

PRÉSENTATION

Jean-Pierre DESCLÉS
Sorbonne Université

L'examen des différentes approches de la syntaxe dégage deux grands paradigmes qui se distinguent parfaitement lorsqu'on examine les représentations formelles construites et utilisées par les linguistes¹.

Un premier paradigme, qualifié de syntagmatique, exploite directement la structure linéairement ordonnée des phrases et des syntagmes qui les décomposent. Ce paradigme a pris son ancrage dans les travaux du distributionalisme (développé par L. Bloomfield, entre autres), ce qui a conduit à des représentations formelles d'organisations hiérarchisées de classes d'équivalence d'unités substituables les unes aux autres au sein d'une même classe (des morphèmes, des mots, des syntagmes, des phrases entretenant des relations paraphrastiques). L'ouvrage *Methods in Structural Linguistics* de Z. Harris (1961) est une mathématisation du distributionalisme nord-américain. Comme on le sait, il a ouvert la voie à la Grammaire Générative de N. Chomsky (avec ses différentes versions successives) qui a réussi à dépasser le cadre trop restreint du distributionalisme nord-américain et celui d'un certain « structuralisme étroit » européen. La Grammaire Générative a travaillé, avec les Grammaires syntagmatiques, sur des représentations structurées comme des « arbres syntagmatiques » (en fait des arborescences doublement orientées où l'ensemble des successeurs immédiats d'un nœud non terminal est totalement ordonné), puis, avec les diverses grammaires transformationnelles, qui mettaient en relation des arbres syntagmatiques

¹ La plupart des auteurs évoqués dans cette présentation ont des références précises données dans les différentes références bibliographiques des articles de ce numéro.

profonds et de surface reliés par des transformations opérant sur les arbres². D'autres approches syntaxiques, comme GPSG (G. Gazdar) ou les TAG (A. Joshi)³ développent des analyses qui opèrent toujours sur et avec des arbres syntagmatiques, auxquels sont souvent ajoutées des informations (par exemple des jeux de traits sémantiques) et des contraintes entre les nœuds des arbres.

Le second paradigme, représenté largement par les Grammaires Catégorielles, s'est ancré sur les réflexions philosophico-logiques de E. Husserl et sur les analyses logico-sémantiques de l'Ecole polonaise, où les notions de catégorème, de syncatégorème et de foncteur jouent un rôle structurant. Cette approche, très différente du paradigme syntagmatique, opère également avec des « arbres d'opérateurs » et tisse des liens pertinents entre les formalismes des Grammaires Catégorielles (GC) et ceux de la Grammaire de dépendance (D. Hays), des Grammaires de dépendance de l'Ecole actuelle de Prague, la théorie des actants et des stemmas de L. Tesnière à la base syntaxique du modèle « sens-texte » de I. Mel'čuk, des représentations applicatives des phrases traitées par la Grammaire Applicative de S.K. Shaumyan et des décompositions entre différents types d'opérateurs et d'opérandes dans la Grammaire d'opérateurs de Z. Harris (1982). La notion d'application d'opérateurs (également appelés foncteurs dans certaines approches) à des opérandes de différentes catégories (ou types), caractérise ce paradigme et traverse différentes modélisations du traitement syntaxique (et parfois aussi de la construction de représentations sémantiques) des langues, et établit, en même temps, des ponts très solides entre d'un côté, la linguistique (théorique, descriptive et traitement automatique des langues), et d'un autre côté, les mathématiques (arithmétique, algèbre et théorie des catégories), la logique (λ -calcul, quantification restreinte analysée par R. Montague, logique combinatoire, déduction naturelle et logique linéaire), l'informatique (programmation fonctionnelle, sémantique des langages de programmation), la philosophie du langage et la sémiotique.

