



HAL
open science

La formalisation : pour quoi faire ?

Béatrice Godart-Wendling

► To cite this version:

Béatrice Godart-Wendling. La formalisation : pour quoi faire ?. De la diachronie à la synchronie et vice versa, Mélanges offerts à Annie Bertin., Presses Universitaires Savoie Mont Blanc, pp.405-421, 2021. hal-03290296

HAL Id: hal-03290296

<https://hal.science/hal-03290296>

Submitted on 19 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LA FORMALISATION : POUR QUOI FAIRE ?

BÉATRICE GODART-WENDLING

LABORATOIRE MODÈLES, DYNAMIQUES, CORPUS –
UMR 7114, UNIVERSITÉ PARIS NANTERRE

Cet article a pour objectif de répondre aux interrogations des linguistes non logiciens portant sur l'intérêt de formaliser les langues. Ce faisant, il entre en écho avec le questionnement de mon amie et collègue Annie Bertin qui, lorsque le temps nous le permettait, se montrait véritablement curieuse vis-à-vis de l'apport effectif de la formalisation pour la linguistique. Afin d'illustrer mon propos, j'examinerai l'approche formelle des grammaires catégorielles, car celle-ci représente actuellement le paradigme dominant en matière de formalisation et son choix – contrairement au mouvement chomskyen – de calculer conjointement la bonne formation syntaxique des phrases et leur interprétation sémantique ne peut que retenir l'attention des linguistes. Mon analyse portera non seulement sur la façon de catégoriser les éléments internes à la phrase, mais également sur les règles mises en jeu. Je montrerai ainsi que l'approche catégorielle s'inscrit dans la continuation de la tradition linguistique en termes de « parties du discours » et soulignera la difficulté à élaborer une grammaire universelle ne recourant qu'à un nombre très réduit de règles. Cette étude analysera aussi le type de calcul sémantique mis en jeu et cette partie de l'article sera l'occasion de montrer les limites de ce que l'approche catégorielle estime être une interprétation de la signification des phrases. Au terme de ce parcours, le linguiste disposera des outils lui permettant d'évaluer si le recours à la formalisation peut lui apporter un éclairage nouveau dans le travail de description et d'analyse des langues qui est le sien.

1. Un idéal de simplicité

Si l'idée et la volonté de représenter les langues naturelles par le biais de leur formalisation ne datent pas de la seconde moitié du XIX^e siècle, comme en témoigne le désir de Leibniz d'établir un *calculus ratiocinator*

couplé à la langue idéale qu'aurait été la Caractéristique Universelle, il demeure toutefois que ce projet ne commença à prendre corps qu'avec les travaux de George Boole¹ (1854) et la critique qui s'ensuivit par Gottlob Frege (1848/1925). Le pas décisif que ces deux logiciens effectuèrent fut de placer les notions de « fonction » et d'« argument(s) » au cœur de l'analyse grammaticale. Sans revenir sur leurs divergences, la forme la plus aboutie que prit cette démarche logiciste est exprimée dans la *Begriffsschrift* (1879) où Frege argumente que « la logique s'est jusqu'ici toujours rattachée trop étroitement à la langue et à la grammaire. En particulier, je crois que le remplacement des concepts *sujet* et *prédicat* par *argument* et *fonction* fera ses preuves à la longue. » (1999 [1879]: 8-9) Exprimée dans le détail, cette nouvelle conception du découpage syntaxique des phrases revient à poser que l' :

on peut envisager de décomposer les propositions affirmatives comme les équations, les inéquations, et les expressions analytiques, en deux parties dont l'une est fermée sur soi et dont l'autre réclame un complément, est insaturée. On analysera par exemple la proposition

« César conquiert les Gaules »

en « César » et « conquiert les Gaules ». La seconde partie est insaturée, elle traîne une place vide avec elle, et ce n'est qu'après avoir rempli cette place par un nom propre ou une expression qui représente un nom propre qu'on voit naître un sens fermé sur lui-même. J'appelle ici encore fonction la dénotation de la partie insaturée. Dans ce cas, l'argument est César (1971 [1891]: 91).

Cette approche, qui revient à coopter dans le domaine de la linguistique les concepts mathématiques de *fonction* et d'*argument*, a pour conséquence de modifier, en les simplifiant, les contenus définitionnels de ces concepts, puisqu'il ne sera par exemple plus question de calculer la limite ou la tangente d'une fonction. Le seul noyau définitionnel conservé est qu'une fonction s'applique à un argument pour donner un résultat, comme le verbe s'applique à un nom pour donner une phrase.

Face à ce nouveau découpage de la phrase, la difficulté est d'établir un formalisme apte à représenter des phrases plus complexes que la simple structure « nom + verbe » [i.e. fonction_{verbe} (argument_{nom})]. L'école polonaise de Lvov-Varsovie, et plus particulièrement l'article « Die syntaktische Konnexität » de Kazimierz Ajdukiewicz (1935), réalisera cette transformation en proposant de noter la fonction grâce à une

