

L'extraction et la modélisation de patrons lexico-syntaxiques pour leur enseignement en FLE : un exemple à partir du verbe *montrer*

Rui Yan et Sylvain Hatier (Grenoble)

Abstract

This study focuses on scientific writing and aims at improving verbal patterns teaching for L2 learners. Using a French corpus of scientific articles in humanities and social sciences (Scientext), we have collected and modeled lexicosyntactic patterns for the verb *montrer* and two of its synonyms. We aim at extracting specific structures for those three semantically close verbs, assuming that their semantic proximity will also be revealed in similar syntactic properties. We thus create a set of verbal patterns, with cooccurrence informations, to combine them to a tool designed to help L2 learners in their scientific writing.

1 Cadre et objectifs de l'étude

Dans le contexte de l'écrit scientifique, le verbe, en tant que pivot de la phrase, contribue à exprimer un grand nombre de fonctions rhétoriques et discursives telles *formuler une hypothèse, faire la comparaison, donner une définition, ou exprimer des opinions*. Sur le plan didactique, ce lexique verbal métascientifique et métadiscursif (cf. Tutin/Grossmann 2014) est largement étudié sous l'angle transdisciplinaire dans le cadre de l'*English for Academic Purposes* (EAP) sous le nom d'« academic vocabulary » ou « sub-technical vocabulary » (cf. Coxhead 2000 ; Paquot 2010). Pour le français, les études similaires dans les champs du *Français sur Objectifs Universitaires* ou *Français sur Objectifs Académiques* sont plus récentes (cf. Cavalla 2008, 2014) sous l'appellation de la « Langue Scientifique Générale » (LSG) (Pecman 2005), ou du « lexique transdisciplinaire des écrits scientifiques » (LST) (Tutin 2007 ; Drouin 2007). Ces dernières années, l'intérêt pour les verbes du LST s'est progressivement accru, avec par exemple quelques études de cas sur les verbes dans le cadre du projet Scientext (<http://scientext.msh-alpes.fr/>, [30.06.2016]) (cf. Grossmann 2014 ; Tutin 2010).

Comme le soulignent Granger et Paquot (2009: 2), la non-maîtrise de ces verbes à fonctions métalinguistique et métadiscursive constitue un handicap certain pour les apprenants, qui ne peuvent alors exprimer leur pensée avec nuance et dans le style approprié¹. Cette difficulté est liée aux difficultés à maîtriser l'utilisation des verbes du LST qui ont tendance à apparaître

¹ “Insufficient knowledge of verbs [is] a serious handicap for learners as it prevents them from expressing their thoughts in all their nuances and couching them in the expected style”.

dans des patrons lexicaux², par exemple, [*résultat/analyse*] *conforter* [*idée/hypothèse*], [*on/nous*] *adopter* [*approche, modèle*] et des constructions syntaxiques préférentielles³ (*il est intéressant de noter que, nous nous proposons d'étudier, résultat permet de dégager, les données sont recueillies*).

De nombreux chercheurs ont constaté des difficultés liées à l'emploi des patrons lexicaux des verbes chez les apprenants (cf. Nesselhauf 2005 ; Hyland 2008 ; Granger/Paquot 2009). Ce constat rejoint une étude précédente (cf. Hatier/Yan 2015) dans laquelle nous comparons un sous-corpus⁴ de Scientext et un corpus composé de mémoires de recherche d'un groupe d'apprenants chinois⁵. Les résultats montrent que par rapport aux experts, les apprenants ont des difficultés à maîtriser certaines constructions, en particulier celles qui servent à structurer le discours (*si l'on considère*), guider le lecteur dans le plan du texte (*comme nous l'avons montré*) et établir les liens entre les éléments textuels (*ceci s'explique par*).

Ainsi, le besoin d'une ressource spécialisée, rendant compte des unités préfabriquées (collocations⁶, patrons) pour faciliter la rédaction scientifique des apprenants, est bien réel (cf. Williams 2008). Nous pouvons citer quelques outils novateurs pour l'anglais : le *Louvain English Academic Purposes Dictionary (LEAD)* (cf. Granger/Paquot 2010) et l'*E-Advanced Learner's Dictionary of Verbs in Science (DicSci)* (cf. Williams/Million/Alonso 2012). En français, nous pouvons mentionner la *Base lexicale du français (BLF)*, destinée à la langue générale (cf. Verlinde/Selva/Binon 2009). Cependant, il n'existe pas encore d'outils spécifiques pour aider les apprenants du FLE dans leur rédaction d'écrits universitaires tels que le mémoire.

