



**HAL**  
open science

# Un modèle lexicographique basé sur le corpus Scientext comparé avec d'autres ressources lexicographiques : l'exemple des constructions verbales

Rui Yan, Agnès Tutin

## ► To cite this version:

Rui Yan, Agnès Tutin. Un modèle lexicographique basé sur le corpus Scientext comparé avec d'autres ressources lexicographiques : l'exemple des constructions verbales. *Texte et corpus*, 2013, 5. hal-02058748

**HAL Id: hal-02058748**

**<https://hal.science/hal-02058748>**

Submitted on 6 Mar 2019

**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.

# **Un modèle lexicographique basé sur le corpus Scientext comparé avec d'autres ressources lexicographiques : l'exemple des constructions verbales**

*Rui YAN & Agnès TUTIN*

Laboratoire LIDILEM - Université Stendhal Grenoble 3

## **Résumé :**

Dans cet article, nous présentons les premiers résultats de nos travaux sur la modélisation des patrons verbaux appartenant au lexique scientifique transdisciplinaire sur un corpus analysé morpho-syntaxiquement composé d'articles de recherche en sciences humaines et sociales pour ensuite une application didactique dans l'aide à la rédaction scientifique des étudiants étrangers. Nous comparons les résultats avec le modèle lexicographique *Les verbes du français* de Dubois et Dubois-Charlier pour déterminer lequel est mieux adapté à notre objectif didactique et évaluer si le modèle lexicographique *LVF* peut être exploité comme base d'amorce pour un repérage semi-automatique des patrons syntaxiques.

## **1. Introduction**

Aujourd'hui, le corpus est considéré comme un outil incontournable d'observation et d'investigation dans la description linguistique de la langue (Tutin, 2007a), ce qui amène à la constitution des corpus représentatifs/spécialisés de la langue pour les besoins de l'étude linguistique, dont on peut citer le projet Scientext<sup>1</sup>. Ce projet intitulé «Corpus et outils pour étudier le positionnement et le raisonnement de l'auteur dans les écrits scientifiques » s'inscrit dans le cadre de l'ANR. Il s'agit d'un corpus spécialisé d'écrits scientifiques variés permet aux chercheurs intéressés par le texte scientifique de consulter des ressources textuelles qui leur permettront d'appliquer plusieurs types de recherches dont l'étude du lexique scientifique pour l'aide à la rédaction des écrits universitaires auprès des étudiants étrangers, ce qui constitue le cadre de notre travail.

---

<sup>1</sup> <http://scientext.msh-alpes.fr>

Les recherches en didactique ont montré que l'acquisition du lexique ne saurait se réduire aux mots isolés, mais qu'il faut, en outre, s'appropriier la structure syntaxique et d'autres éléments lexicaux qui les entourent en introduisant une composante discursive et une composante socio-culturelle (Cavalla, 2008). Cet aspect nous intéresse plus particulièrement et notre étude Inspiré du modèle de Hanks Corpus Pattern Analysis (*CPA*, analyse des patrons basée sur les corpus) s'attache à la description des constructions verbales (les patrons lexico-syntaxiques des verbes) du lexique scientifique transdisciplinaire. Nous montrons dans la présente étude, la description des patrons lexico-syntaxiques sur quatre verbes les plus fréquents dans le corpus Scientext : *considérer, proposer, concerner et obtenir*.

Certes, l'étude du corpus est incontournable pour étudier les propriétés sémantiques et syntaxiques des mots, mais il faut aussi prendre en considération les contributions d'une tradition linguistique. Dans cet article, nous allons observer la description des informations sémantiques et syntaxiques de ces quatre verbes dans la ressource lexicale de référence *Les Verbes français (LVF)* tout en menant une comparaison avec nos premiers résultats relevés du corpus.

Notre objectif est double: d'une part, nous cherchons à déterminer les modèles qui apparaissent les plus adaptés pour une aide à la rédaction scientifique, tout en faisant l'hypothèse que le modèle *CPA*, par sa base empirique sur le corpus, sera particulièrement approprié à cette tâche. D'autre part, nous souhaitons évaluer si le modèle lexicographique *LVF* peut être exploité comme base d'amorce pour un repérage semi-automatique des patrons syntaxiques. Nous envisageons en effet d'automatiser partiellement le repérage des patrons verbaux par des techniques de traitement automatique des langues, qui seront validés manuellement.

Après avoir défini dans un premier temps les patrons lexico-syntaxiques du lexique verbal scientifique transdisciplinaire. Nous aborderons par la suite notre méthodologie, inspirée du modèle *CPA*, pour le repérage et la modélisation des patrons verbaux. Nous expliquerons ensuite la ressource lexicographique *LVF* issue de la tradition linguistique française en faisant une comparaison parallèle avec notre modèle. Après l'analyse comparative des résultats, nous cherchons à déterminer jusqu'à quel point les informations syntaxiques et sémantiques de notre modèle sont plus adaptées pour aide à la rédaction scientifique avant de conclure sur les apports de notre modèle et la perspective de la recherche.

## 2. Les patrons lexico-syntaxiques du lexique verbal scientifique transdisciplinaire

### 2.1. Lexique scientifique transdisciplinaire (LSTS): un lexique omniprésent dans l'écrit scientifique

Les écrits scientifiques renvoient à des pratiques intellectuelles et à des genres codifiés par des communautés de discours en présentant des objectifs et des procédures communs – opérations d'analyse, de raisonnement, d'évaluation (Tutin, 2007a). Partant des besoins des étudiants étrangers, notre étude s'attache aux écrits scientifiques en milieu universitaire (mémoire, thèse, articles scientifiques) qui est défini comme des écrits élaborés dans un contexte académique par des spécialistes d'un domaine – des experts – traitant une problématique clairement posée à laquelle ils tentent d'apporter des réponses (Swales, 1900). Nous abordons le terme « science » dans un sens large incluant les sciences exactes ainsi que les sciences humaines et sociales.

