



HAL
open science

Elaboration d'un nouveau modèle de la langue et lexicographie

Denis Le Pesant

► **To cite this version:**

Denis Le Pesant. Elaboration d'un nouveau modèle de la langue et lexicographie. *Langue(s) & Parole*, 2021, 5, pp.23-42. 10.5565/rev/languesparole.62 . halshs-04074565

HAL Id: halshs-04074565

<https://shs.hal.science/halshs-04074565>

Submitted on 11 May 2023

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Élaboration d'un nouveau modèle de la langue et lexicographie

Denis LE PESANT
MoDyCo, CNRS et Université Paris Nanterre

Résumé

Dans notre article, nous traçons les grandes lignes du modèle de la langue initié par Zellig S. Harris et complété par Maurice Gross. Ce dernier définit ce *modèle* (ou *théorie*) comme étant l'association de la grammaire (en tant que système de règles) et du lexique (en tant que définition extensionnelle du système de règles), d'où la notion de *lexique-grammaire*. La question est pour nous de savoir quel est le type de lexicographie qui convient au modèle du lexique-grammaire : nous montrons que dans ce cadre, les dictionnaires sont nécessairement des bases de données informatisées et que la lexicographie est par conséquent plus qu'une méthodologie de recherche ; elle est aussi la réalisation la plus aboutie du modèle de la langue. Nous nous intéressons aux dictionnaires informatisés de Maurice Gross, de Gaston Gross et de Jean Dubois & Françoise Dubois-Charlier, que nous présentons comme des jalons importants dans l'entreprise de modélisation de la langue française.

Mots clé : lexicographie informatisée, lexique-grammaire, classes d'objets, dictionnaires électroniques de Dubois et Dubois-Charlier.

Abstract

In this paper, we present the model of language initiated by Zellig S. Harris and completed by Maurice Gross. The latter defines this model (or theory) as the association of grammar (as a system of rules) and lexicon (as an extensional definition of the rules system), hence the notion of lexicon-grammar. The question for us is to decide what type of lexicography is suitable for the lexicon-grammar model. We show that, in this context, dictionaries are necessarily computerized databases and that lexicography is therefore more than a methodology of research. It is also the most successful realization of a model of language. We are interested in the computerized dictionaries of Maurice Gross, Gaston Gross and Jean Dubois & Françoise Dubois-Charlier, which we present as important milestones for the French language modeling.

Keywords: computerized lexicography, lexicon-grammar, object classes, electronic dictionaries of Dubois and Dubois-Charlier.

Resumen

En este artículo, describimos las grandes líneas del modelo de la lengua iniciado por Zellig S. Harris y completado por Maurice Gross. Este último define dicho *modelo* (o *teoría*) como la asociación de la gramática (concebida como sistema de

reglas) y el léxico (como definición en extensión del sistema de reglas), de ahí la noción de *léxico-gramática*. La cuestión consiste en saber cuál es el tipo de lexicografía que mejor encaja con el modelo del léxico-gramática. En este marco, los diccionarios son necesariamente bases de datos informatizadas y la lexicografía es, por consiguiente, más que una metodología de investigación; es también la realización más conseguida del modelo de la lengua. Nos interesamos por los diccionarios informatizados de Maurice Gross, de Gaston Gross y de Jean Dubois & Françoise Dubois-Charlier, que presentamos como hitos importantes en la empresa de modelización de la lengua francesa.

Palabras clave: lexicografía informatizada, léxico-gramática, clases de objetos, diccionarios electrónicos de Dubois y Dubois-Charlier.

Introduction

Dans cet article, nous restreignons le thème du numéro (*la lexicographie en tant que méthodologie de recherche en linguistique*) au cas particulier de l'heuristique pratiquée dans les années 1970-1990 par Maurice Gross et son équipe du LADL¹ dans leur entreprise de construction d'un *lexique-grammaire* du français. Ces travaux, qui s'originent des grammaires de Zellig S. Harris, se caractérisent par le fait d'assumer qu'un *modèle* (ou *théorie*) de la langue ne peut pas se limiter à l'établissement d'un corpus de règles appelé « grammaire » : pour que le modèle soit complet, la grammaire a besoin d'être corrélée au lexique. Nous évoquons aussi les recherches de Gaston Gross et de son équipe, le LLI², ainsi que les deux ultimes dictionnaires de Jean Dubois et Françoise Dubois-Charlier, *LVF*³ et le DEM⁴, qui se situent dans la même mouvance théorique.

Le choix de corréliser le lexique à la grammaire implique la conception d'un nouveau type de dictionnaires, à savoir des bases de données

¹ LADL, Laboratoire d'Automatique Documentaire et Linguistique, CNRS et Université Paris 7. Dir : Maurice Gross (1934-2001). L'historique du LADL est disponible sur <http://infolingu.univ-mlv.fr/LADL/Historique.html>.

² LLI (renommé ultérieurement LDI), Laboratoire de Linguistique Informatique, CNRS et Université Paris 13. Dir. : Gaston Gross.

³ LVF, *Les Verbes Français* édité en 1997 par Larousse, ouvrage issu d'une Base de données lexicales. Les deux formes de cette ressource sont disponibles sur <https://www.modyco.fr/fr/base-documentaire/ressources/jean-dubois.html>
<http://rali.iro.umontreal.ca/rali/?q=fr/versions-informatisees-lvf-dem>

⁴ DEM, *Dictionnaire Electronique des Mots*. Base de données lexicales disponible sur <https://www.modyco.fr/fr/base-documentaire/ressources/jean-dubois.html>
<http://rali.iro.umontreal.ca/rali/?q=fr/versions-informatisees-lvf-dem>

informatisées structurées en tables qui mettent en relation les lexèmes avec leurs propriétés linguistiques caractéristiques, ce qui les rend disponibles pour le Traitement Automatique du Langage (TAL). A première vue, ce type de lexicographie est beaucoup plus qu'une méthodologie de recherche, puisque les dictionnaires obtenus, en tant que représentations du lexique, sont surtout là pour révéler et expliciter l'une des deux composantes du modèle de la langue. Il n'en reste pas moins que la méthodologie qui consiste à mettre en regard les propositions de la grammaire avec le plus grand nombre possible de faits expérimentaux remplit ses fonctions traditionnelles de vérification et de perfectionnement du modèle théorique.

1. Notions de grammaire, de lexique et de dictionnaire

La *grammaire* d'une langue donnée peut être considérée, dans une première approche, comme sa *théorie*, c'est-à-dire un *modèle* présenté sous une forme plus ou moins systématique. En Occident, depuis l'Antiquité, la fonction de la *grammaire* (complétée de la *rhétorique* et assise sur la *logique*) est de définir un ensemble de catégories et de principes de base comme les types de phonèmes, les types de morphèmes (« parties du discours »), un schéma de la proposition élémentaire inspiré de l'ancienne logique, à savoir la structure Sujet-Prédicat, les fonctions syntaxiques, les phénomènes de changement de sens (par exemple les « figures du discours »), l'expression des circonstances (temps, modalité, cause, manière ...), etc. Elle définit d'autre part un certain nombre de règles, de conditions (sous forme de restrictions) sur la combinatoire des phonèmes, sur la combinatoire des morphèmes à des fins de construction des mots dérivés, et sur la combinatoire de ce que nous appelons aujourd'hui lexèmes, à des fins de construction du sens. Tout cela constitue l'aspect « distributionnel » de la grammaire. La grammaire définit aussi des types d'alternatives de forme des énoncés tels les changements d'ordre des mots et les pronominalisations ; il s'agit là de l'aspect « transformationnel » de la grammaire.

