

## La méthode du lexique-grammaire

### 3.1. Origines et principes théoriques du lexique-grammaire

La méthode du lexique-grammaire a vu le jour à la fin des années 1960 au L.A.D.L. (Laboratoire d'Automatique Documentaire et Linguistique) de l'Université Paris VII. Son fondateur, Maurice Gross, est un ingénieur polytechnicien fasciné par la linguistique et se fixe un objectif ambitieux : procéder à une description fine du lexique français en vue d'applications informatiques (notamment, la traduction automatique). Le lexique-grammaire puise ses racines dans la grammaire transformationnelle de Zellig S. Harris, que nous allons illustrer rapidement.

#### 3.1.1. La grammaire transformationnelle de Zellig S. Harris

La grammaire distributionnelle de Zellig S. Harris est inspirée du travail de Bloomfield sur les classes formelles finalisé dans le distributionnalisme, courant linguistique né dans les années 1940 et fondé sur le concept de *distribution*. Ce dernier consiste en la détermination des règles régissant la mise en relation des unités structurales de la langue. La notion de distribution l'emporte sur les notions de *fonction* et de *signification*, se configurant ainsi comme la seule relation pertinente entre ces unités. Le concept de distribution en appelle deux autres, ceux d'*environnement* et de *substitution*. Le premier désigne les co-occurrents d'un élément de la langue (c.-à-d., les éléments qui se trouvent à sa gauche et à sa droite dans un énoncé). Le deuxième est une opération qui permet de vérifier si deux éléments de la langue ont la même distribution, les substituant l'un à l'autre. Suivant Bloomfield, Harris affirme que la description du sens ne peut pas être séparée de la description de la forme.

Avec la publication de *Mathematical Structure of Language* (1968 ; traduit par M. Gross en français trois ans plus tard), Harris formalise la notion de *transformation*, sur laquelle il avait déjà commencé à travailler depuis 1952. Le but de la grammaire transformationnelle est d'identifier les combinaisons d'éléments linguistiques qui forment des phrases acceptables. Plus que la construction des phrases, la notion de transformation s'intéresse à découvrir quelles relations les phrases entretiennent entre elles :

« Notre méthode comporte tout d'abord une théorie des relations entre les phrases. Cette approche de la grammaire consiste à se demander en premier lieu, non pas comment les phrases sont segmentées [...], mais comment les phrases sont reliées entre elles. La relation de base qui se trouve établie ici joue entre les schémas de phrases qui exigent les mêmes choix de mots pour fournir des phrases acceptables.

[...] En mettant en lumière entre les schémas de phrases une relation stable qui concerne l'acceptabilité du choix des mots, la grammaire transformationnelle se donne la possibilité de restreindre le problème du choix des mots à l'ensemble fini des phrases élémentaires ; à partir de là, elle étend à l'aide de transformations récursives les acceptabilités de choix de mots à l'ensemble dénombrable des phrases. » (1971 : 55-56)

La notion de transformation sert ainsi de base méthodologique à l'étude de la structure des phrases. Harris est également connu pour ses travaux en analyse du discours,

notamment pour ses travaux sur les *sub-languages* (sous-langages), dont HARRIS et al. (1989) sur le sous-langage de l'immunologie est un exemple.

### 3.1.2. De la grammaire transformationnelle harrissienne au lexique-grammaire

Suite aux études de Harris sur les transformations, deux importants modèles théoriques voient le jour : le générativisme de Noam Chomsky aux Etats-Unis (fin des années 1950) et le lexique-grammaire de Maurice Gross en France (fin des années 1960).

Bâti autour des notions d'*innéité* et de *créativité*, le générativisme chomskyen retravaille la notion harrissienne de transformation et professe la suprématie de la syntaxe, se proposant d'expliquer les irrégularités du lexique par des règles de type syntactico-transformationnel.

Il n'en est pas de même dans le lexique-grammaire mis au point par Maurice Gross, dont le premier pas est la description de 3000 verbes du français régissant une complétive ou une infinitive (GROSS 1968). Gross proclame la différence de base du lexique-grammaire par rapport à toutes les autres études transformationnelles dans son *Méthodes en syntaxe* (1975), à savoir l'étude des phrases simples – et non pas des phrases complexes<sup>106</sup> – et des transformations qu'elles peuvent subir.

Tant les courants de la syntaxe générative (Chomsky) que celui de la sémantique générative (Fillmore, Lakoff) sont ciblés par le linguiste français, qui souligne que l'énonciation de règles sur les faits linguistiques ne peut pas se faire sur des a priori, mais doit suivre des critères strictement empiriques. Particulièrement visé est le traitement des irrégularités :

« Il importe de bien noter que les notions de règle et d'exception sont strictement statistiques. [...] Il est nécessaire d'avoir fait un décompte précis des cas possibles pour arriver à des telles situations. Or [...] ce n'est jamais le cas, ni en grammaire traditionnelle, ni en grammaire transformationnelle.