Toujours dans une perspective didactique, nous nous intéressons au lexique qui décrit les activités scientifiques : de l'exposition du cadre de la recherche aux résultats en passant par la méthodologie et l'argumentation. Ce genre de lexique est non terminologique et fait partie d'un continuum de lexique aux frontières floues entre terminologie et langue générale. Il reçoit plusieurs appellations en français «Vocabulaire Général d'Orientation Scientifique » (Phal, 1970), «Langue Scientifique Générale » (Pecman, 2005), et « lexique transdisciplinaire des écrits scientifiques - LSTS » (Tutin, 2007b).

A la suite de (Tutin, 2007b), nous définissons le LSTS comme « le lexique abstrait du discours commun à la communauté scientifique, largement présenté ou même surreprésenté dans les écrits scientifiques et renvoyant aux objets, concepts et procédures de l'activité scientifique ». Nous illustrons ci-dessous le LSTS (éléments soulignés) par quelques exemples de phrases appartenant aux différents domaines et tirés du corpus Scientext :

- (1) Notre programme de recherche trouve principalement son intérêt dans cet aspect puisqu'il s'agissait de poursuivre en France ce travail initié aux Etats-Unis. [psy-art-356-body]
- (2) Cette modélisation nous permet de dire que dans le premier cas, autotest est considéré comme un acteur et dans le second comme un processus. [lin-com-253-body]

## 2.2. Les patrons verbaux dans les écrits scientifiques

Influencé par le travail pionnier du didacticien Hornby, Hunston et Francis (Hunston & Francis, 2000) prennent un point de vue relativement traditionnel sur la notion de patron qui est considéré comme une configuration syntaxique intégrant des catégories grammaticales et des mots fonctionnels comme préposition. Par exemple un patron typique associé au verbe *counsel* dans la phrase *My advisers counseled me to do nothing* peut être représenté comme *V n to-inf*. Plusieurs verbes peuvent entrer dans ce patron: *ask, tell, advise, urge, recommend*, etc. Il s'agit souvent des synonymes. En effet, pour distinguer les sens de ces verbes, il faut aussi examiner les collocations et les types sémantiques au niveau de différents arguments autour du verbe. En ce sens, la recherche de Hanks est sans doute allée plus loin, selon le linguiste:

*A pattern consists of a valency structure, together with sets of preferred collocation (Hanks, 2013, p. 92).*

Notre intérêt porte sur le verbe, le noyau de la phrase. Sa fonction est d'élaborer une description mobilisant les référents des constituants nominaux. Dans un contexte scientifique, le verbe et les patrons permettent d'organiser le discours, structurer le texte, établir les liens entre les différentes unités lexicales, etc. (Williams, Millon, & Alonso, 2011).

Suite à Hanks, nous définissons un patron lexico-syntaxique du verbe comme « Tous les mot lexicaux (arguments) et grammaticaux (préposition, conjonction) qui sont régulièrement associés à un verbe dans une structure syntaxique et que ce verbe mobilise pour la construction d'un sens ».

Reprenons l'exemple *notre programme de recherche trouve principalement son intérêt dans cet aspect*. Pour le verbe *trouver*, cet exemple rejoint plusieurs lignes de concordances dans Scientext, ce qui nous permet de voir que le verbe a une tendance à choisir comme sujet une personne, celle-ci peut alterner avec le type sémantique étude scientifique. Concernant le complément d'objet, les collocatifs significatifs repérés sont *intérêt* et *compte*, suivis toujours par un complément introduit par la préposition *dans*. De cela, nous pouvons dégager le patron suivant. Ici, le verbe *trouver* prend le sens de *découvrir*:

[[Personne |étude scientifique]] trouve (ADV [[Quantité]]) {intérêt |compte} dans certaines choses<sup>2</sup>

### 3. Des approches sur les patrons et *Corpus Pattern Analysis (CPA)*

#### 3.1. De Hornby à Hanks

Dès le début des années 1930, Hornby a découvert que « le langage en usage est fortement lexicalisé, chaque mot n'est associé qu'à certains patrons lexico-syntaxiques » (Hanks, 2008, p. 2). Et ce n'est qu'avec le corpus et la linguistique du corpus que la nature « patterned » du langage est mise en évidence. Ce changement est marqué avec le projet de *COBUILD Advanced Learner's English Dictionary* et les travaux de Sinclair, selon qui, les patrons et les sens sont étroitement liés. Ses principes (en particulier, l'étude des collocation et des patrons basée sur l'analyse du corpus) ont été développés par plusieurs linguistes tels que Michael Hoey (Hoey, 2005) avec son modèle de *lexical priming* (amorçage lexical), Susan Hunston et Gill Francis (Hunston & Francis, 2000) avec l'approche *pattern grammars* (grammaires des patrons), Hanks (Hanks & Pustejovsky, 2005, Hanks 2008, 2013) avec l'approche *pattern dictionary* dans le cadre du projet *Corpus Pattern Analysis (CPA)*, analyse des patrons basée sur les corpus).

Le phénomène de *pattern* ou *construction* est aussi mis en évidence dans l'approche *construction grammar* (Fillmore, Paul & Mary, 1988, Goldberg, 1995, 2006) qui est plutôt fondée sur l'introspection des linguistes avec l'exemple tiré du corpus. Ces approches portant sur les grammaires des patrons défendent le lien étroit entre lexique et grammaire ou plus précisément syntaxe et proposent que le sens d'un mot est « déterminé par les emplois co-textuels et contextuels antérieurs dans lesquels il apparaît, qu'il s'agisse de l'environnement lexical et des collocations, mais aussi de son environnement sémantique, syntaxique, pragmatique et discursif » (Legallois & Tutin, 2013, p. 8).

Cependant, parmi ces approches, il nous semble que le modèle *CPA* est plus approprié à un objectif didactique ayant pour principales propriétés d'être :

1. basé sur des usages réguliers
2. destiné à la désambiguïisation lexicale dans la mesure où un patron d'un mot

---

<sup>2</sup> a) Double crochets indiquent le type sémantique, ex : [[Personne]], un rôle sémantique peut être précisé après le signe égal, b) Accolades indiquent un élément lexical spécifique, ex : {article}, c) Parenthèses indiquent une option qui peut être absente, mais si elle est présente, elle peut influencer le sens du verbe, ex : (ADV [[Quantité]]).

est composé non seulement d'une structure syntaxique et les catégories grammaticales (sujet, objet, complément, etc.) des arguments (comme *pattern grammars*) mais aussi des collocations significatives. Ces collocations significatives sont généralisées et classées dans un type sémantique qui peut changer le sens du mot.

