

Note de lecture

Gaston Gross. *Manuel d'analyse linguistique : Approche sémantico-syntaxique du lexique*. Villeneuve d'Ascq, Presses Universitaires du Septentrion (coll. « Sens et Structure »), 2012.

Broché, 370 p. ISBN : 978-2-7574-0397-6. 25 €.

Le *Manuel d'analyse linguistique* de Gaston Gross est le legs de trente ans d'expérience de description grammaticale du français. Il fournit une synthèse précieuse du savoir-faire accumulé par l'auteur dans ce domaine, et qui est probablement, en termes de couverture du lexique, sans équivalent dans l'espace francophone. C'est sur ce savoir-faire que s'est construit, à l'université Paris 13, une équipe¹ qui exporte ses compétences en lexicographie numérique auprès d'entreprises et de partenaires de plusieurs pays. Gaston Gross, aujourd'hui professeur émérite, publie ici dans un volume unique des principes et des méthodes de description dont il avait auparavant rendu compte de sous-ensembles plus spécifiques².

L'ouvrage, paru aux Presses Universitaires du Septentrion, se veut abordable par un public assez large. Son prix (25 €) ne le réserve pas à l'achat par des grandes institutions ou des bibliothèques. Il n'exige pas de pré-requis conceptuels autres que ceux du vocabulaire de la grammaire française classique telle qu'elle est enseignée au collège. Il peut donc figurer sur la liste d'achats utiles de tous ceux qui vont être amenés à pratiquer de la description lexicale dans une perspective de traitement automatique — qu'ils soient étudiants de licence ou professionnels, et quelle que soit leur formation de base (informatique ou linguistique).

Un point de vue lexicographique centré sur les emplois

Ce que propose Gaston Gross dans ce *Manuel d'analyse linguistique* est, en résumé, une méthode pour constituer des dictionnaires électroniques. Les dictionnaires électroniques sont à la fois des entrants et des produits du traitement automatique des langues (T.A.L.). C'en sont des entrants, car aucune tâche de TAL ne peut être effectuée sur des textes sans dictionnaire : qu'il s'agisse de traduction automatique ou d'étiquetage en catégories morpho-syntaxiques, voire même de segmentation en unités lexicales, une information sur le lexique doit être convoquée à un moment du processus. C'en sont par ailleurs aussi, et de plus en plus, des produits³.

Au cœur de la problématique du traitement des langues, donc, le dictionnaire doit contenir suffisamment d'information pour décrire la structure d'un texte, avec un niveau de profondeur adéquat à la tâche à effectuer. Il doit permettre par exemple à un système automatique de traduire le français « voulez-vous prendre un café ? » en allemand par la phrase « möchten Sie einen Kaffee trinken ? », mais de ne pas traduire « voulez-vous prendre un bain ? » par « möchten Sie ein Bad trinken ? »

1 Le LADL, devenu LLI, puis aujourd'hui LDI (Lexiques, Dictionnaires, Informatique).

2 Notamment p.ex. dans *Les expressions figées en français : des noms composés aux locutions*. Paris, Ophrys, 1996 ; *Les constructions converses du français*. Genève, Droz, 1989.

3 À titre de miroir de l'évolution de l'activité de la communauté TAL, on peut observer la liste des articles publiés chaque année dans des colloques comme TALN, COLING ou ACL : un nombre toujours croissant d'entre eux (surtout dans le cas d'ACL) décrivent des travaux de fouille de textes (c'est-à-dire, en réalité, d'apprentissage automatique) ; et une partie importante de ceux-ci a pour but d'acquérir de la connaissance lexicale (on parle très souvent, notamment, de désambiguïsation) pour peupler des dictionnaires ou des ontologies. À observer cette évolution, il semblerait que la recherche en TAL a accompli la plus grande partie de son travail historique en aval (dans le domaine de la traduction automatique, par exemple) et qu'elle a déplacé son activité vers l'amont (il s'agit de nourrir les dictionnaires sur lesquels se fondent les systèmes de traduction automatique, car c'est de leur qualité que dépendent à présent les améliorations possibles). Sans entrer dans ce débat, on peut noter que cette évolution semble donner raison à ceux qui, comme Gaston Gross, ont dès le départ considéré la connaissance lexicale comme la question centrale de la discipline.