1 Pour plus de détails, cf. Godart-Wendling (1996).

barre de fraction ; attribuant respectivement au dénominateur le rôle d'indiquer l'argument et au numérateur la tâche d'inscrire le résultat de la fonction lorsque celle-ci est saturée par l'argument qu'elle appelle. Plus précisément, Ajdukiewicz empruntera à la théorie de son condisciple Stanisław Leśniewski l'idée qu'il existe deux catégories sémantiques de base (*der Satz* (la phrase) notée s et *der Name* (le nom) noté n) à partir desquelles on peut établir des catégories de foncteur de plus en plus complexes. Ainsi, un verbe qui forme une phrase s à partir d'un nom n sera représenté par le foncteur : $\frac{s}{n}$. Un article (défini ou indéfini) qui permet de constituer un nom n (en fait un groupe nominal) lorsqu'il a pour argument un nom sera étiqueté grâce à l'index fractionnel $\frac{n}{n}$. Un adjectif qui conduit à former un verbe (i.e. un groupe verbal) lorsqu'il est appliqué à un verbe (pour lequel nous avons vu que sa catégorie est $\frac{s}{n}$) relèvera du foncteur complexe :

$$\frac{\frac{s}{n}}{\frac{s}{n}}$$

À cette catégorisation initiale de chaque mot composant une phrase, Ajdukiewicz adjoint une unique règle, calquée sur le calcul arithmétique de la simplification des fractions², ayant pour objet de vérifier la bonne formation syntaxique et sémantique de la phrase. Cette règle, qui s'applique sur les mots de la phrase réordonnés selon la notation polonaise (le foncteur principal est situé à gauche de ses arguments et cette disposition est réitérée s'il existe des foncteurs secondaires), énonce que si l'on trouve, en lisant toujours la séquence de la gauche vers la droite, une fraction dont le dénominateur est identique à la fraction qui la suit, alors on peut simplifier les deux fractions en les remplaçant par le numérateur de la première fraction. Cette règle permet par exemple de montrer que la séquence « Annie rêve » (réordonnée « rêve Annie », puisque la notation polonaise revient à procéder de la catégorie la plus complexe vers la plus simple) est une phrase syntaxiquement et sémantiquement bien formée :

rêve Annie

$$\frac{\frac{s}{n}}{n}$$

s

L'application de la règle permet en effet d'éliminer le n du dénominateur de « rêve » et le n du numérateur affecté au nom « Annie » et d'obtenir

2 Qui a pour forme $(\frac{b}{a} a = b)$. Soit, par exemple : $\frac{3}{2}$ multiplié par $2 = 3$.

s; signifiant ainsi que l'on est bien en présence d'une *Satz*, c'est-à-dire d'une phrase.

La simplicité du système catégoriel (puisque chaque mot se voit affecté d'une catégorie) est donc à son comble: il suffit de deux catégories sémantiques de base *n* et *s* pour catégoriser tous les mots d'une langue et d'une unique règle de simplification.

Trois remarques doivent toutefois être émises. La première porte sur le fait que cet algorithme a une visée universelle, si bien qu'Ajdukiewicz n'exclut pas qu'il faille définir d'autres catégories de base pour rendre compte de langues qui ne possèderaient pas de nom et/ou de verbe. L'important, pour l'idéal de simplicité qu'incarne cet algorithme, est qu'il puisse s'appliquer sur un nombre réduit de catégories de base propres à la langue considérée.

La seconde précision, qui d'un point de vue épistémologique est importante, est que ce découpage formel en catégories s'inscrit dans la continuation de la tradition grammaticale en termes de parties du discours, car la répartition des mots en catégories s'effectue – conformément aux principes de la Grammaire Pure de Husserl (1962 [1913]) – par le biais de la substitution qui seule est à même, pour ce philosophe, de préserver la « signification unitaire » de la phrase (qui servira de support aux tentatives d'interchangeabilité). Autrement dit, à partir de la phrase « Annie rêve », « rêve » pourra être commuté avec « vrombit » ou « clignote », conduisant ainsi à réunir sous la catégorie $\frac{s}{n}$ tous les verbes intransitifs (les verbes transitifs correspondent, quant à eux, à la catégorie $\frac{s}{mn}$), même si les phrases résultantes semblent sémantiquement suspectes (cf. Godart-Wendling 1999). Il s'ensuit qu'il serait trop simpliste de penser que toute approche formelle constitue une rupture franche vis-à-vis des théories non formalisées qui la précèdent, puisque – comme en témoigne le cas des grammaires catégorielles – il peut s'agir d'une réélaboration sous une forme mathématisée de concepts qu'elle leur emprunte. L'apparition d'une théorie formelle n'introduit donc pas forcément une discontinuité (Kuhn 1962) avec les théories descriptives déjà existantes, car son mode de formation peut se limiter à un changement consistant à représenter autrement une notion ou un concept déjà en usage.

Mais, et ceci sera notre dernière remarque, en quoi consiste la dimension sémantique de cette première grammaire catégorielle qui a pour fondement, ainsi qu'y insiste Ajdukiewicz, qu'« a language is not unequivocally characterized solely by its vocabulary and the rules of its syntax; it is also characterized by the way meaning is coordinated with its words and expressions » (1978 [1934]: 39)? La principale

difficulté que soulève ce type de formalisation réside en effet dans son pari non tenu de calculer conjointement la grammaticalité et la bonne formation sémantique des phrases, car les seules séquences de mots que cette syntaxe réussit à détecter comme étant sémantiquement déviantes sont celles qui relèvent du non-sens (« Peut être Annie si sera cependant rêve » en est un exemple). L'absurde, présent dans les séquences « Annie vrombit, Annie clignote », est par contre jugé sémantiquement cohérent et donc pourvu de sens.³ De fait l'attribution d'une catégorie à un mot demande de connaître au préalable sa signification, car face à une langue inconnue le linguiste-formaliste serait bien en mal de déterminer les liens de dépendance que les mots entretiennent entre eux (il serait donc impossible de déterminer ce qu'il faut noter dans le numérateur et le dénominateur de la fonction) et le réordonnement en notation polonaise qu'exige cet algorithme serait de même entravé, puisqu'il nécessite d'être au clair avec la structure sémantique de la séquence de mots donnée.