Il est donc surprenant de voir se constituer une théorie des règles et des exceptions (Lakoff 1970) sans que son auteur se soit jamais préoccupé d'effectuer un dénombrement quelconque des éléments concernés par ses exemples de règles. De même, Chomsky (1969) a opéré une distinction entre phénomènes "généraux" traités par des transformations, et phénomènes "particuliers", c'est-à-dire "rares", traités au moyen d'une représentation lexicale abstraite. » (1975 :224)

Une autre profonde divergence du lexique-grammaire avec le générativisme concerne la notion de créativité, assimilée par Chomsky au caractère infini du langage :

« Nous considérons que compte tenu des possibilités combinatoires existant au niveau des seules phrases simples, la notion de créativité pourrait très bien trouver une place dans le cadre de processus finis. » (1975 : 18)

---

<sup>106</sup> En 1990, M. Gross définira le lexique-grammaire de la façon suivante : « Nous appelons lexique-grammaire l'ensemble des classes d'équivalence des phrases élémentaires, ensemble qui sert de générateur pour les phrases complexes. » (1990 :45)

GROSS 1981 reprend cet aspect, insistant sur le fait que les mécanismes récursifs<sup>107</sup> dont dispose la syntaxe de la plupart des langues naturelles constituent un ensemble fini. Autrement dit, si une créativité dans la langue existe, elle est circonscrite à la *forme* des phrases, n'en atteint pas le *sens* (1981 : 8).

En outre, bien que la grammaire transformationnelle de Harris soit le point de départ du lexique-grammaire, GROSS indique aussi une différence avec son maître :

« Harris considère que le sens ou l'information sont limités aux phrases noyaux, et que les variations de formes (i.e. les transformations) constituent un « bruit » qu'il est nécessaire d'éliminer pour atteindre le sens. Il met donc l'accent sur l'étude des transformations qui, appliquées aux phrases complexes, permettent de les analyser en phrases simples. Harris obtient donc la description des phrases simples comme une simple conséquence (un résidu) de la description transformationnelle. Malgré certaines difficultés de détermination a priori de la notion de phrase simple, nous avons pris celle-ci comme point de départ. » (1975 : 19)

Le lexique-grammaire se configure donc comme « une vérification expérimentale de la validité de la théorie transformationnelle » (1975 : 221), rigoureusement empirique, accordant une grande importance à la description en extension des données langagières. Le recensement des formes syntaxiques d'une langue ne peut pas s'appuyer sur des suppositions abstraites, mais doit se faire à partir d'inventaires. La méthode adoptée consiste dans la construction d'exemples et de contre-exemples de phrases simples (prenant le soin d'analyser les différentes interprétations en cas de phrases ambiguës), soumis à des tests d'acceptabilité par le biais de l'introspection, évitant de restreindre les sources d'informations à des corpus<sup>108</sup>. L'objectif est de parvenir à une « description exhaustive du français » (1975 :20) selon ses propriétés distributionnelles et transformationnelles.

### 3.1.3. Points de rupture avec la grammaire traditionnelle

Les critiques de Gross ne ciblent pas uniquement le générativisme. De même, le linguiste signale les limites de la théorie grammaticale traditionnelle. Tout d'abord, le fait que cette dernière soit fondée sur l'association entre forme et sens :

« Les grammairiens ont souvent tenté de relier les formes du langage à leur sens. [...] Ils ont constaté que certaines formes réapparaissaient souvent dans le langage, et que ces mêmes formes déclenchaient à chaque fois la même intuition de sens (par exemple, le suffixe *-ment* combiné à des adjectifs divers déclenche l'intuition de manière). Nous dirons que les grammairiens ont alors opéré une attribution *absolue* de sens à certaines formes. La méthode traditionnelle soulève de nombreuses difficultés. Pour toutes ces descriptions, il est aisé de montrer que l'association opérée entre sens et forme n'a aucune généralité, les exemples corrects étant aussi nombreux que les exceptions. » (1975 : 30)

---

<sup>107</sup> Quelques exemples de mécanismes récursifs cités dans l'article: la coordination, la subordination circonstancielle, la subordination complétive, la relativation.

<sup>108</sup> La méthode introspective est préférée au dépouillement de corpus surtout pour l'observation des propriétés transformationnelles et distributionnelles des phrases simples. Toutefois, l'utilisation des corpus dans le lexique-grammaire a progressé au fur et à mesure que la taille des corpus disponibles augmentait et des systèmes de dépouillement de corpus ont vu le jour au sein du L.A.D.L., parmi lesquels le système INTEX (Silberstein 1993). Sur le recours aux corpus en tant que réservoirs d'exemples, voir LAPORTE 2007.

La grammaire traditionnelle, autant que le générativisme, a eu la limite de fournir des descriptions des classes grammaticales en intension et non pas en extension :

« Les grammairiens ont toujours cherché à prédire certaines propriétés à partir d'autres, de nature éventuellement différente. Mais les classes définies par les propriétés n'étant jamais définies en extension, les contre-exemples à ces prédictions sont toujours nombreux et faciles à trouver. » (1975 : 217)

D'une façon générale, on peut dire que les reproches que Gross fait à ses prédécesseurs grammairiens concernent, d'un côté, le manque d'empirisme, d'un autre, la terminologie grammaticale. Tel est le cas de la discussion autour de la dénomination de *pronom* (GROSS 1968), pour en citer un exemple.