L'approche suivie par Gaston Gross est fondée sur le postulat que l'on ne peut séparer lexique, syntaxe et sémantique. En lecteur assidu de Zellig Harris (cité en bibliographie à presque chaque chapitre), il tire les conséquences de l'idée que l'usage d'un mot est défini par les distributions de ses contextes : « les textes ne sont pas constitués de mots interchangeable comme le seraient les briques d'un mur, mais ils forment entre eux des structures élémentaires, qui sont à la base de tout discours. Ces structures élémentaires, nous les appelons des *schémas prédicatifs* [...] Dès qu'un schéma prédicatif a été reconnu à un endroit du texte, on se trouve devant une sélection de mots qui ensemble forment une unité, excluant toutes les autres lectures possibles de ces mots, s'ils s'inséraient dans d'autres schémas » (p. 12). Les propriétés sémantiques et syntaxiques d'un mot sont donc indissociables, et « décrire une langue, c'est faire le recensement organisé de l'ensemble des emplois qu'elle comporte » (p. 8). Ce livre propose « un traitement intégré du lexique, de la syntaxe et de la sémantique » (p. 23).

Dans cette perspective, la syntaxe et la sémantique sont intégrées au lexique. Ce principe peut paraître évident à une génération d'étudiants et de chercheurs en TAL élevés aux grammaires à adjonction d'arbres (TAG) ; mais il faut souligner que Gaston Gross a raison depuis longtemps ; qu'il a longtemps eu raison contre la pensée dominante en linguistique formelle⁴ ; enfin, que le fait que son approche soit fondée sur des principes maintenant largement reconnus comme valables par la communauté des chercheurs, n'indique pas que ceux-ci aient percolé jusqu'aux applications réellement utilisées dans le vaste monde extérieur. À titre d'illustration, un test effectué aujourd'hui (en novembre 2013 : table ci-dessous) sur les versions de base des trois systèmes de traduction automatique les plus utilisés, montre que la distinction des différents emplois de *prendre* (un café, un bain ...), ou d'*abattre* (pour reprendre l'exemple donné par Gaston Gross au début de son ouvrage) est loin d'être devenue un problème trivial, dont la solution serait depuis longtemps intégrée.

	<i>J'ai pris un café</i>	<i>J'ai pris un bain</i>	<i>J'ai abattu un arbre</i>	<i>J'ai abattu un soldat ennemi</i>
Systran	Ich habe einen Kaffee getrunken	Ich habe ein Bad genommen	Ich habe einen Baum abgerissen	Ich habe einen feindlichen Soldaten abgerissen.
Reverso	Ich habe Kaffee genommen	Ich habe ein Bad genommen	Ich habe einen Baum niedrigerissen	Ich habe einen feindlichen Soldaten niedrigerissen
Google Translate	Ich nahm einen Kaffee	Ich nahm ein Bad	Ich einen Baum zu fällen	Ich schoss ein feindlicher Soldat ⁵

Le chapitre 2 montre, sans faire le choix d'un formalisme de représentation particulier, l'importance de la notion d'emploi de prédicat. Celle-ci définit des catégories qui permettent de rassembler commodément des faisceaux de traits linguistiques empiriquement corrélés, et relevant de différents « niveaux » de description dans la partition traditionnelle du système des langues en lexique, morphologie, syntaxe et sémantique : le choix de l'un des sens d'un verbe, la structure de la phrase déployée autour de lui, le nombre de ses compléments, les prépositions et flexions de chacun d'entre eux, le degré de figement de certains, les transformations activables ou au contraire inhibées⁶. C'est en ces termes que Gaston Gross constate le fait que le sens d'une unité linguistique dépend de son contexte d'emploi (« le sens n'est pas premier ni "isolable" mais [...] il est en connexion avec bien

4 En 1993, les grammaires d'unification étaient encore des « nouvelles grammaires », pour reprendre le titre d'un ouvrage d'Anne Abeillé.