Complémentairement aux grammaires, à partir de l'invention de l'imprimerie, ont été publiés des *dictionnaires* de la langue générale. Ce sont des relevés du plus grand nombre possible de lexèmes (mots simples, mots composés et locutionnels, et mots-phrases), chaque lexème se voyant attribuer une définition sémantique (qui mobilise au moins implicitement les notions d'hyponymie, de synonymie et d'antonymie), ainsi qu'une

définition grammaticale, c'est-à-dire l'assignation à chaque entrée d'un ensemble de propriétés linguistiques empruntées à la *grammaire* : partie du discours, propriétés syntaxiques, appartenance éventuelle à un champ dérivationnel, informations diachroniques, diatopiques et diastratiques (ces dernières sous forme de règles dites « de bon usage »), etc. Une partie des informations sont illustrées d'exemples empruntés à des corpus ; ces exemples permettent au lecteur d'être mis en présence de nombreuses propriétés linguistiques laissées implicites dans les définitions.

Historiquement, les dictionnaires d'une époque donnée sont conçus à des fins pédagogiques, culturelles et, bien souvent aussi, de divertissement. Mais comme ils reprennent la plupart des catégories et des termes des grammaires de leur temps, ils se révèlent en être les corrélats. C'est grâce à l'existence des dictionnaires que nous prenons conscience que les théories de la langue que sont les grammaires sont des modèles incomplets. Certes les grammaires instituent des catégories et des règles et elles les définissent ; mais c'est seulement par des définitions *en compréhension*. Or, on a besoin que ces catégories et règles soient définies, complémentirement, *en extension* : on a besoin de connaître leur *extension lexicale*.

2. Les deux composantes inséparables d'un modèle de la langue (ou théorie de la langue) sont la grammaire et le lexique

La définition en *extension* des catégories et des règles de la *grammaire* d'une langue, c'est ce qu'on appelle le *lexique* de cette langue. Soit par exemple une règle de la grammaire du français qui stipule (par définition *en compréhension*) que certains verbes, tel *étonner* mais non pas *manger*, peuvent admettre en position sujet une proposition subordonnée de forme « (*le fait*) que P ». La définition *en extension* de cette « règle » ne sera rien d'autre que la liste exhaustive des verbes et locutions verbales qui admettent un sujet d'une telle forme. Ainsi, la conception de la *grammaire* comme *modèle* de la langue se révèle inexacte : ce qui est véritablement le *modèle* (ou la *théorie*) de la langue, c'est l'ensemble associant la grammaire au lexique et réciproquement.

Revenons aux dictionnaires. Certes on peut avoir l'impression que ce sont avant tout des sortes de discours sur la langue consistant à énumérer les mots par ordre alphabétique, à indiquer leur sens, c'est-à-dire en fait à

les associer à leur(s) référent(s) (objets du monde physique et concepts), à illustrer leurs emplois par des exemples et parfois à les traduire dans une langue étrangère. Mais plus fondamentalement, la fonction d'un dictionnaire est de *représenter* une réalité aussi abstraite que la *grammaire* d'une langue, à savoir son *lexique*.

Force est de constater que la forme et le contenu du *lexique* d'une langue donnée à une époque donnée sont beaucoup plus dépendants de l'évolution de la langue elle-même que ne l'est l'évolution, très lente, de son corrélat, la *grammaire*. Car il y a une discordance majeure entre le rythme de l'histoire de la langue, qui fait partie de l'histoire des sociétés, et celui de l'histoire des grammaires, qui fait partie de l'histoire des idées. Dans le domaine qui nous occupe, de l'Antiquité à nos jours, il y a eu certes jusqu'à présent des évolutions considérables dans l'histoire des grammaires, mais le paradigme est resté intact dans ses principes : le modèle en cause reste celui de ce qu'on appelle la *grammaire traditionnelle*. Il s'en est suivi que les représentations du lexique, à savoir les dictionnaires, reposent sur le même paradigme : la grammaire traditionnelle et son corrélat, la lexicographie traditionnelle, sont les deux composantes de la théorie grammaticale « normale » de notre époque.

C'est un héritage prestigieux qu'il est hors de question de répudier, mais qui pourrait être dépassé. Il y a eu au XX^e siècle deux tentatives de changement radical de paradigme, avec l'émergence des grammaires de Z. Harris et de N. Chomsky. C'est peu dire que ces dernières n'ont pas fait l'unanimité des chercheurs. Ce seul fait prouve qu'il n'y aura pas eu en grammaire, du moins jusqu'à présent, de changement de paradigme par rapport à la science « normale ». L'explication de ce phénomène ne fait pas partie des objectifs de cet article.

3. L'invention du lexique-grammaire (Maurice Gross, 1975)

Dans son premier livre, *Méthodes en syntaxe*, Maurice Gross (Gross, 1975) présente son *lexique-grammaire* des verbes à « construction complétive »⁵. Ils sont répartis entre 19 classes : verbes de mouvement à

⁵ Il s'agit pour l'essentiel des verbes à proposition subordonnée complétive sujet ou complément. La présence des auxiliaires aspectuels dans la Classe 1 (par ex. dans « Max *commence* à travailler ») pourrait surprendre, les auxiliaires ne pouvant (par définition) avoir de compléments. En réalité, à cette époque Gross n'utilise pas la notion de *prédicat* mais celle d'*Opérateur*, qui chez Harris

complément infinitif (ex. *sortir fumer une cigarette*, Table 3) ; verbes transitifs directs à sujet phrastique (ex. *Que P tranquillise Marie*, Table 5) ; verbes à complétives prépositionnelles en *à* (ex. *remédier à ce que P*, Table 7) et en *de* (*se faire un devoir de ce que P*, Table 8) ; etc. Les classes sont présentées sous la forme de matrices binaires mettant en relation chaque entrée avec ses propriétés syntaxiques, dont la Figure 1 ci-dessous donne une idée.

Les tables elles-mêmes, avec leurs commentaires, occupent plus que la moitié de l'ouvrage. Le reste constitue une sorte de « discours de la méthode ». Nous nous limiterons pour le moment à la courte introduction du livre (Gross, 1975, 9-11). L'auteur s'y interroge sur le statut théorique de la notion de transformation et le juge être « *loin d'être clair* ». On comprend assez vite en lisant la suite que ce sont les propositions de la grammaire générative qui sont visées, notamment par ces lignes :

L'approche très répandue que nous critiquons et que nous nous proposons de remplacer consiste essentiellement à effectuer des observations isolées dans des régions différentes de la structure linguistique. Ces observations ne sont jamais systématiques, et les « trous » qui les séparent ne sont pas explorés empiriquement. Les constructeurs de modèles laissent à l'induction le soin de combler ces trous. Cette foi dans l'induction ne repose sur aucune base scientifique. (...). D'après cette croyance, une théorie (un modèle) pourrait prédire un nombre de phénomènes beaucoup plus importants que ceux qui ont servi de base à la construction de la théorie. Il ne semble pas qu'il ait jamais existé de telles situations. (Gross, 1975, 9)

Gross soutient ensuite que « *toute construction théorique a toujours été précédée d'un long travail d'accumulation systématique des données* », et illustre cette thèse par l'exemple de ce que Kepler, pour la formulation des lois qui portent son nom, doit à l'accumulation de données astronomiques effectuée avant lui par Tycho Brahé. Un extrait de ce texte attire plus particulièrement notre attention :

Une théorie ne fait, en tout cas dans un premier temps, que reformuler la classification de ces données. (Gross, 1975, 9-10)

Cette phrase fait écho à une idée soutenue au tout début du XX^e siècle par le physicien et philosophe des sciences Pierre Duhem (1861-

englobe, d'une part, les prédicats, dotés d'une structure argumentale, d'autre part, les auxiliaires aspectuels et modaux et les verbes supports, qui ne sont pas de nature prédicative.