### 3.2. *Corpus Pattern Analysis (CPA)*

Le *CPA* est à la fois une approche sur les patrons et une lexicographie de corpus qui s'inscrit la linguistique du corpus de Sinclair. Le premier modèle lexicographique du projet est intitulé « A pattern Dictionary of English Verbs (PDEV)<sup>3</sup> ». Ce dictionnaire basé sur le corpus BNC consiste à analyser les sens normaux des verbes et de les distinguer des exploitations (usages créatifs, originaux des mots). Chaque sens (implicature<sup>4</sup>) est associé à un patron et non pas au verbe parce que selon Hanks, le sens des mots n'existe qu'en contexte et il consiste en un groupe d'éléments sémantiques.

Pattern : [[Activity | Institution]] swallow [[Resource = Money]] {(up)}

Implicature : [[Activity| Institution]] completely uses up [[Resource = Money]]

Sur le plan sémantique, les groupes lexicaux occupant les arguments sont généralisés dans des types sémantiques entre la double crochets, les rôles sémantiques et les collocations sont parfois précisés pour affiner l'analyse sémantique des arguments. D'ailleurs, les compléments des verbes tels que infinitifs, complétives, compléments circonstanciels sont décrits dans les patrons.

En outre, sauf l'anglais de la langue générale, *CPA* est aussi adapté en espagnol (Alonso & Renau, 2013) et en italien (Jezek & Frontini, 2010). D'ailleurs, Alonso et Renau ont montré que l'approche *CPA* pourrait être appliquée dans les langues de spécialité comme environnement.

En ce qui concerne notre travail, nous nous focalisons sur les usages normaux des verbes dans un contexte scientifique afin d'observer leur comportement lexico-syntaxique dans les écrits scientifiques.

---

3 Accessible en ligne <http://deb.fi.muni.cz/pdev/>

4 Le sens est représenté dans l'implicature, une paraphrase du patron employant un verbe différent et une phraséologie différente.

## **4. Modélisation des patrons lexico-syntaxiques dans le corpus Scientext**

### **4.1. Corpus d'analyse**

Notre corpus d'analyse est un sous-corpus de Scientext composé d'articles de revues préalablement sélectionnées. Les textes ont été analysés avec le logiciel Syntex et annotés morpho-syntaxiquement avec annotation des parties textuelles (introduction, développement, conclusion, résumé, etc.)

Plus précisément, le corpus d'analyse comporte environ trois millions de mots provenant de dix disciplines des sciences humaines et sociales : anthropologie, économie, géographie, histoire, linguistique, sciences de l'éducation, sciences politiques, sciences de l'information, sociologie et psychologie.

### **4.2. Extraction des verbes du LSTS**

Partant d'une liste des verbes préétablie avec les taux de fréquence par discipline, nous nous limitons aux verbes de sens pleins et riches (les verbes auxiliaires, les verbes supports, les verbes modaux sont exclus). Nous avons combiné les deux critères telles que la fréquence (la moyenne du taux de fréquence de dix disciplines au-dessus de 0.1%) et la répartition disciplinaire (répartis dans les dix disciplines). Pour cette étude préliminaire, nous étudions les quatre verbes les plus fréquents extraits du corpus : *considérer, proposer, concerner et obtenir*.

### **4.3. Identification des patrons lexico-syntaxiques**

Concernant le verbe, l'identification d'un patron lexico-syntaxique se fait en trois étapes : la première repose sur l'observation et du comportement syntaxique du verbe, il s'agit de regrouper des lignes de concordances qui présentent la même structure syntaxique pour établir le premier profil syntagmatique du verbe. La deuxième étape consiste à extraire les collocatifs significatifs appartenant à chaque structure syntaxique par des techniques de traitement automatique de langues. Ces collocatifs seront par la suite placés dans les groupes lexicaux. La troisième étape est la généralisation et la distribution des rôles et types sémantiques pour chaque groupe lexical, ce qui est essentiel pour la désambiguïsation des sens. A la fin, une fois que le patron est établi, il faut y associer un sens. Chaque patron est associé à un sens. Ils sont interdépendants.

### 4.3.1. proposer

Le verbe *proposer* provient du latin *proponere* « poser devant », nous avons retenu quatre patrons les plus fréquents relevant de trois sens. Le premier sens est de « faire connaître quelque chose à qqn, soumettre à son choix, à son adhésion, en laissant la liberté de rejet ». Deux patrons y sont associés :

1. Patron : {[[Personne]] |Article| Travail} propose (ADV [[Logique |temps |lieu]])  
[[Quelque chose d'abstrait = noms d'artefacts scientifiques et noms de processus scientifiques]]  
Implicature : {[[Personne]] |Article| Travail} fait connaître (ADV [[Logique |temps |lieu]]) [[Quelque chose d'abstrait = noms d'artefacts scientifiques et noms de processus scientifiques]] par écrit
2. Patron : [[Quelque chose d'abstrait]] est proposé par [[Personne]] pour atteindre un objectif  
Implicature : [[Quelque chose |Abstrait]] est soumis comme choix par [[Personne]]

Il s'agit d'un usage courant du verbe mais dans un contexte scientifique, nous avons retenu deux constatations : d'une part, le sujet est souvent l'auteur représenté par *on* ou *nous*, d'ailleurs, le sujet peut aussi être un article ou un travail scientifique. On sait bien que c'est l'auteur qui fait ce travail scientifique qui propose, ainsi, celui-ci peut être considéré comme une alternance du sujet – auteur. D'autre part, l'objet direct relevant de quelque chose d'abstrait désigne souvent un nom d'artefacts scientifique (*modèle, approche, idée, démarche, etc.*) ou un nom de processus (*modification, élaboration, etc.*). D'ailleurs, nous avons repéré des constructions intéressantes sur le plan didactique comme *l'approche que nous proposons, le projet proposé*. Quant au deuxième patron, il n'est qu'une variation au passif du premier.