5 Dans les traductions données par Google Translate, les verbes *fällen* et *schießen* sont corrects, mais les phrases sont agrammaticales. L'ajout d'un point final à la fin de la phrase en français donne des résultats différents mais pas moins surprenants. La traduction automatique statistique (dans sa version de base en tout cas) a les défauts de ses qualités : elle trouve les verbes les plus fréquemment co-occurents, mais assemble des tronçons de texte dont la forme grammaticale n'est pas contrôlée à l'échelle de la phrase.

6 Cette compilation de la structure grammaticale dans le lexique — y compris des transformations — permet, on le sait par ailleurs, de résoudre le problème de la complexité d'analyse du modèle générativiste.

d'autres propriétés d'une structure phrastique », p. 43).

L'un des fondements de la méthode de Gaston Gross est la notion de « classes d'objets » ; elle constitue la matière du chapitre le plus important de l'ouvrage (chap. 4). En résumé, pour éviter de produire « ich habe einen feindlichen Soldaten abgerissen », il faut identifier les paradigmes de compléments d'objet correspondant aux différents emplois du verbe *abattre*, et noter que ce verbe se traduit par *abreißen* s'il s'agit de démolir une construction, par *fällen* s'il s'agit d'un arbre, et par *erschießen* s'il s'agit d'un homme (p. 35).

Une description exhaustive du français

Au-delà de ces principes fondamentaux, les qualités remarquables de l'ouvrage de Gaston Gross sont sa finesse et son exhaustivité. Il met en garde dès l'abord contre une vision naïve de la structure du lexique, qui permettrait une description à peu de frais du fonctionnement de la langue sous forme de règles simples, en posant des équations entre catégories morpho-syntaxiques et sémantiques (le verbe exprime les prédicats, le nom exprime les entités, etc.) : « du point de vue de leur fonction dans le cadre de la phrase », note-t-il très justement, « les catégories grammaticales majeures ne constituent pas des ensembles disjoints ». Il établit ainsi l'existence de prédicats nominaux et de prédicats adjectivaux, dont l'étude détaillée fait l'objet de plusieurs chapitres (chap. 5 et 6 respectivement, le chap. 8 étant consacré à l'étude spécifique de leurs verbes support). S'il n'entre pas dans l'objectif de cet ouvrage de donner des descriptions extensives de certains cas d'usage⁷, on y trouve en revanche un large panorama des catégories de phénomènes décrits et des questions qui surgissent lors de leur examen.

De même, de nombreuses pages sont consacrées, avec à chaque fois un souci de complétude, à l'analyse du fonctionnement des prépositions, des déterminants, des adverbes, des différents types de prédicat (événements, actions, états), et de l'expression des subordonnées et compléments circonstanciels. L'auteur développe notamment une réflexion très intéressante sur la base de l'intuition que la différence fondamentale entre compléments et circonstanciels peut être définie comme suit : le complément est l'argument d'un prédicat de premier ordre, alors que le circonstanciel est un prédicat de second ordre, dont les arguments sont eux-mêmes des prédicats (p. 52-53). Cette vision est très éclairante pour ceux qui se confrontent à la problématique de la correspondance entre une structure syntaxique et une structure sémantique.

Gaston Gross consacre enfin son plus long chapitre (chap. 10) aux phénomènes de figement. Ceux-ci sont en effet au cœur de son entreprise : non seulement parce qu'ils constituent un thème important, voire fondateur, dans son cheminement intellectuel et dans celui de ceux qu'il a formé ; mais aussi parce qu'en fondant, à la suite de Harris, la description linguistique sur les régularités observées dans les distributions, on commence par rejeter le fantasme du signe linguistique idéalement isolé et indépendant, librement combinable avec tout le reste du lexique, pour identifier les différents emplois en fonction de la restriction de leurs contextes d'usage (restriction des paradigmes), et l'on finit par se confronter aux cas extrêmes de constructions où la liberté de combinaison tend à devenir nulle (cas limite de la restriction des paradigmes) : les constructions figées. L'auteur note très justement qu'on a affaire, avec le figement, à « un fait massif, qui doit être considéré comme une propriété définitionnelle des langues naturelles et qui a totalement échappé à la tradition grammaticale » (p. 27).