1916), selon lequel les théories scientifiques ne sont pas autre chose que des instruments de classification des expériences⁶. Il évoque par exemple « *la théorie idéale et parfaite [qui] nous donnerait, au sens propre du mot, la classification naturelle des lois* » (cf. Roux, 2016). Les lois dont il s'agit ici, sont des « lois expérimentales », c'est-à-dire des *expériences* (réalisées selon des protocoles, à partir d'hypothèses, de méthodes, avec des grilles d'interprétation, etc.).

La thèse selon laquelle on ne peut formuler légitimement une théorie qu'en disposant au préalable d'un échantillon de faits expérimentaux aussi large que possible ne relève pas d'un empirisme naïf qui prétendrait que quand on ne dispose au départ d'aucune théorie dans un domaine donné, il suffit d'examiner un maximum de faits relevant de ce domaine et d'en déduire une théorie. Ce n'est pas le cas pour Gross : il connaît intimement deux théories concurrentes, celle de Harris et celle de Chomsky (cf. à ce sujet Chevalier & Encrevé, 2006, 251-261)⁷, qui se sont imprégnés des premiers travaux distributionalistes (notamment de Bloomfield, Boas et Sapir), et qui ont eux-mêmes déjà travaillé de façon approfondie sur plusieurs langues occidentales, sémitiques et amérindiennes. Pour ce qui est de la théorie du français, Gross ne part pas de rien : en dehors de ce qu'il a appris de ses maîtres américains, il a assimilé l'héritage remarquable des grammaires et des dictionnaires traditionnels⁸. Au départ de son entreprise, il applique cette méthode : accumuler le maximum de données expérimentales afin de les « classer » (pour reprendre la formule de Duhem) au sein du cadre général des grammaires de Harris.

Comme attendu, au fil de l'élaboration de l'inventaire, des faits et des problèmes nouveaux se manifestent, ainsi que des solutions nouvelles à des problèmes traditionnels.

C'est l'objet de la partie III de Gross (1975, 107-159) : la composante transformationnelle de la grammaire peut rendre compte des dérivations

⁶ Nous avons personnellement, lors d'un cours auquel nous assistions vers 1986, entendu Maurice Gross faire sur ce thème l'éloge de Pierre Duhem.

⁷ Comme il le révèle dans une interview donnée à J.-Cl. Chevalier et P. Encrevé, Gross fait deux séjours aux États-Unis, le premier en 1961 à Harvard où, en compagnie de M.-P. Schützenberger, il aide Chomsky à donner une forme mathématique correcte à ses grammaires, le second en 1964 à l'université de Pennsylvanie, où, s'étant éloigné de Chomsky, il aide à son tour Harris à mathématiser ses propres grammaires.

⁸ En tant que bibliophile, M. Gross collectionnait les grammaires anciennes.

suffixales, ce qui, dans les nominalisations notamment, met en lumière le rôle des verbes supports (nommés à l'époque « opérateurs ») tels *faire*, *pousser*, *porter*, *avoir*, *être* ; des « transformations nouvelles » sont décrites aux pages 134-143 : par exemple la transformation qui sera dite de *neutralité* par Boons, Guillet & Leclère (1976), et qui met par exemple en relation d'homologie la construction transitive de certains verbes avec une construction intransitive (ex. *on cuit des pieds de porc* vs *des pieds de porc cuisent*), ou les transformations dites de *restructuration* qui mettent en relation, par exemple pour le verbe *admirer* (Table 12), la structure « *admirer les qualités de quelqu'un* » avec la structure « *admirer quelqu'un pour ses qualités* ».

Dans cette entreprise, les apports en sémantique sont considérables. Gross y consacre plusieurs développements. Par exemple, aux pages 30-33, au moyen d'un exemple opposant la phrase « *La décoratrice enjolive la vitrine de minijupes claires* » (sens « volontaire » du verbe) avec la phrase « *La décoratrice enjolive la vitrine de sa minijupe claire* » (sens « involontaire »), il se démarque des linguistes qui excluent le sens des descriptions, en montrant qu'« *il est possible d'effectuer diverses opérations sur le sens dans de bonnes conditions expérimentales* ». Dans un autre passage, aux pages 146-149, M. Gross traite des « emplois figurés des verbes ». Aux pages 218-221, consacrées aux « propriétés sémantiques », il constate *a posteriori* l'homogénéité sémantique de certaines tables (notamment les Tables 2, 3 et 12), ce qui l'amène à formuler l'hypothèse que « *ce seraient des propriétés sémantiques qui détermineraient les propriétés syntaxiques* », avec le commentaire suivant :

Cette idée est loin d'être neuve, mais nos tables, par leur extension, permettent de lui donner un contenu précis et elles nous permettent d'aller plus loin que la grammaire traditionnelle. (M. Gross, 1975, 219)⁹

Dans un écrit ultérieur, *Les bases empiriques de la notion de prédicat sémantique*, M. Gross (1981, 20-21) explicite sa conception du sens. En se réclamant de plusieurs écrits de Harris, il se déclare partisan de la théorie de l'isomorphisme de la structure syntaxique et de la représentation sémantique.

Enfin, le lexique, à condition qu'il soit le plus extensif possible, assure une fonction de *vérification* des propositions de la grammaire, selon les termes de Gross :

⁹ Ces considérations amènent Gross à formuler une hypothèse (p. 221) sur le phénomène surprenant de l'apprentissage d'une multitude des structures syntaxiques par de très jeunes enfants.

Notre étude peut être considérée comme une vérification expérimentale de la validité de la théorie transformationnelle (Gross, 1975, 221)¹⁰.

4. Problème du traitement syntaxique des noms non-prédicatifs et apport dans ce domaine des travaux de Gaston Gross et de Jean Dubois & Françoise Dubois-Charlier

Dans son article sur *La notion harrissienne de transformation*, Maurice Gross évoque en 1990 le traitement des noms dans le cadre du lexique-grammaire. Les *noms prédicatifs* ont à cette époque déjà fait l'objet de nombreux travaux, et Gross fait la liste d'ouvrages de ses collaborateurs sur divers types de constructions nominales à verbe support. « *Tous ces travaux, écrit Gross, démontrent la possibilité de représenter ces mots dans un lexique-grammaire* ». Et d'ajouter :

*Seuls semblent échapper à cette description les substantifs **concrets** (e.g. chaise, notaire, homme des bois, pomme de terre), qui ne donnent pas lieu à des phrases élémentaires syntaxiquement significatives, et présentent d'autres problèmes de représentation (M. Gross, 1990, 49).*

Les noms non prédicatifs, c'est-à-dire des noms qui n'ont pas de structure argumentale, sont donc absents des lexiques-grammaires du LADL.

Cette lacune est comblée par les travaux, indépendants l'un de l'autre, d'une part, de Gaston Gross et son équipe, le LLI, d'autre part, de Jean Dubois et Françoise Dubois-Charlier avec leur *Dictionnaire Electronique des Mots* (DEM). Il est à souligner que ces trois auteurs ont été proches du LADL et se réclament explicitement des travaux de Z. Harris et de M. Gross.