Dans cette perspective, le présent article a pour objectif d'examiner les propriétés sémantiques et syntaxiques d'un exemple de verbe représentatif du genre des écrits scientifiques et académiques, le verbe *montrer* (et ses synonymes *démontrer* et *indiquer*), afin d'en proposer une modélisation linguistique qui sera intégrée dans Dicorpus, outil d'aide à la rédaction universitaire en ligne (cf. Tutin/Falaise 2014). Nous souhaitons, en donnant accès à une ressource lexicographique comprenant des informations sur les patrons verbaux, que les apprenants puissent intégrer les spécificités du comportement lexico-syntaxique des verbes dans l'écrit scientifique.

Nous présentons dans un premier temps la méthode d'extraction automatique des cadres de sous-catégorisation verbaux⁷ effectuée d'après le corpus analysé syntaxiquement. Ces cadres sont ensuite regroupés et analysés sémantiquement et syntaxiquement pour mettre en évidence les patrons lexico-syntaxiques des verbes en suivant le modèle *Corpus Pattern*

² Le concept de *patron* (de l'anglais *pattern*) s'inscrit dans la linguistique contextualiste. Il consiste en une structure syntaxique intégrant des collocations privilégiées.

³ Nous entendons par construction syntaxique préférentielle une configuration syntaxique récurrente pour un verbe donné (intégrant une construction particulière telle le passif impersonnel, un verbe support tel *permettre*, etc.).

⁴ Composé d'articles de recherche de revue en sciences humaines et sociales.

⁵ Nous envisageons d'intégrer à cette étude un corpus d'écrits d'étudiants natifs (corpus de littéracie avancée) afin de le comparer aux corpus d'experts et de non-natifs.

⁶ Au sens donné par Tutin (2014): « expressions récurrentes, essentiellement binaires, dont les éléments entretiennent une relation syntaxique comme *faire une hypothèse* ».

⁷ Ensemble des relations de dépendances impliquant le verbe étudié.

Analysis (CPA, analyse des patrons basée sur les corpus) de Hanks (2013). Pour l'identification des acceptions verbales, nous prenons pour référence la ressource lexicographique *Les Verbes français (LVF)* de Dubois et Dubois-Charlier (1997), qui intègre informations sémantiques et syntaxiques. La modélisation des patrons est ensuite validée par un examen manuel des occurrences dans le corpus. Dans un deuxième temps, nous analysons les patrons aux niveaux sémantique et syntaxique avant leur intégration dans Dicorpus, un outil d'aide à la rédaction scientifique.

2 Aspects méthodologiques

2.1 Le choix des verbes étudiés

Nous avons choisi de nous intéresser à des verbes très fréquents et partageant une certaine proximité sémantique. Nous avons opté pour le verbe *montrer* qui, dans notre corpus de travail de 5 millions de mots, est le second verbe du LST le plus fréquent (après *permettre*) avec 3053 occurrences (soit 630 occurrences par million de mots). Ce verbe se caractérise par une fonction marquée de démonstration scientifique et s'insère dans des patrons lexico-syntaxiques variés : *nous montrons que, comme le montre le tableau suivant, ce résultat nous montre que*, etc. Dans un deuxième temps, les synonymes du verbe *montrer* ont été choisis en consultant le *LVF*⁸ et le dictionnaire des synonymes CRISCO (www.crisco.unicaen.fr/des/, [30.06.2015]).

Parmi les synonymes de *montrer*, nous avons sélectionné les verbes *indiquer* et *démontrer* en raison de leur appartenance au LST et de leur haute fréquence dans le corpus de travail (2^e et 3^e synonymes de *montrer* les plus fréquents avec respectivement 901 et 315 occurrences). Ces trois verbes, partageant certains traits sémantiques et syntaxiques, sont fréquemment utilisés dans des procédés de démonstration scientifique.

