

Lexique-grammaire et sémantique

Dans les grammaires de Zellig Harris, l'extension lexicale des règles syntaxiques n'est pas spécifiée. Soit par exemple dans Harris (1976, 64) la construction de type *prédicat à deux arguments nominaux*, qu'il appelle construction *Onn*. Dans cet ouvrage, il n'entrait pas dans le projet de l'auteur d'énumérer les prédicats entrant dans telle ou telle construction et d'en faire la classification : on n'y trouve que cinq exemples de prédicats *Onn*, à savoir *manger, porter, penser, être près de* et *être le père de*. Cette tâche fut en revanche entreprise par Maurice Gross et son équipe, le LADL (Université Paris 7). Pour chaque construction générale, ces chercheurs ont travaillé à définir en extension la classe des prédicats (verbaux, adjectivaux et nominaux) qui peuvent en être la tête (autrement dit, ils les ont énumérés) ; et ils ont spécifié, d'une part les différentes formes syntaxiques (SN, SV infinitif, Proposition au subjonctif, SP en *à*, en *de*, etc.) que peuvent prendre les arguments des prédicats, d'autre part les alternatives de forme (transformations) caractérisant chaque type de construction syntaxique. Ces différentes spécificités définissent des sous-classes au sein de la catégorie générale. Les règles de syntaxe ayant été lexicalisées de la sorte, le LADL a produit un ensemble de grammaires, appelées lexiques-grammaires¹. Ces derniers se présentent sous la forme de tables, dans un format propice à des applications au TAL (Traitement Automatique des Langues). D'autre part, un des membres du LADL, Max Silberstein, a développé la plate-forme d'ingénierie linguistique INTEX (devenue ultérieurement NooJ) qui exploite la notion de *grammaire locale*, due à Gross, à des fins d'implémentation des lexiques-grammaires pour des tâches de linguistique de corpus.

Les lexiques-grammaires « historiques » du LADL ont été élaborés, pour diverses langues, au cours des 30 dernières années du 20^{ème} siècle. L'objet principal de cet article est de montrer, contre un préjugé assez largement répandu, que la préoccupation sémantique est centrale dans les grammaires de Maurice Gross et, en-deçà, dans celles de Zellig Harris. Nous présentons ensuite d'autres travaux, ceux de Gaston Gross et de Jean Dubois & Françoise Dubois-Charlier qui, tout en s'inscrivant dans la même tradition, assument le projet de décrire le sens plus explicitement que ne l'avait fait le LADL. Nous évoquons enfin notre propre *thésaurus syntactico-sémantique des mots d'affect* et son implémentation à des fins d'annotation sémantique automatique.

¹ Gross (1975) est le premier lexique-grammaire important à avoir été publié. Les ressources élaborées au LADL sont aujourd'hui disponibles sur le site du LIGM, Equipe Informatique Linguistique (<http://infolingua.univ-mlv.fr/>).

1. Le problème de la relation syntaxe-sens dans les lexiques-grammaires du LADL

La part plus ou moins grande part prise par la sémantique dans une théorie syntaxique peut être évaluée assez précisément à partir de la proportion de notions propres à cette théorie relevant respectivement de la syntaxe et de la sémantique. Par exemple, une théorie syntaxique où ne figureraient pas les notions d'*agent*, *patient*, *humain*, *animé*, *processif*, *statif*, *mouvement*, *événement* etc. devrait être considérée comme étant dépourvue de composante sémantique.

Ce semble être justement le cas des premiers lexiques-grammaires du LADL, tant y sont rares les notions fondamentales relevant de la sémantique ; de fait, dans Gross (1975), qui étudie le régime des constructions complétives, la seule notion sémantique est l'opposition *humain/non humain*. De là vient le préjugé courant selon lequel la théorie du lexique-grammaire serait une théorie syntaxique dépourvue de composante sémantique.

Pourtant la question de la sémantique est loin d'être absente dans cet ouvrage (cf. Gross (1975, 30-33), qui reprend Gross (1972)). Gross y affirme qu'il ne fait pas partie des grammairiens qui « ont complètement exclu le sens des descriptions », et démontre, à partir de paires d'exemples de phrases comparées « dans de bonnes conditions expérimentales », qu'il est possible d'introduire des notions sémantiques fines telles que celle d' *agent volontaire* (Gross préfère l'appeler « *actif* »). « Il y aurait là, conclut Gross (1975, 33), un progrès notable et donc un espoir sérieux de localiser avec précision un certain nombre d'éléments de sens, ce qui est une condition nécessaire au développement d'études plus complexes ».

Avant d'arriver à ce stade évolué de la description linguistique, il parut nécessaire et prioritaire d'entreprendre des inventaires exhaustifs des constructions syntaxiques et de leurs propriétés. C'est ce à quoi se sont attelés les chercheurs du LADL. Ces travaux ont mis en évidence un certain nombre de phénomènes qui, sans le parti-pris de l'exhaustivité, seraient restés négligés, parmi lesquels le caractère massif des expressions à verbe support et des expressions figées.