En ce qui concerne l'objet de la description linguistique, Gross tient à se démarquer des lignes-guides établies par Saussure (les élèves de Saussure, plutôt) dans le *Cours de Linguistique Générale*, surtout de la séparation entre l'étude synchronique et l'étude diachronique de la langue. Il serait difficile d'isoler un état de langue relatif à une époque donnée, car des traces d'un état de langue d'une autre époque pourraient y persister. Tel est par exemple le cas des éléments du français classique demeurant dans le français contemporain littéraire ou cultivé. Voulant trouver une analogie dans une autre science en faveur de la non-séparation des deux descriptions, GROSS se sert de l'étude des fossiles en biologie et de la relation que cette dernière entretient avec l'étude des animaux vivants :

« [...] La distinction saussurienne diachronie-synchronie [...] des données linguistiques est acceptée par la quasi unanimité des linguistes contemporains, et pourtant la justification qu'en ont donné les auteurs du *Cours de Linguistique Générale* attribué à Saussure est loin d'être convaincante. [...] « Saussure » n'opère que par analogie. Il introduit une classification des sciences dans laquelle « les sciences opérant sur des valeurs » sont privilégiées, la linguistique et l'économie politique sont de telles sciences. [...] Il est possible que cette classification ait été justifiée par quelqu'auteur du dix-neuvième siècle, mais elle apparaît aujourd'hui comme totalement arbitraire. (1975 : 226)

[...] Il est par contre une analogie qui nous semble présenter un intérêt plus direct du point de vue de cette opposition, et qui était disponible à l'époque où Saussure donnait son cours. Il s'agit de l'analogie avec la classification biologique. Les problèmes apparaissent comme entièrement parallèles. D'une part se pose le problème de décrire des animaux vivants, de l'autre des langues vivantes. Mais de la même manière que la biologie est concernée par les fossiles, la linguistique est concernée par les langues mortes. [...] La seule différence qui existe entre l'étude des fossiles et celle des animaux vivants est d'ordre strictement pratique, certainement pas théorique. [...] La situation est exactement la même en linguistique. Alors que le locuteur d'une langue vivante peut construire de nombreux exemples significatifs, le linguiste étudiant une langue morte est réduit à n'utiliser que du corpus, c'est-à-dire des données très rares. En particulier, il est pratiquement privé de toutes les informations de non-acceptabilité.

[...] « Saussure » a également essayé de donner un argument purement linguistique en faveur de la séparation des descriptions. [...] Autrement dit, lorsque l'on décrit un état de langue, on n'a pas à faire appel à d'autres états de la langue. Or, [...] dans de nombreuses descriptions synchroniques, il s'introduit des formes qui ont souvent une interprétation diachronique naturelle. » (1975 : 227)

« Nous considérons donc que l'objet de la linguistique n'a aucune raison d'être constituée d'un état synchronique de la langue. [...] Un état de langue comporte simultanément des niveaux diachroniques et dialectaux variés. » (1975 : 228)

### 3.1.4. Le pivot théorique du lexique-grammaire : la phrase simple

L'expression *phrase simple* est récurrente dans le discours théorique et méthodologique de Maurice Gross. L'insistance de Gross sur les phrases simples se comprend mieux lorsqu'on découvre le pivot théorique de cette méthode, qui considère que « les entrées du lexique ne sont pas des mots, mais des phrases simples » (1981 : 48). Ce qui pourrait sembler une affirmation quelque peu forte, mais qui en réalité

« [...] n'est en contradiction avec les notions traditionnelles de lexique que de façon apparente. En effet, dans un dictionnaire, il n'est pas possible de donner le sens d'un mot sans utiliser une phrase, ni de contraster des emplois différents d'un même mot sans les placer dans des phrases. » (*Ibid.*)

La combinatoire syntaxique de la phrase élémentaire devient l'objet de la méthode lexico-grammaticale sous un nouveau jour. Traditionnellement, les phrases en linguistique sont formalisées en termes de fonctions syntaxiques ou **prédicats** et de variables sémantiques ou **arguments**. L'hypothèse sous-jacente à ce type de description est que le verbe est souvent la partie du discours ayant la fonction de prédicat. Or, la théorie dérivationnelle formulée par HARRIS 1964 a avancé l'hypothèse que la fonction prédicative peut aussi être assumée par des noms et des adjectifs : c'est justement la vérification de cette hypothèse sur des bases empiriques qui donne la propulsion aux travaux du lexique-grammaire et à une nouvelle représentation des unités lexicales en termes de prédicats et arguments<sup>109</sup>, qui s'organisent dans la langue selon le procédé de **cooccurrence** ou **sélection**.

