



HAL
open science

Règle(s) et métarègles dans l'approche catégorielle de Lambek

Béatrice Godart-Wendling

► **To cite this version:**

Béatrice Godart-Wendling. Règle(s) et métarègles dans l'approche catégorielle de Lambek. Verbum (Presses Universitaires de Nancy), 2019, XL (2), pp.221-236. hal-02115512

HAL Id: hal-02115512

<https://hal.science/hal-02115512>

Submitted on 18 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.

RÈGLE(S) ET MÉTARÈGLES DANS L'APPROCHE CATÉGORIELLE DE LAMBEK

Béatrice GODART-WENDLING

Laboratoire Modèles, Dynamiques, Corpus

CNRS – UMR 7114

Université Paris Nanterre

RÉSUMÉ

Cet article examine le choix théorique des grammaires catégorielles d'élaborer des syntaxes ne mettant en œuvre qu'un minimum de règles algébriques. Après avoir retracé l'histoire de la constitution progressive des règles (du calcul syntaxique de Lambek à la prolifération des règles dans les années 80), l'examen se porte sur le dernier modèle de Lambek (la grammaire des prégroupes) et procède à un examen critique de la décision méthodologique d'user d'un modèle alliant l'action de deux règles universelles à l'application locale de métarègles spécifiques à chaque langue étudiée.

ABSTRACT

This article analyzes categorial grammars's theoretical choice to develop syntaxes using a minimum of algebraic rules. After having retraced the history of the progressive constitution of the rules (from Lambek's syntactic calculus to the proliferation of rules in the 80s), the study focuses on Lambek's last model (the pregroup grammar) and proceeds to a critical analysis of the methodological decision to use a model combining the action of two universal rules with the local application of metarules specific to each studied language.

INTRODUCTION

Paradigme concurrent des différentes versions de la grammaire de Chomsky, le courant des grammaires catégorielles non combinatoires présente la propriété remarquable d'avoir systématiquement recouru à des modèles algébriques susceptibles de formaliser les syntaxes des langues

naturelles à l'aide d'un nombre très restreint de règles¹. Face à la variété et la complexité des faits de langue, ce choix théorique s'est cependant heurté à des difficultés, voire à des impasses, qui ont contraint les théoriciens à mettre en œuvre différentes solutions. Ainsi, la première stratégie a consisté jusqu'à la fin des années 80 à enrichir la grammaire de nouvelles règles. Ce travail de « nettoyage » (au sens kuhmien du terme) a toutefois connu un coup d'arrêt lorsque la prolifération des règles a donné lieu à des problèmes de surgénération. L'attention s'est alors tournée vers de l'adoption de modèles mathématiques différents qui ont alors permis de réduire de façon drastique le nombre de règles.

L'évolution de l'approche catégorielle de Joachim Lambek témoigne de ce processus. Dans son article de 1958, il définit un *calcul syntaxique* ne mettant en jeu que cinq règles. Ce calcul servira de base de travail – ou de « calcul originel » comme le dénomme Michael Moortgat (2002 : 111) – à tous les théoriciens des grammaires catégorielles qui, pendant une trentaine d'années, tenteront de le parfaire en lui adjoignant de nouvelles règles. La prolifération des règles et des systèmes catégoriels² qui s'en est suivie a conduit Lambek à abandonner ce modèle en 1999 et à se tourner vers la *théorie des prégroupes* qui permet, sur la base de deux règles, de vérifier la bonne formation syntaxique des phrases. Mais, parallèlement à ce mouvement de restriction du nombre de règles³, est apparue la notion de « métarègle » au statut ambigu, puisque – comme le montrera cet article – l'application des métarègles a pour conséquence que la grammaire ne peut plus, comme le voulait Lambek, être complètement intégrée dans le dictionnaire des types affectés aux mots de la langue étudiée.

Après avoir rappelé les règles constitutives du calcul syntaxique de Lambek (1958) et souligné l'essor en matière de règles qui s'en est suivi, cet article analysera le rôle des métarègles dans la grammaire des prégroupes de Lambek et les conséquences qu'elles induisent sur la construction de la grammaire.

¹ Cet article ne traitera pas des grammaires catégorielles recourant à des combinateurs. Pour une approche de ce courant, le lecteur pourra se reporter aux articles de Desclés et Biskri dans ce volume et se référer également aux écrits de Shaumyan (1982, 1987), Steedman (1988, 1991), Desclés (1990), Desclés et Biskri (1995) et Desclés *et al.* (2016).

² Pour une classification de ces différents systèmes, *cf.* Moortgat (1988 : 40-41).

³ Michael Moortgat suivra la même trajectoire, puisqu'après avoir élaboré, à la fin des années 80, de nouvelles règles permettant de traiter de phénomènes langagiers restés récalcitrants, il optera pour un nouveau cadre théorique – *la logique multimodale des types* – où le noyau grammatical ne compte plus que trois règles (dites de « résiduaction ») ; *cf.* Moortgat (1996, 1997, 1999, 2001, 2002).