2. Quand la machine s'emballe

2.1. Le passage d'une unique règle à cinq règles

Si d'un point de vue grammatical la syntaxe unidirectionnelle⁴ d'Ajdukiewicz s'avère déjà assez puissante, puisqu'elle permet de rendre compte des liens que toutes les parties du discours entretiennent entre elles tout en rendant justice aux phrases plus complexes qui contiennent des conjonctions de coordination ou de subordination (catégorisés $\frac{S}{SS}$ puisqu'elles forment, à partir de deux phrases, une nouvelle phrase), les syntaxes catégorielles ultérieures se donneront pour but de respecter l'ordre de surface des phrases (abandon du réordonnement induit par la notation polonaise) et de rendre compte d'une plus grande variété de faits empiriques (la possibilité que les mots soient pluri-catégorisés, les liens de dépendance discontinus ou croisés, les ellipses...).

Ainsi il reviendra à Yehoshua Bar-Hillel (1953) de supprimer la détermination intuitive des catégories en recourant à la méthode distributionnelle définie par Zellig Harris (1951). Dans ce cadre, les catégories de foncteur ne marqueront plus, dans leur dénominateur,

3 Ce type de théorie n'a bien évidemment pas vocation à rendre compte des énoncés figurés.

4 Ce qualificatif d'« unidirectionnel » provient du fait que l'algorithme (*i.e.* la règle de simplification) procède toujours de la gauche vers la droite.

des liens de dépendance syntactico-sémantique, mais la trace de l'environnement syntaxique immédiat du mot considéré.⁵ Il s'ensuit qu'une expression qui peut se présenter dans des contextes syntaxiques différents se verra attribuer une liste de catégories (appelée « dictionnaire ») correspondant à ses différents environnements. De plus, afin de respecter l'ordre d'apparition des mots intervenant dans les séquences à tester, Bar-Hillel proposera un algorithme bidirectionnel adjoignant à la règle gauche-droite d'Ajdukiewicz ($\frac{x}{y}$ suivi de y se simplifie en x) sa converse droite-gauche ($\frac{y}{x}$ précédé de x se simplifie en y) et définira une méthode déductive de détermination de la structure des phrases et des catégories affectées aux mots, telle que :

- a)- il faut tout d'abord lister sous chaque mot de la phrase les catégories syntaxiques possibles.
- b)- Puis appliquer, n'importe où dans la phrase et tant que cela est possible, une des deux règles de simplification.

Grâce à cette procédure, les deux règles procèdent à des simplifications successives qui éliminent des possibilités de catégorisation et de structuration de la phrase. Il devient également possible de déterminer les « constituants immédiats » qui participent à la structuration des phrases. Ainsi, l'application des règles conduit à mettre mécaniquement à plat que les constituants de la phrase « Poor John sleeps » sont « poor John » et « sleeps » (et non « poor » et « John sleeps »), si bien que la structure globale de cette séquence de mots correspond au parenthésage [(Poor John) sleeps].

Bien que ce modèle théorique soit plus pertinent que celui d'Ajdukiewicz pour l'étude des langues naturelles (puisque'il ne bouleverse plus l'ordre des mots), son auteur (1960) en arriva toutefois à la conclusion qu'il était particulièrement problématique, car il engendrait un très grand nombre de dérivations dont certaines étaient « très peu plausibles » ou « fort douteuses » syntaxiquement et/ou sémantiquement. Ainsi, la phrase simple « Paul pensait que John dormait profondément » qui *a priori* ne présente pas d'ambiguïté donne lieu à quatre possibilités de dérivations (1960: 81); mettant ainsi en évidence que la procédure de Bar-Hillel engendre des cas de « fausses ambiguïtés ». Si l'on joint, à cette prise

5 Ce faisant, la dimension sémantique n'intervient plus dans la détermination des catégories.

de conscience de la prolifération des dérivations, la nécessité d'étendre sans fin la liste des catégories affectées aux mots afin de pouvoir rendre compte de tous leurs contextes d'usage, on comprend que Bar-Hillel ait pu écrire, avant d'adhérer au modèle transformationnel chomskyen, que cette façon de formaliser les langues devait être abandonnée puisqu'elle produisait un nombre infini de dérivations concurrentes.

Le mathématicien Joachim Lambek prit alors le relais en modifiant la procédure d'attribution des catégories et en adjoignant aux deux règles déjà existantes trois nouvelles règles (*l'associativité*, la *montée* et la *division*) ayant pour fonction soit d'éliminer les cas de fausse ambiguïté soit d'agir sur la (ou les) catégorie(s) (appelées «type(s)» dans cette théorie) ayant été initialement affectée(s) aux mots. Abandonnant la méthode distributionnelle, Lambek définit les types grâce à une procédure récursive mettant en œuvre «trois opérations formelles : la multiplication, la division à gauche (notée \backslash) et la division à droite (notée $/$)» (1961 : 166). Nul n'est besoin d'entrer dans le détail de cette procédure, car les types simples ou plus complexes qui étiquettent les mots se comprennent aisément. Ainsi, «Annie rêve» se catégorisera :

Annie rêve

n n\s

et grâce à la règle de simplification des fractions, on obtient toujours s .