Le deuxième sens du verbe est « communiquer un conseil par écrit » lié au patron suivant :

3. Patron : {[[Personne]] |Article| Travail} propose (ADV [[Logique |temps |lieu]]) de faire quelque chose  
Implicature : {[[Personne]] |Article| Travail} communique par écrit (ADV [[Logique |temps |lieu]]) un conseil de faire quelque chose pour bien mener une étude scientifique.

Cet usage pourrait relever non seulement de l'écrit scientifique mais aussi de la langue générale. Nous avons remarqué la même alternance lexicale au niveau de sujet comme le premier patron.

Enfin, le troisième sens consiste en un emploi pronominal « se fixer comme but ou projet de faire quelque chose » auquel est associé le quatrième patron :

4. Patron : {[[Personne]] |Article| Travail} se propose de faire quelque chose

Implicature : [[Personne |Article| Travail]] a pour objet de faire quelque chose dans le but de bien mener une étude scientifique

De même, cet usage peut se rencontrer aussi bien dans la langue générale que dans l'écrit scientifique.

#### 4.3.2. considérer

Le verbe *considérer* vient du latin *considerare* dont le sens d'origine est d'«examiner attentivement par les yeux ou par la pensée». Dans scientext, nous avons retenu quatre patrons relevant de trois sens. Sous le premier sens « donner une estimation ou un jugement » sont regroupé deux patrons :

1. Patron : [[Personne]] (peut) (ADV [[Logique]]) considérer [[Quelque chose d'abstrait |phénomène]] comme [[d'autre chose]]

Implicature : [[Personne]] (peut) (ADV [[Logique]]) donner son jugement sur [[Quelque chose d'abstrait |phénomène]]

2. Patron : [[Quelque chose d'abstrait |phénomène]] (peut) (ADV [[Logique]]) être considéré comme [[d'autre chose]]

Implicature : [[Quelque chose d'abstrait |phénomène]] (peut) (ADV [[Logique]]) être vu comme [[d'autre chose]], selon le jugement de quelqu'un

Ici, trois constats peuvent être mentionnés : le premier porte sur le sujet qui est souvent représenté par *nous* ou *on*, le deuxième se rapporte au verbe modal *pouvoir* qui est souvent précédé du verbe *considérer*. Enfin, les objets entrant dans la classe sémantique « quelque chose d'abstrait » ou « phénomène » renvoie aux outils, méthodes, notions, objets scientifiques, etc. Alors que le deuxième patron n'est qu'une variation au passif du premier.

Dans le deuxième sens soit l'indication d'une opinion, nous n'avons repéré qu'un

seul patron.

3. Patron : [[Personne]] (peut) (ADV [[Logique]]) considérer [complétive introduite par *que*]

Implicature : [[Personne]] (peut) (ADV [[Logique]]) avoir une opinion sur un fait scientifique

La complétive introduite par *que* correspond à une opinion sur un fait scientifique. D'ailleurs, le verbe est souvent précédé du verbe modal *pouvoir*. Ici, le verbe porte le même sens que les verbes *croire* et *penser*.

Le troisième sens est « prendre en considération », il correspond au quatrième patron relevé:

4. Patron : Si/Lorsque [[Personne]] considère (ADV [[Logique| quantité]]) [[quelque chose d'abstrait]]

Implicature : Si/Lorsque [[Personne]] envisage (ADV [[Logique| quantité]]) [[quelque chose d'abstrait]] par un examen attentif ou critique.

Il s'agit plutôt d'une construction associée au contexte scientifique. Les noms qui se trouvent à la place de l'objet sont variés mais appartenant à une classe sémantique commune « quelque chose d'abstrait » comme *notion*, *indicateur*, *élément*, *caractéristique*, etc.

#### 4.3.3. concerner

Pour le verbe *concerner*, nous avons identifié trois patrons qui activent deux sens différents. Le premier sens est « avoir rapport à, toucher » qui s'associe au patron suivant :

1. Patron : [[Quelque chose]] concerne (ADV [[Logique]]) [[Personne | Quelque chose d'autre]]

Implicature : [[Quelque chose]] se rapporte à [[Personne | Quelque chose d'autre]]

2. Patron : [[Quelque chose | Personne]] est concerné par quelque chose d'autre

Implicature: [[Quelque chose | Personne]] est touché par quelque chose d'autre

Etant donné la variété des choix du verbe en matière de sujet, il semble impossible de généraliser le type de sujet qui peut être un nom observable de l'activité scientifique comme *objet*, *phénomène*, *exemple*, nom de support comme *tableau*, *figure*, nom de processus scientifique tel que *description*, nom de caractérisation ou même nom d'artefacts scientifiques. La même question se pose aussi pour le complément d'objet, le

verbe peut avoir comme objet non seulement une chose mais aussi une personne.

Le deuxième patron est une variation au passif du premier patron. Nous avons remarqué que le passif est plus fréquent. En plus, la forme du participe passé est très fréquente sous forme d'un groupe nominal comme dans l'exemple (3) :

(3) *Les principaux territoires **concernés** par cette libération des terres seraient la Chine, l'Afrique sub-saharienne et l'Amérique latine, et notamment le Brésil. [geo-art-23-body].*

Le deuxième sens « relative à, au sujet de » est lié à une locution prépositive représenté dans le troisième patron :

3. Patron : pour ce qui concerne [[Quelque chose]]...

Implicature : au sujet de [[Quelque chose]]...

Ce genre de locutions omniprésent dans l'écrit scientifique est destiné à établir les liens entre les différents éléments référentiels. Voir l'exemple (4) :

(4) ***Pour ce qui concerne** les revues en libre accès, les articles en sont téléchargés quatre fois plus que ceux qui ne sont pas libres d'accès. De ce fait ces revues offrent une plus grande visibilité aux travaux des chercheurs. [infcom-art-232-body].*

Grâce à la locution *pour ce qui concerne*, l'auteur réussit à lier le fait que *les articles en sont téléchargés quatre fois plus que ceux qui ne sont pas libres d'accès* à quelque chose telle que *les revues en libre accès*.

#### 4.3.4. obtenir

Concernant le verbe *obtenir*, nous avons identifié cinq patrons appartenant au trois sens. Le premier sens « parvenir à se faire accorder » est activé par trois patrons.