Choisir la phrase simple comme unité minimale du lexique entraîne aussi d'autres conséquences, dont la plus évidente est la justification de la notion de verbe support (*Vsup*), indispensable dans l'étude des prédicats nominaux<sup>110</sup>. On peut rapprocher les verbes supports des *Funktionsverben* étudiés par certains linguistes allemands (surtout, Peter von Polenz, à l'origine de cette dénomination) dans les années 1960-70 (GIRY-SCHNEIDER 1987 : 5). Un verbe support est un verbe sémantiquement « vide » qui a le rôle d'actualiser, dans une phrase simple, un mot prédicatif qui n'est pas un verbe. Des exemples typiques de verbes supports sont *être Prép*, *avoir*, *donner*, mais aussi *faire*, étudié en détail par GIRY-SCHNEIDER (1979, 1987). Nous empruntons à VIVÈS (1993 : 10) la suite d'exemples ci-dessous, pour bien mettre en évidence comment la prédication touche des catégories lexicales autres que le verbe et la structure des constructions à verbe support. Ainsi, étant donné les phrases simples (à peu près sémantiquement synonymes) :

- (1) *Luc admire ce tableau*
- (2) *Luc a de l'admiration pour ce tableau*
- (3) *Luc est en admiration devant ce tableau*
- (4) *Luc est admiratif devant ce tableau*

nous dirons que l'élément prédicatif déterminant les autres arguments de la phrase est : en (1), le verbe *admirer*, en (2) et (3) le substantif *admiration*, en (4) l'adjectif *admiratif*. Comme nous pouvons le constater, les verbes *avoir* et *être* des exemples (2) à (4)

---

<sup>109</sup> « Une phrase simple est composée d'un élément prédicatif et de ses arguments. » (VIVÈS 1993 : 8)

<sup>110</sup> « Étudier la syntaxe d'un mot prédicatif, c'est étudier la syntaxe d'une phrase simple. » (VIVÈS 1993 : 10)

n'apportent aucune information sémantique susceptible de modifier le sens de la phrase : leur rôle est l'actualisation des éléments prédicatifs.

Certains noms prédicatifs mettent en jeu aussi d'autres verbes supports, considérés comme des extensions de verbes supports et qui ne sont pas sémantiquement vides : un verbe support ne peut être défini comme tel a priori, c'est sa combinaison avec un nom spécifique qui lui confère le statut de verbe support. Ainsi, le verbe *écrire* peut assumer le rôle de verbe support dans

(1) La lettre de Marie à Paul [est arrivée]

= La lettre que Marie a (faite+écrite) à Paul [est arrivée] (GIRY-SCHNEIDER 1987 : 21).

L'élargissement de la fonction prédicative aux noms et aux adjectifs et l'accent mis sur les contraintes syntaxiques impliquent une nouvelle attitude à l'égard du lexique, dont l'étude ne peut pas être séparée de l'étude de la grammaire (d'où la dénomination lexique-grammaire). En raison de cela, le lexique se voit attribuer une importance dans la syntaxe qu'il n'avait pas dans des modèles théoriques précédents, en particulier depuis Bloomfield. D'après LAMIROY, les travaux du lexique-grammaire ont démontré que le lexique n'est pas le « locus de l'irrégularité », contrairement à la grammaire, considérée comme « la composante par excellence des régularités » :

« Les travaux du lexique-grammaire ont précisément montré que des règles relevant de la grammaire scolaire aussi élémentaires que la règle du passif, par exemple, sont moins bien 'régulières' qu'il n'y paraît à première vue, précisément à cause de la variation lexicale à laquelle elles sont soumises. Et inversement, un des grands mérites du modèle à mon sens est d'avoir démontré que le lexique s'organise malgré tout en ensembles sémantico-syntaxiques relativement homogènes, identifiés à force de soumettre des masses d'entrées lexicales à des dépouillements syntaxiques systématiques. » (1998 : 11)

Nous venons de voir que la syntaxe revêt une importance capitale dans le lexique-grammaire. Mais qu'en est-il du traitement du sens ? La phrase simple, étant l'unité minimale du lexique, est en même temps l'unité minimale de sens. La valeur sémantique d'une unité lexicale ne peut pas être déterminée si cette dernière n'est pas insérée dans une phrase : l'environnement syntactico-distributionnel d'un mot détermine aussi son sens. Les différences de distribution permettent ainsi de traiter la polysémie, un des « problèmes » traditionnellement associés au sens, par le biais de la syntaxe<sup>111</sup>. Nous verrons cela lorsque nous aborderons l'analyse de nos corpus d'étude.

### 3.2. Description du lexique du français : les travaux du L.A.D.L.

Nous avons vu plus haut (§3.1.2.) que l'étude de la combinatoire syntaxique du français a commencé en 1968 avec l'analyse de 3 000 verbes régissant une complétive ou une infinitive, reprenant en partie les travaux faits par Harris sur l'anglais. Prenant des distances par rapport à la description linguistique par arbres syntaxiques, chère au générativisme, Gross adopte un système de représentation par matrices binaires, qui

---

<sup>111</sup> Il faut préciser que, malgré la priorité accordée à la syntaxe dans les descriptions, des notions de sens sont prises en considération dans la démarche lexico-grammaticale lorsqu'elles présentent une certaine reproductibilité. C'est le cas, par exemple, des informations sémantiques contenues dans quelques intitulés des tables de lexique-grammaire, comme nous le verrons dans le paragraphe suivant.

deviendront les célèbres « tables » du lexique-grammaire (voir §2.2.1.). Pour chaque verbe sont identifiés les arguments nécessaires à sa description, notés  $N_0$ ,  $N_1$ ,  $N_2$ . Dans la phrase simple choisie comme construction de base pour représenter une entrée lexicale  $N_0$  indique la position sujet et  $N_1$ ,  $N_2$  les autres compléments.