De fait l'apport de cette nouvelle façon de représenter les types est qu'elle attribue à la fonction elle-même (dotée maintenant de la double directionnalité \backslash et $/$) le rôle d'indiquer l'ordre des mots. La conception même de la notion de règle se trouve également précisée dans ce modèle connu sous le nom de «calcul syntaxique», car comme s'en expliquera Lambek quelques années plus tard, il s'agit de défendre l'idée que «la production et la reconnaissance du discours impliquent des calculs – bien que ce soit à un niveau subconscient – et que les règles grammaticales, que le locuteur a maîtrisées, même s'il est incapable de les expliciter, ressemblent aux axiomes et aux règles d'inférence de la logique mathématique» (2001 : 21). Dans cette perspective, la nouvelle règle dite de *l'associativité*:

$$(x \backslash y) / z \leftrightarrow x \backslash (y / z)$$

contribue à réduire les possibilités de calcul, puisqu'elle signifie qu'une fonction possédant deux arguments de part et d'autre peut s'associer avec eux dans n'importe quel ordre. La création de cette règle a donc pour conséquence de rendre caduque la problématique en termes de constituants immédiats et d'éliminer ainsi un certain nombre de

structures faussement ambiguës. Mais la véritable innovation de ce calcul fut de créer des règles permettant d'assigner aux mots des types flexibles. La règle de *montée* :

$$x \rightarrow y / (x \setminus y) \text{ et } x \rightarrow (y / x) \setminus y$$

introduite dans le but de rendre compte de l'interchangeabilité dans certains contextes entre un nom (tel que *John*) et un pronom (sous ses formes *he* et *him*) est ainsi représentative de ces règles donnant lieu à des changements de types, puisqu'elle autorise qu'une expression de type n puisse endosser, en fonction de son contexte, les types $s/(n \setminus s)$ et $(s/n) \setminus s$ normalement affectés aux pronoms.

Mais ce faisant, Lambek attribue à la notion de «type» des propriétés que ne possédait pas la notion de «partie du discours». En effet, chaque expression de la langue se voit dans le calcul de Lambek affectée de son type syntaxique le plus probable, et celui-ci est dans un premier temps stocké dans le dictionnaire des types. Considéré dans le contexte d'une phrase, ce type provisoire pourra alors évoluer sous l'action des règles et endosser successivement de nouveaux types si son type de départ bloque la dérivation. Il s'ensuit l'idée nouvelle qu'un type peut être pensé sous la forme d'une «famille de types équivalents» reliés entre eux grâce aux règles du calcul. En conceptualisant ainsi la notion de type comme une structure instanciée, celle-ci devient alors apte à gérer – contrairement à la notion de partie du discours – la structure des phrases sans avoir à définir de composant purement structurel. En effet, on obtient ce que les années 80 appelleront une conceptualisation «flexible» de la structure des phrases, où les phrases ne sont plus associées à une unique structure rigide en constituants, mais à un ensemble de représentations simultanées (et sémantiquement équivalentes) qui reflètent les différentes réductions syntaxiques que l'on peut opérer sur les phrases. La conséquence d'une telle approche est que le contenu informationnel de la structure syntaxique en constituants est inséré dans la structure interne des types de telle façon que les règles de transition de types préservent cette information. La notion de «structure syntaxique autonome en constituants» devient donc inutile dans un système catégoriel flexible et est supplantée par ce que l'on appellera «l'autonomie des types». Si un des intérêts théoriques de cette perspective est, ainsi que l'explique Moortgat (1988a: 57), qu'elle permettra aux grammaires ultérieures de se dispenser d'ajouter des «principes de réajustement non triviaux» pour gérer la relation entre les dimensions syntaxique et sémantique, elle conduit également à souligner comment le processus graduel de formalisation a conduit à éloigner définitivement la notion de type de celle de partie du discours qui avait permis son émergence.

Si l'on obtient ainsi dès la fin des années 50 une gestion particulièrement originale de l'ordre des mots, la difficulté dont sont porteuses ces règles est qu'elles ne permettent qu'une approche strictement concaténative respectant l'ordre linéaire des expressions composant les phrases. En conséquence, elles se révèlent inopérantes à rendre compte des phénomènes de discontinuité, d'extraction, de dépendances croisées ou nichées en langues. En effet, en ne mettant en jeu que des types pouvant être construits à l'aide des trois opérations que sont le produit et les deux slashes, ces règles s'avèrent juste aptes à résoudre des structures en attente d'un type situé à droite ou à gauche. Elles ne permettent pas de démontrer la bonne formation grammaticale de phrases mettant notamment en jeu :

- un verbe à particule séparable (*He looked it up*) ou un morphème discontinu (la négation *ne... pas* en français) ;
- des phénomènes d'extraction (*the book I filed without reading*) ;
- des constructions à vide parasite (*Myrtille aime sortir, mais pas Arthur*) ;
- les dépendances croisées des verbes dans les complétives en néerlandais (... , *omdat Alice de Soepschildpad wil plagen*) ou la structure nichée des compléments dans les subordinées allemandes (... , *ob ich dich heute sehe*).

2.2. L'explosion du nombre de règles

En s'attaquant à ces cas récalcitrants tout en conservant les règles du calcul syntaxique de Lambek, les années 80 s'inscrivirent dans un processus cumulatif des connaissances témoignant de la constitution d'un « paradigme » catégoriel (cf. Kuhn 1962 : 139). La difficulté à résoudre pour ce qui fut appelé les grammaires catégorielles « généralisées » est de rendre compte des phénomènes de discontinuité en langue tout en maintenant un principe d'adjacence qui pose que les règles catégorielles « peuvent seulement s'appliquer à des entités qui sont linguistiquement réalisées et adjacentes » (Steedman 1985 : 533). La raison du maintien de ce principe est que les théoriciens souhaitent pouvoir calculer conjointement la grammaticalité et la bonne formation sémantique des phrases en adoptant une sémantique compositionnelle où la signification d'une expression est déterminée par la signification de ses parties et par la façon dont elles sont combinées.