1. DU CALCUL SYNTAXIQUE DE LAMBEK À LA PROLIFÉRATION DES RÈGLES

Le *calcul syntaxique* s'inscrit dans la continuation des travaux d'Ajdukiewicz (1935) et de Bar-Hillel (1953)⁴, ainsi que le souligne Lambek dans l'introduction de son article :

The second part of this paper is concerned with a development of the technique of Ajdukiewicz and Bar-Hillel in a mathematical direction. We introduce a calculus of types, which is related to the well-known calculus of residuals. (1958 : 154)

D'Ajdukiewicz et de Bar-Hillel, Lambek conservera l'idée que la catégorisation des mots d'une langue peut être obtenue à partir des types⁵ de base. Aux types **n** (name) et **s** (sentence), utilisés par ces prédécesseurs, Lambek adjoindra tout d'abord le type de base **n*** pour catégoriser les noms au pluriel, puis ajoutera, dans son article de 1959, les trois nouveaux types de base **i**, **p** et **q** pour rendre compte respectivement de l'infinitif, du participe présent et du participe passé des verbes intransitifs. C'est en fonction de ces types de base que les types composées seront formés grâce à

la définition récursive : si x et y sont des types, alors x/y (qui se lit *x sur y*) et $y \backslash x$ (*y sous x*) sont des types⁶ (1958 : 155).

Ainsi, la séquence *Jean travaille* correspondra à la suite de types ' $n \ n \ s$ '. Lambek maintiendra également, comme première règle de son calcul, la règle d'*application*, telle qu'elle a été successivement formulée⁷ par Ajdukiewicz ' $(x/y) \ y \rightarrow x$ ' et par Bar-Hillel ' $y \ (y \backslash x) \rightarrow x$ '. Cette règle⁸ signifie qu'« une expression de type x/y suivie par une expression de type y produit une expression de type x , et il en va de même d'une expression de type $y \backslash x$ lorsqu'elle précède d'une expression de type y »⁹ (Lambek 1958 : 155-

⁴ Pour une présentation historique et une explication de ces deux syntaxes catégorielles, cf. Casadio (1988), Desclés (1990), Gardies (1975), Godart-Wendling (2002)

⁵ Sous l'influence des travaux de Alonzo Church et de Haskell Curry, Lambek abandonnera l'appellation de « catégorie » au profit de celle de « type ».

⁶ « the recursive definition : If x and y are types, then so are x/y (read *x over y*) and $y \backslash x$ (read *y under x*) »

⁷ Pour des raisons de simplicité, nous avons utilisé – comme Lambek – l'orientation des slashes pour exprimer les deux versants de cette règle. Mais cet usage n'est pas conforme à la notation d'Ajdukiewicz qui emploie une barre horizontale et à celle de Bar-Hillel qui définit deux types de parenthèses pour indiquer si la simplification doit s'opérer à gauche ou à droite.

⁸ Pour expliciter cette règle, nous restons fidèle à la terminologie de Lambek qui n'utilise pas les termes d'« opérateur » et d'« opérande » employés par le courant des grammaires catégorielles combinatoires.

⁹ « an expression of type x/y when followed by an expression of type y produces an expression of type x , and so does an expression of type $y \backslash x$ when preceded by an expression of type y »

156). Appliquée à la séquence de types ' $n\ n\ s$ ' assignés à *Jean travaille*, cette règle permet de montrer qu'il s'agit d'une phrase de type s syntaxiquement bien formée. Lambek adjoindra à la règle d'application quatre nouvelles règles (l'*associativité*, la *composition*, la *division*¹⁰ et la *montée*¹¹) qui ont pour fonction de gérer l'ordre des mots ou de permettre à un type de s'adapter à son contexte en se transformant en un autre type qui rendra la dérivation effectuable grâce à l'intervention de la règle d'application¹². La règle d'associativité ' $(x\backslash y)/z \Leftrightarrow x\backslash(y/z)$ ' établit qu'une fonction qui possède deux arguments (situés de par et d'autre, conformément à l'ordre linéaire syntagmatique) peut s'associer avec eux dans n'importe quel ordre. Dans sa forme la plus explicite ' $x\backslash y/z$ ', cette règle signifie qu'il est indifférent syntaxiquement de commencer à combiner, par exemple, un verbe transitif avec son objet ou à l'inverse avec son sujet. Cette règle permet ainsi de rendre compte du fait que la séquence *Jean aime Jeanne* peut être décomposée selon deux groupements différents, mais équivalents ; à savoir '*Jean (aime Jeanne)*' où *aime* est de type $(n\backslash s)/n$ et '*(Jean aime) Jeanne*' où *aime* a, cette fois-ci, le type $n\backslash(s/n)$. L'utilisation de cette règle conduit donc à éliminer un certain nombre de structures faussement ambiguës qui posaient problème dans la grammaire de Bar-Hillel (1964 : 78-79), car celle-ci ne pouvait montrer que les deux possibilités de structuration ne donnaient pas lieu à deux interprétations distinctes de *Jean aime Jeanne*. La *règle de composition*, qui comprend les deux alternatives ' $(x/y)(y/z) \rightarrow x/z$ ' et ' $(x\backslash y)(y\backslash z) \rightarrow x\backslash z$ ' liées à la directionnalité de la fonction, a été introduite dans le calcul syntaxique afin de rendre compte des problèmes de catégorisation liés aux deux formes du pronom (*he* et *him*) en anglais¹³. Cette règle offre la possibilité théorique d'appliquer deux fonctions au dernier argument trouvé sans avoir à prendre en compte l'ordre de surface des mots de la phrase. La *règle de division*, qui s'exprime ' $x\backslash x \rightarrow (y\backslash x)\backslash(y\backslash x)$ ', et dont l'introduction a été motivée pour rendre compte des deux possibilités de typage de l'adverbe *here* dans la phrase *John works here*¹⁴, ouvre la voie à la création de règles qui permettent des *expansions* de types, alors que toutes les autres règles (que nous avons déjà passées en revue) réalisent des *contractions* qui simplifient le calcul. Enfin, la *règle de montée*, qui engendre également des expansions, puisqu'elle établit que ' $x \rightarrow y\backslash(x/y)$ ' et ' $x \rightarrow (y/x)\backslash y$ ', a été introduite dans le but de rendre compte de la substitution

¹⁰ Cette règle sera également proposée par Peter Geach (1970) qui la découvrira en toute méconnaissance du calcul syntaxique de Lambek.