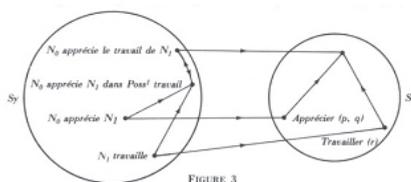
Certaines généralisations sémantiques n'en sont pas moins esquissées dès Gross (1975) dans des commentaires informels : il y est relevé par exemple que la Table 2 (*N0 V V-infinitif*) et la Table 3 (*N0 V V-infinitif*) comportent, respectivement, des « verbes de mouvement » et des « verbes causatifs de mouvement » ; et que la majorité des verbes de la Table 4 (*que P V NI*) « correspondent à un sentiment "déclenché" par N0 et "éprouvé" par N1 » (Gross 1975, 170). Ainsi se dévoile, de façon encore floue, un phénomène de correspondance entre les formes syntaxiques et l'intuition de leur sens². D'autre part, dans les lexiques-grammaires ultérieurs, vont être introduites de plus en plus de notions sémantiques ; c'est ainsi que figurent dans les tables de Guil-

² On peut rapprocher cette analyse de celle de Halliday (1985) qui observe que les constructions en *N0 V that NI* expriment toujours une pensée ou une déclaration.

let/Leclère (1992) des descripteurs linguistiques tels que « Nom abstrait », « Résultat », « Instrument », « Attache ».

2. L'article de Maurice Gross intitulé « Les bases empiriques de la notion de prédicat sémantique »

Dans Gross (1981), l'auteur fait la synthèse des résultats déjà obtenus par le LADL depuis dix ans (cf. entre autres Boons/Guillet/Leclère (1976); Daladier (1978); Gross (1975); Giry-Schneider (1978); Picabia (1978)). Le seul titre de l'article est une preuve de l'importance de la préoccupation sémantique dans les travaux du LADL. L'auteur y précise sa position sur la relation entre la forme syntaxique et le sens. A la page 20 de l'article, figure la figure suivante :



Le cercle de gauche figure la forme syntaxique. L'exemple est celui des deux constructions reliées par une transformation : *N₀ apprécie N₁ dans son travail* et *N₀ apprécie le travail de N₁* ; elles sont le produit de la composition des deux phrases élémentaires *N₀ apprécie N₁* et *N₁ travaille*. Quant au cercle de droite, il figure la forme sémantique homologue. Le commentaire de Gross est le suivant :

Le diagramme de la figure 3 suggère l'existence d'un morphisme entre *Sy* et *Se*, c'est-à-dire d'une redondance importante entre les deux ensembles, au point que l'un d'entre eux pourrait être inutile. *Se* pourrait en effet être éliminé, si les descriptions se faisaient toutes en termes d'un ensemble de phrases simples ; par convention, les unités de base du sens seraient portées par ces phrases simples. La description sémantique consisterait donc à décomposer les phrases complexes selon les phrases simples de base, elle ne différerait donc guère de la description syntaxique [...]. Cette position est celle de Harris 1968, 1976, 1978.

Ce qui est dit là, ce n'est pas que la description du sens doit être éliminée au profit de la seule description syntaxique, mais que les deux représentations *Sy* et *Se* seraient identiques, en tant que mettant en jeu la même opération, si on procédait à la décomposition de l'entité syntactico-sémantique complexe qu'est la phrase selon les entités syntactico-sémantiques simples qui la composent. Or c'est justement cette méthode de décomposition que pratiquent Harris et Gross. Ces considérations sont fondées sur une théorie de la compositionnalité du sens couplée à une théorie de l'ho-

momorphisme de la syntaxe et du sens (cf. son emploi du mot *morphisme*) sur laquelle on reviendra dans la Section 4.

3. La question du sens chez Zellig Harris

Contrairement, encore une fois, à un préjugé assez répandu, la question du sens est centrale, elle aussi, dans l'œuvre de Zellig Harris. Dans un livre qui a contribué à le faire connaître en France (Harris (1968)) grâce à la traduction qu'en a faite Catherine Fuchs (cf. Harris (1971)), figure une des opinions auxquelles Gross (1981) fait allusion dans l'extrait que nous avons cité en Section 2 :

Presque tout ce qu'on peut dire de la signification d'une phrase peut être obtenu directement à partir des significations et des positions occupées par les composantes φ , K^3 . Aussi est-il très peu besoin d'ajouter à cette théorie des transformations de base une théorie sémantique (Harris 1971, 232).

Dans un autre passage de l'ouvrage (Harris 1971, 230), Harris écrit :

Cette corrélation entre structure et vocabulaire va de pair avec une corrélation entre structure et signification. Chaque sous-classe de mots comprenant un opérateur ou une classe d'arguments particuliers possède un type de signification correspondant aux relations syntaxiques (*modalités* dans φv ; *connaissance, sentiments, etc.* dans φs ; etc.).

Il est intéressant de relever qu'Harris évoque des notions sémantiques particulières en les associant à des constructions syntaxiques homologues. En effet, « φv » est défini comme étant la classe des « opérateurs »⁴ portant sur un verbe », et il se trouve regrouper des auxiliaires aspectuels et modaux. Et « φs », défini comme étant la classe des « opérateurs portant sur une phrase », se trouve regrouper des verbes de connaissance et de sentiments tels que *to surprise, intrigue, know, wonder*, mais aussi des verbes de communication comme *to tell, ask, require*⁵.