La stratégie qui sera alors mise en œuvre pour ce travail de « nettoyage » (Kuhn, 1962 : 46) consistera à augmenter le nombre de catégories de

base et/ou à définir de nouvelles règles. Concernant le domaine strict des règles, trois voies ont été explorées : 1)- accroître le pouvoir récursif de la grammaire ; 2)- augmenter les possibilités de directionnalité et/ou autoriser des permutations ; 3)- jouer sur les processus de contraction ou d'expansion⁶ afin de permettre à la fonction de s'appliquer à un nombre variable d'arguments.

Ainsi, s'inspirant de la règle récursive de *montée*, qui autorise de nombreuses dérivations pour une même séquence en permettant à un foncteur et à son argument d'être alternativement « montés » tout en laissant le reste du schéma dérivationnel inchangé, Anthony Ades et Mark Steedman (1982) ont créé la règle récursive de « composition généralisée »⁷ qui, en conférant à la grammaire « a full context-sensitive power », offre la possibilité de résoudre les phénomènes de dépendances non liés. Une autre voie entreprit de conjuguer l'action des règles de composition et de division, mais le problème auquel toutes ces extensions donna lieu fut une extension excessive du pouvoir formel de la grammaire, si bien que celle-ci ne réussit plus à discriminer les séquences déviantes grammaticalement de celles qui étaient acceptables par les locuteurs.⁸

La solution consistant à élaborer des règles qui généreraient la directionnalité de façon à formaliser l'ordre variable des mots et les phénomènes de discontinuité a été essentiellement travaillée par Michael Moortgat (1988a) et Mark Steedman (1991a). Ainsi, Moortgat a proposé une forme « disharmonique⁹ de composition » (qui s'énonce : $(y/z) (y/x) \rightarrow x/z$) pour gérer les problèmes de discontinuité et de « bracketing paradoxes » et Steedman a réfléchi sur des variantes directionnelles de la règle de montée

-
- 6 Les règles définies par Ajdukiewicz et Bar-Hillel ne réalisent que des contractions, puisqu'elles compactent deux catégories pour n'en donner qu'une seule. Par contre, la règle de montée de Lambek ($x \rightarrow y/(x)y$) et $x \rightarrow (y/x)y$) réalise une expansion du type de départ en le complexifiant.
- 7 Afin de ne pas alourdir inutilement l'exposé, nous avons fait le choix de ne pas indiquer la représentation formelle des règles qui nécessiteraient pour leur compréhension de nombreux commentaires. Notre but n'est pas ici d'analyser le contenu de ces règles, mais de préciser les nouveaux phénomènes linguistiques qu'elles permettent de résoudre et les conséquences théoriques qui résultèrent de leur élaboration.
- 8 Pour n'en donner qu'un exemple, Marie-Ange Légeret (1992 : 111) a montré que la règle de *composition mixte* est trop puissante, puisqu'elle reconnaît comme étant une phrase la structure inacceptable en anglais « article – verbe intransitif – nom commun ».
- 9 Une règle est dite « disharmonique » lorsqu'elle met en œuvre des foncteurs qui ont des slashes orientés selon une directionnalité opposée. L'intérêt de ce genre de règles est qu'elles permettent d'unir des foncteurs dont les connecteurs n'ont pas même directionnalité.

pour résoudre les problèmes d'ellipse. Emmon Bach (1984) s'est orienté, quant à lui, vers l'élaboration de nouvelles règles (appelées «wrapping rules») autorisant la permutation de mots non nécessairement adjacents. Mais toutes ses solutions se sont heurtées, comme pour l'approche usant de la récursivité, à la trop grande puissance du système qu'elles engendrent, car celui-ci attesta que presque toutes les possibilités d'ordre des mots étaient grammaticales.

Le choix de créer de nouvelles règles de «contraction» et d'«expansion» a également rencontré le même problème de «sur-génération». L'idée directrice était de ne plus respecter la propriété de comptage du calcul syntaxique de Lambek qui oblige chaque fonction à s'appliquer à tous ses arguments, mais en ne le faisant qu'une seule fois. Ainsi, un verbe intransitif ne portera que sur un seul nom, alors qu'un verbe transitif devra s'appliquer à deux expressions nominales. Dans ce cadre, une règle d'expansion permettra à une fonction ne portant que sur un argument de s'appliquer à deux arguments et une règle de contraction autorisera, par contre, une fonction nécessitant deux arguments du même type à n'être satisfaite que par un seul ; ces modifications offrent la possibilité de résoudre les problèmes d'ellipses («gapping») que l'on trouve, par exemple, dans l'énoncé *Myrtille nourrit les tigres et Clara le chat des voisins*.