¹¹ Cf. également Montague (1974) pour une utilisation de cette règle.

¹² Sur ce sujet, cf. Moortgat (1988 : 23).

¹³ Pour plus de détails, cf. Lambek (1958 : 161-162) ou Godart-Wendling (2002 : 63).

¹⁴ En effet, dans la structuration « *(John works) here* », l'adverbe de phrase est de type $s\backslash s$; alors que dans le découpage « *John (works here)* », *here* (en tant qu'adverbe de verbe) est de type $(n\backslash s)\backslash(n\backslash s)$. La règle de division autorise le passage du type $s\backslash s$ à $(n\backslash s)\backslash(n\backslash s)$.

pouvant être opérée entre un nom (tel que *John*) et un pronom (sous ses formes *he* et *him*). Cette règle autorise donc qu'une expression de type *n* puisse également disposer, en fonction de ses contextes, des types *s/(n's)* et *(s/n)s* normalement affectés aux pronoms.

En proposant ainsi des règles effectuant des changements de types, le calcul syntaxique de Lambek allait profondément modifier la définition de la notion de « type », puisque celle-ci allait être pensée sous la forme d'une famille de types reliés entre eux par une relation d'ordre¹⁵. Enfin, l'ajout de ces quatre règles a eu pour conséquence de faire évoluer la façon de concevoir la représentation de la structure des phrases, de telle façon qu'on en vint à penser qu'elle était « flexible », puisqu'il devenait possible de poser des équivalences de structuration ou de modifier les types à certaines étapes de la dérivation.

Mais ce modèle mathématique souffrait cependant de ne pouvoir représenter les faits de langue présentant des discontinuités¹⁶, des ellipses, des dépendances croisées¹⁷... et la nécessité, qui se fit sentir dans les années 80, d'étendre la capacité descriptive de l'approche catégorielle¹⁸, a conduit les théoriciens 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 :

- accroître le pouvoir récursif de la grammaire ;
- augmenter les possibilités de directionnalité et/ou autoriser des permutations ;
- jouer sur les processus de contraction ou d'expansion 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 »¹⁹ qui, en conférant à la grammaire « a full context-sensitive power », offre la possibilité de résoudre les phénomènes de dépendances non liées. Une autre voie entreprit de con-

¹⁵ Une relation d'ordre se caractérise par le fait d'être réflexive, antisymétrique et transitive.

¹⁶ Une des difficultés dont sont porteuses les règles des grammaires catégorielles est le « principe d'adjacence » : une fonction ne peut s'appliquer qu'à un argument qui lui est contigu, entraînant ainsi une impossibilité à gérer les phénomènes linguistiques impliquant une discontinuité.

¹⁷ Comme ceci est le cas dans les subordonnées en néerlandais.

¹⁸ Pour plus de détails sur ce sujet, cf. Godart-Wendling (2007).

¹⁹ 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.

juguer l'action des règles de composition et de division, mais le problème auquel toutes ces extensions donnèrent 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 locuteurs²⁰.

La solution consistant à élaborer des règles qui gèreraient 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 « disharmonique²¹ de composition » (qui s'énonce : '(y/z) (y/x) → x/z') pour gérer les problèmes de discontinuité et de « bracketing paradoxes » ; Steedman a réfléchi sur des variantes directionnelles de la règle de montée pour résoudre les problèmes d'ellipse. Emmon Bach (1979a, 1979b, 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, de la trop grande puissance du système qu'elles engendrent²², 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 »²³. 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 cadre²⁴, une règle d'expansion permettra à une fonction ne portant que sur un argument de s'appliquer à deux

²⁰ Sur ce sujet, cf. les articles critiques de Uszkoreit (1986), Zeevat (1988), Bouma (1989) et Wood (1989).

²¹ 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.

²² Pour une preuve formelle de cette difficulté, cf. Moortgat (1988b : 90).

²³ Wood explicite ainsi ce problème récurrent : « The difficulty is to extend the system just far enough, without allowing wild overgeneration. Rules which allow some sort of permutation risk collapsing into total 'permutation closure', licensing any possible order of elements (thus LP). Rules which allow deletion ('contraction') or copying ('expansion') (thus LPCE) are comparably profligate. The problem is how to define a system at the appropriate level between L and LPCE (see Moortgat 1988b : 45-46), but without resorting to *ad hoc*, externally stipulated constraints, preserving the 'free algebra' » (1993 : 60-61).

²⁴ Pour une étude plus détaillée de ces règles structurales, voir Johan van Benthem (1987).

arguments et une règle de contraction²⁵ 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 *Michel nourrit les tigres et Christian le chat des voisins*.