Le *LVF* couvre « 25 610 entrées verbales simples représentant 12 310 verbes différents dont 4 188 à plusieurs entrées » (Dubois/Dubois-Charlier 1997: 1). Ce modèle propose une classification des verbes selon leurs propriétés sémantiques et syntaxiques et montre l'adéquation entre les constructions syntaxiques et l'interprétation sémantique par les classes sémantico-syntaxiques.

Après avoir observé les concordances à l'aide du Lexicoscope⁹ (cf. Kraif/Diwersy 2012), nous avons repéré deux principaux sens, recensés dans le *LVF*, présents dans l'écrit scientifique. Ces acceptions sont classées dans les catégories C2a.1 et C4b.32¹⁰, illustrées dans le tableau ci-dessous :

Classe générique	Classe sémantico-	Sous-classe syntaxique	Sous-type syntaxique	Constructions syntaxiques	Verbes concernés
------------------	-------------------	------------------------	----------------------	---------------------------	------------------

⁸ La nouvelle version électronique (avril 2013) est accessible en ligne : <http://rali.iro.umontreal.ca/rali/?q=fr/lvf> [30.06.2015].

⁹ Outil d'interrogation de corpus comparable au Sketch Engine, accessible en ligne : <http://phraseotext.u-grenoble3.fr/lexicoscope/> [30.06.2015].

¹⁰ À chaque lettre/chiffre correspond un niveau de classement (ex : classe sémantique générique, sous-type syntaxique). Pour chaque entrée verbale (un sens y est associé), cinq niveaux de classement sont établis.

	syntaxique				
C : Verbes de communication	C2 : dire/demander qc	C2a : dire que, dire qc à qn C2a	C2a.1 : dire qc, que + indicatif à qn	Transitif direct à sujet humain et suivi par une complétive ou un objet chose (à, instrument, moyen)	<i>démontrer, indiquer, montrer, révéler, etc.</i>
C : Verbes de communication	C4 : dire ou montrer qc	C4b : montrer qch, sujet nom de comportement	C4b.3 : montrer un sentiment, que	Transitif direct à sujet chose et suivi par une complétive ou un objet chose	<i>démontrer, indiquer, montrer, etc.</i>

Tableau 1 : Hiérarchie de 5 niveaux de classes du *LVF* : deux entrées du verbe *montrer*

Comme le montre le tableau, les verbes *démontrer* et *indiquer* sont associés, dans le *LVF*, aux mêmes classes syntaxiques et sémantiques que le verbe *montrer*. Nous devons cependant tenir compte du fait que le *LVF* est une ressource des verbes français pour la langue générale. Afin de prendre en compte les spécificités de l'écrit scientifique, nous adaptons les informations sémantiques et syntaxiques à l'aide des patrons lexico-syntaxiques relevés dans la phase de modélisation.

2.2 Corpus

Notre corpus d'analyse, issu du projet Scientext, est composé de 500 articles de recherche en français (Tran 2014), de 10 disciplines des sciences humaines et sociales. Ces articles proviennent de numéros récents de revues avec comité de relecture, dont l'évaluation par l'AERES¹¹ est A ou B. Ce corpus de 5 millions de mots, a été balisé en parties textuelles, mis au format XML-TEI, puis annoté à l'aide de l'analyseur en dépendances XIP (Aït-Mokhtar/Chanod/Roux 2002). Le corpus est ainsi enrichi en lemmes, catégories syntaxiques, traits morpho-syntaxiques et relations syntaxiques. Nous avons également procédé à un post-traitement du corpus en définissant des grammaires locales en vue de l'amélioration de l'annotation en dépendances. Nous souhaitons ainsi améliorer l'analyse à travers notamment la propagation du sujet pour les structures avec verbes de contrôle, ou lorsque le sujet est en situation de coréférence¹² ou de coordination.

¹¹ Agence d'évaluation de la recherche et de l'enseignement supérieur.

¹² Considérons la phrase « Nous analysons la préposition qui introduit le nom ». Si l'analyseur syntaxique définit une relation de coréférence entre qui et préposition et une relation sujet entre *introduit* et *qui*, alors nous propageons la relation sujet entre *introduit* et *préposition*.