1. Patron : [[Personne]] obtient [[quelque chose d'abstrait]]

Implicature : [[Personne]] parvient à se faire accorder [[quelque chose d'abstrait]]

2. Patron : Pour obtenir [[quelque chose d'abstrait]]

Implicature : Afin de se faire donner [[quelque chose d'abstrait]]

3. Patron : [[Activité scientifique| quelque chose d'abstrait |objet]] permet (à [[Personne]]) d'obtenir [[quelque chose d'abstrait]]

Implicature: [[Activité scientifique| quelque chose d'abstrait |objet]] permet (à [[Personne]]) de se faire donner [[quelque chose d'abstrait]]

Pour le premier patron, Le verbe choisi comme sujet une personne souvent représentée par *on* ou *nous*, et objet quelque chose qui peut être soit un objet comme *diplôme*, soit quelque chose d'abstrait comme *solution*. Concernant le deuxième « Pour obtenir [[quelque chose |abstrait]] », le verbe *obtenir* à l'infinitif précédé par la préposition *pour* indique un but. Ici, quelque chose d'abstrait peut renvoyer à *analyse*, *vision*, *image*, etc. Concernant le troisième patron, les noms qui se trouve au sujet sont assez variés, mais l'objet renvoie souvent à quelque chose d'abstrait comme *taux*, *niveau*, *dynamique*, *résultat*, *prévision*, etc.

Le deuxième sens est « décrocher » associé au patron suivant :

4. Patron : [[Personne| organisme]] obtient [[statistique]]

Implicature : [[Personne]|organisme] parvient à décrocher [[statistique]]

Au niveau de sujet, on retrouve le pronom *on* ou *nous*, qui peut alterner avec un organisme. Quant à l'objet, la classe sémantique statistique regroupe des collocatifs comme *score*, *note*, *pourcentage*, *taux*, *suffrage*, *voix*, etc.

Dans le corpus, nous avons remarqué que le verbe *obtenir* ont une grande tendance à choisir le nom *résultat* comme objet, ce qui active un sens différent soit « réussir à atteindre un résultat ». Un seul patron est identifié dans ce sens :

5. Patron : [[Personne]] obtient {résultat, effet}

Implicature : [[Personne]] parvient à {résultat, effet}

Il est intéressant à noter que dans ce sens, nous avons aussi repéré des alternances syntaxiques : *le résultat est obtenu* et *le résultat que [[Personne]] obtient*.

## **5. Comparaison entre CPA adapté et LVF**

### **5.1. Description des ressources : Les Verbes français (LVF)**

Dans le *LVF*, J. Dubois et F. Dubois-Charlier proposent la classification des verbes français autour des propriétés sémantiques et syntaxiques. C'est un modèle lexicographique des verbes français exhaustif couvrant « 25 610 entrées verbales simples représentant 12310 verbes différents dont 4 188 à plusieurs entrées » (Dubois &

Dubois-Charlier, 1997, p. 1). Il existe en version papier et est accessible en ligne<sup>5</sup>.

Basé sur les « méthodes classiques de la grammaire distributionnelle et transformationnelle », il montre l'adéquation entre les constructions syntaxiques et l'interprétation sémantique par les classes syntactico-sémantiques, ce qui rejoint le principe du modèle *CPA* où les patrons lexico-syntaxiques et les sens sont fortement liés.

Sur le plan sémantique, la description se fait à l'aide d'une suite de caractères alphabétiques dénommée « opérateur » qui « interprètent sémantiquement les schèmes syntaxiques ». A chaque entrée verbale correspond un opérateur complexe (composé des opérateurs simples) différent qui contribue à la construction des classes syntactico-sémantiques des verbes. Par exemple, l'opérateur de l'entrée *considérer* 02 (un emploi parmi les cinq) « ger.mens QUE tel » veut dire « avoir telle pensée que ». Il indique que l'entrée *considérer* 02 appartient au sous-type syntaxique P1g.1 qui se lit « avoir activité mentale portant sur tel objet ». Chaque alphabet ou chiffre indique un niveau de classement (cinq niveaux de classement). Ainsi, ce sous-type syntaxique relève de la sous-classe syntaxique P1g « prendre qc comme objet d'activité mentale » qui à son tour appartient à la classe sémantico-syntaxique P1 « avoir tel sentiment, telle activité consciente » et la classe générale « verbes psychologiques ».

Sur le plan syntaxique, chaque variante syntaxique est exprimée à l'aide d'un « schème » qui comporte une association de lettres et de chiffres entre crochets. A chaque entrée verbale peut correspondre un ou deux schèmes de construction (dont l'un relève souvent de l'emploi pronominal). Par exemple, le schème de l'entrée *considérer* 02 « T1500 » se lit « transitif direct à sujet humain, et suivi par une complétive ou un infinitif ».

Les opérateurs et les schèmes font ressortir les propriétés sémantiques et syntaxiques des verbes et constituent deux éléments primordiaux du *LVF*.

## **5.2. Comparaison avec le *LVF* – exemple du verbe *proposer***

Nous passons maintenant à une comparaison parallèle en examinant les quatre verbes dans le *LVF*. Dans cet article, nous montrons la comparaison par l'exemple du verbe *proposer*.

---

<sup>5</sup> <http://rali.iro.umontreal.ca/Dubois/>

Domaine	classe	opérateur	sens	Exemple	schème	
1	COM	C2c	dic pour décision A qn	soumettre	On p~ à P un prix pour sa voiture.	<a href="#">T13a8</a>
2	SOC	D1a.2	dat qn A poste	présenter comme candidat	On p~ P à ce poste. On se p~ pour diriger, comme secrétaire.	<a href="#">T11a0</a> , <a href="#">P10a</a>
3	LIT	C2b.2	dic offre A qn DE	suggérer, offrir à	On p~ à P de l'aider, que la voiture vienne le prendre.	<a href="#">T15a8</a>
4	ENS	D3b.1	dat sujet A qn	donner à traiter	On p~ aux candidats un sujet difficile, un test.	<a href="#">T13a0</a>
5	LIT	C2b.2	dic conseil A qn DE	conseiller	On p~ à P qu'il reste à dîner, de rester à dîner.	<a href="#">T15a8</a>
(se) 6	PSY	C2a.2	dic projet à soi DE	se promettre de	On se p~ de rédiger un article.	<a href="#">P1400</a>

**Figure 1 : Les 6 emplois du verbe *proposer* répertoriés dans le *LVF***

Comme on peut le constater dans la figure 1, les deux entrées **proposer 02 et 04** relèvent de la classe générique « D – don, privation » alors que les autres appartiennent à la classe générique « C – communication ». Dans le corpus Scientext, les usages de « donner un poste » et « donner un sujet à traiter » sont absents, nous nous concentrons donc aux entrées de la classe « C – communication ».