Le problème principal qu'a rencontré le paradigme catégoriel au début des années 90 a été le caractère inconciliable de toutes ces approches qui s'étaient développées en parallèle. En effet, ainsi que l'expose Wood, 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 principes 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).

Dans ce contexte, on comprend que le renouvellement de la perspective catégorielle se soit effectué par l'adoption de modèles algébriques différents destinés à permettre la formalisation des syntaxes des langues à partir d'un nombre très réduit de règles. A cet égard, la grammaire de prégroupes, présentée en 1999 par Lambek, est la plus exemplaire, puisqu'elle rejoint, tout en innovant, le projet catégoriel originel en faisant reposer tout l'édifice grammatical sur l'application de deux règles de simplification.

2. LA GRAMMAIRE DE PRÉGROUPES

Lambek n'a appliqué son approche en termes de prégroupes qu'à deux types de phénomènes linguistiques :

- rendre compte de l'ordre des mots dans les subordinées allemandes, qui se caractérise par la place en position finale du verbe conjugué²⁶ (SOV) : *ob ich dich sehe*²⁷ (Lambek 2000b),
- traiter de l'ordre des clitiques en français (Bargelli et Lambek, 2001) et en italien (Casadio et Lambek, 2001).

²⁵ Cf. les travaux de Steedman (1987 et 1988) et Szabolcsi (1987).

²⁶ Rappelons que pour les propositions principales le verbe intervient toujours en deuxième position (SVO) : *ich sehe dich* (je te vois).

²⁷ Littéralement : 'si je te vois'.

Concernant le français, proposer une analyse qui atteste de la grammaticalité de séquences telles que « *la lui donner / lui en donner* » et rejette les formes « ** lui la donner / * en lui donner* ».

Pour chacun de ces trois cas, des métarègles – spécifiques à la langue considérée – ont été définies, si bien qu’une des difficultés relevées par Lambek est le caractère non universel des métarègles. L’article sur l’ordre des mots en allemand se clôt en effet sur ce constat :

Unfortunately, the metarule proposed here for German won’t account for Dutch subordinate clauses. (2000b : 30).

L’idée qui régit la grammaire de prégroupes reste en soi la même que celle qui animait le calcul syntaxique et même la grammaire bidirectionnelle de Bar-Hillel²⁸, puisqu’elle consiste à soutenir que :

to each French word there are assigned one or more types so that the sentencehood of a string of French words can be checked by a simple calculation. (Bargelli et Lambek, 2001 : 62)

Lambek définit la grammaire de prégroupes comme une structure algébrique engendrée par un ensemble partiellement ordonné²⁹ de types de base. Ces types, que nous listons ci-dessous, prennent en compte un certain nombre de propriétés morphologiques, énonciatives et syntaxiques de la langue française :

- s_j** = les phrases déclaratives (l’indice j indique le temps verbal et les modes spécifiques du Français, tels que pour le présent, $j = 1$; pour l’imparfait, $j = 2$; pour le futur, $j = 3$; pour le conditionnel, $j = 4$; pour le subjonctif, $j = 5$; pour le passé simple, $j = 6$ et pour le subjonctif passé, $j = 7$).
- q_j** = questions d’ordre total (l’indice j permettant de marquer le temps).
- t_j** = les subordinées (l’indice j permettant de spécifier le temps).
- π_k** = les pronoms personnels (l’indice k , qui va de 1 à 6, représente les six pronoms de conjugaison ($je = 1$, $tu = 2$, etc.)).
- o** = l’objet direct
- ω** = l’objet indirect
- i** = l’infinitif d’un verbe intransitif
- \bar{i}** = l’infinitif des verbes transitifs ou ditransitifs
- $\bar{\bar{i}}$** = l’infinitif des verbes modaux
- n** = les noms propres
- m** = les noms de masse
- c** = les noms comptables

²⁸ Rappelons que dans la syntaxe d’Ajdukiewicz, chaque mot ne se voyait assigner qu’une seule catégorie.

²⁹ Car muni d’une relation réflexive, transitive et antisymétrique que Lambek note \leq et qui peut être lue comme signifiant « est réductible à ». Sur l’ordonnement partiel des types de base, cf. Godart-Wendling et Joray (2011).

- p** = les noms au pluriel
m, c et s = les différents déterminants (i.e. *beaucoup de, de nombreux, bien des, un grand nombre de*) des noms de masse, des noms comptables et des noms au pluriel.
p₁ = participe présent
p₂ = participe passé
a = adjectif
λ = locatif

La principale caractéristique de la structure de prégroupes engendrée par les types de base est la suivante : chaque type de base a possède un adjoint gauche a^s et un adjoint droit a^d , tels que les séquences $a^s a$ et aa^d se réduisent à un élément 1 qui peut toujours être éliminé. Pour chaque élément a , nous avons donc les deux règles de réduction suivantes³⁰ :

$$\begin{aligned}
 (\text{Rg}) \quad a^s a &\leq 1 \\
 (\text{Rd}) \quad aa^d &\leq 1
 \end{aligned}$$

Les adjoints permettent de former les types dérivés à partir des types de base et le rôle des règles (Rg) et (Rd) revient à assurer la simplification catégorielle lorsqu'une expression d'un type dérivé est correctement appliquée aux arguments qu'elle exige. Ainsi, on peut démontrer que la séquence *il devait dormir* est une phrase au passé, c'est-à-dire une phrase de catégorie s_2 , car :

- *il* est de type π_3 , étant donné qu'il s'agit d'un pronom à la troisième personne ;
- *devait* a le type $(\pi_3^d s_2 i^s)$ car il permet la formation d'une phrase, lorsqu'il dispose d'un pronom à la troisième personne au singulier qui le saturera par la droite et d'un verbe à l'infinitif qui le complètera par la gauche ;
- *dormir* est de type i , puisqu'il s'agit de l'infinitif d'un verbe.

