fred W., 1973, "On Arguments from Asterisks". *Foundations of Language*, Vol. 10, N° 3, p. 365-376. <http://www.jstor.com/stable/25000725>
- Huot Hélène, 1981, *Constructions infinitives du français. Le subordonnant DE*. Genève-Paris Droz.
- Jakobson Roman, 1956, *Fundamentals of Language*, The Hague, Mouton & Co. En partie traduit dans : *Essais de linguistique générale*, Paris Minuit, 1963.
- Jakobson Roman, 2020, *Lo Sviluppo della semiotica e altri saggi*, saggio introduttivo di Umberto Eco, saggio conclusivo di Nunzio La Fauci, traduzioni di Andrea La Porta, Emilio Picco e Ugo Volli, Milano Bompiani, Campo aperto.
- Johnsen Laure Anne, 2019, *La Sous-détermination référentielle et les désignateurs vagues en français contemporain*, Berne Peter Lang. <https://doi.org/10.3726/b13008>
- Kleiber Georges, 1987, « Mais à quoi sert donc le mot *chose* ? », *Langue française*, 73, p. 109-128. <https://doi.org/10.3406/lfr.1987.6431>
- Labov William, 1975, *What is a linguistic fact?*, Lisse The Peter de Ridder Press. Traduction française : *Marges linguistiques* 1, 2001, p. 25-68.
- La Fauci Nunzio, 2005, *Facettes de linguistique rationnelle*, Pisa Edizioni ETS.
- La Fauci Nunzio, 2016, « Sul neutro. Grammatica e linguistica », in : Benedetti M., Bruno C., Dardano P., Tronci L. (a cura di), *Grammatiche e grammatici. Teorie, testi e contesti*,

Atti del XXXIX Convegno della Società Italiana di Glottologia (Siena, 23-25 ottobre 2014), Roma Il Calamo, p. 9-38.

Laks Bernard, 2013, “Why is there variation instead of nothing?”, *Language Sciences*, 39, p. 31-53. Version française : <https://bernardlaks.files.wordpress.com/2018/04/pourquoi-y-a-t-il-de-la-variation-plutot-que-rien.pdf>

Lauwers Peter, 2010, « Les locutions en *sous* comme constructions », *Le Français moderne* 78/1, 3-27.

Lauwers Peter, 2014, « Les pluriels ‘lexicaux’ : typologie quantifiée des déficits de dénombrabilité », *Langue française*, 183, 117-132

Lazard Gilbert, 2006, *La Quête des invariants interlangues : la linguistique est-elle une science ?* Paris Champion.

Le Hang, Vial Loïc, Frej Jibril, Segonne Vincent, Coavoux Maximin, Lecouteux Benjamin, Allauzen Alexandre, Crabbe Benoit, Besacier Laurent & Schwab Didier, 2020, “FlauBERT: Unsupervised Language Model Pre-training for French ”, LREC, Marseille.

Leeman Danielle, 2010, « Description, taxinomie, systématique : un modèle pour les emplois des verbes français », *Langages*, 179-180, p. 5-29. DOI : 10.3917/lang.179.0005.

Legallois Dominique, 2016, « La notion de Construction », dans : *Encyclopédie grammaticale du français* (EGF). En ligne : encyclogram.fr.

Legros Georges, 2009, « ‘Une des femmes qui ont bien mérité de la langue française.’ Petite glane entre norme, grammaire et usage », dans : Willems Martine (dir.), *Pour l’amour des mots : Glanures lexicales, dictionnairiques, grammaticales et syntaxiques. Hommage à Michèle Lenoble-Pinson*, Bruxelles Presses de l’Université Saint-Louis, p. 199-209. DOI : <https://doi.org/10.4000/books.pusl.23209>.

Levshina Natalia, 2015, *How to Do Linguistics with R : Data Exploration and Statistical Analysis*, Amsterdam J. Benjamins.

Mann William & Thompson Sandra, 1988, “Rhetorical Structure Theory: Toward a functional theory of text organisation”, *Text*, 8-3, p. 243-281.