Le Laboratoire de Linguistique Informatique (LLI) fut fondé vers 1993. Le projet fondamental de Gaston Gross était d'intégrer les classes de noms non prédicatifs à un dictionnaire de type *lexique-grammaire*. La notion théorique mobilisée à cette fin fut celle de *sélection lexicale*, qui découle de celle de *distribution du prédicat* : étant donné un prédicat

¹⁰ De même Dubois et Dubois-Charlier, au tout début de LVF, présentent en quelque sorte l'ouvrage comme une vérification de l'hypothèse de l'*isomorphisme* de la forme syntaxique et du sens en déclarant : « *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* ».

quelconque, il y a des restrictions sur le choix des éléments qui constituent sa structure argumentale. À partir des classes de prédicats, on peut définir les classes de mots qu'ils sélectionnent. Mais le chemin inverse est possible : on peut définir des classes d'arguments en prenant en compte les prédicats qui les sélectionnent de façon *appropriée*.

Ces classes d'arguments sont appelées par Gaston Gross *classes d'objets* ; et les prédicats qui sélectionnent de façon appropriée les éléments des classes d'objets furent appelés leurs *prédicats appropriés*. Par exemple les classes de noms d'écrits, tels *article, ouvrage, roman*, ont pour prédicats appropriés des verbes tels que *lire, déchiffrer, traduire*, et *écrire, rédiger, composer*. Sur les travaux du LLI, on pourra consulter entre autres Gaston Gross (1992 et 1994), Le Pesant (1994), Mathieu-Colas (1998), et Le Pesant & Mathieu-Colas (1998).

Revenons au *Dictionnaire Electronique des Mots* (DEM). Il s'agit d'une base de données lexicales à laquelle Jean Dubois et Françoise Dubois-Charlier ont travaillé jusque peu de temps avant leur mort, intervenue respectivement en 2015 et 2016. C'est une entreprise colossale, puisque la base de données compte près de 150.000 entrées, mais elle est largement inachevée. Elle n'en constitue pas moins une « matière première » exceptionnelle pour l'élaboration de ressources lexicales nouvelles plus spécifiques (cf. sur ce point l'article de (Lapalme et Le Pesant, 2020)). Le DEM présente les quatre caractéristiques suivantes :

1. contrairement aux dictionnaires du LADL, il incorpore les mots non prédicatifs, notamment les noms concrets ;
2. il incorpore tous les registres de langue ainsi que les termes de spécialité ;
3. il incorpore, en les mettant sur le même plan que les mots simples, les locutions et les mots-phrases, que ce soient des interjections ou des pragmatèmes ;
4. il est connecté à l'autre dictionnaire informatisé de Dubois & Dubois-Charlier, LVF, au moyen d'un champ spécifique, le champ « OP1 » (cf. *infra* Figures 3 et 4) ; ce dernier met en relation des classes de noms non-prédicatifs du DEM avec des classes de verbes prédicatifs de LVF.

La quatrième caractéristique fait du DEM un outil privilégié pour l'accomplissement du projet de Gaston Gross : définir les *classes d'objets*, c'est-à-dire pour l'essentiel les noms non prédicatifs (ceux qui figurent dans le DEM), en les mettant en relation avec leurs *prédicats appropriés* (ceux qui figurent dans LVF). On développera ce point *infra* dans la Section 6.

5. Structure possible d'un lexique général de la langue corrélé à une nouvelle grammaire de tradition harrissienne

Il y a eu plusieurs étapes dans l'élaboration des grammaires de Harris. Les deux principales versions de la théorie sont *Mathematical Structures of Language* et *A Grammar of English on Mathematical Principles* publiées respectivement en 1968 et 1982. Entre ces deux étapes figure les *Notes du cours de syntaxe* (Harris, 1976) qui sont la traduction en français, par Maurice Gross, d'un cours tenu à l'université de Paris-Vincennes par Harris en 1973-1974.

Les grammaires de Harris, dans leur aspect « distributionnaliste », posent une distinction entre les *Opérateurs* (cf. *supra*, note 5) et les *arguments* (analogues respectivement aux *fonctions* et aux *variables* en mathématiques), les *arguments* pouvant être eux-mêmes des *Opérateurs*, comme dans une phrase complexe telle que « Marie croit que Pierre est parti ». Les différentes distributions possibles des prédicats et des arguments sont énumérées. Dans leur aspect « transformationnaliste », ces grammaires distinguent les trois principaux types d'opérations suivants : les *réductions* (effacements et pronominalisations), les *permutations* (changements d'ordre des mots) et les *attachements* (application au prédicat des morphèmes de temps, modes, aspects ; application des prépositions, conjonctions et adverbes ; nominalisations, autres affixations). Comme l'indique Maurice Gross dans sa préface à Harris (1976, 6), le formalisme de Harris est de type algébrique (axiomatique), de telle façon que « la minimalisation des concepts de base conduit nécessairement à une atomisation des règles ».

Quand on considère les travaux de Harris et qu'on les met en relation avec ceux de ses « héritiers », M. Gross, J. Dubois & F. Dubois-Charlier et G. Gross, on a beau reconnaître dans cet ensemble les catégories de la grammaire traditionnelle, notamment les parties du discours et les règles de remaniement de la forme de la phrase comme les permutations, les effacements et les pronominalisations, dans les faits

tout apparaît « sous un jour différent » ; on assiste pour ainsi dire à une métamorphose. On découvre ainsi, entre autres, les faits suivants :

- les mots simples ont perdu leur statut de lexèmes prototypiques, puisque les locutions, dans de vastes régions du lexique comme celui des adverbes, se révèlent largement prédominantes¹¹ ;
- le verbe a perdu son statut de prédicat prototypique : les prédicats peuvent aussi bien être des noms, des adjectifs, des adverbes, des prépositions et des conjonctions ;
- la structure de la phrase élémentaire n'est plus de type *SN-SV*, mais de type *Prédicat-argument(s)* ;
- la catégorie des connecteurs circonstanciels, temporels, argumentatifs, etc. se révèle être composée non seulement de prépositions, de conjonctions et d'adverbes, mais aussi de verbes¹², d'adjectifs, de noms (accompagnés en général de verbes-supports), voire de flexions verbales¹³ ;
- de nombreuses entrées consistent en un radical doté d'une structure argumentale, sur lequel on peut construire par affixation (y compris par affixation 0) des verbes, des adjectifs, des noms et des adverbes¹⁴ ; en d'autres termes la morphologie dérivationnelle est intégrée à la composante transformationnelle de la grammaire ;
- la catégorie des verbes supports qui, en étant porteurs de flexions verbales, permettent de « conjuguer » les adjectifs et les noms prédicatifs comme s'ils étaient des verbes, se révèle, loin d'être une catégorie verbale marginale, rassembler plus de 300 entrées¹⁵ ;

¹¹ Par exemple, dans sa *Syntaxe de l'adverbe* qui date de 1987, Maurice Gross présente un lexique-grammaire de 6.400 adverbes. Ils sont répartis entre 16 classes : 1 classe de 520 adverbes simples, tel *soudainement*, et 15 classes de 5.880 locutions adverbiales, telle *en désespoir de cause*.

¹² Par exemple la locution verbale impersonnelle *il n'empêche que* est un des connecteurs argumentatifs de la concession.

¹³ Par exemple, le préfixe itératif *re-* a des homologues dans le domaine des adverbes et des auxiliaires aspectuels.