Nous obtenons ainsi la simplification suivante :

$$\begin{array}{ccc}
 il & devait & dormir \\
 \pi_3 & (\pi_3^d s_2 i^s) & i \\
 (\pi_3 \pi_3^d) & s_2 & (i^s i)
 \end{array}$$

Les règles de réduction peuvent alors opérer des simplifications qui conduisent à montrer que toute la séquence est de type s_2 , c'est-à-dire une phrase au passé : $1 \leq s_2 \leq 1$.

Une telle grammaire présente l'avantage de ne plus mettre en jeu un système complexe de règles syntaxiques, puisque la vérification de la bonne formation des phrases repose sur les deux règles de réduction, le reste de

³⁰ Pour une meilleure lisibilité, nous avons adapté les indices au français en notant a^l par a^s et a^r par a^d .

l'information grammaticale étant inscrit dans le lexique (par l'indication de leurs types).

Le traitement de l'ordre préverbal des clitiques en français a toutefois conduit Daniele Bargelli et Joachim Lambek à complexifier ce modèle algébrique en lui adjoignant des métarègles dont le statut n'est jamais précisé³¹. Pour rendre compte des phrases (*je vous les offre* /* *je les vous offre*), ces deux auteurs procèdent préalablement à une analyse des infinitifs des verbes (de type i) afin d'obtenir une analyse des formes telles que « *vous les offrir* ». Le résultat obtenu leur permet alors de complexifier leur analyse en considérant des formes verbales composées qui se conjuguent avec les auxiliaires *avoir* ou *être*. C'est à ce stade de leur réflexion qu'intervient l'introduction d'une première métarègle qui établit que³² :

Métarègle I. Si l'infinitif d'un verbe V (pris sans ses compléments) a le type ix^g , alors son participe passé a le type p_2x^g pour la plupart des verbes, incluant tous les verbes transitifs, et le type p'_2x^g pour un groupe déterminé de verbes intransitifs et pour tous les verbes réflexifs. L'infinitif passé du premier est formé avec *avoir* de type ip_2^g , et celui du deuxième avec *être* est de type ip'_2^g .

Cette métarègle permet de proposer des résolutions syntaxiques pour des séquences telles que : « *avoir mangé une pomme* », « *la lui avoir donnée* », « *y être allé* », « *être mis sur la table* », ... (Bargelli et Lambek, 2001 : 71), puisque son rôle est de changer le type du participe passé en fonction de l'auxiliaire avec lequel il se conjugue (p_2 pour *avoir*, p'_2 pour *être*). On obtient ainsi :

avoir mangé une pomme et *y être allé*
 (ip_2^g) (p_2o^g) o $(i^{egg}i^g)$ (ip'_2^g) $(p'_2\lambda^g)$

Néanmoins, les propriétés qui ressortent de la lecture de cette métarègle est qu'elle est à la fois très descriptive et cependant très vague, puisqu'elle contient des expressions telles que « pour la plupart des verbes » et « un groupe déterminé » qui entravent la possibilité même de la mettre en œuvre par des locuteurs non natifs du français³³. Mais, si l'on adopte maintenant un point de vue méthodologique visant à soupeser ce choix d'user de « prescriptions » n'entrant pas dans le cadre des règles, on peut interpréter cette approche comme témoignant d'une volonté d'appliquer une stratégie *locale*

³¹ En effet, il en va de même dans Lambek (2000b) et Casadio et Lambek (2001).

³² **Metarule 1.** If the infinitive of the (non-extended) verb V has type ix^l , then its past participle has type p_2x^l for most verbs, including all transitive verbs, and type p'_2x^l for a select group of intransitive verbs and for all reflexive verbs. The composite past of the former is formed with *avoir* of type ip_2^l , that of the latter with *être* of type ip'_2^l (2001 : 70).

³³ Précisons, de plus, que pour les natifs, elle ne présente aucun intérêt.

restreinte au groupe verbal et qui n'interviendrait donc pas au niveau global du traitement de la phrase³⁴.

La seconde métarègle, définie par Bargelli et Lambek, concerne les différents types devant être assignés à un verbe selon qu'il se trouve dans une phrase déclarative ou interrogative. On obtient ainsi (2001 : 73)³⁵ :

Métarègle II : Si le verbe pris avec ses compléments a le type ix^g , \bar{ix}^g ou $\bar{\bar{ix}}^g$, sa forme finie V_{jk} a le type $\pi_k^d s_j x^d$ dans une phrase déclarative.

Métarègle II (suite) : V_{jk} ($j = 1$ à 4) a le type $qj^g \pi_k^g$ dans les interrogatives directes avec inversion du sujet. Les interrogatives directes peuvent aussi être formées sans inversion à partir d'une phrase déclarative en préfixant *est-ce que* de type qs^g .