Mais le problème principal qu'a rencontré le paradigme catégoriel au début des années 1990 a été le caractère inconciliable de toutes ces approches qui s'étaient développées en parallèle. En effet, ainsi que l'expose Wood (1993), les analyses alternatives des mêmes phénomènes linguistiques se sont révélées mutuellement exclusives et l'idée utopique d'une grammaire catégorielle qui cumulerait les principaux acquis non contradictoires donnerait lieu à un modèle inapplicable et trahissant les principaux fondamentaux ayant donné naissance à cette approche mathématisée de la grammaire :

Thus although the sum total of these developments, and the total coverage of the specific linguistic descriptions [...] may seem impressive, they do not add up to their apparent total. Some are deliberate counters to others ; some are, often by luck, compatible, but generally ignore each other. A single Categorical Grammar which incorporated all these extensions, even one which avoided internal contradictions, would be far too large, complex and cumbersome for real use, a long way from the elegant simplicity of the original 'core' grammar. (Wood 1993 : 61)

2.3. Peut-on vraiment revenir à un petit ensemble de règles ?

Parallèlement à ce travail de nettoyage réalisé dans les années 1980, des recherches logiques mirent en évidence que le calcul syntaxique de 1958 pouvait être pensé comme étant la première « logique sous-structurelle » sensible aux « ressources » (i.e. aux mots de la phrase) et à leur ordre d'apparition (cf. Schroeder-Heister et Dosen 1994). L'idée fut alors de créer des extensions du calcul syntaxique. Aussi le renouveau actuel du paradigme catégoriel – sous les formes de la logique linéaire, bilinéaire, etc. – ne peut-il être interprété en termes de « révolution », puisque les nouveaux modèles ne sont ni incompatibles, ni incommensurables avec le calcul syntaxique originel de Lambek.

Or si ces nouveaux systèmes se caractérisent – comparativement aux grammaires catégorielles généralisées – par une réduction drastique du nombre de règles (chaque modèle n'en compte que deux ou trois), on constate que ces approches font appel à des notions de « métarègles » ou de « postulats structurels » dont le statut méthodologique est *a priori* problématique, puisqu'ils soulèvent la question de leur articulation avec la notion de « règle ». La grammaire des prégroupes de Lambek (2000a) est à cet égard exemplaire, puisqu'elle adjoint aux deux règles de réduction que comporte cette nouvelle syntaxe de nombreuses métarègles qui ont pour défaut d'être *ad hoc* car non généralisables à d'autres langues. Ainsi l'analyse des dépendances nichées dans les subordinées allemandes effectuée par Lambek (2000b) le conduit à définir une métarègle permettant de rendre compte de l'ordonnement des mots dans celles-ci. En pratique, cette métarègle autorise le verbe conjugué à endosser trois types différents en fonction de son contexte phrastique (déclaratif, interrogatif ou subordinée)¹⁰ et le problème qui ressort de cette analyse est que ces métarègles sont à chaque fois spécifiques à la langue considérée, si bien que le modèle perd son caractère d'universalité. De même, l'analyse syntaxique de l'ordre préverbal des clitiques en français (Bargelli et Lambek, 2001), qui nécessite, quant à elle, l'intervention de deux métarègles permettant respectivement de modifier le type du participe passé en fonction de son auxiliaire et d'assigner au verbe des types différents en fonction de son contexte phrastique, rend patent que l'action *ad hoc* des métarègles est d'assigner aux expressions des types spécifiques¹¹, si bien que la question se pose du lieu (le dictionnaire ou la

10 SVO pour les phrases affirmatives (*I sehe dich*), VSO pour les interrogatives (*Sehe ich dich ?*) et SOV pour les subordinées (... , *ob ich dich sehe*).

11 L'analyse des clitiques en italien conduit à définir des métarègles différentes de celles proposées pour le français (cf. Casadio et Lambek, 2001).

syntaxe?) où les métarègles doivent intervenir. En effet, une possibilité rejetée par Lambek (2000b : 29) serait de lister dans le lexique tous les types qui peuvent être affectés aux mots en fonction de ces métarègles, permettant ainsi à la grammaire de n'être régie que par les deux règles de réduction. En choisissant de faire coexister les métarègles avec la règle, le but poursuivi par Lambek – élaborer un modèle catégoriel où la grammaire serait totalement intégrée dans le dictionnaire des types affectés aux mots de la langue étudiée – se trouve de nouveau en échec, puisque les métarègles n'opèrent pas les changements de type à l'intérieur du dictionnaire. De plus, la nécessité d'ajouter des métarègles conduit à perdre le modèle mathématique fondateur en termes de pré-groupe, puisque celui-ci signifie par sa définition même qu'il ne peut y avoir qu'un mode de construction pour les types composés, à savoir celui qui est généré à partir de l'ensemble des types de base grâce à la mise en œuvre de deux sortes d'adjoints (*cf.* Godart-Wendling, 2005). L'idéal de simplicité propre à l'approche catégorielle s'avère donc irrémédiablement perdu, puisque de très nombreuses métarègles, distinctes de langue en langue, nécessitent d'être définies et appliquées pour vérifier la bonne formation syntaxique des séquences de mots à tester.

3. En guise de conclusion : qu'en est-il de la sémantique ?

Jusqu'à ce point de l'exposé, et le lecteur l'aura remarqué, peu de remarques ont été consacrées à la dimension sémantique, alors que la visée d'Ajdukiewicz de définir une procédure qui calculerait la grammaticalité des séquences tout en rendant compte du sens dont elles sont porteuses, constitue un des enjeux les plus intéressants de l'approche catégorielle (*cf.* Bach 1988), puisque sa réalisation permettrait de modéliser le fonctionnement de la compétence langagière humaine. En effet, face à un énoncé écrit ou oral, les locuteurs d'une langue donnée ne soupèsent pas tout d'abord si celui-ci est grammatical pour s'interroger par la suite sur sa signification, mais évaluent dans un même temps – plus ou moins consciemment (*cf.* Lecomte, 2008 : 24) – sa correction syntaxique et le sens qu'il véhicule. Aborder ce point soulève toutefois la question délicate de déterminer si le modèle se doit de reproduire l'activité langagière des humains. Francis Renaud (2005) constate en effet que « la coupure que le physicien fait facilement entre l'objet qu'il décrit et sa description formelle a du mal à s'imposer au linguiste qui oscille entre sa position d'observateur extérieur à la langue et de locuteur s'observant par introspection ». Autrement dit, le problème est : pourquoi faudrait-il

que les calculs que fait le linguiste en tant que linguiste soient ceux qu'il fait en tant que locuteur ?