Les études entamées par Gross sur les complétives et les infinitives sont étendues à d'autres types de phrases : citons en particulier la thèse de BOONS sur les constructions transitives à compléments locatifs (1974) et les travaux sur la structure des constructions intransitives (BOONS, GUILLET et LECLÈRE 1976). L'attention de Gross et de ses collègues ne se focalise pas uniquement sur la syntaxe du verbe, mais également sur la syntaxe du nom, comme LABELLE 1974 et GIRY-SCHNEIDER 1978, respectivement sur les constructions nominales à verbe opérateur *avoir* et sur les groupes nominaux à verbe support *faire*. L'étude du nom sous l'angle de la prédication constitue une nouveauté par rapport à de nombreux modèles théoriques en linguistique.

Au début des années 1980, outre la combinatoire de la phrase libre, Gross étend ses recherches aux éléments figés (ou composés) du français. L'étude des phrases figées est une conséquence directe de l'étude sur la combinatoire de la phrase simple, qui a permis de constater que de nombreuses séquences lexicales ont un sens non-compositionnel et que le phénomène du figement intéresse toutes les catégories lexicales. Une catégorie de séquences figées à laquelle Gross se consacre particulièrement est celle des structures adverbiales. Son analyse de la syntaxe de l'adverbe naît du constat que cette catégorie ait été plutôt délaissée par les grammairiens (GROSS 1988). Outre les séquences adverbiales, d'autres études lexico-grammaticales sont consacrées aux séquences adjectivales.

Nous avons vu qu'à la base du lexique-grammaire il y a des applications informatiques concernant les langues naturelles. Toutefois, l'attitude de Gross à l'égard des ordinateurs a toujours été plutôt prudente. Tout en reconnaissant le rôle fondamental des machines dans l'exploration des textes, l'organisation et le traitement des données linguistiques, il place la description syntaxique sur le plan des tâches non automatisables.

Depuis 2000, l'équipe du L.A.D.L. a été intégré au sein du LIGM (Laboratoire Informatique Gaspard Monge) de l'Université paris-Est Marne-la-Vallée<sup>112</sup>, où elle a continué ses travaux même après la mort de Maurice Gross (2001). Dans une étroite collaboration entre linguistes et informaticiens, d'importants projets ont été accomplis, notamment concernant l'exploitation informatique des tables (voir §3.2.1.) et la mise au point de logiciels et de programmes pour le TAL (Traitement Automatique des Langues), dont l'exemple le plus significatif est le logiciel Unitex.

Commencés sur le français, les travaux du lexique-grammaire ont été élargis dès les années 1970 à d'autres langues, comme l'italien, le portugais, l'espagnol, le grec moderne, le roumain, le coréen, le hongrois, le malgache<sup>113</sup>. Chaque année, un colloque général du lexique-grammaire a lieu et se veut un moment d'échange entre les différentes équipes.

---

<sup>112</sup> Université qui, depuis 2009, fait partie du pôle universitaire Paris-Est, avec l'ancienne Université Paris XII.

<sup>113</sup> Nous renvoyons aux bibliographies suivantes pour la liste complète des publications ayant trait au lexique-grammaire : 1) LECLÈRE, Christian et SUBIRATS-RÜGGERBERG, Carlos, 1991, « A bibliography of studies on lexicon-grammar », *Linguisticae Investigationes* XV : 2, Amsterdam/Philadelphia, John Benjamins, pp. 347-409 ; 2) LECLÈRE, Christian, 1998, « Travaux récents en lexique-grammaire », Travaux

### 3.2.1. Formalisation des données du lexique-grammaire : tables, dictionnaires électroniques et grammaires locales

Les données lexicales analysées dans les études lexico-grammaticales ont été formalisées principalement dans des tables, des dictionnaires électroniques et plus récemment, dans des grammaires locales.

Comme nous l'avons dit plus haut, les tables ont la forme de matrices binaires. Chaque table regroupe une classe d'éléments lexicaux de la même catégorie présentant des traits communs et décrit les propriétés syntaxiques qu'ils peuvent posséder (c'est-à-dire, des types de phrases dans lesquelles ils peuvent entrer). Les entrées lexicales font l'objet des lignes, tandis que les propriétés syntaxiques sont listées dans les colonnes. L'intersection entre une ligne et une colonne renseigne sur l'acceptabilité d'une forme par le biais des signes '+' et '-', selon que le mot entre ou pas dans un certain type de phrase. Les propriétés définitoires sont de deux types : distributionnelles et transformationnelles. Les premières se réfèrent aux caractéristiques que peuvent assumer les arguments d'un mot prédicatif. Dans cette catégorie entrent, par exemple, la différenciation entre noms humains (*Nhum*) et noms non humains (*N-hum*), la notation *Nnr*<sup>114</sup>, et les compléments prépositionnels. Pour ce qui concerne les propriétés transformationnelles, elles désignent les équivalences entre phrases rendues explicites par un lien de paraphrase. Tel est le cas des réductions de prépositions, des réductions de complétives, du passif, de la pronominalisation, de l'extraposition, pour n'en citer que quelques exemples<sup>115</sup>. Lorsqu'un même mot possède des structures syntaxiques différentes, mais véhiculant des sens différents, l'entrée est dédoublée.