¹⁴ Par exemple, sur le radical *calm-*, on construit *calme* (N), *calme* (Adj), *calmer*, *calmement*.

¹⁵ Par exemple, la relation entre un nom d'humain et des noms de sentiments tels que *doute* et *remords* peut être exprimée par recours aux verbes supports généraux des noms de sentiments *avoir*

- chaque classe de noms inanimés concrets (classes d'objets) entretient avec une ou plusieurs classes de prédicats une relation de sélection (des arguments par le prédicat) : les noms concrets peuvent donc être définis dans le cadre de la relation *prédicat-argument(s)*.

Les grandes parties d'un lexique corrélé aux grammaires de tradition harrissienne évoquées dans cet article seraient donc :

1. Les prédicats *connecteurs*¹⁶ : connecteurs circonstanciels, coordonnants, argumentatifs ; toutes les parties du discours sont concernées (avec des regroupements morphologiques, tel *à cause de, causer, cause, pour cause*).
2. Les prédicats autres que connecteurs, qu'on peut répartir entre deux sous-catégories :
 - 2a. les prédicats faisant partie d'une famille morphologique ; par exemple, la famille *Surprendre* réunit les lexèmes suivants : *surprenant ; riche en surprises ; agréablement surpris, désagréablement surpris ; surprendre ; surprendre agréablement, surprendre désagréablement ; créer la surprise ; surprise ; bonne surprise, mauvaise surprise ; à la surprise générale ; à la surprise de, à la grande surprise de ; être la surprise de la vie (de qqn)*.
 - 2b. les prédicats n'appartenant pas à une famille morphologique, tels les verbes *estomaquer, interloquer, méduser*.
3. Des rassemblements de plusieurs familles des types 2a. et 2b, au sein d'une Classe syntactico-sémantique commune ; par exemple la

et *éprouver* (ex. *Paul éprouve un remords*). Mais à ces verbes supports généraux peuvent être substitués les verbes suivants (à l'actif et au passif), qui dans leur emploi métaphorique à valeur principalement intensive perdent leur qualité de prédicats pour acquérir celle de verbe-support : *assaillir, consumer, crucifier, démanger, dévorer, envahir, habiter, étreindre, hanter, harceler, miner, opprimer, poursuivre, ronger, saisir, submerger, tarauder, tenailler, titiller, torturer*. (Ex. *Paul est rongé par le remords*). On doit aux travaux de Gaston Gross la prise de conscience de la richesse du vocabulaire des verbes supports.

¹⁶ Dans les grammaires de Harris, l'entité fondamentale qui s'offre à la description linguistique n'est pas la phrase élémentaire (comme c'est le cas chez Maurice Gross), mais le *discours*, qui peut d'ailleurs avoir la forme d'une phrase élémentaire voire d'une interjection. C'est la raison pour laquelle nous accordons dans notre esquisse la première place aux *connecteurs*, susceptibles de relier entre eux des discours ou des suites de discours, eux-mêmes composés de phrases simples ou complexes.

Classe SURPRISE regroupe les familles et lexèmes individuels suivants, outre *Surprendre* : *Abasourdir, Aburir, avoir le souffle coupé, avoir les jambes coupées, croire rêver, Décoiffer, estomaquer, Ébahir, Ébaubir, Éberluer, Ébouriffer, Émerveiller, Épater, Époustoufler, Étonner, être interdit, être sur le cul, Faire sensation, faire l'effet d'une bombe, Foudroyer, Frapper, Halluciner, Impressionner, interloquer, méduser, miraculeux, pantelant, scier, Sidérer, Souffler, Stupéfaire, Stupéfier, Suffoquer, tomber à la renverse, tomber de haut*¹⁷.

4. Les Opérateurs non prédicatifs : auxiliaires aspectuels, modaux (plusieurs parties du discours sont concernées), verbes supports des noms et des adjectifs, opérateurs causatifs.
5. Les lexèmes non prédicatifs : il s'agit notamment des noms inanimés concrets formant des *classes d'objets* (mises en relation avec des classes de prédicats appropriés).
6. Les interjections, pragmatèmes et autres *Phrases Préfabriquées des Interactions* (PPI) (cf. (Tutin, 2019) et (Le Pesant, 2019)¹⁸.

6. Quel type de lexicographie permet-il de représenter un modèle de la langue de type harrissien ?

Il est difficile d'imaginer, pour la représentation d'un lexique tel que celui dont nous venons d'esquisser la forme, une autre forme de dictionnaire que celle de dictionnaires tabulaires et informatisés. Ce trait peut s'expliquer en partie par la spécialité d'origine de Maurice Gross, l'informatique théorique. Mais plus fondamentalement cette forme lexicographique est incontournable quand il s'agit de produire des inventaires lexicaux comportant plusieurs dizaines de milliers d'entrées et enregistrant plusieurs dizaines de types de propriétés linguistiques : principalement morphologiques, syntaxiques, sémantiques, de domaine

¹⁷ Dans cette liste les noms de familles morphologiques sont affectés d'une initiale majuscule. L'initiale minuscule signale un verbe qui n'appartient pas à une famille morphologique.

¹⁸ La notion de PPI (*Phrase Préfabriquée des Interactions*) est due à Agnès Tutin. Cette catégorie de « phrases figées » se répartit entre quatre groupes : les *pragmatèmes* ou formules de la vie quotidiennes, les PPI réactives, les PPI métadiscursives et les PPI situationnelles (qui commentent le contexte d'énonciation). C'est un des mérites des chercheurs du LADL de les avoir recensées par milliers dès le début des années 1980. Elles figurent également en bon nombre dans le DEM de Dubois & Dubois-Charlier.

pragmatique et de registre. Le format *tabulaire informatisé* est nécessaire non seulement pour le stockage de données lexicales riches et complexes, mais aussi pour le travail des utilisateurs qui ont besoin de supports logiciels pour effectuer des recherches selon des critères multiples.

C'est la nécessité de permettre l'accès à la multiplicité des dimensions du lexique qui explique que les résultats des travaux du LADL, qui interviennent assez tôt dans l'histoire des sciences computationnelles, ont pris d'emblée la forme de tables de matrices de traits binaires. Les premières tables du lexique grammaire furent conçues par M. Gross au début des années 1970. Voici un extrait de la Classe 4 des verbes à construction complétive décrite dans Gross (1975) :

	NO =: Nhum	NO =: le fait Qu P	NO =: V1-inf W	<ENT> Ppv	Ppv =: Neg	<ENT>V	Neg	Aux =: avoir	Aux =: être	NO V ant W	NO être V- ant W	NO être V- able W	NO être V- eux W	NO être V- (eur+ teur) W	N1 =: Nabs	N1 =: le fait Qu P	N1 se V de ce Qu P	N1 se V auprès de N3hu m de ce Qu P	N1 être Vpp de ce Qu P	[passif par]	[passif de]	NO V N1 contre N2hu m	<OPT>
1						abasourdir	+	-	+	+	-	-	-	-	-	-	-	+	+	-	-	Cette nouvelle a abasourdi Max	
2	+	+	+	<E>	-	abattre	+	-	+	-	-	-	-	-	-	+	-	+	+	-	-	Que Paul ait dit cela a abattu Max	
3	+	+	+	<E>	-	abêtir	+	-	+	+	-	-	-	-	-	+	-	+	+	-	-	Que Paul soit toujours là abêtit Maria	
4	+	+	+	<E>	-	abimer	+	-	+	-	-	-	-	+	-	-	-	-	+	-	-	Que Paul ait révélé cela a abimé Marie	
5	+	+	+	<E>	-	abrutir	+	-	~	~	~	~	~	~	~	~	~	~	~	~	~	Que Max ait tant parlé a abrutit Luc	
6	~	~	~	<E>	-	abuser	+	-	+	+	-	-	-	+	-	-	-	-	+	-	-	Que Luc ait fait triché a abusé Max	
7	+	+	+	<E>	-	accaparer	+	-	+	+	-	-	-	+	-	-	-	-	+	-	-	Ces tâches accaparent Léa	
8	+	+	+	<E>	-	accomplir	+	-	+	+	-	-	-	-	-	+	-	-	+	-	-	Le travail a accompli Max	
9	+	+	+	<E>	-	accrassiner	+	-	~	~	~	~	~	~	~	~	~	~	~	~	~	Cette situation a accrassiné Marie	
10	~	~	~	<E>	-	accrassiner	+	-	~	~	~	~	~	~	~	~	~	~	~	~	~	Ce spectacle a	