Plus tard, Harris (1990, 19) revient sur la question du sens dans un passage de son autobiographie intellectuelle publiée dans le volume édité par Daladier (1990) sur *Les grammaires de Harris et leurs questions* :

La grammaire des opérateurs révèle une relation plus fine entre la structure d'une phrase et son contenu informatif (telle que Carnap et les Positivistes du cercle de Vienne semblent en avoir eu l'intuition). La base de cela est dans le fait de spécifier et d'ordonner les événements linguistiques non équiprobables. Certaines de ces ruptures d'équiprobabilité, qui donnent lieu à des structures, sont porteuses d'information (dans un sens apparenté à celui de la théorie mathématique de l'information).

³ Les symboles φ et K désignent respectivement les prédicats et les phrases élémentaires.

⁴ Par « opérateur », il faut entendre la même chose que « prédicat ».

⁵ Harris (1971, 83) fait une esquisse de la classification des « φs ». La note de bas de page due à C. Fuchs indique que ce passage a inspiré un mémoire de M. Gross élaboré à l'Université de Pennsylvanie et intitulé « Analyse transformationnelle des constructions verbales françaises ». Or ce mémoire est manifestement un précurseur de Gross 1975 (cf. *supra* Section 1).

Dans cet extrait, Harris enracine sa théorie de l'homologie de la structure syntaxique et de la structure syntaxique dans la double tradition du positivisme logique (dans le même article, il fait allusion aux entretiens qu'il a eus avec Carnap à Chicago après la deuxième guerre mondiale) et des théories de l'information de la fin des années 1940 aux USA (Shannon). Il fait plus qu'affirmer le principe de cette homologie. Il en explique la cause : c'est dans la structure syntaxique elle-même (la combinatoire des lexèmes) que s'incarne le sens.

La prégnance de la question du sens dans les recherches de Harris se manifeste dans le fait qu'il y a consacré son dernier ouvrage, publié peu avant sa mort : *Language and Information*. Harris (1988, 57-64) y rappelle sa théorie bien connue de l'opposition entre *information* et *meaning*.

Sur l'*information*, il écrit (nous citons l'extrait dans la traduction due à Amr Ibrahim et Claire Martinot (cf. Harris (2007, § 3.1)) :

L'essentiel de l'information est le produit de deux contraintes [...] : la mise en ordre partielle qui est la relation prédicative de l'opérateur à l'argument, et les disparités dans les probabilités d'occurrence des mots dans le cadre de la mise en ordre partielle (ce qui distingue le sens des mots).

Sur le sens (*meaning*), Harris (1988, 63) écrit (nous citons la traduction de Harris (2007, § 3.2)) :

Nous pouvons aller de l'environnement du mot à sa signification : si nous connaissons les significations des mots avec lesquels un mot donné apparaît, nous pouvons estimer assez précisément le sens de ce mot.

En d'autres termes, la notion d'*information* s'applique à la phrase (lieu de la relation *prédicat-argument*), tandis que celle de *meaning* s'applique aux mots pris individuellement (étant pour la plupart polysémiques, ils activent tel ou tel de leurs sens potentiels en fonction des mots qui les environnent).

4. Sous-jacents aux grammaires de Zellig Harris et de Maurice Gross : la théorie de l'homomorphisme syntaxe-sémantique et le principe de compositionnalité

Dans un écrit consacré aux grammaires de Montague, Partee/Hendrik (1997), cités par Rivenc/Sandu (2009, 107) évoquent de la façon suivante la théorie de l'*homomorphisme syntaxe-sémantique* et le principe de *compositionnalité* :

Théorie de l'homomorphisme syntaxe-sémantique : *La syntaxe est une algèbre, la sémantique est une algèbre ; et il y a un homomorphisme envoyant les éléments de l'algèbre syntaxique sur les éléments de l'algèbre sémantique.*

Principe de compositionnalité (dit aussi «Principe de Frege») : *La signification d'une expression est fonction des significations et de la combinaison syntaxique des parties composantes de cette expression.*

Ces deux principes nous paraissent conformes aux approches de Z. Harris et de M. Gross telles qu'elles ont été présentées ci-dessus, et ce jusque dans le choix des termes, à ceci près que Gross parle de *morphisme* au lieu d'*homomorphisme*. Et sous réserve de recherches plus approfondies, nous constatons l'existence d'une identité d'approche de la relation syntaxe-sémantique entre celles des grammaires de Harris-Gross et celles de Montague.

5. Après le LADL : deux exemples de travaux d'élaboration de classes syntactico-sémantiques

Après la disparition de Maurice Gross, beaucoup de ses collaborateurs et élèves, en France et à l'étranger, ont continué à travailler dans la mouvance des grammaires de Harris. C'est le cas par exemple des chercheurs réunis par Andrée Borillo, à l'Université de Toulouse-Le Mirail, dont les travaux en linguistique de l'espace (cf. entre autres Borillo (1998)) prolongent ceux, pionniers et fondamentaux, de Boons (1985) et Guillet/Leclère (1992). C'est aussi le cas des chercheurs impliqués dans le Projet BFQS (cf. Lamiroy et al. 2010), qui exploite en partie les grands inventaires réalisés au LADL dans le domaine des expressions figées. Nous évoquons plus particulièrement dans cette section deux autres projets d'envergure, dus respectivement à Gaston Gross et à Jean Dubois & Françoise Dubois-Charlier. Ils se caractérisent par une intention plus explicitement orientée que par le passé vers l'élaboration de classes sémantiques.