Dans les faits, la période 1935-1970 se caractérisera pour les grammaires catégorielles par l'échec et même l'abandon du projet de ce calcul conjoint syntaxe-sémantique. Il faut attendre les travaux de Richard Montague (1970a et b) mettant en correspondance chaque règle de la grammaire avec une règle sémantique calculant la signification résultante dans l'optique la théorie des mondes possibles pour que cet objectif commence à se concrétiser. Cette approche permet alors à la fin des années 1980 d'établir un calcul compositionnel des phrases basé sur un lexicalisme radical tel que « l'interprétation sémantique est contrôlée par les types » (cf. Moortgat 1988a : 2).

Toutefois, un des problèmes qui émane du principe de compositionnalité est l'utilisation qui en est faite pour corroborer la plausibilité psychologique de l'approche catégorielle. Ainsi Steedman (1985 : 525) tire profit de la compositionnalité pour argumenter que les grammaires catégorielles posséderaient des « qualités explicatives en matière de psychologie » et Carpenter (1997 : 33) y voit même la preuve que ce type de syntaxe permet de rendre compte de la productivité du langage humain ; occultant ainsi non seulement les usages non littéraires dont la fréquence dans les langues est élevée, mais également les mécanismes d'inférence contextuelle que l'être humain mobilise pour comprendre la signification d'un énoncé.

Cette difficulté se retrouve dans les justifications apportées pour légitimer le lexicalisme radical. En effet, alors que celui-ci résulte de l'universalité des règles catégorielles qui conduit à conférer au lexique le rôle d'exprimer – grâce à la variation des types affectés aux entrées lexicales – la spécificité de chaque langue, son introduction est présentée comme étant « d'un grand intérêt pour l'apprentissage », car :

les grammaires lexicalisées présentent le grand avantage de ne pas avoir de règles à apprendre, puisque celles-ci sont fixes ; il suffit alors de déterminer l'entrée lexicale associée à un nouveau mot ou au nouvel usage d'un mot déjà présent dans le lexique. (Retoré 2000 : 5)

À cette confusion entre « apprentissage » au sens des informaticiens et celui – toujours en voie d'exploration – de l'être humain, s'ajoute la difficulté que la sémantique formelle s'avère toujours axée sur la portée référentielle des expressions ; niant par là même le rôle de la compréhension humaine, des dimensions communicative, socio-culturelle, historique, ... dans la structuration des significations.

Bibliographie

- ADES Antony et Mark STEEDMAN, « On the Order of Words », *Linguistics and Philosophy*, vol. 4, 1982, pp. 517-558.
- AJDUKIEWICZ Kazimierz, « Sprache une Sinn », *Erkenntnis* IV, 1934, pp. 100-138, trad. angl. dans Giedymin Jerzy (Dir.), *Kazimierz Ajdukiewicz : The Scientific World-Perspective and Others Essays, 1931-1963*, Dordrecht, D. Reidel Publishing Company, 1978, pp. 35-66.
- AJDUKIEWICZ Kazimierz, « Die syntaktische Konnexität », *Studia Philosophica* 1, 1935, pp. 1-27. (traduit "Syntactic Connection" par Weber H., *Polish Logic, 1920-1939*, in McCall Storrs (Dir.), Oxford, Clarendon, 1967, pp. 207-231).
- BACH Emmon, « Some generalizations of categorial grammars », *Varieties of Formal Semantics*, Landman F. et Veltman F. (Dir.), Dordrecht, Fori, 1984, pp. 1-23.
- BACH Emmon, « Categorial Grammars as Theories of Language », in Oehrle Richard *et al.* (Dir.), 1988, pp. 17-34.
- BARGELLI Daniele et Joachim LAMBEK, « An Algebraic Approach to French Structure », in *Logical Aspects of Computational Linguistics*, de Groote P., Morrill G. et Retoré C. (Dir.), Berlin et Heidelberg, Springer-Verlag, 2001.
- BAR-HILLEL Yehoshua, « A Quasi-Arithmetical Notation for Syntactic Description », *Language* 29(1), 1953, pp. 47-58.
- BAR-HILLEL Yehoshua, « Some Linguistic Obstacles to Machine Translation », *Advances in Computers*, vol. 1, 1960, pp. 197-207.
- BOOLE George, *An Investigation of the Laws of Thought on which are founded the mathematical theories of logic and probabilities*, London, Walton and Maberley, 1854. Trad. fr. par Souleymane Bachir Diagne, *Les lois de la pensée*, Paris, Vrin, 1992.
- CARPENTER Bob, *Type-Logical Semantics*, The MIT Press, 1997.
- CASADIO Claudia et Joachim LAMBEK, « An Algebraic Analysis of Clitic Pronouns in Italian », *Logical Aspects of Computational Linguistics*, de Groote P., Morrill G. et Retoré C. (Dir.), 2001, pp. 110-124.
- FREGE Gottlob, *Posthumous Writings, 1948/1925* (« Boole's logical Calculus and the Concept-script », pp. 9-46 et « Boole's logical Formula-language and my Concept-script », pp. 47- 52), The University of Chicago Press, 1979.
- FREGE Gottlob, *Idéographie*, 1879, trad.fr. de Corinne Besson, Paris, Vrin, 1999.
- FREGE Gottlob, « Fonction et concept », 1891, in *Écrits logiques et philosophiques*, trad.fr. de Claude Imbert, Paris, Seuil, 1971.