C'est également lorsqu'on confronte le modèle de Gaston Gross au cas des figements qu'on en découvre le mieux, nous semble-t-il, les limites — liées à ce qu'il implique de tracer des frontières entre catégories, sans pouvoir exprimer la liberté qu'a la langue de faire varier de façon continue la

⁷ Au contraire de l'approche de la *Lexicographie Explicative et Combinatoire* de l'école Sens-Texte, développée par Igor Mel'čuk, et ayant brillamment essaimé, pour le français, à partir du Canada francophone, sous l'impulsion d'Alain Polguère. C'est sans doute l'école qui a le plus d'affinité intellectuelle avec l'approche de Gaston Gross, par l'accent mis sur la description de l'emploi.

contrainte de restriction d'emploi liée au contexte.

Les limites de l'approche

Une conception statique du lexique

Malgré sa perception très fine des variations d'usage, Gaston Gross finit en effet toujours, dans l'optique de la constitution de dictionnaires électroniques, par résoudre les questions de description des emplois par le découpage en catégories étanches. Le découpage se fait certes au rasoir — ce qui est intellectuellement bien plus satisfaisant que lorsqu'il se fait à la hache —, mais il y a malgré tout découpage ; et cela même impose une simplification arbitraire aux possibilités d'usage de la langue.

Pour faire comprendre notre propos, commençons par préciser un postulat qui nous semble fondamental : les différents emplois du verbe *abattre* évoqués plus haut ne sont pas simplement des homonymes. La galaxie de signification qu'ils constituent n'a pas pour seule unité celle d'une étymologie commune ; elle constitue un ensemble modulaire et ramifié de sens potentiels, activés ou désactivés dans le contexte. C'est cette cohésion qui fait que le verbe *abattre* constitue une unité linguistique de la langue française, perçue comme telle par les locuteurs du français ; et le fait que cette unité se traduise parfois mieux en allemand tantôt par *erschießen*, tantôt par *fällen*, et tantôt par *abreißen*, n'enlève rien à ce fait. Cette unité est rompue par la méthode de Gaston Gross, ce qu'il assume par exemple en écrivant qu'« on ne peut pas parler d'un verbe *abattre* unique ». On fait donc l'impasse sur les sèmes intrinsèques de l'unité linguistique, renonçant par avance à toute approche sémasiologique.

On pourrait considérer qu'il s'agit là d'un moindre mal, et que la perception de cette unité est superflue dans la perspective d'une description synchronique des emplois manifestés dans des textes contemporains. Cependant, cette solidification des compartiments de la description impose d'autres contraintes : elle empêche de reconnaître d'éventuelles subdivisions plus fines ; elle empêche de percevoir le lien, dynamique, qui peut encore se faire entre différents sens ; elle empêche, enfin, de pouvoir reconnaître un emploi s'il n'est pas appliqué très strictement à la catégorie paradigmatique (classe d'objets) qui lui a été assignée à l'avance par la description. L'usage vivant de la langue ne connaît pas, lui, ces cloisons étanches. En résumé, l'approche proposée constitue une description sémantique *statique*.

Justifions cette critique. La définition des classes d'objets, chez Gaston Gross, est fondée sur la constitution de paradigmes dans le contexte de certains emplois de certains prédicats (p. 74-75). Ainsi, la classe des « arbres » est celle qui peut émerger dans le contexte de la position de complément d'objet du verbe *émonder*, du verbe *élaguer*, ou d'un certain emploi du verbe *tailler*. Cycliquement, le sens d'un certain emploi d'un prédicat est déterminé par les classes d'objet (« une classe d'objets est un ensemble de substantifs qui, dans une position argumentale donnée, détermine le sens d'un prédicat », p. 34). Le niveau de granularité auquel doit être arrêtée l'identité d'un certain emploi d'un prédicat — ou, de façon converse, d'une certaine classe d'objets — n'est pas déterminé par une règle générale, qui trouverait son propre fondement en-dehors du lexique. Pourquoi alors ne pas assumer pleinement les conséquences de l'idée distributionnaliste, en reconnaissant que tout paradigme, d'une manière générale, est créé par le contexte ; et que ce paradigme est d'autant plus limité que le contexte impose de contraintes sémantiques extérieures ? En somme, des paradigmes très génériques (ceux que Gaston Gross appelle des *hyperclasses*, comme par exemple les « concrets ») peuvent être définis par des contextes très peu contraints (exemple du complément d'un verbe générique comme *toucher*, dans un environnement contextuel très pauvre : « je touche un ... ») ; des paradigmes un peu plus restreints comme les « arbres » peuvent être définis par des contextes un peu plus contraints (« j'ai pris ma hache pour abattre un ... ») ; enfin, des paradigmes encore plus limités pourraient être construits sur la base de contextes plus restrictifs (« je vais faire une confiture avec tous les fruits que j'ai ramassés sur les