Martin Louis, Muller Benjamin, Ortiz Suárez Pedro Javier, Dupont Yoann, Romary Laurent, de la Clergerie Eric, Seddah Djame & Sagot Benoît, 2020, « CamemBERT: a tasty French language model », in: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 7203-7219, Association for Computational Linguistics.

Marchello-Nizia Christine, Combettes Bernard, Prévost Sophie & Scheer Tobias (éds), 2020, *Grande Grammaire historique du français* (GGHF), 2 volumes, Berlin et Boston, Walter de Gruyter.

- Martinon Philippe, 1927, *Comment on parle le français*. Paris Larousse.
- Meillet Antoine, 1929-1930, Compte rendu de la *Grammaire des fautes* de H. Frei, *BSL* 30, p. 145-149.
- Mellet Sylvie, 2008, « Introduction », dans : Mellet Sylvie (dir.), *Concession et dialogisme. Les connecteurs concessifs à l'épreuve des corpus*, Berne Peter Lang, p. 1-16.
- Mondada Lorenza, 2000, « Les effets théoriques des pratiques de transcription », *LINX*, 42, p. 131-146. <https://doi.org/10.4000/linx.902>
- Pike Kenneth L., 1967, *Language in relation to a unified theory of the structure of human behaviour*, The Hague Mouton.
- Pusch Claus, 2020, « Je dirais que la grosse majorité sont là. Questions d'accord(s) avec les noms collectifs dans les données OFROM », *Studia linguistica Romanica*, 4, publication électronique : <https://studialinguisticaromanica.org/index.php/slr/article/view/39/56>
- Rastier François, 2013, *Apprendre pour transmettre : l'éducation contre l'idéologie managériale*, Paris PUF.
- Rastier François, 2015, *Saussure au futur*, Paris Les Belles Lettres.
- Saha Anik, Gittens Alex & Bulent Yener, 2023, "Word Sense Induction with Knowledge Distillation from BERT", [arXiv:2304.10642](https://arxiv.org/abs/2304.10642).
- Sandfeld Kristian, 1970, *Syntaxe du français contemporain. Les pronoms*, Paris Champion.
- Sandfeld Kristian, 1977, *Syntaxe du français contemporain. Les propositions subordonnées*, Genève Droz.
- Saussure Ferdinand de, 1879 [1878], *Mémoire sur le système primitif des voyelles dans les langues indo-européennes*, dans : Bally Charles & Gautier Léopold (éds), 1922, *Recueil des publications scientifiques de Ferdinand de Saussure*, Genève Sonor [réimpression Genève Slatkine, 1984], p. 1-268.
- Saussure Ferdinand de, 1916, *Cours de linguistique générale*, éd. par Bally Charles & Sechehaye Albert ; 2^e éd. 1922 ; 3^e éd. 1931. Éd. critique par De Mauro Tullio, Paris Payot, 1972 (= CLG/DM). Éd. critique et synoptique par Engler Rudolf, t. I, 1968 ; t. II, Appendice, 1974, Wiesbaden Otto Harrassowitz (= CLG/E).
- Saussure Ferdinand de, 2002, *Écrits de linguistique générale* (= ÉLG), Paris Gallimard.
- Schütze Carson T., 2016 (= 1996), *The empirical base of linguistics: Grammaticality judgments and linguistic methodology*, Berlin Language Science Press.
DOI: 10.17169/langsci.b89.100

Schütze Carson T., 2020 (= 2019), “Acceptability ratings cannot be taken at face value”, in: Schindler Samuel, Drożdżowicz Anna & Brøcker Karen (eds), *Linguistic intuitions: Evidence and method*, Oxford OUP, ch. 11 (3rd Draft 2019.07.11), lingbuzz/004862

Van Wettere Niek, 2018, *Copularité et productivité. Une analyse contrastive des verbes attributifs issus de verbes de mouvement en français et en néerlandais*, thèse de l’Université de Gand.

Zeldes Amir, 2012, *Productivity in Argument Selection. From Morphology to Syntax*, Berlin & Boston De Gruyter.