Mais il ne s'agit pas d'une rupture avec les travaux du LADL, même si la dénomination de *lexique-grammaire* n'est pas conservée par ces auteurs : leurs classifications restent élaborées à partir de critères purement syntaxiques. Une attention plus marquée que par le passé est portée sur les sous-ensembles de classes syntaxiques ayant un plus ou moins haut degré de cohésion sémantique. On cherche alors à déterminer les facteurs syntaxiques qui sont corrélés aux agrégats sémantiques. Dès lors, il devient possible d'introduire massivement, et ce en restant globalement dans le cadre théorique du lexique-grammaire, des notions telles que *noms de moyens de transport*, *de métiers*, *d'aliments*, *de maladies*, *verbes de déplacement*, *verbes de communication*, *prédicats de sentiment*, etc.

5.1. Les travaux de Gaston Gross sur les classes d'objets

Les travaux du LADL sont axés sur les prédicats et négligent (en les remettant à plus tard) les travaux sur les noms non prédicatifs. C'est à des tâches de classification de ces derniers que se sont principalement attelés Gaston Gross et son équipe, le LLI (cf. entre autres Gaston Gross (1989), (1992), (1994) (1995) et (1996)).

La classification des prédicats est principalement fondée sur des critères de *sous-catégorisation* : elle prend en compte principalement les formes syntaxiques des arguments que les prédicats sous-catégorisent. En revanche, la classification des noms non prédicatifs du LLI est fondée sur des critères de *sélection lexicale*. En effet, elle prend

en compte les prédicats qui les sélectionnent en propre. Ces prédicats sont appelés par Gaston Gross *prédicats* (ou *opérateurs*) *appropriés* ; quant aux classes syntactico-sémantiques définies par un ensemble de prédicats appropriés, elles sont appelées *classes d'objet* (cf. Le Pesant/Mathieu-Colas (1998)). Les données sont présentées sous forme de tables, à des fins de traitement informatique ultérieur.

Prenons l'exemple de la classe d'objets des noms de *voies de communication* (environ 500 noms) élaborée par Mathieu-Colas (1998). Les différentes catégories de noms de voies de communication (navigables, ferrées, terrestres non ferrées) sont définies par leurs prédicats appropriés. Par exemple, les noms de *voies terrestres* ont pour prédicats appropriés, quand ils sont en position sujet, des verbes comme *zigzaguer*, *contourner*, *passer (par)* et des adjectifs intransitifs comme *accidenté*, *asphalté*, *balisé*, *barré*, ou transitifs comme *accessible à*, *bordé de*, *impraticable pour*, *interdit à*. Cette méthode permet de définir des sous-catégories telles que, pour les noms de voies terrestres, les noms de *voies internes à un édifice*, de *voies urbaines*, *places*, *tunnels*, *ponts*, etc. On obtient de la sorte une classification croisée complexe : par exemple les locutions adjectivales *à double sens* et *à sens unique* sont appropriées aussi bien à la classe des noms de voies urbaines qu'à celles des tunnels et des ponts. En revanche, *percer* n'est approprié qu'aux noms de *tunnels* et *jeter sur* (un cours d'eau) n'est approprié qu'aux noms de *ponts*.

5.2. Les dictionnaires syntactico-sémantiques de Jean Dubois & Françoise Dubois-Charlier

Depuis plus de vingt ans, Jean Dubois et Françoise Dubois-Charlier développent des ressources lexicographiques électroniques, dont certaines sont en cours de révision et sur le point d'être mises à la disposition du public. Ce sont principalement :

- le dictionnaire des *Verbes français* (LVF)⁶, 25 609 entrées (Dubois/Dubois-Charlier (1997) ; disponible sous deux formes : forme textuelle et base de données électroniques) ;
- le dictionnaire des *Locutions en français* (LEF), environ 10 000 entrées (Dubois/Dubois-Charlier (2004) en version textuelle ; version électronique en cours de réalisation) ;
- la base de données des *Locutions verbales* (LOCV), 3 510 entrées (à paraître prochainement) ;
- le *Dictionnaire électronique des mots* (DEM), environ 140 000 entrées, à paraître prochainement et dont un aperçu a été présenté par Dubois/Dubois-Charlier (2010). DEM incorpore les entrées des ressources ci-dessus et contient de nouvelles entrées (noms, adjectifs, adverbes, prépositions, conjonctions, interjections).

Les travaux de Dubois/Dubois-Charlier se situent explicitement dans la mouvance du LADL. La ressource la plus connue actuellement, LVF, a été élaborée dans les locaux du LADL et elle est le produit de la collaboration entre Jean Dubois et un membre éminent du LADL, Alain Guillet (cf. aussi les mots de remerciement

⁶ Sur la structure de LVF, on pourra se reporter à François et al. (2007). D'autre part, sur l'ensemble des travaux les plus récents de Dubois & Dubois-Charlier, on consultera le site <http://talep.lif.univ-mrs.fr/Fondamental/>.

adressés par les auteurs de LVF à Maurice Gross et à Alain Guillet dans la 2^{ème} page de la version textuelle de LVF).