L'entrée **proposer 01** est regroupé sous la classe C2c « décrire qc à qn au moyen de ». Le schème syntaxique (T13a8) présente la construction transitive directe à sujet humain et objet chose, le complément d'objet direct chose peut être suivi par la préposition *à* qui introduit un complément d'objet indirect humain. La valeur d'opérateur *dic pr décision A qn* s'entend « communiquer qc pour décision à qn ». Cette entrée correspond au première patron du verbe *proposer*. Cependant, il faut noter deux différences : tout d'abord, le *LVF* n'allant pas plus loin, décrit tout simplement le complément d'objet direct comme chose ; alors que dans un contexte scientifique, le complément d'objet direct dans le premier patron est spécifié comme une abstraction qui renvoie à un nom d'artefacts scientifique (modèle, approche, idée, démarche, etc.) ou un nom de processus (modification, élaboration, etc.) :

(5) *Christopher Wlezien a dès lors **proposé une explication** mieux adaptée aux variations du public américain, la théorie thermostatique. [spo-art-36-body].*

Ici, le nom *explication* peut être considéré comme un nom de processus scientifique. Ensuite, dans l'entrée **proposer 01**, le complément d'objet indirect humain désigne une personne précisée qui a le droit d'accepter ou refuser la proposition, ce qui n'est pas le cas dans le premier patron où le complément d'objet indirect humain est non précisé, et par défaut les lecteurs qui lisent l'article. Cet aspect reflète la différence entre la langue

générale et la langue destinée à décrire les activités scientifiques.

Les entrées **proposer 03** et **proposer 04** appartiennent à la classe C2b. Pour **03** sa valeur d'opérateur *dic offre A qn D* se lit « communiquer un offre à qn (impersonnel infinitif en de) » alors que **04** *dic conseil A qn D* se traduit comme « communiquer un conseil à qn (impersonnel infinitif en de) ». Elles partagent le même schéma syntaxique (T15a8) qui présente une construction transitive directe à sujet humain et objet complétive ou infinitif, ce dernier peut être suivi par la préposition *à* qui introduit un complément d'objet indirect humain. Dans l'écrit scientifique, il est impossible pour l'auteur de communiquer une offre aux lecteurs, ce qui n'étonne pas que l'emploi de **proposer 03** soit absent dans notre corpus. Par contre, l'auteur peut communiquer un conseil par écrit. Ainsi, l'entrée **proposer 04** correspond au troisième patron « {[[Personne]] |Article| Travail} propose (ADV [[Logique |temps |lieu]]) de faire quelque chose ». Néanmoins, dans ce patron, le complément d'objet indirect est non précisé. D'ailleurs, ce qui est proposé à faire doit contribuer à bien mener une recherche scientifique.

L'entrée **proposer 06** concerne un emploi pronominal, ce qui est expliqué dans son schéma syntaxique P1400 (type pronominal, sujet humain et objet complétive ou chose). Elle correspond au quatrième patron « {[[Personne]] |Article| Travail} se propose de faire quelque chose » dans la mesure où elle active le sens que la personne se fixe comme projet de faire quelque chose. Dans l'écrit scientifique, ce projet consiste à bien mener une recherche scientifique.

## 6. Résultats et analyses

### 6.1. Correspondance partielle

Sur le plan sémantique, dans le *LVF*, les opérateurs permettent de relever les préférences de sélection pour les arguments comme humain, chose, abstrait, etc. et d'interprètent sémantiquement les schèmes syntaxiques. Mais ils ne suffisent pas pour désambiguïser le sens. Par exemple, pour l'entrée verbale obtenir 02, l'opérateur « grp abstrait DE » ne peut se distinguer de celui pour obtenir 06 « grp abstrait » tandis que les sens sont différents: dans le premier cas, il veut dire « se faire donner pour » comme dans la phrase *On obtient un rendez-vous de P pour son fils*; dans le deuxième, le sens se rapporte à la statistique « décrocher » comme le montre la phrase *Le parti obtient dix pour cent*. Dans ce cas, il faut recourir à la rubrique sens pour faire la distinction.

Cette ambiguïté est levée dans le *CPA* par la prise en considération des collocatifs, des types et des rôles sémantiques. Par exemple, pour l'entrée verbale obtenir 06, cet emploi se trouve aussi dans Scientext, l'approche *CPA* nous permet de préciser le sujet comme [[Personne| organisme]] et l'objet comme [[statistique]]. C'est en ce sens que le patron faire ressortir le sens automatiquement, ici, c'est bien le sens « décrocher » qui est activé.

Sur le plan syntaxique, dans le *LVF*, les constructions sont relativement restreintes à l'emploi pronominal tandis que dans un contexte scientifique, comme le montre nos résultats, les verbes peuvent avoir des constructions assez variées comme la relativation, la passivation, construction impersonnelle, pronominal, etc.

De manière générale, à travers l'étude des quatre verbes, la comparaison montre une correspondance entre les deux modèles. D'une part, le *LVF* répertorie presque tous les emplois usuels des verbes français, mais pour l'adapter à la langue de spécificité comme l'écrit scientifique, ces informations lexicographiques ne sont pas suffisamment précisées: la description syntaxique (en particulier, les tournures appartenant à l'écrit scientifique) n'est pas complète et la description sémantique par le biais des opérateurs ne permettent pas parfois à la désambiguïsation du sens. D'autre part, les patrons dégagés dans Scientext couvrent un sous-ensemble d'emplois des verbes, mais présentent certaines spécificités du comportement sémantique/ syntaxique liées au genre de l'écrit scientifique.