branches d'un ... »). Pour le dire dans les termes de François Rastier, le contexte crée un taxème dont il impose les sèmes génériques, et les distinctions internes à ce taxème, celles qui permettent encore un choix paradigmatique, en sont les sèmes spécifiques.

De même, pourquoi limiter la notion de classe à la catégorie des substantifs (passage de la p. 34, cité plus haut), alors même que l'on note que les différents emplois d'un substantif peuvent être délimités par des classes de contexte (distinction des différents emplois du mot *livre*, p. 84) ?

Le fait que les classes ne soient définies que pour les substantifs, l'emploi même du terme de classe d'*objet*, trahit, selon nous, la conception qui imprègne implicitement l'approche de Gaston Gross : celle de la langue comme ontologie⁸. Cette conception réifie le taxème (le paradigme créé dynamiquement par un contexte donné) : il en fait une *classe*, c'est-à-dire une sorte de contenant rigide, existant indépendamment des situations d'emploi, et s'imposant, en tant que surensemble, dans la définition des unités qui en font partie — au lieu de rester un regroupement à géométrie variable d'éléments constitutifs élémentaires de la signification des unités. Or dans l'usage réel de la langue, il n'y a pas toujours une classe des < sièges > : il y a un paradigme des choses sur lesquelles on s'assoit (qui contient certes la chaise, mais aussi, à l'occasion, la table, ou dans d'autres contextes le gros rocher plat), et un paradigme des choses que l'on achète dans un magasin de mobilier (qui contient lui aussi la chaise, mais également le placard). De même, il y a les choses qu'on attelle (les chevaux et les bœufs), les choses qu'on enfourche (les chevaux et les vélos), et les choses dont on répare la roue (les vélos et les camions).

Cette fixation du regard sur le niveau de description de la classe a des conséquences. Elle oblige Gaston Gross à constater parfois des « intersections entre classes »⁹, et à remettre en circulation, pour en rendre compte, un artefact descriptif comme la métaphore (4-8.6). On sait que l'idée de trope est d'un tel manque d'utilité pour la description synchronique de la langue qu'elle a provoqué un désintérêt général pour les études rhétoriques après que Fontanier les eut assimilées à leur description. Or le partitionnement du lexique en classes oblige à les réinventer. En effet, si (1) la catégorisation des arguments d'un prédicat ne peut se faire que par le biais des classes ; (2) certains arguments de certains prédicats recouvrent différentes classes (les véhicules et les animaux, par exemple, pour le complément d'objet d'*enfourcher*) ; (3) tout en coupant transversalement chacune de ces classes (on enfourche des vélos et pas des trottinettes, des chevaux et pas des chats) ; (4) que ces classes, enfin, ne peuvent être fusionnées sous peine de perdre l'information lexicale dont elles sont par ailleurs porteuses ; alors il ne reste effectivement plus qu'à faire appel à la vieille roue de secours de la métaphore pour expliquer les choses. Le problème pratique de ce choix est qu'il faudrait, pour être complet, inventorier une quantité illimitée de métaphores — pour rendre compte en somme des cas de tous les mots qui ont différents contextes d'emploi.

La rigidification de la notion de *paradigme* en notion de *classe* est un glissement sans doute difficile à éviter dans la perspective d'une description statique du système de la langue dans un dictionnaire électronique, où il faut bien ranger l'information quelque part¹⁰. Gaston Gross est conscient, lorsqu'il laisse parler son intuition linguistique, de la gradualité de la détermination des emplois. Il le dit

8 Bien qu'il s'en défende en déclarant que les classes ont une définition distributionnelle et non conceptuelle (« la construction des classes se fonde sur des propriétés grammaticales et non sur une classification présyntaxique », p. 75).