La question de la relation entre la forme syntaxique et le sens est posée d'emblée au début de l'introduction de LVF. Cette classification de près de 26 000 emplois verbaux est présentée comme une vérification de l'hypothèse de l'*homomorphisme syntaxe-sémantique* (Dubois / Dubois-Charlier (1997, 3) parlent d'*adéquation* entre la syntaxe et l'interprétation) :

La classification des verbes français repose sur l'hypothèse qu'il y a adéquation entre les schèmes syntaxiques de la langue et l'interprétation sémantique qu'en font les locuteurs de cette langue : à la différence syntaxique entre *venir à Paris* et *venir de Paris* correspond une différence sémantique entre la destination et l'origine.

A partir de là l'apport, dans l'analyse syntaxique, de termes relevant de la sémantique devient massif. La catégorie des verbes est divisée en 14 classes dites génériques : C (verbes de communication) ; D (verbes de don) ; E (verbes d'entrer-sortir) ; F (verbes de frapper) ; etc. A l'intérieur des classes *génériques* sont distinguées 54 classes *syntactico-sémantiques* :

Tabelle (*Les Verbes français* 1997)

Les classes syntactico-sémantiques se répartissent elles-mêmes en 248 *sous-classes syntaxiques*, subdivisées à leur tour en un certain nombre de *sous-types syntaxiques*. Les critères de classification sont syntaxiques. La rubrique « Construction » est fondamentale : elle code les propriétés de transitivité-intransitivité, de pronominal-non pronominal, de forme des arguments (SN +/- humain ; SN non restreint ; complétive ; infinitif ; forme de la préposition ; type de complément (locatif, temps etc.). La rubrique « Opérateur » représente l'interface syntaxe-sémantique : les opérateurs « interprètent

sémantiquement les schèmes syntaxiques » (Dubois / Dubois-Charlier (2007, 7).

Prenons l'exemple, au sein de la *classe générique C* (Communication), de la *classe syntactico-sémantique C2*. Au sein de cette dernière figure la *sous-classe syntaxique C2b*. Cette dernière est définie par la construction T15a8 (T = « verbe transitif » ; 1 = « sujet humain » ; 5 = « complétive ou infinitif » ; a = « complément prépositionnel en à » ; 8 = « complément instrumental éventuel »)⁷.

Voici un extrait de la classe syntactico-sémantique C2b (où figurent également les champs «Domaine» (littérature, langue, droit, religion, sociologie, psychologie), «Sens» (c'est-à-dire synonymes ou parasyonymes) et «Phrase» (c'est-à-dire des exemples de phrases) :

Mot	Dom.	Class.	Constr.	Opérateur	Sens	Phrase
commander 04	LIT	C2b	T15a8	dic ordre A qn D	ordonner	On c-à P de venir,qu'il vienne demain.
conseiller 01	LIT	C2b	T15a8	dic conseil A qn D	recommander, suggérer	On c-à P de partir,le départ,qu'il parte.
déconseiller 01	LIT	C2b	T15a8	dic ordre A qn D nég	dissuader qn de	On d-à P de partir,le départ,qu'il avertisse son ami.
défendre 05	LIT	C2b	T15a8	dic ordre A qn D nég	interdire	On d-à P de sortir,toute sortie.Cet ali- ment est d~.
demander 02	LIT	C2b	T15a8	dic ordre A qn D	exiger de qn que	On d-à P le silence,de faire silence,qu'on fasse silence.

⁷ La notation « 8 » prédit souvent le phénomène de déplacement métonymique d'un complément instrumental en position sujet. Dans le cas de la classe C2b, on a par exemple l'équivalence suivante : *On m'a ordonné de rentrer par un simple e-mail = Un simple e-mail m'a ordonné de rentrer.*

Mot	Dom.	Class.	Constr.	Opérateur	Sens	Phrase
dicter 02	LIT	C2b	T15a8	dic ordre A qn D	obliger, contraindre	On d~à P d'avouer. La situation d~à P sa démis- sion.
dire 03	LIT	C2b	T15a8	dic ordre A qn D	ordonner	On d~à P de venir, qu'il se rende en cet endroit.
édicter 02	LANT	C2b	T15a8	dic ordre A qn D	ordonner autoritaire- ment	Le père é~à P qu'on ne se lève pas de table avant la fin.
enjoindre	LIT	C2b	T15a8	dic ordre A qn D	commander expressément	On e~au visi- teur de partir aussitôt.

Le format de la base de données LVF est approprié au TAL (Traitement Automatique des Langues), comme le montrent Sabatier/Le Pesant (2013) qui décrivent une tâche d'implémentation d'un extrait de la classe P1c (verbes psychologiques de manifestation de tel sentiment pour ou contre quelqu'un ou quelque chose) sur la plate-forme d'ingénierie linguistique NooJ (cf. Silberztein (2013)), en vue de tâches d'annotation syntaxique et sémantique de corpus.

6. Présentation d'un dictionnaires syntactico-sémantique transcatégoriel

La plupart des lexiques-grammaires sont séparés les uns des autres par les barrières de parties du discours: les lexiques de verbes, de noms, d'adjectifs etc. sont distincts les uns des autres, sans interconnexions systématiques. Or il existe de larges zones du lexique où se manifestent des phénomènes de morphologie dérivationnelle autour d'un même radical, avec préservation de constantes tant syntaxiques que sémantiques entre les différents éléments du champ dérivationnel.