## **6.2. Intérêt du modèle *CPA* pour les applications didactiques**

Selon les résultats obtenus du corpus, il nous semble que le modèle *CPA* est approprié pour une aide à la rédaction scientifique dans la mesure où les patrons font ressortir des éléments argumentatifs liés à l'écrit scientifique qui sont néanmoins absents dans un modèle lexicographique destiné à la langue générale comme le *LVF*. Nous présentons ci-dessous certains éléments argumentatifs relevés :

### Alternance du type sémantique

Par exemple, dans le contexte où le verbe indique un apport scientifique de l'auteur comme *proposer*, on constate qu'il existe une alternance régulière au niveau du sujet entre [[Personne]] et {Article} ou {Travail} :

(6) *Christopher Wlezien a dès lors proposé une explication mieux adaptée aux variations du public américain, la théorie thermostatique. [spo-art-36-body] ([[Personne]])*

(7) *Cet article propose un modèle microéconomique simple permettant d'expliquer le rôle du revenu dans l'évolution de la pollution. [eco-art-102-summary] ({Article} ou {Travail})*

Dans les articles scientifiques français, l'auteur a tendance de s'effacer devant l'article qu'il construit. L'emploi des métonymiques du type *cet article* permet à l'auteur de se manifester implicitement. Ce constat rejoint Tutin (2010) dans une étude portant sur la présence auctoriale et du positionnement de l'auteur :

*« Une autre façon pour l'auteur de ne pas se mentionner explicitement est de recourir à des expressions métonymiques comme ce travail ou cette étude. L'auteur s'efface ainsi devant l'ouvrage qu'il construit. L'emploi de ces expressions n'empêche pas en principe de mettre en avant un positionnement affirmé à l'aide du verbe. (Tutin, 2010 : 36).*

Si l'on examine l'exemple du verbe *proposer*, un effet obtenu est que le positionnement de l'auteur est teinté d'une couleur moins subjective et présentant donc une dimension argumentative.

### Alternances syntaxiques : le passif et les phrases relatives

Les alternances syntaxiques permettent de produire des constructions variées. Nous constatons que certains types d'alternances syntaxiques sont plus fréquents pour certains verbes. Prenez le verbe *obtenir*, le passif est plus fréquent que l'actif:

(8) *Ce résultat est obtenu dans un cadre théorique où échouerait une politique de discrimination intertemporelle pure. [eco-art-74-introduction]*

### Associations des mots : On+pouvoir

L'association du type on peut (pourrait) considérer implique une dimension négociative et dialogique dans la mesure où elle atténue la responsabilité de l'auteur en impliquant les pairs et ouvrant un dialogue avec la communauté des lecteurs, comme le montre l'exemple (9) :

(9) *Sur le plan symbolique, **on pourrait considérer** que la figure d professeur renvoie à une forme hiérarchique plus explicite que la figure de l'enseignant. [sed-art-28-body]*

Le point de vue défendu par l'auteur est aussi celui de la communauté entière , et de ce fait, comme l'indique Tutin (2010: 36), l'opinion défendue gagne en validité scientifique : elle n'est plus subjective mais partagée.

#### Connecteurs argumentatifs

Comme le montre l'exemple (10), des adverbes tels que *dès lors*, *également* et *donc* apparaissent souvent après le verbe:

(10) *On **proposera donc** un schéma, pour illustrer et synthétiser cette dynamique de la conservation volontaire des ruines urbaines traumatiques dans une perspective d'atténuation du risque (fig. 1). [geo-art-188-body].*

Ces adverbes peuvent être considérés comme des connecteurs argumentatifs qui servent à établir des liens entre des phrases et des éléments lexicaux.

#### Locutions et constructions

Les locutions et constructions associées régulièrement aux verbes constituent un élément essentiel pour produire un texte scientifique argumentatif. Concernant les quatre verbes étudiés, nous avons obtenu par exemple des locutions comme *pour ce qui concerne* et des constructions telles que *si l'on considère*, *pour obtenir*, *permettre d'obtenir*, *le résultat que l'on obtient*.

On voit donc que les patrons lexico-syntaxiques repérés dans le corpus, qui doivent par la suite être affinés et complétés avec des statistiques par la technique TAL, permettent de mettre en évidence certains éléments tant sémantiques que syntaxiques liés au fonctionnement et aux spécificités de l'écrit scientifique. Même si le nombre des verbes étudiés ne nous permet pas de généraliser les éléments linguistiques observés, il nous semble intéressant de les souligner dans la présente

étude et d’approfondir ces points en augmentant le nombre des verbes étudiés dans la poursuite de l’étude des patrons lexico-syntaxiques.

### 6.3. Récapitulatif

Même si ces deux modèles ne sont pas fondés sur les mêmes théories linguistiques (approche distributionnelle et transformationnelle de *LVF*, approche patron de *CPA* adapté à notre étude), ils présentent des points communs, d’une part, les deux modèles portant sur le verbe se basent sur le principe que le sens et la syntaxe sont fortement liés, d’autre part, les propriétés sémantiques et syntaxiques sont associées et contribuent ensemble à la description des unités lexicales. Si le *LVF* est exhaustif par sa couverture exceptionnelle des entrées des verbes, notre modèle inspiré du *CPA* se montre adapté et spécialisé à l’écrit scientifique en mettant en évidence les préférences des verbes du LSTS au niveau des arguments et des structures syntaxiques. Nous présentons ci-dessous un tableau récapitulatif de la comparaison :