- GODART-WENDLING Béatrice, «De l'influence du principe de «dualité» sur le traitement algébrique des phrases», *Logique et Analyse*, vol. 39, n° 155-156, 1996.
- GODART-WENDLING Béatrice, «Mais que fait Bar-Hillel quand “le soleil siffle”?»», *History of Linguistics 1996*, Cram David, Andrew Linn et Elke Nowak (Dir.), vol. 2, *From Classical to Contemporary Linguistics*, Amsterdam/Philadelphia, John Benjamins, 1999, pp. 321-329.
- GODART-WENDLING Béatrice (Dir.), *Les grammaires catégorielles. Langages* 148, 2002.
- GODART-WENDLING Béatrice. «Analyse du traitement de la quantification en Grammaire Catégorielle: logique non linéaire versus grammaire de pré-groupe», in *La Quantification dans la logique moderne*, Joray Pierre (Dir.), Paris, L'Harmattan, collection «Épistémologie et Philosophie des Sciences», 2005, pp. 313-342.
- GODART-WENDLING Béatrice et Pierre JORAY, «When Categories go back to Parts of Speech», *History of Linguistics 2008*, Hassler Gerda (Dir.), Amsterdam/Philadelphia, John Benjamins, 2001, pp. 399-408.
- HARRIS Zellig, *Methods in Structural Linguistics*, Chicago, University of Chicago Press, 1951.
- HUSSERL Edmund, *Logische Untersuchungen*, 2^e éd. Halle, Max Niemeyer, 1913. Trad. fr. *Recherches logiques*, Paris, PUF, 1962.
- KUHN Thomas, *The Structure of Scientific Revolutions*, Chicago, University of Chicago Press, 1962.
- LAMBEK Joachim, «The Mathematics of Sentence Structure», *The American Mathematical Monthly* 65/3, 1958, pp. 154-170.
- LAMBEK Joachim, «On the Calculus of Syntactic Types», in Jakobson Roman (Dir.), *Structure of Language and its Mathematical Aspects*, Providence, R.I., American Mathematical Society, 1961, pp. 166-178.
- LAMBEK Joachim, «Pregroups: a new algebraic approach to sentence structure», *Recent Topics in Mathematical and Computational Linguistics*, Martin-Vide Carlos (Dir.), Bucharest, Editura Academici Romane, 2000a, pp. 182-195.
- LAMBEK Joachim, «Type grammar meets German Word Order», *Theoretical Linguistics*, vol. 26, 2000b, pp. 19-30.
- LECOMTE Alain, «La «grammaire pure»: une grammaire des normes», *La logique & les normes, Hommage à J.-L. Gardies. Recherches sur la philosophie et le langage* 25, Lardic J.-M. (Dir.), Paris, Vrin, 2008, pp. 111-137.
- LEGERET Marie-Ange, «Categorial Grammar with Type-Raising and Composition», *Word Order in Categorial Grammar / L'ordre des*

- mots dans les grammaires catégorielles*, Alain Lecomte (Dir.), Clermont-Ferrand, Adosa, 1992, pp. 105-112.
- MONTAGUE Richard, «English as a Formal Language», *Linguaggi nella Società e nella Tecnica*, Visentini B. et al. (Dir.), 1970a, pp. 189-224. Réédité in Thomason R. (Dir.), pp. 188-221, 1974.
- MONTAGUE Richard, «Universal Grammar», *Theoria*, vol. 36, 1970b, pp. 373-398. Réédité in Thomason Richmond (Dir.), pp. 222-246, 1974.
- MOORTGAT Michael, *Categorical Investigations*, Dordrecht, Foris, 1988a.
- MOORTGAT Michael, «Mixed composition and discontinuous dependencies», in Oehrle et al. (Dir.), 1988b, pp. 319-348.
- OEHRLER Richard, Emmon BACH et Deirdre WHEELER, *Categorical Grammars and Natural Language Structures*, Dordrecht, D. Reidel, 1988.
- RENAUD Francis, *Temps, durativité, ténacité*, Louvain-Paris, Peeters, 2005.
- RETORE Christian, «Systèmes déductifs et traitement des langues : un panorama des grammaires catégorielles», *Rapport de recherche INRIA 3917*, 2000.
- SCHRODER-HEISTER Peter et Kosta DOSEN (Dir.), *Substructural Logics*, Oxford, Clarendon Press, 1993.
- STEEDMAN Mark, «Dependency and Coordination in the Grammar of Dutch and English», *Language* 3, 1985, pp. 523-568.
- STEEDMAN Mark, «Type-raising and directionality in combinatory grammar», *Proceedings of ACL91*, Berkeley, CA., Association for Computational Linguistics, 1991, pp. 187-194.
- THOMASON Richmond (Dir.), *Formal Philosophy: Selected Papers of Richard Montague*, New Haven, Yale University Press, 1974.
- WOOD M. McGee, *Categorical Grammars*, London and New York, Routledge, 1993.