C'est le cas par exemple d'une série de radicaux de couleur (*-blanc-*, *-noir-*, *-vert-* etc.) qui ont en commun la même structure dérivationnelle (ex. *blanc*, *blanchâtre*, *blanchissant*; *blancheur*; *blanchir*). Prenons un autre exemple dans le domaine des radicaux verbaux. Soit le radical du verbe *écoeuré*; on peut représenter simultanément ses diathèses et ses dérivations de la façon suivante :

qq est *écoeuré*, *s'écoeuré* (de qqch, de V-infinitif, de ce que P<subj>, 0)

l'écoeurément de qq (de qqch, de V-infinitif, de ce que P<subj>, 0)

qq est écoeuré par qq, qqch / qq, qqch écoeur qq ;
qq, qqch est écoeurant

Les propriétés morphosyntaxiques ci-dessus ne sont pas spécifiques au radical *-écoeur-* ; elles caractérisent toute une classe morpho-syntaxique de prédicats : *affliger, affoler, agacer, crisper, désespérer, effarer, effrayer, émouvoir, étonner, exaspérer, irriter, préoccuper, révolter*, etc. On voit alors se dessiner, en même temps qu'une classe morpho-syntaxique, une classe sémantique de mots d'affects. Elle est cependant peu raffinée. D'autres propriétés doivent être convoquées pour délimiter des sous-ensembles sémantiquement plus cohérents ; par exemple la propriété d'acceptation d'un complément en *contre* permet de rassembler les prédicats de colère *-exaspér-*, *-irrit-* et *-révolt-*.

Il est montré dans Eshkol/Le Pesant (2005) que cette méthode de classification syntactico-sémantique permet non seulement de transgresser les barrières de parties du discours, mais aussi : (1) de factoriser les descriptions linguistiques ; (2) de mobiliser un plus grand nombre de critères syntaxiques définitionnels de catégories sémantiques, puisqu'aux propriétés spécifiques aux verbes s'ajoutent des propriétés spécifiques des adjectifs et des noms. Le dictionnaire prend dès lors la forme d'un thésaurus de classes syntactico-sémantiques.

Nous avons défini dans Le Pesant (2011) les principes d'un *Thésaurus des mots d'affects*. Il s'agit d'un gros vocabulaire (plus de 3000 mots) dont 1/3 sont des locutions. A cet effectif, il faut ajouter plusieurs centaines de collocations de manifestations physiques des sentiments, telles *pâlir de rage, trembler de colère, écarquiller les yeux de surprise*. L'ensemble de la catégorie se caractérise par la présence nécessaire d'un argument nominal animé correspondant au rôle thématique de *l'individu affecté (experier)* ; les deux autres rôles thématiques éventuels sont celui de la *cause* et celui de *l'objet de l'affect*. Il y a trois grandes catégories syntactico-sémantiques de mots d'affects :

- Catégorie 1 : Mots d'états psychologiques (notamment les mots d'humeurs). Il s'agit de verbes, noms et adjectifs intransitifs (ex. *mélancolique, optimiste*), ou intransitifs à complément de cause en *de* (ex. *jubilation*) ; en position sujet figure le rôle de *l'experier*.
- Catégorie 2 : Mots d'émotions (causées par quelque chose ou quelqu'un). Il s'agit de verbes, noms et adjectifs transitifs directs (ex. *décourager, stimuler*) ; il existe un complément d'agent en *par* ; à la voix active, en position sujet figure le rôle de l'agent ou de la cause, le rôle de *l'experier* étant en position de complément d'objet ; dans la même catégorie on trouve des verbes transitifs indirects à sujet agentif ou causatif (ex. *plaire à*).
- Catégorie 3 : Mots de *sentiments* (pulsions et répulsions). Il s'agit de verbes, noms et adjectifs transitifs directs ou indirects (ex. *aimer qq, se passionner pour qqch, se repentir de qqch*) ; en position sujet figure le rôle de *l'experier* ; en position de complément figure le rôle de *l'objet* (du désir ou de la répulsion).

Il existe en outre des catégories « hybrides », c'est-à-dire des combinaisons de plusieurs des catégories définies ci-dessus, telles :

- Catég. 1 + Catég. 2. Ex. *consterner* (Je suis consterné d'avoir entendu cela ; je suis consterné par cela)
- Catég. 2 + Catég. 3. Ex. *enthousiasmer* (Cela m'enthousiasme ; je m'enthousiasme pour cela)
- Catég. 1 + Catég. 3. Ex. *être en colère* (d'avoir vu cela ; contre quelqu'un)
- Catég. 1 + Catég. 2+ Catég. 3. Ex. *irriter* (être irrité d'avoir vu cela ; Pierre m'irrite ; je suis irrité contre Marie)

Beaucoup de catégories de propriétés syntaxiques sont convoquées, telles les contraintes de sous-catégorisation et de sélection lexicale sur les arguments (SN +/- animé, SV-infinitif, Proposition subordonnée au subjonctif), la forme des prépositions, les formes spécifiques de verbes supports ou d'opérateurs causatifs, et les collocations appropriées (de réactions physiologiques, de réactions verbales, de refoulement et de défolement).