Figure 1 : Table 4 des verbes simples (V_4lgt.xls) du lexique-grammaire

Cette figure représente une version du lexique-grammaire du LADL « modernisée » quant à son support, par Elsa Tolone. Les tables, diffusées sur un site de l'Université de Marne-la-Vallée, sont accessibles à partir de :

<http://infolingu.univ-mlv.fr/DonneesLinguistiques/Lexiques-Grammaires/Telechargement.html>

Des tris peuvent être pratiqués de façon croisée ou non en fonction des propriétés syntaxiques définies en tête de colonne : existence ou non d'un dérivé adjectival suffixé en *-ant* ; existence ou non d'un emploi pronominal complément de forme *se V de ce que P* ; existence ou non d'un complément supplémentaire en *contre* ; etc.

Dubois & Dubois-Charlier ont adopté un format différent, sans matrice de traits binaires, pour LVF et le DEM. On le trouvera décrit dans l'article de François, Le Pesant et Leeman (2007). Voici une image qui représente les 33 premières lignes (dans l'ordre alphabétique) de LVF+1¹⁹ :

A	B	C	D	E	F	G	H	I	J	K	L
ID	MOT	DOMAINE	MAINE en dv	LANGU	GIONALIS	OPERATEUR	CLASSE	SENS	PHRASE	CONSTRUCTION	A_able
1	abaisser 01	LOC	locatif, lieu			(#) [rid] bas qc	T3c	baisser	On a~ le rideau de fer, le store. Le rideau du magasin	T1308 - P3008	1
2	abaisser 02	TEC	technique			(#) [fire] qc VERS bas	E3f.3	incliner, pencher	On a~ la manette, le levier. La manette s'a~ vers le ba	T13g0 - P30g0	-
3	abaisser 03	QUA	quantitatif			(#) [f.mvt] moins hauteur	M3c	baisser	On a~ le mur d'un mètre. Le mur s'a~ de beaucoup.	T1306 - P3006	-
4	abaisser 04	MON	monnaie, argent			(#) [f.mvt] moins valeur	M4b.1	baisser	On a~ les prix, les revenus de dix pour cent. Les prix	T1306 - P3006	-
5	abaisser 05	MED	médecine			(#) [rid] bas quantité	T4b	faire descendre	Le malade a~ la fièvre avec l'aspirine. La fièvre s'a~.	T1308 - P3008	-
6	abaisser 06	PSYt	psychologie littéraire			(#) [m.e.état] mauvais qn abstrai	R2a.3	avilir, humilier	On a~ P, son orgueil en le blâmant. On s'a~ en public	T1108 - P1000	-
7	abaisser 07 (s')	PSY	psychologie			(#) [f.mvt] humble A+infinif	M2b.2	s'humilier à	On s'a~ à demander une faveur, à cette demande.	P10a0	-
8	abaisser 08 (s')	VEH	véhicule motorisé			(voie) [lc qp] vers bas	L3a.1	descendre vers	La route s'a~ vers la rivière.	P3001	-
9	abaisser 09 (s')	PSY	psychologie			(#) [f.mvt] A état vil	M2b.2	s'avilir, tomber jusqu'à	On s'a~ au niveau de cet escroc.	P10a0 - T11a0	-
10	abalourdir (s')	PSYt	psychologie littéraire			(#) [rid] balourd (adjectif)	T2b	abêtir, abrutir	On s'a~ avec un tel film. On a~ P avec ce livre. La télé	P1000 - T9106	-
11	abandonner 01	DRO	droit, administration			(#) [dat] qc A qn	D2a	laisser, léguer	On a~ ses biens à ses enfants, à une fondation.	T13a0	-
12	abandonner 02	MAR	marine			(#) [dat] mouvement A qc	D3a.2	laisser aller	On a~ sa barque à un fort courant.	T13a8 - P30a8	-
13	abandonner 03	EQU	équitation, hippologie			(#) [dgrp] rênes	S3b.1	lâcher, laisser aller	On a~ les rênes de l'attelage.	T1300	-
14	abandonner 04	PSY	psychologie			(#) [dgrp] abstrait	S4b.1	renoncer à	On a~ un projet, une idée. Le projet est a~.	T1300	-
15	abandonner 05	SOC	sociologie			(#) [dgrp] abstrait	S4b.1	quitter, lâcher	On a~ ses études. On n'a~ pas après un premier éch	T1300 - A10	-
16	abandonner 06	SPO	sports			(#) [dgrp] lutte	S2a.1	s'avouer vaincu	On a~ le match. Le boxeur a~ au premier round.	T1300 - A10	-
17	abandonner 07	COM	commerce			(#) [li.mut] CONTRE/POUR	T3f.2	quitter	On a~ un appartement pour une maison.	T13k0	-
18	abandonner 08	SOC	sociologie			(#) [dgrp] lieu	S3h	quitter, inhabité	On a~ ce village devenu désert. Le village est a~.	T1307	-
19	abandonner 09	LOC	locatif, lieu			(#) [dgrp] qn qp	S2d.1	quitter, laisser	On a~ un chien sur la route, un enfant dans la voiture.	T1101	-
20	abandonner 10	SOC	sociologie			(#) [dgrp] abstrait	S4b.1	renoncer à	On a~ ses responsabilités. Après cet échec, on a~.	T1300 - A10	-
21	abandonner 11	SOC	sociologie			(qc) [dgrp] qn	S2d.1	quitter, manquer à	Ses forces, son énergie, la chance ont a~ P.	T3100	-
22	abandonner 12 (s')	PSY	psychologie			(#) [ger.mens] A qc/infinif	P1h.1	se laisser aller à	On s'a~ à un grand désespoir, à désespérer.	P10a0	-
23	abandonner 13 (s')	LANT	langue, pard littéraire			(#) [loq] confiance	C1a.2	s'épancher	On s'a~ dans la détresse.	P1006	-
24	abasourdir 01	PAT	pathologie			(#) [f.som] étourdi	H1b.1	étourdir, assourdir	On a~ P avec la radio. Le bruit a~ P.	T1108	-
25	abasourdir 02	PSY	psychologie			(abstrait) [f.sent] vif à qn	P2a.1	stupéfier	La nouvelle a~ P. On reste a~ après cette annonce.	T3100	-
26	abat-carre	PEA	peuserie, tannerie			(#) [rag] qc avec abat-carre	R3f.1	tailler un méplat	L'ouvrier a~ une peau de mouton.	T1308 - P3000	-
27	abâtardir 01	LITt	littérature littéraire			(#) [rid] bâtard abstrait	T4b	dégénérer, s'altérer	On a~ la littérature avec ces publications. Le roman s	T1308 - P3008	-
28	abâtardir 02 (s')	BIOT	biologie littéraire			(qc) [d] bâtard	T3a	dégénérer	La race s'a~. est a~. Un métissage a~ cette race.	P3000 - T3308	-
29	abattre 01	OBJ	objet			(#) [fire] qc en bas	E3f.3	renverser, faire tomber	On a~ les arbres sur le sol. Le chêne est a~ par la ten	T13g8	1
30	abattre 02	MIL	militaire			(#) [ict] mauvais aéronef	F3c.1	descendre	On a~ un avion avec un missile.	T1308	-
31	abattre 03	CUI	cuisine, pâtisserie			(#) [ict] animal à mort	F1b.1	tuer	On a~ les boeufs avec une masse dans un abattoir.	T1208	-
32	abattre 04	PATf	pathologie familial			(#) [ict] qn à mort	F1b.1	descendre	Le tueur a~ P d'une balle à la sortie de l'hôtel.	T1108	-
33	abattre 05	PSY	psychologie			(abstrait) [f.sent] vif à qn	P2a.1	déprimer, accabler	Ce coup du sort a~ P. La fièvre a~ P. On est a~.	T3100	-
34	abattre 06	POI	politique			(#) [ict] total on abstrait	F2a.1	renverser	On a~ le gouvernement par un vote	T1108	-