L'élaboration des tables s'est étalée sur une quarantaine d'années, rencontrant ainsi des changements de support électronique (cartes perforées, bandes magnétiques, disquettes) et de format, suivant les progrès de l'informatique. Dans une première version, les tables étaient disponibles sur papier et sur cartes perforées. En 2002, 60% des tables informatisées sont rendues disponibles en ligne sur le site du LIGM : il s'agit de la version 2. Les tables ont connu d'importantes modifications et améliorations depuis 2008, qui sont précisément détaillées dans la thèse de TOLONE (2011), suite à laquelle la totalité des tables a été mise en ligne<sup>116</sup>. Les ressources actuellement disponibles pour le français sont les suivantes : 67 tables de verbes distributionnels simples, 81 tables de noms prédicatifs simples et composés, 69 tables d'expressions figées (surtout verbales et adjectivales) et 32 tables d'adverbes, pour un total de plus de 78 000 entrées.

---

de linguistique, n°37, Bruxelles, De Boeck & Larcier s.a., pp. 155-186 ; 3) <http://infolingu.univ-mlv.fr/> (< Bibliographie).

<sup>114</sup> « La notation *Nnr* ne correspond pas à une classe de substantifs, elle spécifie une position syntaxique comme pouvant recevoir un substantif d'une nature sémantique quelconque, une complétive ou une infinitive. » (GROSS 1975 : 50)

<sup>115</sup> « Une propriété définitoire peut aussi indiquer qu'un élément de la table entre dans deux constructions, qui sont généralement reliées par un lien de paraphrase. Cette deuxième construction est appelée propriété transformationnelle, car elle est déductible de la première par une redistribution, la première étant la construction de base. » (TOLONE 2011 : 67)

<sup>116</sup> La version des tables actuellement téléchargeable est la 3.4..

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	<TO> NO = Num	NO = Neg	NO = le fait Qu P	NO = VO-inf W	<ENT> P	Ent = Neg	<ENT> V	Neg	Aux = avoir	Aux = être	NI V	NI = Qu P	Qu P = VO-inf W	NI = Qu P = Aux VO-inf W	
1	1+														
2	2+														
3	3+														
4	4+														
5	5+														
6	6+														
7	6+														
8	7+														
9	8+														
10	9+														
11	10+														
12	11+														
13	12+														
14	13+														
15	14+														
16	15+														
17	16+														
18	17+														
19	18+														
20	19+														
21	20+														
22	21+														
23	22+														
24	23+														
25	24+														
26	25+														
27	26+														
28	27+														
29	28+														
30	29+														
31	30+														
32	31+														
33	32+														

Figure 1 : un extrait de la table V\_12.

Parallèlement à l'élaboration des tables, durant les années 1980 un système de dictionnaires électroniques a été mis au point au L.A.D.L., le système DELA. A la différence des dictionnaires traditionnels, les dictionnaires électroniques – à ne pas confondre avec les dictionnaires usuels sur support électronique, comme le *TLFi* – sont conçus pour être lus et exploités par des machines (des analyseurs syntaxiques, par exemple), donc les informations qu'ils contiennent doivent être explicites<sup>117</sup>. À chaque mot sont associées des étiquettes contenant ces informations (genre, nombre, catégorie grammaticale, plus mode et temps pour les verbes). Le premier dictionnaire électronique du système à être confectionné a été le dictionnaire DELAS, dictionnaire des formes simples. Par le biais de programmes informatiques, un dictionnaire des formes fléchies (DELAF) a été élaboré à partir du DELAS. Après le recensement des formes simples a été construit un dictionnaire des formes composées (DELAC)<sup>118</sup>. Ces dictionnaires ont été successivement utilisés par des logiciels comme INTEX et Unitex.

Le développement des logiciels que nous venons de citer a permis de formaliser les données lexicales dans des grammaires locales, qui sont des transducteurs représentables sous forme de graphes d'automates finis (GROSS 1997). La formalisation par automates finis, jugée trop simpliste par CHOMSKY (1956), a été en revanche adoptée par Gross comme un instrument particulièrement convenable à la description syntaxique.

<sup>117</sup> « Les dictionnaires électroniques sont destinés à des utilisations informatiques. Ils ont la forme de tables de données formalisées auxquelles des programmes divers ont accès. [...] Un dictionnaire électronique consacré à l'analyse de textes devra comporter des indications grammaticales codées et formatées utilisables par un analyseur syntaxique. [...] Un dictionnaire électronique syntaxique doit comporter des informations *systematiques* sur les possibilités de sujets et de compléments qui s'offrent pour chaque verbe, ainsi que des indications formelles sur les positions et les formes diverses que peuvent prendre les sujets et compléments dans des phrases quelconques. » (GROSS 1989 : 5)

<sup>118</sup> Nous citons également le dictionnaire phonétique DELAP (LAPORTE 1988) et le dictionnaire des racines DELAR (COURTOIS 1988).