Ainsi, cette métarègle permet de différencier l'ordre spécifique des mots des phrases déclaratives de celui caractérisant les interrogatives avec les deux simplifications suivantes :

<i>Il</i>	<i>mange</i>	<i>une pomme</i>
π_3	$\pi_3^d s_1 o^g$	o
$\pi_3 \pi_3^d$	s_1	$o^g o$
1	s_1	1
s_1		

c'est-à-dire une phrase déclarative au présent.

<i>Mange-t-il</i>	<i>une</i>	<i>pomme ?</i>
$(q_1 o^g \pi_3^d)$	π_3	o
$q_1 o^g$	$\pi_3^d \pi_3$	o
$q_1 o^g$	1	o
q_1	$o^g o$	
q_1	1	
q_1		

c'est-à-dire une phrase interrogative au présent.

La fonction de cette métarègle est donc double : 1°) différencier le type à assigner au verbe en fonction du contexte phrastique dans lequel il s'insérera, et 2°) spécifier l'ordre des mots caractéristique des phrases déclaratives et interrogatives. Le problème qui reste à résoudre concerne l'articulation

³⁴ Desclés et Biskri (1995) ont une toute autre utilisation de la notion de métarègle, puisqu'ils l'emploient pour décider, au cours d'un calcul syntaxique, s'il faut composer les types ou déclencher un changement de type.

³⁵ **Metarule II.** If the extended verb V has type ix^l , $\bar{1}^l$ or $\bar{\bar{1}}^l$, its finite form V_{jk} has type $\pi_k^l s_j x^l$ in a declarative sentence.

Metarule II (continued) : V_{jk} ($j = 1$ à 4) has type $qj^g \pi_k^l$ in inverted direct questions. Direct questions can also be formed without inversion from a declarative sentence by prefixing *est-ce que* of type qs^l .

que les métarègles entretiennent avec la règle algébrique. En effet, il peut sembler logique de penser que les métarègles interviennent au niveau du dictionnaire des types affectés à chaque mot, et qu'elles permettent de sélectionner, avant l'application des règles algébriques de simplification, le type qui sera adéquat pour que la dérivation grammaticale soit menée à bien. Autrement dit, nous ne serions plus du tout dans un modèle « à la Bar-Hillel » où la règle de simplification détermine la structure de la phrase et la catégorie pertinente de chaque mot (parmi la liste de catégories spécifiées dans le dictionnaire reflétant les différents environnements où le mot peut être employé), mais dans une approche syntaxique qui use de règles mathématiques destinées à vérifier la grammaticalité et qui délèguerait aux métarègles le rôle d'établir la liste des types dans le dictionnaire. Or, de fait, il n'en est pas ainsi, car Lambek (2000b) – traitant à l'aide de la création d'une métarègle de l'ordre des verbes dans les différents types de phrases en Allemand – précise dans le *Postscript* de son article :

I hope to have shown that the three principal word orders one finds in German sentences (declarative, interrogative and subordinate) can be predicted if appropriate types have been assigned to finite and non-finite verb-forms. In principle, these could all be listed in the dictionary (Chomsky's lexicon); but, in practice, only the infinitives are listed there and the types of the other forms are to be derived with the help of a metarule. (Lambek, 2000b : 29).

On obtient donc un modèle reposant sur deux règles algébriques et un dictionnaire de types ne contenant qu'une information minimale qui sera transformée par les métarègles, afin que le typage des mots entre en cohérence avec le contexte strict (immédiat) dans lequel ils doivent intervenir (cas de la métarègle I qui spécifie le type du participe passé en fonction de l'auxiliaire *être* ou *avoir* avec lequel il se conjugue) ou le contexte phrasique (cas de la métarègle II qui modifie le type du verbe en fonction de la forme déclarative ou interrogative de la phrase).

3. CONCLUSION

L'architecture globale que dessine l'action conjuguée des règles et des métarègles dans la grammaire des prégroupes s'avère donc assez complexe. En effet, les deux règles sur lesquelles repose ce modèle agissent à deux niveaux très distincts de la grammaire, puisque : 1°) elles ont pour tâche d'engendrer, grâce aux adjoints, les types composés et 2°) elles opèrent toutes les simplifications permettant de vérifier la bonne-formation syntaxique des séquences de mots considérés. Les métarègles s'insèrent dans ce système pour engendrer également de façon locale des types composés et, ce faisant, elles se révèlent, d'un point de vue méthodologique, très problématiques, car elles conduisent, de par le grand nombre de métarègles qu'il faudrait spécifier de façon *ad hoc* pour rendre compte de la complexité de

chaque langue³⁶, à faire perdre au modèle ses propriétés de simplicité et d'économie qui le caractérisaient. De plus, le but poursuivi par Lambek d'élaborer un modèle catégoriel où la grammaire serait totalement intégrée dans le dictionnaire des types affectés aux mots de la langue étudiée n'est plus réalisable, puisque les métarègles n'opèrent pas les changements de type à l'intérieur du dictionnaire. Enfin, et cette conséquence est de loin celle qui remet le plus en cause cette approche catégorielle, la nécessité d'ajouter des métarègles conduit à s'éloigner du modèle mathématique fondateur en termes de prégroupes, 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 engendré à partir de l'ensemble des types de base grâce à la mise en œuvre des deux sortes d'adjoints.