Figure 2 : Image de LVF+1 sur tableur Excel

Dans ce format, plusieurs classes (cf. le champ « Classe ») peuvent être rassemblées sur la même table, mais les propriétés syntaxiques (cf. le champ « Construction ») et sémantiques (cf. le champ « Opérateur ») sont codées et renvoient potentiellement à des grammaires locales. Par exemple, dans la ligne de l'entrée *abandonner* #1 (Ex. *abandonner* qqch à qqn), le code « T3a0 » dans le Champ *Construction* renvoie à la description « verbe à sujet *humain* transitif direct à complément indirect en *à* », et le code correspondant dans le Champ *Opérateur* spécifie qu'on a affaire à un verbe de *don* (code « dat »).

¹⁹ LVF+1 est une version de LVF corrigée par Paul Sabatier, sous la supervision des auteurs.

Les tables des dictionnaires informatisés de Dubois & Dubois-Charlier (LVF et DEM) peuvent être exploitées au moyen de tableurs OpenOffice ou Microsoft-Excel, qui permettent des tris. Elles ont été ultérieurement converties en format XML par Guy Lapalme (Université de Montréal). Ce dernier a produit récemment un nouvel outil de consultation, en format JSON. Cette interface permet une consultation combinée de LVF, qui est un dictionnaire extrêmement raffiné de verbes prédicatifs simples associés à leurs éventuels déverbaux, et du DEM, qui est pour l'essentiel un dictionnaire de verbes locutionnels, et de noms, d'adjectifs, d'adverbes, de conjonctions et d'interjections tant simples que locutionnels. L'outil de consultation, dans sa version la plus récente est accessible à : <http://rali.iro.umontreal.ca/rali/?q=fr/versions-informatisees-lvf-dem>. On trouvera des précisions sur son utilisation dans l'article de (Lapalme et Le Pesant, 2020).

Supposons qu'on souhaite obtenir la liste des noms d'écrits littéraires figurant dans le DEM. Un sondage sur quelques noms tels que *roman*, *poème* etc. suggère que ce sont des noms appartenant au Domaine LIT (littérature) et qu'ils sont codés « écrire N » dans le champ « CONT ». Grâce à la partie « Dictionnaire électronique des mots » de l'interface de consultation, on obtient la liste de 277 noms d'écrits littéraires, dont les 18 premiers dans l'ordre alphabétique sont ceux qui figurent sur cette image :

Les verbes français

Dictionnaire électronique des mots

Mot cat gnr type littérature Précision CONT écrire N OP OP1 R4a1 JSON

Mot*	no	cat	gnr	type	Domaine	CONT	OP	OP1	Sens
abstract		N	M non-anime	LIT		écrire N	text	R4a1	résumé d un texte
acrostiche		N	F non-anime	LIT		écrire N	text	R4a1	pièce d vers à initiale
acte	2	N	M non-anime	LIT		écrire N	text	R4a1	division d pièce
addenda		N	M non-anime	LIT		écrire N	text	R4a1	ajouts à texte
allégorie		N	F non-anime	LIT		écrire N	text	R4a1	symbole, personnificatio
amphibolie		N	F non-anime	LIT		écrire N	text	R4a1	équivoque, ambiguïté
anagramme		N	M non-anime	LIT		écrire N	text	R4a1	mot p mut place lettres
anecdote		N	F non-anime	LIT		écrire N	text	R4a1	histoire isolée
annales		N	F non-anime	LIT		écrire N	text	R4a1	histoire événementielle
annexe	3	N	F non-anime	LIT		écrire N	text	R4a1	doc annexé à texte
annuaire		N	M non-anime	LIT		écrire N	text	R4a1	liste annuelle abonnés
anthologie		N	F non-anime	LIT		écrire N	text	R4a1	recueil textes choisis
antistrophe		N	F non-anime	LIT		écrire N	text	R4a1	strophe en réponse
aphorisme		N	M non-anime	LIT		écrire N	text	R4a1	sentence, proverbe
apocryphe		A	M non-anime	LIT		écrire N	text	R4a1	écrit faux, suspect
apologie		N	F non-anime	LIT		écrire N	text	R4a1	genre panegyrique, éloge
apologue		N	M non-anime	LIT		écrire N	text	R4a1	genre récit moral court
apoptegme		N	M non-anime	LIT		écrire N	text	R4a1	sentence

277 entrées, dont 27 expressions

Rechercher dans les résultats :

Figure 3 : Image du DEM en format JSON

Le Champ « OP1 » renvoie à la Classe R4a1 de LVF. On peut dès lors faire « basculer » la requête vers la partie « Les verbes français » de l'interface de consultation, pour obtenir la liste des verbes de la Classe R4a1 dans le Domaine LIT. On obtient une liste de 37 verbes (dont voici les 10 premiers), accompagnés de leurs éventuels déverbaux (dans le Champ Noms) :

• Les verbes français ° Dictionnaire électronique des mots

Mot Lex littérature Précision Classe R4a.1 Opérateur Construction Chercher JSON Aide