### 3.3. Lexique-grammaire et discours spécialisés : état de l'art

Jusqu'à présent, nous avons passé en revue les travaux du lexique-grammaire portant sur la syntaxe de la « langue générale ». Nous nous proposons dans ce qui suit de faire le point sur les rapports entre lexique-grammaire et terminologie.

Quelle est la place accordée à la terminologie dans le discours théorique et dans les descriptions du lexique-grammaire ? En ce qui concerne le premier aspect, Gross s'exprime sur les termes dans deux articles, l'un consacré aux expressions figées, l'autre à la construction de dictionnaires électroniques (GROSS 1982, 1989), mais de façon très marginale. La terminologie en tant que science n'y est pas citée et aucune des problématiques liées au statut théorique du terme (§I.2.) n'est abordée par le linguiste français. Ce qui est tout à fait compréhensible, si l'on considère sa démarche strictement empirique et surtout la priorité accordée à la description des unités non techniques de la langue. Une autre contribution portant sur la terminologie, combinant des aspects théoriques et pratiques, est le *Rapport final à la Mission Interministérielle pour le Développement de l'Information scientifique et technique*, en collaboration avec Diane TREMBLAY<sup>119</sup> (1985). Les auteurs y étudient en détail le contenu d'une banque de données terminologiques, ayant trait au domaine du nucléaire<sup>120</sup>.

Illustrant une première classification des phrases figées du français, Gross mentionne aussi quelques exemples de vocabulaires techniques (le vocabulaire culinaire), qui obéissent aux mêmes dynamiques que le vocabulaire ordinaire, car « il n'est guère possible de séparer les vocabulaires techniques du vocabulaire ordinaire » (GROSS 1982 : 173). La continuité entre vocabulaire ordinaire et vocabulaires techniques est particulièrement visible dans la tendance de ces derniers à la composition :

« [...] La composition des noms [...] : le vocabulaire scientifique et technique fait largement usage de ce procédé. En fait, l'introduction de mots nouveaux construits de toutes pièces comme *radar* ou *laser* est plutôt rare, et le procédé général consiste à composer des mots existants comme dans

*canon à électron, mémoire de masse, génie génétique, changement de vitesse, verre de contact, poêle à frire.* » (Ibid.)

Le procédé de la composition explique ainsi les dimensions des dictionnaires techniques, où les termes composés sont bien plus nombreux que les termes simples (1982 : 174).

Particulièrement attentif au procédé de la composition est le *Rapport* de 1985, qui s'attache à décrire en détail les types de structures syntaxiques des noms composés du domaine étudié :

« [...] On observe que dans la plupart des domaines, la terminologie scientifique et technique est constituée de mots composés : *écran cathodique, terminal graphique couleur, accélérateur de particules*, etc.

---

<sup>119</sup> « Le présent rapport porte sur les méthodes de traitement automatique des termes scientifiques et techniques en vue de leur incorporation dans une banque de terminologie multilingue et de la maintenance du contenu d'une telle banque. » (1985 : 4)

<sup>120</sup> Toutefois, l'ouvrage ne manque pas de références au lexique des assurances. Ce type de lexique, qui n'est pas aussi technique que l'énergie nucléaire, sert à montrer la continuité avec la langue courante.

[...] Cette observation constitue une constante langagière fondamentale, et elle nous conduit à adopter une approche linguistique à la formation des termes composés. En conséquence, nous ne traiterons ici que des procédés de formation des termes qui font appel à la grammaire de la langue ordinaire : adjonctions d'épithètes ou de compléments de noms pour l'essentiel. Ces procédures affectent tous les domaines et toutes les langues. » (1985 : 5)

GROSS souligne que le phénomène de la composition a souvent été délaissé en lexicographie, principalement pour trois raisons : 1) le nombre des mots composés, bien plus élevé que les mots simples ; 2) la difficulté à définir le mot composé ; 3) les procédures d'inclusion dans un dictionnaire (1985 : 6-8). Au moins pour ce qui concerne les points 1) et 3), il s'agit de contraintes liées en partie au support des dictionnaires traditionnels, difficultés qui peuvent être surmontées par l'outil informatique.