RÉFÉRENCES

- ADES A., STEEDMAN M. (1982). On the Order of Words. *Linguistics and Philosophy* 4, 517-558.
- AJDUKIEWICZ K. (1935). Die syntaktische Konnexität. *Studia Philosophica* 1, 1-27. Trad anglaise: In: Giedymin (ed.), 1978, p. 118-139.
- BACH E. (1979a). Montague Grammar and Classical Transformational Grammar. In : S. Davis, M. Mithun (eds), *Linguistics, Philosophy, and Montague Grammar*. Austin : University of Texas, 3-49.
- BACH E. (1979b). Control in Montague Grammar. *Linguistic Inquiry* 10, 515-531.
- BACH E. (1984). Some Generalizations of Categorical Grammars. In : F. Landman, F. Veltman (eds), *Varieties of Formal Semantics*. Dordrecht : Foris, 1-23.
- BACH E. (1988). Categorical Grammars as Theories of Language. In : R.T. Oehrle *et al.* (eds), 1988, 17-34.
- BARGELLI D., LAMBEK J. (2001). An Algebraic Approach to French Sentence Structure. In : Ph. de Groote, G Morrill, Chr. Retoré (eds), *Logical Aspects of Computational Linguistics*. Berlin & Heidelberg : Springer-Verlag.
- BAR-HILLEL Y. (1953). A Quasi-Arithmetical Notation for Syntactic Description. *Language* 29, 47-58. Repr. in: Bar-Hillel 1964, 61-74. Trad. franc. 1968, *Langages* 9, 9-22.
- BAR-HILLEL Y. (1964). *Language and Information: Selected essays on their theory and application*. Reading (Massachusetts), Palo Alto, Don Mills (Ontario) : Addison-Wesley.

³⁶ En effet, les métarègles définies pour le français sont, de l'aveu même de Lambek, inopérantes pour les autres langues. Elles ne possèdent donc pas la propriété d'universalité dont sont détentrices les règles algébriques de simplification.

- BENTHEM J van. (1987). Semantic type change and syntactic recognition. *University of Amsterdam Department of Mathematics Report*, 87-105.
- BENTHEM J. van. (1988a). The Lambek calculus. In : R.T. Oehrle *et al.*, 1988, 35-68.
- BENTHEM J. van. (1988b). New trends in categorial grammar. In : Buszkowski *et al.* (eds.), 1988, 23-33.
- BOUMA G. (1989). Efficient processing of flexible categorial grammar. *Proceedings of EACL89*, Manchester, Association for Computational Linguistics, 19-26.
- BUSZKOWSKI W, MARCISZEWSKI W & VAN BENTHEM J. (eds) (1988). *Categorial Grammar*. Amsterdam & Philadelphia : John Benjamins.
- CASADIO C. (1988). Semantic Categories and the Development of Categorial Grammars. In : R.T. Oehrle *et al.*, 95-123.
- CASADIO C, LAMBEK J. (2001). An Algebraic Analysis of Clitic Pronouns in Italian. In : Ph. de Groote, G. Morrill, Ch. Retoré (eds), *Logical Aspects of Computational Linguistics*, 110-124.
- CASADIO C, LAMBEK, J. (2002). A tale of four grammars. *Studia Logica* 71(3), 315-329.
- CHURCH A. (1941). *The Calculi of Lambda Conversion*. Princeton University Press.
- CURRY H.B. (1951). La théorie des combinateurs : la logique combinatoire et les antinomies. *Rendiconti di matemat. e delle sue applic.* 10, 347-359, 360-370.
- DESCLÉS J.-P. (1990). *Langages applicatifs, langues naturelles et cognition*. Paris : Hermès.
- DESCLÉS J.-P, BISKRI I. (1995). Logique combinatoire et linguistique : grammaire catégorielle combinatoire applicative. *Mathématiques et sciences humaines* 132, 39-68.
- DESCLÉS J.-P, GUIBERT G., SAUZAY B. (2016). *Logique combinatoire et lambda calcul : des logiques d'opérateurs*. Toulouse : Cépaduès.
- DOWTY D. (1988). Type raising, functional composition, and non-constituent conjunction. In : R.T. Oehrle *et al.* (eds.), 1988, 153-197.
- GARDIES J.-L. (1975). *Esquisse d'une grammaire pure*. Paris : Vrin.
- GEACH P. (1970). A Program for Syntax. *Synthese* 22, 3-17.
- GIEDYMIN J. (ed.). (1978). *Kazimierz Ajdukiewicz : The scientific world-perspective and other essays, 1931-1963*. Dordrecht : D. Reidel.
- GODART-WENDLING B. (1999). Histoire de la notion de catégorie dans les premières grammaires catégorielles. In : M. Fernandez Rodriguez, F. Garcia Gondar & N.Vasquez Veiga (eds), *Actas del Congreso Internacional de la Sociedad Española de Historiografía Lingüística*. Madrid, Arco/Libros, S.L., 335-346.
- GODART-WENDLING B. (2002). (ed.). Les grammaires catégorielles. *Langages* 148. Paris : Larousse.