La prise en compte de divers types d'antonymies permet en outre d'envisager une structuration de l'ensemble en forme d'ontologie :

Catégorie 1 : Mots d'humeurs		
Groupe 1	Bonne humeur : gaieté, allégresse, euphorie ...	Mauvaise humeur : morosité, être maussade, ...
Groupe 2	Bon moral : énergie, tonus, optimisme ...	Mauvais moral : langueur, déprime, pessimisme, ...
Groupe 3	Insouciance, nonchalance	Etre soucieux, pensif,
Groupe 4	Sérénité : paix, tranquillité ...	Agitation : fébrilité, excitation ...
Groupe 5	/	Sentiment d'abandon, sentiment de solitude

Catégorie 2 ou combinaison C1 + C2 : Mots d'émotions causées		
Groupe 1	Joie	Tristesse
	Plaisir	Déplaisir ; Répugnance ; Souffrance
	Bonheur	Malheur
	Satisfaction	Frustration
Groupe 2	Surprise : bonne/mauvaise surprise	
Groupe 3	Soulagement	Déception

Groupe 4	Sentiment de sécurité	Sentiment d'étrangeté ; Inquiétude ; Trouble ; Effroi
Groupe 5	Réconfort	Blessure morale
Groupe 6	Séduction	Agacement ; Enervement
Groupe 7	Excitation	Ennui
Groupe 8	Emotion ; Etat de choc	

Catégorie 3 : Mots de sentiment (pulsion-répulsion)		
Groupe 1 Polarité supplémentaire extraversion/introversion	Amour (des autres) Contentement de soi	Haine (des autres) Haine de soi (entre autres : remords)
	Admiration (des autres) Admiration de soi (fierté) Estime de soi/des autres	Mépris (des autres) Mépris de soi (honte) Mésestime, sous-estimation
	Gratitude (envers autrui) Bonne conscience, orgueil (se féliciter de qc)	Ressentiment (envers autrui) Mauvaise conscience, remords (s'en vouloir de qc)
	Confiance (en autrui) Confiance en soi	Défiance (envers autrui) Manque de confiance en soi
Groupe 2	Sympathie Amitié, affection	Antipathie Inimitié
Groupe 3 De l'intérêt au désir	Intérêt	Indifférence
	Goût, Inclination	Aversion
	Désir, Aspiration	Répulsion
	Nostalgie, regret	/
Groupe 4	Espoir	Crainte
Groupe 5	Sentiment de révolte	Sentiment de résignation
Groupe 6 : En lien avec l'empathie	Empathie	Pitié ; Etre envieux
Groupe 7 : Autres	/	Colère ; Peur ; Jalousie sexuelle

Nous sommes en train d'implémenter notre thésaurus des mots d'affects dans NooJ (cf. Silberztein (2013)), à des fins de tâches d'annotation sémantique automatique. Voici un exemple d'annotation dans un corpus au moyen de balises XML :

The image shows a screenshot of a text document with XML annotations overlaid. The text is in French and describes a scene from 'L'Assommoir'. The annotations are in blue and red. A blue box highlights the phrase 'Les Lorilleux', with the XML tag <EXPER>Les Lorilleux</EXPER> above it. Another blue box highlights the word 'passaient', with the XML tag <VERB CATEG="défolement">passaient</VERB> above it. A red box highlights the word 'leur', with the XML tag <NOM CATEG="Affect" CLASS="colère" SSCLASS="rage">leur</NOM> above it. Other annotations include <CLASS="colère" SSCLASS="rage">rage</CLASS> above 'rage' and <OBJ>sur le rôti</OBJ> above 'sur le rôti'.

Dans cet extrait de *L'Assommoir* de Zola figure l'annotation suivante :

<EXPER>Les Lorilleux</EXPER><VERB CATEG="défolement">passaient</VERB>
leur <NOM CATEG="Affect" CLASS="colère" SSCLASS="rage">leur</NOM> <OBJ>sur
le rôti</OBJ>.

Grâce aux dictionnaires et grammaires locales NooJ dérivés du *Thésaurus des mots d'affect* et projetés sur le corpus, le Groupe Nominal « Les Lorilleux » a été reconnu comme assumant le rôle thématique de l'*experier* ; le mot « passaient » est reconnu comme étant un *verbe de défolement* ; le mot « rage » a été reconnu comme étant un nom d'*affect* appartenant à la classe *colère* et à la sous-classe des mots de *rage* ; et le Groupe Prépositionnel « sur le rôti » a été reconnu comme assumant le rôle thématique de l'*objet* (de la rage).

Conclusion

Le regain d'intérêt aux Etats-Unis pour l'œuvre de Zellig Harris, dont témoigne Barsky (2011), laisse espérer que soient mieux connues sa théorie de l'homomorphisme syntaxe-sens et sa conception de la combinatoire syntaxique en tant que support de l'information (cf. Harris (1988)). Harris s'inscrit dans une tradition philosophique et logique marquée par les travaux de Frege et Russel, de Wittgenstein et des positivistes logiques du Cercle de Vienne. Nous avons essayé de montrer dans cet article que les recherches de Maurice Gross et de ceux qui se réclament de la théorie du lexique-grammaire s'inscrivent à leur tour dans cette tradition, tout au moins dans le domaine de la question de la relation syntaxe-sens.