L'analyse des séquences composées permet à Gross d'appliquer à la terminologie quelques méthodes fondamentales de la linguistique transformationnelle, insistant sur la composante linguistique des termes et dénonçant l'idée reçue que ces derniers sont des mots de contenu lexical fixe :

« D'une part, on connaît les phénomènes de raccourcissement des termes, lorsqu'ils sont employés répétitivement dans des textes. Il peut s'agir d'abréviations plus ou moins arbitraires qui vont du sigle au mot-clé, mais il existe de nombreuses réductions qui obéissent aux règles générales de la syntaxe. C'est ainsi que l'on observe l'application

- de règles de réduction des conjonctions :

*le taux d'intérêt et le taux d'amortissement*  
= *le taux d'intérêt et d'amortissement*

- de règles de pronominalisation :

*le taux d'intérêt et le taux d'amortissement*  
= *le taux d'intérêt et celui d'amortissement*

Mais ils existent d'autres possibilités de variations dues à la présence de parties variables combinées à des parties fixes de termes. Considérons les formes :

*le taux d'intérêt de N*  
= : *le taux d'intérêt des banques centrales*  
= : *le taux d'intérêt de 12%*

elles se présentent avec une partie fixe *taux d'intérêt* à laquelle on adjoint un complément de nom variable de la forme *de N*. Toutefois, les deux formes montrent qu'il existe deux catégories différentes de compléments. Cette intuition est confirmée par la possibilité de combiner les deux compléments dans un même groupe nominal :

(1) *le taux d'intérêt de N de N*  
= : *le taux d'intérêt de 12% des banques centrales*

En fait, ces deux compléments s'analysent en termes de phrases simples à verbes supports, qui sont ici :

*Le taux d'intérêt est de 12%*

*Les banques centrales (ont, pratiquent) un certain taux d'intérêt*

Les verbes supports sont ici *être de* et *avoir*, ou sa variante *pratiquer*. En conséquence, la forme "pronominal" leur taux d'intérêt ne peut avoir qu'une interprétation, celle où leur = les banques centrales. [...] Ces verbes support [...] constituent l'outil de construction de la quasi-totalité des noms composés. » (1985 : 9-10)

La particularité de cette étude, surtout si l'on considère l'année de publication, est l'attention consacrée à la variation des déterminants des noms, qui, d'après les auteurs, pourrait avoir des « conséquences imprévisibles sur la taille du lexique » à décrire dans une base de données (1985 :11). Une partie considérable de l'étude est donc consacrée aux déterminants, notamment aux déterminants figés, ce qui n'est pas courant en terminologie.

Centré sur la typologie des structures syntaxiques des noms composés et de leurs comportements en discours, le *Rapport* jette les bases de la formalisation de la construction des dictionnaires électroniques des mots composés, dont nous avons déjà parlé (voir §3.2.1.). Il se veut également l'occasion pour réitérer la continuité entre langue courante et lexiques spécialisés (1985 : 58).

Toutefois, pour des raisons pratiques, il faut expliciter la différence entre vocabulaire courant et vocabulaires scientifiques et techniques. Comment traduire cette exigence ? C'est à ce propos que le linguiste français opère la distinction entre le noyau du lexique-grammaire et les lexiques satellites. Ces derniers correspondent à des situations variées et sont constitués, d'un côté, du vocabulaire scientifique et technique, d'un autre, des vocabulaires et formes syntaxiques dialectaux ou vieillis (1989 : 18). Rien n'empêche à un même mot de faire partie du noyau du lexique-grammaire et d'avoir pourtant des emplois dans un ou plusieurs lexiques satellites. De nouveau, la continuité entre vocabulaire courant et vocabulaire spécialisé ressort, suggérant déjà que les termes peuvent être décrits à l'instar de toutes les autres unités lexicales<sup>121</sup>.

Sans doute le petit nombre de travaux sur les lexiques spécialisés dans la bibliographie du lexique-grammaire s'explique-t-il par la priorité de décrire le « noyau du lexique-grammaire ». Néanmoins, il existe des études sur des secteurs d'activités variés. Ils ont une approche sensiblement différente par rapport aux travaux terminographiques traditionnels, et ils présentent des pistes de réflexion intéressantes pour la terminologie.

Toujours pour la langue française, nous signalons la grammaire des noms prédicatifs de maladie (LABELLE 1986), l'étude sur les constructions à verbes supports du lexique du sport, de la musique et des activités intellectuelles (PIVAUT 1989), les grammaires locales construites sur les bulletins boursiers (NAKAMURA 2005). Actuellement, un projet sur le discours de la médiation est en cours de réalisation, dans une perspective contrastive français-italien (CONENNA et DE GIOIA), qui, comme cette thèse, se voudrait une contribution à l'exploitation du lexique-grammaire en terminologie.

### *Pour résumer*

Dans ce chapitre, nous avons présenté la méthode du lexique-grammaire, une méthode de description linguistique inspirée de la grammaire transformationnelle de Zellig S. Harris. Nous avons rapidement introduit cette dernière (§3.1.1.), avant d'analyser le passage au lexique-grammaire (§3.1.2.), dont nous avons exposé les points de rupture avec la grammaire traditionnelle (§3.1.3) et le pivot théorique, la phrase simple (§3.1.4.).

Ensuite, nous avons passé en revue les travaux du L.A.D.L. (§3.2.), qui ont été formalisés en tables, en dictionnaires électroniques et en grammaires locales (§3.2.1.). Le

---

<sup>121</sup> Nous rappelons que cette continuité est mise en relief dans certains modèles terminologiques nés en réaction à la TGT et postérieurs à l'article de GROSS (notamment, GAMBIER 1991, CABRÉ 2003).

dernier point de cette présentation a été un état de l'art des travaux du lexique-grammaire sur les discours spécialisés (§3.3.).