	<b>LVF</b>	<b>Ressources lexicales des patrons lexico-syntaxiques inspiré du modèle CPA</b>
<b>Entrée</b>	Verbe et sens normaux de la langue générale	Verbe et sens normaux dans l’écrit scientifique
<b>Ressources</b>	les données de la tradition lexicographique française	Corpus Scientext
<b>Type de liens</b>	Description de propriétés sémantiques et syntaxiques	Description des patrons lexico-syntaxiques
<b>Interdépendance entre le lexique, le sémantique et la syntaxe</b>	Adéquation entre les constructions syntaxiques et l’interprétation sémantique	Chaque patron apparaît avec un groupe d’éléments lexicaux restreint, et associé à un sens particulier
<b>Propriétés syntaxiques</b>	1. Schème syntaxique défini par la nature des constituants et les paradigmes lexicaux des constituants 2. Alternances syntaxiques pauvres	1. Patron lexico-syntaxique composé des mots lexicaux (arguments et ses valeurs) et grammaticaux (préposition, conjonction, etc.) 2. Relativisation, passivation
<b>Propriétés sémantiques</b>	« opérateur » explicite mais pas suffisamment précisé pour la désambiguïsation, niveau de généralisation reste très vague (quelque chose, quelqu’un, abstrait, etc.)	Types et rôles sémantiques, niveau de généralisation relativement affiné
<b>Description lexicographique</b>	Domaine, classe sémantico-syntaxique, opérateur (sémantique), sens, conjugaison, exemple (un pour chaque sens), construction syntaxique, dérivation	Patron syntaxique, types et rôles sémantiques des arguments, sens, commentaires, indices supplémentaires (expressions), exemple (plusieurs exemples tirés du corpus et de divers disciplines)

**Tableau 1 : comparaison des caractéristiques d’une ressource lexicale inspirée du CPA avec le LVF**

## 7. Conclusion

De manière générale, notre étude préliminaire montre que *LVF* reste une ressource lexicale très riche pour les verbes français, mais cela ne signifie pas qu'il peut remplacer un dictionnaire spécialisé des constructions verbales comme aide à l'écrit scientifique. Parce que ce dernier présente des constructions syntaxiques intéressantes permettant non seulement de mieux articuler et structurer le texte comme *si l'on considère*, mais aussi exprimer la position de l'auteur comme *on peut considérer que*, et des préférences de sélection, en particulier en matière des groupes nominaux, plus adaptées et spécifiques au genre de l'écrit scientifique. D'ailleurs, des alternances (le passif et les phrases relatives) et des constructions lexico-syntaxiques fréquentes sont aussi présentes. Nous avons présenté une correspondance partielle entre le *LVF* et notre ressource lexicographique inspiré du modèle *CPA*. A travers cette comparaison, notre hypothèse du départ est confirmée, notre ressource inspiré du *CPA* et basé sur le corpus Scientext est appropriée pour une aide à la rédaction scientifique. Cela n'empêche pas que le *LVF* étant une ressource à grande couverture, riche en informations sémantiques peut être exploité comme une base d'amorce dans notre étude.

Pour la suite, nous souhaitons augmenter le nombre des verbes étudiés pour que le résultat soit plus représentatif et envisager un repérage semi-automatique des patrons, en particulier le repérage des collocatifs. D'ailleurs, nous envisageons de classer les catégories sémantiques des noms scientifiques, ce qui permettra de préciser le niveau de généralisation.

## Bibliographie

Alonso, A. & Renau, I. (2013). Corpus Pattern Analysis in determining specialised uses of verbal lexical units. *Terminalia* 7: 26-33.

Cavalla, C. (2008). Les écrits universitaires : Un français spécifique pour les apprenants étrangers. In O. Bertrand & I. Schaffner (Éd.), *Le français de spécialité : Enjeux culturels et linguistiques*, 93-104. Paris: Ecole Polytechnique.

Dubois, J., & Dubois-Charlier, F. (1997). *Les Verbes français*. Larousse.

François, J., Le Pesant, D., & Leeman, D. (2007). Présentation de la classification des Verbes Français de Jean Dubois et Françoise Dubois-Charlier. *Langue française*, 153(1),

3-19.

Hadouche, F., & Lapalme, G. (2010). Une version électronique du LVF comparée avec d'autres ressources lexicales. *Langages*, 179-180(3-4), 193-220.

Hanks, P. (2008). *Lexical Patterns: from Hornby to Hunston and beyond*. *Barcelona: Iula, Documenta Universitaria*. Consulté à l'adresse [http://www.euralex.org/elx\\_proceedings/Euralex2008/004\\_Euralex\\_2008\\_Plenary\\_Patri ck%20Hanks\\_Lexical%20Patterns\\_from%20Hornby%20to%20Huston%20and%20Beyond.pdf](http://www.euralex.org/elx_proceedings/Euralex2008/004_Euralex_2008_Plenary_Patri ck%20Hanks_Lexical%20Patterns_from%20Hornby%20to%20Huston%20and%20Beyond.pdf)

Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations* (MIT Press.). MIT Press.

Hoey, M. (2005). *Lexical priming: a new theory of words and language*, London: Routledge.

Hunston, S. & Francis, G. (2000). *Pattern Grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins.

Jezek, E. & Frontini, F. (2010). From Pattern Dictionary to Patternbank, in G.-M. de Schryver (a cura di), *A Way with Words: Recent Advances in Lexical Theory and Analysis*. Kampala: Menha Publishers.

Legallois, D. & Tutin, A. (2013). Présentation: Vers une extension du domaine de la phraséologie. *Langages*, 189(1), 3-25.

Pecman, M. (2005). Les apports possibles de la phraséologie à la didactique des langues étrangères. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, Vol. 8, n° 2.

Phal, A. (1970). *Vocabulaire général d'orientation scientifique (V.G.O.S.) - Part du lexique commun dans l'expression scientifique*. Paris Didier.

Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings: Cambridge Applied Linguistics*, Cambridge University Press.

Tutin, A. (2007a). Autour du lexique et de la phraséologie des écrits scientifiques. *Revue française de linguistique appliquée, Lexique et écrits scientifiques*(XII(2)), 5-14.

Tutin, A. (2007b). Modélisation linguistique et annotation des collocations : une application au lexique transdisciplinaire des écrits scientifiques. In S. Koeva, D. Maurel, & M. Silberztein (Éd.), *Formaliser les langues avec l'ordinateur : de INTEX à Nooj* (Vol. 8, p. 189-215). Besançon: Presses Universitaires de Franche Comté.

Williams, G., Millon, C., & Alonso, A. (2011). Collocational networks and their application to an E-Advanced Learner's Dictionary of Verbs in Science (DicSci), (Actes de la 2e conference eLexicography in the 21st century: New applications for new users, Bled (Slovenie)). Consulté à l'adresse <http://www.trojina.si/elex2011/Vsebine/proceedings/eLex2011-2.pdf>