2.3 Lexique Scientifique Transdisciplinaire (LST)

Dans de précédents travaux (cf. Hatier 2013), nous avons extrait une liste de 1312 mots du LST : 274 adjectifs, 202 adverbes, 493 noms et 342 verbes. Dans un premier temps, nous avons mis en place une méthode d'extraction automatique basée sur les critères de répartition dans le corpus d'analyse et de sur-représentation par rapport à un corpus de contraste diversifié (mêlant texte de fiction, journalistique, sous-titres et oral transcrit, et comportant environ 120 millions de mots). Cette liste a, dans un deuxième temps, été filtrée puis validée manuellement. Nous l'avons par la suite enrichie par une classification sémantique en classes et sous-classes transcategorielles. Ce travail a été effectué notamment sur la base du *Dictionnaire électronique des mots (DEM)* de Dubois et Dubois-Charlier (2010). Les classes et sous-classes sémantiques correspondent à des ensembles de co-hyponymes (mots partageant un même hyperonyme) répondant à des tests lexico-syntaxiques. Ces classes ont été élaborées en rapprochant les mots partageant des cooccurrents privilégiés. Ainsi, la sous-classe *document* de la classe *communication* a pour test d'appartenance : *Ce N présente*. Le test est ainsi validé par les membres de la sous-classe : *article, ouvrage, texte*, etc.

Un premier travail manuel d'identification des sens présents dans le corpus nous a permis de passer d'une liste de lemmes du LST à une liste ayant pour entrée des acceptions. Ainsi, pour le lemme *développement*, nous créons deux entrées correspondant aux deux acceptions recensées en corpus : *exposé détaillé* et *croissance*. En effet, la polysémie concerne également les langues spécialisées, comme le notent Condamines et Rebeyrolle (1996). Les regroupements en classes et sous-classes sémantiques ont été effectués en se fondant sur les informations sémantiques présentes dans le *DEM* et en rapprochant les mots ayant un profil combinatoire proche, obtenus à l'aide de l'outil le Lexicoscope¹³. Ainsi, les membres de la sous-classe *document (article, ouvrage, texte)* sont fréquemment sujets des verbes *présenter*, *aborder*, objets du verbe *publier* et déterminés par *ce*. Nous pouvons alors intégrer à notre corpus des annotations sémantiques sous forme de traits de classes et sous-classes sémantiques. L'objectif est ici de pouvoir identifier, dans les cadres de sous-catégorisation verbale que nous extrayons, si certaines classes sémantiques apparaissent de façon préférentielle en tant que sujet ou objet des verbes étudiés. De plus, en nous concentrant sur des verbes appartenant aux mêmes classes sémantiques, nous voulons vérifier si cette proximité sémantique se retrouve dans des constructions syntaxiques communes, à la suite de Messiant, Gábor, et Poibeau (2010) qui notent que « la syntaxe peut aider à faire émerger des classes lexico-syntaxiques ». Ainsi, les verbes *montrer*, *démontrer* et *indiquer*, qui appartiennent à la même sous-classe *révélation* dans notre classification sémantique, devraient également partager un certain nombre de propriétés syntaxiques.

2.4 Extraction des cadres de sous-catégorisation

Le fait de travailler sur des verbes fréquents (de 315 à 3053 occurrences) nous permet d'analyser un nombre important de constructions. Pour chaque occurrence d'un verbe, nous récupérons l'intégralité des relations de dépendances l'impliquant, qu'il soit gouverneur ou dépendant. Considérons la phrase suivante :

¹³ Cet outil permet d'avoir, pour un mot donné, l'ensemble de ses cooccurrents les plus significatifs.

Les résultats semblent confirmer nos hypothèses.