Mot*	no	gr	lex	Domaine	Classe	Opérateur	Const.	Sens	Phrases	Adjectifs	Noms	Déverbal
accoucher	2	1aZ	5	LIT	R4a.1	(#) [fab] oeuvre de pensée	N1b	rédigier, pondre	On a- d'un livre difficile.		accouchement	
bâtir	6	2aZ	5	LIT	R4a.1	(#) [fab] écrits	T1306 P3000	construire	On b- une phrase, un article, un roman.			
commettre	3	5Z	5	LIT	R4a.1	(#) [fab] écrits	T1306 P3000	composer	On a c- des poèmes.			
compiler	1	1aZ	3	LIT	R4a.1	(#) [fab] qc en réunissant	T1306 P3000	réunir	On c- un recueil avec des textes divers.	compilable incompilable	compilation compilateur	
composer	3	1aZ	5	LIT	R4a.1	(#) [fab] oeuvre de pensée	T1306 P3000	écrire, rédiger	Cet auteur c- de bons romans.		composition compositeur	
confectionner	2	1aZ	5	LIT	R4a.1	(#) [fab] oeuvre de pensée	T1306 P3000	écrire, rédiger	On c- un rapport.		confectionne confection	
construire	4	6cZ	5	LIT	R4a.1	(#) [fab] oeuvre pensée	T1306 P3000	produire, écrire	Cet auteur c- un roman, une phrase.		construction	
créer	8	1bZ	5	LIT	R4a.1	(#) [fab] rôle	T1306 P3000	représenter 1ère fois	L'acteur c- un personnage. Le metteur en scène c- une pièce.		création	
développer	4	1aZ	5	LIT	R4a.1	(#) [fab] oeuvre de pensée	T1306 P3000	écrire, composer	On d- une idée, un paragraphe. Son sujet se d- sur trois pages.		développement	
écrire	3	6dZ	5	LIT	R4a.1	(#) [fab] graphie du mot	T1306 P3000	orthographe	On é- ce mot avec deux "r". Ce mot s'é-, est é- avec deux "p". ..		écrit	
écrire	4	6dZ	5	LIT	R4a.1	(#) [fab] oeuvre de pensée	T1306	composer, rédiger	Cet auteur é- un livre.		écriture	

37 verbes Rechercher dans les résultats:

Figure 4 : Image de LVF en format JSON

Les données lexicales peuvent également être implémentées sur un logiciel d'ingénierie linguistique comme NooJ (cf. Silberztein, 2015) à des fins de tâches de traitement de corpus, notamment l'annotation syntactico-sémantique. On en trouvera un exemple dans Sabatier et Le Pesant (2013, 165-168). Un autre exemple d'annotation syntactico-sémantique de corpus est présenté en détail dans Lapalme et Le Pesant (2020), à partir de données empruntées à la classe C2b (verbes de communication) de LVF.

7. Conclusion. Dans la mouvance des travaux de Zellig Harris, la lexicographie peut-elle être considérée comme une « méthodologie pour la recherche d'un nouveau modèle de la langue » ?

Du point de vue heuristique, la lexicographie se situant dans la mouvance des travaux de Harris peut assurément être considérée comme une méthodologie pour la recherche d'un nouveau modèle de la langue : les

données lexicales, représentées de façon aussi exhaustive et systématique que possible dans des dictionnaires informatisés, permettent la vérification, la correction et l'enrichissement de la théorie. Mais d'un autre point de vue, l'entreprise de classification des données lexicales, leur organisation par classes et sous-classes selon des critères syntaxiques et sémantiques, ainsi que le recours à une lexicographie informatisée disponible pour le TAL, sont bien plus qu'une méthodologie de la recherche : ces réalisations font partie du *modèle* même de la langue, de sa *théorie*, en somme de sa *grammaire* au sens large du terme. Nous croyons que, dans le cadre théorique général des *grammaires distributionnelles et transformationnelles* d'inspiration harrissienne, le *modèle* (ou *théorie*) d'une langue donnée ne saurait avoir d'autre format que celui d'un *lexique-grammaire informatisé*.

Les travaux de Zellig Harris, Maurice Gross, Gaston Gross et J. Dubois & F. Dubois-Charlier sont loin de former un corpus accompli : ils appellent des recherches ultérieures, tant sont nombreux les aspects restés à l'état d'ébauche, par exemple la question des verbes supports et celle de l'essence prédicative d'une partie des prépositions et des conjonctions. Parallèlement, les dictionnaires informatisés que plusieurs de ces chercheurs ont conçus ont besoin d'être dépassés. Le plus « totalisateur », le DEM de Dubois & Dubois-Charlier, a l'inconvénient d'être resté dans un état d'inachèvement important. Nous n'avons donc dans cet article pas eu d'autre ambition que celle d'esquisser, à partir de l'exemple des quelques lexiques-grammaires existants, la forme idéale que pourrait prendre un lexique-grammaire informatisé du français.

Références bibliographiques

- BOONS, J.-P., GUILLET, A. LECLÈRE, C., *La structure des mots simples en français. Constructions intransitives*, Genève, Droz, 1976.
- CHEVALIER, J.-Cl., ENCREVÉ, P., *Combats pour la linguistique. De Martinet à Kristeva*, Paris, ENS Éditions, 2006.
- DUBOIS, J., DUBOIS-CHARLIER, F., *Les Verbes Français*, Paris, Larousse, 1997.
- DUBOIS, J., DUBOIS-CHARLIER, F., La combinatoire lexico-sémantique dans le Dictionnaire Électronique des Mots. Les termes du domaine de la musique à titre d'illustration, *LANGAGES*, 2010, **179-180**, 31-56.
- FRANÇOIS, J., LE PESANT, D., LEEMAN, D., Présentation de la Classification des Verbes Français, de Jean Dubois et Françoise Dubois-Charlier, *LANGUE FRANÇAISE*, 2007, **153**, 3-19.

- GROSS, G., Forme d'un dictionnaire électronique, in CLAS, A., SAFAR, H., *L'environnement traductionnel*, Sillery, Presses de l'Université de Québec, 1992, 255-271.
- GROSS, G., Classes d'objets et description des verbes, *LANGAGES*, 1994, **115**, 15-30.
- GROSS, G., Pour une véritable fonction « synonymie » dans un traitement de texte, *LANGAGES*, 1998, **131**, 103-114.
- GROSS, M., *Méthodes en syntaxe. Régime des constructions complétives*, Paris, Hermann, 1975.
- GROSS, M., Les bases empiriques de la notion de prédicat sémantique, *LANGAGES*, 1981, **63**, 1981, 7-52.
- GROSS, M., La notion harrissienne de transformation, *LANGAGES*, 1990, **99**, 39-56.
- HARRIS, Z., *Notes du cours de syntaxe*, Paris, Le Seuil, 1976.
- LAPALME, G., LE PESANT, D., Deux dictionnaires informatisés de Jean Dubois et Françoise Dubois-Charlier, leurs ultimes travaux, *LINX*, 2020, **80**.
- LE PESANT, D., Les compléments nominaux du verbe *lire*, *LANGAGES*, 1994, **115**, 31-46.
- LE PESANT, D., Suggestions méthodologiques et outils de traitement de corpus pour l'étude des Phrases Préfabriquées des Interactions, *CAHIERS DE LEXICOLOGIE*, 2019, **114**, 93-118.
- LE PESANT, D., MATHIEU-COLAS, M., Introduction aux classes d'objets, *LANGAGES*, 1998, **131**, 6-33.
- MATHIEU-COLAS, M., Illustration d'une classe d'objets : les voies de communication, *LANGAGES*, 1998, **131**, 77-90.
- ROUX, S., Lire *La Théorie physique* aujourd'hui, in DUHEM, P., *La théorie physique, son objet, sa structure*, 2^e édition, 1914, Paris, CNRS Éditions, 2016.
<https://books.openedition.org/enseditions/6910?lang=fr#ftn1>
- SABATIER, P., LE PESANT, D., Chapitre 5. Les dictionnaires électroniques de Jean Dubois et Françoise Dubois-Charlier et leur exploitation en TAL, in GALA, N., ZOCK, M., *Ressources Lexicales.*, Amsterdam, John Benjamins, 2013, 153-186.
- SILBERZTEIN, M., *La formalisation des langues. L'approche de NooJ*, London, Iste Editions, 2015.
- TUTIN, A. Phrases préfabriquées des interactions : quelques observations sur le corpus CLAPI, *CAHIERS DE LEXICOLOGIE*, 2019, **114**, 63-91.