9 Section 4-8.6.3 : « une classe d'objets [...] accapare et pour ainsi dire "vole" des prédicats strictement appropriés à une autre classe d'objets. Comme on l'a vu, les < deux roues > sont comparés à des moyens de transport animaux : comme eux, ils prennent [...] des verbes qui désignent un mouvement (*enfourcher*) ou une position (*être juché sur*). » (p. 90-91).

10 À moins de faire une révolution qui consisterait à faire exploser les classes, *contenants* des unités linguistiques, et à redistribuer intégralement l'information dont elles sont porteuses dans des atomes syntaxico-sémantiques *contenus* dans les unités linguistiques (dont les paradigmes seraient des recombinaisons dynamiques). François Rastier a fait cette révolution sur le plan théorique, mais ces idées n'ont jamais débouché, à notre connaissance, sur la constitution de ressources linguistiques fondées dessus.

explicitement en introduisant la notion de figement : celle-ci, explique-t-il dans un sous-chapitre (10-1.6), n'est pas une alternative binaire ; un figement peut incorporer une certaine « liberté lexicale », et ces variantes sont même « plus fréquentes que le figement total » (p. 202). Les possibilités de commutation, note-t-il, sont « plus ou moins importantes », et « l'absence de paradigme n'est qu'un cas limite » (*ibid.*). Pourtant, malgré ces précautions oratoires, il ne peut éviter le piège de la catégorisation, et finit même par expliquer la notion de paradigme par celle de classe, ce qui est une belle inversion de la hiérarchie épistémique entre fait observé et artefact descriptif¹¹ !

Une illustration, peut-être anecdotique, mais nous semble-t-il tout à fait éclairante, de cette impossibilité technique à prendre réellement en compte la gradualité, se trouve dans l'exemple donné p. 206 pour justifier la nécessité de catégoriser certaines expressions comme figées : « Les structures figées, identiques [aux syntagmes libres] en termes de séquences de catégories, ont quant à elles, en position argumentale, non pas des classes d'objets mais des éléments isolés, qui ne sont susceptibles d'aucun paradigme [...]. On pourrait objecter à cette remarque que certaines locutions ont des débuts de paradigmes : *perdre (la boule, la boussole, le Nord, les pédales, etc.)*, mais ces phrases sont synonymes, alors que le sens de chacun de ces mots ne permet pas de prédire cette interprétation [...] ».

C'est pourtant bien le contraire : ce paradigme, à s'y pencher d'un peu plus près, est parfaitement cohérent, et « le sens de chacun de ces mots » *permet*, justement, « de prédire cette interprétation » ! Il a une interprétation tout à fait naturelle, il n'est pas propre à la langue française, et il est d'ailleurs tout à fait productif¹². Gaston Gross illustre par cet exemple le contraire de ce qu'il affirme, et il est intéressant de comprendre pourquoi. La seule chose qui le conduit, selon nous, à affirmer qu'ici la combinatoire est figée, est la volonté de décrire chaque emploi comme correspondant à une signification unique, littérale et concrète — ce que l'auteur formulerait sans doute plutôt comme : liée à une classe d'objets et une seule. Il faudrait ainsi que le syntagme libre « perdre la boussole » — puisque *boussole* rentre dans la classe des « outils » — ne puisse vouloir dire que : « égarer un objet utilitaire doté d'une aiguille magnétique, après l'avoir fait tomber de sa poche », et il faudrait confier à un syntagme figé de même forme de surface le sens de : « ne plus être capable d'un total discernement »¹³. Les deux emplois seraient disjoints ; ils seraient pour ainsi dire quasiment des homonymes. Pourtant, si l'on considère le cas de « perdre le Nord », il paraît assez évident qu'il n'y a qu'un seul et même emploi (« ne plus savoir où se trouve une certaine direction, qui est la direction de référence »), avec une continuité entre des emplois dans des contextes plus ou moins concrets (selon qu'on cherche effectivement son chemin dans une forêt, ou que l'on hésite sur la manière de se diriger dans la vie). On voit mal ce qui justifierait de scinder les deux emplois, à plus forte raison de les considérer comme n'ayant pas de lien entre eux. Pour la boussole, c'est le même phénomène, à une métonymie près¹⁴. Alors, pourquoi isoler artificiellement cette série d'emplois du fonctionnement normal et naturel du reste de la langue (qui permet en effet, de manière tout à fait ordinaire, d'employer des verbes concrets pour décrire des situations moins concrètes), et les enfermer dans le carcan d'un figement opaque ?