Selon les résultats de l'analyse syntaxique automatique, le verbe *confirmer* est ici concerné par trois relations de dépendances :

1. il est dépendant dans la relation [obj]¹⁴ où il a pour gouverneur *semblent* ;
2. il est gouverneur dans la relation [obj] où il a pour dépendant *hypothèses* ;
3. il est gouverneur dans la relation [subj], calculée via les grammaires locales, où il a pour dépendant *résultats* ;

Nous obtenons alors une représentation linéaire des relations (associées à la catégorie du cooccurrent) impliquant le verbe dans la phrase. Le verbe *confirmer* aurait alors pour représentation de la phrase 1 :

[a pour sujet NOM] [est objet de VERBE] *confirmer* [a pour objet NOM]

Sur le même principe que Messiant, Gábor et Poibeau (2010), nous acquérons ainsi des cadres « sans a priori, pour faire émerger du corpus [ceux] correspondant à l'usage ». Nous adoptons ainsi une méthodologie « corpus-driven » (Tognini-Bonelli 2001), reposant sur les données issues du corpus et sur des traitements semi-automatiques. Lors de cette phase, nous identifions les mots les plus fréquemment sujet et/ou objet des verbes analysés. Nous repérons également les classes sémantiques les plus présentes en tant que sujet ou objet du verbe étudié.

Nous procédons ensuite, comme le décrit Kupsc (2007), à une phase de factorisation des cadres en regroupant certaines relations de dépendance sous une même étiquette. Ainsi, en partant des résultats de l'analyse syntaxique, nous aboutissons à un cadre de sous-catégorisation comme illustré ci-après :

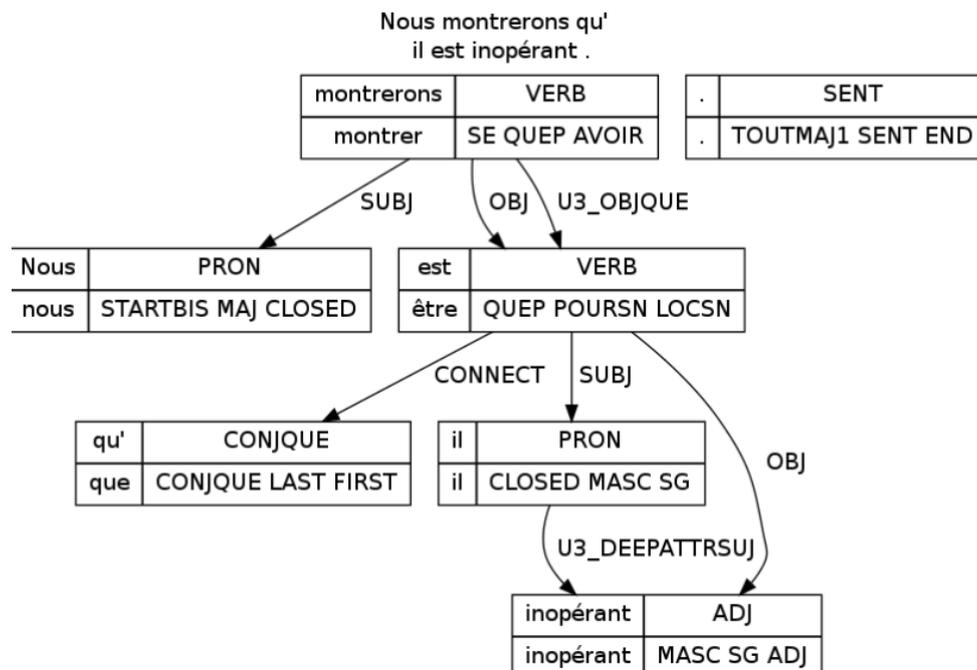


Figure 1 : Sortie de l'analyse en dépendances

¹⁴ [obj] réfère aux étiquettes de relations de dépendances issues de l'analyseur.

Le cadre de sous-catégorisation sera comme suit :

(SUBJ) montrer_VERB(COMPLETIVE_QUE)

Le tableau ci-dessous présente, pour l'exemple du verbe *montrer* (dans le sens de *prouver*, *indiquer*), 3 autres cadres associés à ses compléments les plus fréquents (le chiffre précédant le cooccurrent étant le nombre de fois où le mot apparaît dans cette relation, dans le cadre indiqué).