Université Paris Ouest Nanterre & MoDyCo (CNRS)

Denis LE PESANT

Références

- Barsky, Robert F, 2011. *Zellig Harris. From American Linguistics to Socialist Zionism*, Cambridge, MIT Press.
- Boons, Jean-Paul, 1985. «Preliminaires à la classification des verbes locatifs». *Lingvisticae Investigationes* 9-2, Amsterdam, John Benjamins, 195-267.
- Boons, Jean-Paul/Guillet, Alain/Leclère, Christian, 1976. *La structure des phrases simples en français. Constructions intransitives*. Genève, Droz.
- Borillo, Andrée, 1998. *L'espace et son expression en français*. Paris: Ophrys.
- Daladier, Anne, 1978. *Problèmes d'analyse d'un type de nominalisation en français et de certains groupes nominaux complexes, Thèse de 3^{ème} cycle*. Paris, LADL-Université Paris 7.
- Daladier, Anne, 1990 (ed.). *Les grammaires de Harris et leurs questions, Langages* 99.
- Dubois, Jean/Dubois-Charlier, Françoise, 1997. *Les Verbes Français*. <www.modyco.fr>.
- Dubois, Jean/Dubois-Charlier, Françoise, 2004. *Locutions en français*. <www.modyco.fr>.
- Eshkol, Iris/Le Pesant, Denis, 2005. «Trois petites études sur les prédicats de communication verbaux et nominaux», *Langue Française* 153, 20-32.
- François, Jacques/Le Pesant, Denis/Leeman, Danielle (ed.), 2007. «Le classement syntactico-sémantique des verbes français», *Langue Française* 153.
- Giry-Schneider, Jacqueline, 1978. *Les nominalisations en français. L'opérateur faire dans le lexique*, Genève, Droz.
- Gross, Gaston, 1989. *Les constructions converses du français*, Genève, Droz.
- Gross, Gaston, 1992. «Forme d'un dictionnaire électronique», *L'environnement traductionnel*, Sillery (Canada), Presses de l'Université du Québec.
- Gross, Gaston, 1994. «Classes d'objets et description des verbes», *Langages* 115, Paris, Larousse.
- Gross, Gaston/Kiefer, Ferenc, 1995. «La structure événementielle des substantifs», *Folia Linguistica*, XXX/1-2, Berlin, Mouton de Gruyter.
- Gross, Gaston, 1996. *Les expressions figées en français : noms composés et autres locutions*, Gap-Paris, Ophrys.
- Gross, Maurice, 1972. «On Grammatical Reference», *Generative Grammar in Europe*, Dordrecht, Reidel, 203-217.
- Gross, Maurice, 1975. *Méthodes en syntaxe*, Paris, Hermann.
- Gross, Maurice, 1981. «Les bases empiriques de la notion de prédicat sémantique», *Langages* 63, 7-52.
- Guillet, Alain/Leclère, Christian, 1992. *La structure des phrases simples en français. Constructions transitives locatives*. Genève-Paris, Droz.
- Halliday, Michael, 1985. *An Introduction to Functional Grammar*. London, Arnold.
- Harris, Zellig, 1968. *Mathematical Structures of Language*, New York, Wiley.
- Harris, Zellig, 1971. *Structures mathématiques du langage*, Paris, Dunod.
- Harris, Zellig, 1976. *Notes du cours de syntaxe*, Paris, Le Seuil.
- Harris, Zellig, 1988. *Language and Information*, New York, Columbia University Press.
- Harris, Zellig, 2007. *La langue et l'information*, Paris, CRL.
- Lamiroy, Béatrice et al., 2010. *Les expressions verbales figées de la francophonie. Belgique, France, Québec et Suisse*, Paris, Ophrys.

- Leeman, Danielle / Sabatier, Paul, 2011. «Empirie, théorie, exploitation : le travail de Jean Dubois sur les verbes français», *Langages*, 179-180.
- Le Pesant, Denis / Mathieu-Colas, Michel, 1998. «Introduction aux classes d'objets», *Langages* 131, 6-33.
- Le Pesant, Denis, 2011. «Vers un thésaurus syntactico-sémantique des mots d'affect», *Cahiers de Lexicologie* 99, 2011-2, Paris, Classiques Garnier, 117-132.
- Mathieu-Colas, Michel, 1998. «Illustration d'une classe d'objets: les voies de communication», *Langages* 131, 77-90.
- Partee, Barbara / Hendrik, 1997. «Montague Grammars», *Handbook of Logic and Language*, Cambridge, MIT Press.
- Picabia, Lélia, 1978. *Les constructions adjectivales en français*. Genève, Droz.
- Rivenc, François / Sandu, Gabriel, 2009. *Entre logique et langage*, Paris, Vrin.
- Sabatier, Paul / Le Pesant, Denis, 2013. «Chapitre 5. Les dictionnaires électroniques de Jean Dubois & Françoise Dubois-Charlier et leur exploitation en TAL», *Ressources Lexicales*, Amsterdam, John Benjamins, 153-186.
- Silberztein, Max, 2013. *NooJ Manual*. <www.nooj4nlp.net>.