Une réflexion franco-centrée

Une autre limite de l'ouvrage de Gaston Gross est que les réflexions qu'il y fait sont parfois assez

11 Ce que Saussure appelle « l'axe paradigmatique », selon Gaston Gross, « ne traduit rien d'autre que cette réalité des langues naturelles qui veut qu'en position d'arguments on ait affaire non à des unités mais à des classes de mots » (p. 202). « Ainsi, après le verbe *manger*, on peut avoir la classe des « aliments » et non pas seulement le mot *steak* ».

12 Cette même grande dynamique métaphorique autour de l'idée de « perdre des repères spatiaux » produit régulièrement de nouvelles manifestations, comme le très à la mode « elle ne sait plus où elle habite ».

13 On s'est efforcé de trouver une paraphrase qui n'emploie pas le verbe *perdre*.

14 C'est-à-dire le fait de dire « perdre la boussole » pour « perdre la direction que permet d'indiquer l'instrument *boussole* ». Ce fait entre typiquement dans le cadre que propose de dresser Gaston Gross, en p. 92 de ce même ouvrage, pour traiter la métonymie (« établir une relation précise entre deux classes d'objets »). Ceci réglé, « perdre la boussole » est rigoureusement équivalent à « perdre le Nord ».

centrées sur le français, et pas forcément généralisables. C'est le cas par exemple lorsqu'il parle des effacements d'arguments, en disant que le sujet n'est pas effaçable (p. 54) alors que les compléments d'objet le sont (p. 55-56). Il existe des langues où le sujet n'est pas obligatoire (l'italien par exemple), alors qu'il existe des langues où un complément par défaut est très fréquemment manifesté, même lorsqu'il est sémantiquement vide — en chinois, par exemple, où *manger* (sans plus de précision) se dit « chī fàn » (litt. *manger du riz*)¹⁵.

Cela étant, cet aspect franco-centré de l'ouvrage est parfaitement explicite et clairement assumé. La première phrase de l'introduction précise bien que le premier objectif de cet ouvrage est de « présenter une description des mécanismes du français contemporain » (p. 7), et l'on peut considérer que cet objectif est brillamment atteint. Un second objectif, mentionné aussitôt après, est de « proposer une méthode d'analyse de la langue destinée au traitement automatique, vu du côté de la linguistique » ; et il se trouve, pour cet objectif, que Gaston Gross laisse au lecteur qui voudrait s'inspirer de son travail le soin de déterminer à quel niveau d'abstraction il peut généraliser les méthodes descriptives appliquées au français. Il est impossible de reprocher à Gaston Gross la myopie qui affecte parfois certains auteurs, lorsqu'ils croient donner des vérités générales sur le fonctionnement du langage, alors qu'ils surgénéralisent en fait des propriétés de leur langue. Gaston Gross ne décrit que le français, il le dit clairement, et on ne peut pas le soupçonner de le faire par méconnaissance d'autres langues (sa maîtrise de l'anglais et de l'allemand, notamment, transparait dans de nombreux exemples qu'il donne ici ou là). Il n'a tout simplement pas poussé systématiquement le travail consistant à préciser, dans chacune de ses descriptions, ce qui est propre au français et ce qui ne l'est pas.

Linguistique ou traitement automatique des langues ?