Cadre - Exemple	Fréquence	Sujet	Objet / Verbe de la complétive
(SUBJ) montrer_VERB(COMPLETIVE_QUE) <i>Nos analyses montrent qu'il y a élaboration des gestes [...]</i>	632	82 : <i>résultat</i> (N) 54 : <i>analyse</i> (N) 37 : <i>étude</i> (N) 23 : <i>travail</i> (N) 17 : <i>tableau</i> (N)	155 : <i>être</i> (V) 48 : <i>pouvoir</i> (V) 37 : <i>avoir</i> (V) 10 : <i>exister</i> (V) 9 : <i>faire</i> (V)
(SUBJ)montrer_VERB(OBJ) <i>Les résultats montrent un effet significatif de la présence de l'appareil [...]</i>	118	11 : <i>analyse</i> (N) 7 : <i>résultat</i> (N) 5 : <i>exemple</i> (N) 5 : <i>figure</i> (N)	7 : <i>importance</i> (N) 7 : <i>existence</i> (N) 5 : <i>différence</i> (N) 5 : <i>effet</i> (N)
Comme le montrer_VERB(SUBJ) <i>Comme le montre le tableau, les résultats [...]</i>	14	2 : <i>analyse</i> (N) 2 : <i>donnée</i> (N)	

Tableau 2 : Cadre de sous-catégorisation du verbe *montrer*

Les classes sémantiques issues de notre classification nous permettent d'ajouter un trait au niveau des actants. Ainsi, le deuxième cadre a notamment pour sujet *figure* et *tableau* (sous-classe sémantique 'communication/support'), et pour objet *importance* et *rôle* (sous-classe sémantique 'qualité/importance'). Une spécification sémantique de ce cadre serait alors :

(SUBJ *communication/support*) montrer_VERB (OBJ *qualité/importance*)

Ce type de cadre sémantico-syntaxique permet d'envisager une étude des fonctions rhétoriques associées aux verbes présents dans l'écrit scientifique. Granger et Paquot (2009) notent d'ailleurs que ces verbes ont tendance à apparaître dans des structures « routinisées ».

Nous utilisons également le Lexicoscope pour avoir les co-occurents les plus fréquents tous cadres confondus, et observons, à l'aide du même outil, les concordances correspondant aux cadres afin de vérifier en corpus leur validité. Ces cadres de sous-catégorisation permettent alors de passer à l'étape de modélisation des patrons lexico-syntaxiques.

2.5 Le modèle Corpus Pattern Analysis

Le *Corpus Pattern Analysis* de Hanks (2013) est à la fois une approche sur les patrons et un modèle lexicographique basé sur corpus qui s'inscrit dans la continuité de la linguistique de

corpus de Sinclair. Ici, le patron est considéré comme une configuration syntaxique intégrant un ensemble de collocations préférées.

A pattern is a statement of the clause structure (valency) associated with a meaning of a verb—together with the typical semantic values of each argument. Different semantic values of arguments (subject, object, prepositional object) activate different meanings of the verb.

(Hanks 2013 : 92).

L'approche des patrons défend le lien étroit entre lexique et grammaire et considère 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 : 8).

Le premier modèle lexicographique *CPA* consacré à l'anglais est le projet *Pattern Dictionary of English Verbs (PDEV)*¹⁵. Ce dictionnaire basé sur le British National Corpus (BNC) propose une analyse des sens standard des verbes en les distinguant des usages marqués (usages créatifs, originaux des mots).

Pattern : [[Activity | Institution]] swallow [[Resource = Money]] {(up)}
 Implicature¹⁶ : [[Activity| Institution]] completely uses up [[Resource = Money]]

Sur le plan sémantique, les groupes lexicaux intervenant en tant qu'arguments sont généralisés dans des types sémantiques entre doubles crochets, les rôles sémantiques et les collocations sont parfois précisés pour affiner l'analyse sémantique des arguments. Par ailleurs, les compléments des verbes tels qu'infinitifs, complétives ou compléments adverbiaux sont décrits dans les patrons.

En plus de l'anglais de langue générale, le *CPA* est aussi adapté à l'espagnol (Campos/Araque 2013) et à l'italien (Jezek/Frontini, 2010). Par ailleurs, Campos et Araque ont montré que l'approche *CPA* pouvait aussi être appliquée dans les domaines de spécialité comme l'environnement.