Les Grammaires Catégorielles (GC) remontent aux travaux plus sémantiques de S. Leśniewski, et aux travaux nettement plus syntaxiques de K. Ajdukiewicz puis, légèrement plus tard, à ceux de Y. Bar-Hillel, regroupés dans ce qu'on appelle les Grammaires AB, souvent présentées dans les manuels de linguistique, par exemple celui de J. Lyons⁴. L'idée principale

² Remarquons que le « Programme minimaliste » actuel de Chomsky présente d'autres caractéristiques qui le distinguent nettement des différentes présentations antérieures et successives de la Grammaire Générative où les arbres syntagmatiques (et les « formes logiques » de la logique classique) ont toujours tenu un rôle central.

³ A. Abeillé, *Les nouvelles syntaxes. Grammaires d'unification et analyse du français*. Paris : Armand Colin, 1993.

⁴ J. Lyons, *Linguistique générale. Introduction à la linguistique théorique*. Paris : Librairie Larousse, 1970, 175-178.

des Grammaires AB est la suivante : associer aux catégories syntaxiques usuelles (phrase, noms, verbes, adjectifs, adverbes, prépositions ...) de la grammaire traditionnelle, des notations réglées, appelées types syntaxiques, et calculer, au moyen de ces types, la bonne formation syntaxique de séquences de mots, chacun d'entre eux étant catégorisé par un type syntaxique qui a pris une forme analogue à un nombre fractionnaire 'p/q' ; le calcul s'effectue par des simplifications tout à fait analogues aux simplifications des fractions numériques. A partir de 1960, la « vague » importante de la Grammaire Générative et des grammaires formelles, présentées comme des systèmes de réécriture (avec une hiérarchie pertinente, dite de Chomsky, de la complexité des grammaires : markovienne, non contextuelle, contextuelle, récursive), a submergé l'approche calculatoire de la syntaxe de la GC, qui n'a enregistré, après 1965, aucune contribution importante pendant plus de vingt ans. L'intérêt pour les GC s'est pourtant réveillé autour des années 1990, avec notamment le développement du « Calcul de Lambek » qui a su exploiter les propriétés algébriques du calcul sur les types catégoriels, en étendant le calcul initial d'Ajdukiewicz-Bar-Hillel (AB) et en analysant des phrases plus complexes que le simple formalisme catégoriel AB ne réussissait pas à capter. Il est alors apparu de plus en plus clairement que les types syntaxiques des GC n'étaient que des variations notationnelles du système des « types fonctionnels » que le logicien A. Church avait déjà formulé pour l'étude mathématique des fonctions entre ensembles. Les types syntaxiques (puis également, dans certaines études, les types sémantiques) peuvent désormais être appréhendés sous la forme de types fonctionnels dont les instances sont des opérateurs (des unités linguistiques comme les verbes, les conjonctions, les déterminants adjectivaux et adverbiaux) qui s'appliquent à des opérands (comme des noms, des phrases ou d'autres opérateurs linguistiques) ayant des types compatibles avec ceux des opérateurs. Une troisième étape du développement des GC est ensuite apparue avec d'un côté, une étude plus systématique des liens entre les GC et les calculs modaux et surtout avec les nouvelles contraintes apportées par la logique linéaire de Jean-Yves Girard, et d'un autre côté, avec une complexification des GC où l'on s'autorise des compositions et des transformations de types syntaxiques en relation directe avec la Logique Combinatoire typée du logicien américain Haskell B. Curry ; cette logique doit être pensée comme une logique d'opérateurs quelconques de différents types fonctionnels, composables ou transformables par des opérateurs abstraits, appelés combinateurs, et non pas comme un système sémiotique très général de combinaisons de symboles quelconques. La reconnaissance des liens entre GC et certains combinateurs aboutit à des grammaires catégorielles étendues.

Ce numéro spécial portant sur les Grammaires Catégorielles présente les avancées actuelles de ce second paradigme d'étude de la syntaxe des langues. Les auteurs sont des informaticiens-linguistes, spécialistes du TALN, et des logiciens intéressés par l'analyse logique des langues.