La question la plus fondamentale que l'on se pose, en fin de compte, à la lecture du *Manuel d'analyse linguistique* de Gaston Gross, est de savoir s'il s'agit d'un livre de linguistique ou d'un livre de TAL. On a le sentiment que l'auteur voudrait qu'il soit avant tout un livre de linguistique : cela transparait dans le titre, ou dans la première phrase du texte présenté au lecteur en quatrième de couverture (ce livre est une présentation « [...] d'une méthode d'analyse des mécanismes qui sont à la base du fonctionnement du français »). Et en effet, l'ouvrage est intégralement consacré à la description de phénomènes de syntaxe ou de sémantique, sans s'inféoder à une quelconque école de représentation formelle. Il parle de prédicats et d'arguments, pas d'algorithmes d'analyse ou de génération. Il évalue ses résultats par la complétude de sa couverture du lexique, pas par des taux de précision ou de rappel.

Cependant, considéré comme un livre de pure description de la langue, le *Manuel d'analyse linguistique* pourrait prêter le flanc à un certain nombre de critiques, comme la faible justification du nombre des hyperclasses¹⁶, ou du niveau de spécificité choisi dans le découpage des emplois des unités ; la vision statique qu'il impose des paradigmes linguistiques ; l'incapacité à représenter la gradualité des phénomènes de figement (alors même que celle-ci est affirmée en préambule) ; ou la difficulté à gérer les recouvrements de classes sans faire appel à l'encombrant mécanisme de la métaphore. Nous avons explicité ces critiques dans la section précédente.

En réalité, l'ouvrage de Gaston Gross est bien un livre qui traite essentiellement de phénomènes linguistiques, mais l'application informatique, qu'il ne perd jamais de vue, est le fondement qui lui impose ses contraintes de faisabilité (qui se déclinent en l'occurrence en contraintes de finitude et de complétude, dans l'état de l'art actuel des formalismes de représentation de l'information lexicale et syntaxique). La formulation la plus honnête de ce qu'est fondamentalement cet ouvrage est donc celle qu'a choisie l'auteur dans la première phrase de son introduction, déjà citée plus haut : une « méthode

15 Le plus intéressant, au-delà des contraintes syntaxiques de telle ou telle langue, est ce que nous apprennent sur leurs *topoi* culturels les valeurs par défaut qui y sont cristallisées, en creux ou en plein : en chinois, *manger du riz*, c'est tout simplement manger ; et en français, comme le fait remarquer Gaston Gross, *boire*, c'est boire du vin (p. 57).

16 Pourquoi leur nombre est-il fixé précisément à six (p. 73) ? Pourquoi pas huit, douze, ou vingt-quatre ?

d'analyse de la langue destinée au traitement automatique, vu du côté de la linguistique » (p. 7).

Conclusion

Le *Manuel d'analyse linguistique* est une contribution majeure aux travaux de description du français. Il apporte aux spécialistes de lexicographie électronique des méthodes de modélisation de la langue où la syntaxe est intégrée à la description du vocabulaire, et qui ne laisse dans l'ombre aucune partie importante du système de la langue française.

Les limites de l'approche de Gaston Gross sont d'une part qu'il est fortement centré sur le français, et que l'auteur ne donne pas toujours les clés pour pouvoir généraliser ses méthodes à d'autres langues ; et d'autre part qu'il reflète un état de la pensée linguistique qui donne une vision statique du lexique. Ces limites ne sont pas un défaut spécifique au travail de Gaston Gross, mais reflètent un état de l'art en description linguistique formelle ; et elles n'enlèvent rien au caractère monumental de ce qu'il faut appeler une somme dans le domaine de la description et de la modélisation de la langue. Il ne nous semble pas qu'il sera raisonnable, à l'avenir, de travailler sur un projet de linguistique informatique impliquant une description de la langue française sans se référer à l'ouvrage de Gaston Gross.

Par sa présentation, claire et didactique, sans pré-requis conceptuels ou terminologiques, ce livre se positionne comme un ouvrage de référence. Il n'est pas lié à un formalisme ou à une technologie, et s'ancre donc plus profondément que l'écume des vagues d'avancées techniques que le TAL réalise ces dernières années — sa bibliographie est d'ailleurs constituée d'ouvrages bien plus que d'articles. Il mérite donc, à notre avis, de devenir une référence durable présente dans la bibliothèque de tout spécialiste de modélisation du lexique et de la syntaxe.

Pascal Vaillant

LIM&Bio

Université Paris 13, Sorbonne Paris Cité

74, rue Marcel Cachin

93017 Bobigny cedex

vaillant@univ-paris13.fr