Nous nous inspirons du modèle *CPA* pour la modélisation des patrons car ce modèle est particulièrement approprié à notre étude dont le principal objectif est didactique : d'une part, il est basé sur des usages réguliers ; d'autre part, il est destiné à la désambiguïté lexicale. L'approche de patrons nous permettra de prendre en compte non seulement les constructions syntaxiques privilégiées, mais aussi les caractéristiques sémantiques des verbes. Par ailleurs, dans ce modèle, la notion de construction est beaucoup plus large que la notion classique de sous-catégorisation.

3 Modélisation des patrons lexico-syntaxiques

La modélisation des patrons se fait en trois étapes. Dans un premier temps, nous regroupons les cadres de sous-catégorisations du verbe qui présentent les mêmes propriétés sémantiques

¹⁵ Accessible en ligne www.pdev.org.uk/#browse?q=&f=C [30.06.2015].

¹⁶ Le sens est représenté dans l'implicature, une paraphrase du patron employant un verbe différent et une phraséologie différente.

et syntaxiques. Nous associons ensuite les cadres de sous-catégorisations aux sens correspondants dans le *LVF*. Enfin, nous présentons chaque patron lexico-syntaxique, intégrant structure syntaxique et collocations privilégiées, dans une base de données en ligne – Dicorpus¹⁷ – qui permet de tirer profit des descriptions linguistiques pour une transposition didactique.

Prenons l'exemple du verbe *montrer* (3053 occurrences). Les deux premiers cadres de sous-catégorisations extraits sont :

(1) (SUBJ ±hum) montrer_VERB (COMPLETIVE_QUE) (632 occurrences)

a) **Les résultats montrent que** le facteur qui influence plus particulièrement la performance des enfants est l'alignement de la position finale avec un axe implicite du dispositif (vertical ou horizontal). (psycho_32_XipEmolex)

b) **Trois études internationales récentes montrent bien que** la manière de classer la France et son école maternelle parmi d'autres systèmes d'accueil de la petite enfance repose sur ces logiques très hétérogènes. (scienceseducation_9_XipEmolex)

(2) (SUBJ +hum) montrer_VERB (COMPLETIVE_QUE) (175 occurrences)

c) **Nous montrons qu'**il est alors inutile de conserver des transferts en seconde période de vie. (economie_23_XipEmolex)

Dans le premier cadre, le verbe *montrer* est suivi par une complétive *que*. Le trait ± *humain* du sujet n'est pas précisé¹⁸, les noms les plus fréquents sont *résultat* (82), *analyse* (54), *étude* (37), *travail* (23), *exemple* (18), *recherche* (18), *auteur* (16). Ici, comme le montre l'exemple b), certains noms tels *étude*, *recherche* et *travail* représentant implicitement l'auteur par un emploi métonymique se distinguent des autres noms d'observables de l'activité scientifique comme *résultat* et *exemple* ou des noms du processus scientifique comme *analyse*. Ainsi, nous choisissons de regrouper une partie du premier cadre (ayant pour sujet *étude*, *recherche*, *travail*, *auteur*) avec le deuxième cadre par leur proximité sémantique. Nous constatons que le cadre 1a) (ex. *les résultats montrent que*) correspond ainsi au sens 'mettre en évidence' relevé dans le *LVF*.

Dans le deuxième cadre, le verbe sélectionne un sujet humain et il est suivi d'une complétive en *que*. Le sujet se réalise majoritairement sous forme de pronoms : *nous* (79), *il* (59), *on* (33) et *je* (4), qui représentent essentiellement l'auteur de l'article. Nous associons à ce deuxième cadre le sens 'communiquer en faisant savoir' présenté dans le *LVF*. D'après les cadres analysés ci-dessus, nous pouvons présenter le patron¹⁹ suivant :

Quelqu'un (on, nous, etc.) montre que (175 occurrences).

SENS : En s'appuyant sur l'analyse et des données, quelqu'un donne des arguments pour faire savoir et justifier certains faits scientifiques en suivant un développement argumentatif.

ALTERNANCE LEXICALE : (*étude*, *travail*, *recherche*, *article*, etc.) *montre que...* (78 occurrences)

¹⁷ Dictionnaire d'apprentissage d'aide à la rédaction scientifique.