L'article « Brève généalogie des Grammaires Catégorielles » retrace l'histoire du formalisme des GC depuis son enracinement philosophico-logique jusqu'à la reconnaissance des liens étroits entre les GC étendues et les combinateurs de la logique combinatoire aboutissant à la Grammaire Catégorielle Combinatoire (GCC) de M. Steedman et à la Grammaire Catégorielle Combinatoire et Applicative (GCCA) de Biskri et Desclés. Avec la discussion linguistique et leur traitement par des GC de plusieurs exemples bien choisis du français, avec subordination et coordination, Ismaïl Biskri montre, dans sa contribution, comment les types syntaxiques se composent et se transforment dans des calculs syntaxiques détaillés de phrases complexes. Dans leur article, Juyeon Kang et Jungyeul Park utilisent le formalisme catégoriel étendu (notamment pour prendre en compte l'approche catégorielle des quantificateurs) pour résoudre des problèmes qui se posent dans des secteurs industriels (comme l'extraction de connaissances ou la formulation de « langues contrôlées ») qui sortent du champ de la seule recherche théorique et universitaire. La présentation générale des GC est largement et heureusement complétée par l'article de Béatrice Godart-Wendling qui présente, entre autres, le dépassement théorique et descriptif du « Calcul de Lambek », en présentant les liens qui ont été établis par Joachim Lambek entre la théorie algébrique des « pré-groupes » et les GC. La contribution de Christian Retoré présente les rapports formels entre le Calcul de Lambek et la logique, appréhendée sous la forme technique du calcul des séquents (une extension de la déduction naturelle) et de la logique linéaire, pour rejoindre, dans certaines analyses linguistiques prises comme exemples illustratifs, certains principes du programme minimaliste de N. Chomsky. Ce rapprochement témoigne que l'opposition entre le paradigme syntagmatique et le paradigme catégoriel n'est peut-être pas aussi tranchée qu'elle semble paraître. I. Biskri, pour l'arabe, et J. Kang et J. Park, pour le coréen, montrent que le formalisme des GC étendues s'applique de façon pertinente à des langues non indo-européennes, ce qui tend à justifier le caractère universel de l'analyse en termes d'opérateurs et d'opérandes de différents types des expressions linguistiques.

Les types fonctionnels de Church sont essentiellement des types d'opérateurs appliqués à des opérandes, qui peuvent être, par ailleurs, des opérateurs et pas seulement des opérandes absolus. Ils constituent le formalisme de base des GC, qui, elles, ajoutent, au type de chaque opérateur, une orientation vers la position syntagmatique de l'opérande attendu, afin de pouvoir relier plus directement la représentation applicative « opérateur/opérande » aux organisations syntagmatiques des phrases qui ordonnent les unités linguistiques selon un ordre total. Les catégories syntaxiques, dont les rôles opératoires sont décrits par les types fonctionnels, entretiennent des rapports analogiques non triviaux dans ce qui est appelé « l'isomorphisme de Curry-Howard » (déjà constaté par Curry en 1958), avec le calcul propositionnel positif (sans négation). Cela conduit à voir dans les types syntaxiques des

sortes de spécifications propositionnelles, chargées de contrôler la bonne correction syntaxique. Cet isomorphisme « types fonctionnels – propositions (positives) » renforce les relations, qui doivent être encore être sérieusement approfondies, entre l’analyse linguistique, l’analyse logico-philosophique et le traitement automatique des langues. Ce dernier doit chercher à fonder ses procédures d’analyse sur des conceptualisations théoriques hautement argumentées du point de vue de la linguistique en vue de comprendre la réelle complexité de l’activité langagière que les systèmes sémiotiques des diverses langues manifestent. En présentant des analyses syntaxiques précises de phrases de plus en plus complexes, le formalisme des GC étendues tend à jouer un rôle de « guide syntaxique » qui, par le biais de calculs directement implémentables sur des machines informatiques, fournissent des sorties qui sont, en même temps, des entrées pour les processus de construction de représentations sémantiques. Les articles de ce numéro spécial de la revue *Verbum* contribuent, nous l’espérons, à cet effort interdisciplinaire où linguistes, logiciens et informaticiens doivent savoir s’écouter et échanger leurs techniques, pour être capables de mieux identifier les problèmes à résoudre et collaborer à la construction d’instruments d’analyse efficaces du langage et des langues.

Paris, 5 décembre 2018