- GODART-WENDLING B. (2007). La preuve de Gaifman. Réflexion sur la méthode de construction des grammaires catégorielles. In : D. Kibbee (ed.), *History of Linguistics 2005*. Amsterdam & Philadelphia : John Benjamins, 432-440.
- GODART-WENDLING, B., JORAY P. (2011). When categories go back to parts of speech. In : G. Hassler (éd.), *History of Linguistics 2008*. Amsterdam & Philadelphia : John Benjamins, 399-407.
- HARRIS Z. (1951). (*Methods in*) *Structural Linguistics*. Chicago : Chicago Press.
- JACOBSON P. (1990). Raising as function composition. *Linguistics and Philosophy* 13, 423- 75.
- JAKOBSON R. (1961). *Structure of Language and its Mathematical Aspects*. Providence, R.I. : American Mathematical Society.
- KARLGRÉN H. (1978). Categorical grammar – a basis for a natural language calculus ? *Studia Logica* 37, 65-78.
- KARTTUNEN L. (1989). Radical Lexicalism. In : M.R. Baltin, A.S. Kroch (eds), *Alternative Conceptions of Phrase Structure*. Chicago : University of Chicago Press, 43-65.
- KUHN T. (1962). *The Structure of Scientific Revolutions*. Chicago : University of Chicago Press. Trad. franç. *La structure des révolutions scientifique*, Paris : Flammarion, 1983.
- LAMBEK J. (1958). The Mathematics of Sentence Structure. *American Mathematical Monthly* 65, 154-170.
- LAMBEK J. (1959). Contributions to a Mathematical Analysis of the English Verb-Phrase. *Journal of the Canadian Linguistic Association* 5, 83-89.
- LAMBEK J. (1961). On the Calculus of Syntactic Types. In : Jakobson (1961), 166-178.
- LAMBEK J. (1976). A Mathematician Looks at French Conjugation. *Linguistic Analysis* 2, 203-214.
- LAMBEK J. (1987). Note. In : M. Moortgat, R.T. Oehrle, M. McGee Wood (eds), *Categorical Grammar Newsletter* 1, 19.
- LAMBEK J. (1988). Categorical and categorial grammars. In : R.T. Oehrle *et al.* (eds), 1988, 297-318.
- LAMBEK J. (1999). Type Grammar Revisited. In : A. Lecomte *et al.* (eds), *Logical Aspects of Computational Linguistics*. Berlin & Heidelberg : Springer.
- LAMBEK J. (2000a). Pregroups : a new algebraic approach to sentence structure. In : C. Martin-Vide, G. Paun (eds), *Recent Topics in Mathematical and Computational Linguistics*. Bucharest : Editura Academici Romane, 182-195.
- LAMBEK J. (2000b). Type grammar meets German Word Order. *Theoretical Linguistics* 26, 19-30.
- LAMBEK J. (2001). Type Grammars as Pregroups. *Grammars* 4, 21-39.
- MONTAGUE R. (1974). *Formal Philosophy. Selected Papers of Richard Montague*. Richmond H. Thomason (ed.), New Haven & London : Yale University Press.

- MOORTGAT M. (1988a). Mixed composition and discontinuous dependencies. In : R.T. Oehrle *et al.*, (eds), 1988, 319-348.
- MOORTGAT M. (1988b). *Categorial Investigations : logical and linguistic aspects of the Lambek calculus*. Dordrecht : Foris.
- MOORTGAT M. (1996). Multimodal Linguistic Inference. *Journal of Logic, Language and Information* 5, 349-385.
- MOORTGAT M. (1997). Categorial type logics. In : J. van Benthem, A. ter Meulen (éds), *Handbook of Logic and Language*. Elsevier/MIT Press, Chapitre 2, 93-177.
- MOORTGAT M. (1999). Meaningful Pattern. In : J. Gerbrandy, M. Marx, M. de Rijke & Y. Venema (eds), *JFAK Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*. Amsterdam : UAP, 93-177.
- MOORTGAT M. (2001). Structural equations in language learning. In : Ph. de Groote, G. Morrill, C. Retoré (eds.), *Logical Aspects of Computational Linguistics*. LNAI 1001, Berlin: Springer, 1-16.
- MOORTGAT M. (2002). Quelques thérapies en logique des types pour le problème de Sapir. *Langage* 148, 111-123.
- OEHRLE R.T., BACH E., WHEELER, D. (eds) (1988). *Categorial Grammars and Natural Languages Structures*. Dordrecht : Reidel.
- RIVENC F., SANDU G. (2009). *Entre logique et langage*. Paris : Vrin.
- SHAUMYAN S.K. (1982). The goals of linguistic theory and applicative grammar. *Mathématiques et sciences humaines* 77, 7-42.
- SHAUMYAN S.K. (1987). *A Semiotic Theory of Natural Language*. Bloomington : Indiana University Press.
- STEEDMAN M. (1988). Combinators and Grammars. In : R.T. Oehrle *et al.*, (eds), 1988, 417-442.
- STEEDMAN M. (1991). Type-raising and directionality in combinatory grammar. In : *Proceedings of ACL91*. Berkeley, CA : Association for Computational Linguistics, 71-78.
- USZKOREIT H. (1986). Categorial Unification Grammars. In : *Proceedings of Coling 86*. Bonn : Association for Computational Linguistics, 187-194.
- WOOD M. McGee (1989). *A categorial syntax for coordinate constructions*. Doctoral dissertation, University College London, 1988. University of Manchester, Department of Computer Science, Technical Report UMCS-89-2-1.
- WOOD M. McGee. (1993). *Categorial Grammars*. London & New York : Routledge.
- ZEEVAT H. (1988). Combining categorial grammar and unification. In : U. Reyle, Ch. Rohrer (eds), *Natural Language Parsing and Linguistic Theories*. Dordrecht : Reidel, 202-229.