¹⁸ Nous intégrons le trait +hum lorsque le sujet se réalise sous la forme de *je*, *on*, *nous*.

¹⁹ Nous situant dans une perspective didactique, nous avons adopté un formalisme simplifié.

ALTERNANCE SYNTAXIQUE : *en montrant que*

EXPRESSIONS : *nous tentons de montrer que, nous essayons de montrer que, nous cherchons à montrer que, notre objectif est de montrer que, il s'agit de montrer que*

Comme illustré dans l'exemple ci-dessous, ce patron intègre préférentiellement un sujet pronominal. De plus, il est fréquent qu'un adverbial de lieu/de relation logique soit mobilisé pour renvoyer à un élément péri-textuel (*tableau, figure, graphique*), ou à une portion du texte (*paragraphe, annexe, etc.*) (voir l'exemple 1).

- 1) **On montre** toutefois, **dans l'annexe 1**, dans certaines configurations simples, **que** la matrice des dérivées partielles est de plein rang sous l'hypothèse nulle. (eco.xml-s10285)

Les patrons lexico-syntaxiques des verbes analysés seront intégrés comme entrées dans l'outil Dicorpus.

4 Analyse et résultats

4.1 Fréquence des constructions

Le tableau ci-dessous présente pour les trois verbes les fréquences des constructions. Nous posons le seuil de fréquence minimale à 5 occurrences afin de proposer aux apprenants non natifs des constructions dont l'usage est attesté en corpus. Entre parenthèses est noté le nombre d'occurrences de la construction pour le verbe en question. Le pourcentage est la proportion de cette construction parmi l'ensemble des occurrences du verbe dans le corpus.

Construction	Montrer	Démontrer	Indiquer
1. Sujet chose _ que <i>Le tableau montre que...</i>	21,5 % (670)	7,4 % (24)	31,3 % (284)
2. Sujet humain (métonymie) _ que <i>L'étude démontre que...</i>	12,9 % (403)	10,7 % (35)	4,2 % (38)
3. Sujet chose _ Objet chose <i>Cela indique la différence...</i>	5,3 % (166)	6 % (20)	29 % (263)
4. Comme le _ Sujet chose <i>Comme le démontre l'auteur...</i>	2,44 % (76)	3,7 % (12)	8 % (34)
5. Sujet humain _ Objet chose <i>On montre l'importance...</i>	0,9 % (30)	3,1 % (10)	-
6. Comme le _ Sujet humain <i>Comme nous l'avons montré...</i>	0,8 % (26)	3,1 % (10)	3,3 % (30)
7. En (_ participe présent) que <i>En montrant que...</i>	0,7 % (22)	-	-
8. Pour (_ infinitif)	0,4 % (12)	-	-

<i>Pour montrer...</i>			
9. Il est (_ participe passé) que <i>Il est montré que...</i>	0,3 % (10)	-	-
10. Sujet chose être (_ participé passé) [passif] <i>L'effet est démontré...</i>	-	12,9 % (42)	5,5 % (50)
11. Comme (_ participe passé) <i>Comme indiqué...</i>	-	-	1 % (9)
12. Rien ne _ que <i>Rien n'indique que...</i>	-	-	0,9 % (8)

Tableau 3 : Fréquences des constructions verbales

Nous remarquons que la majorité des constructions est commune à l'ensemble des verbes étudiés. Ainsi, les constructions 4 et 6, introduites par *comme*, bien que peu fréquentes, sont partagées par les trois verbes :

- 2) En effet, **comme l'indiquent Pruitt et Adlin** (2006), cette technique ne saurait remplacer les autres méthodes de conception utilisées en ergonomie des logiciels. (psycho_24_XipEmolex)
- 3) **Comme on l'a démontré ailleurs**, l'épaisseur historique dans laquelle s'inscrit leur migration semble avoir permis l'activation de réseaux efficaces pour l'encadrer (Sintès, 2003, 2005). (geo.xml-s1085)
- 4) Or en fait, cette éventualité est tout à fait possible, **comme le montrent des exemples...** (linguistique_13_XipEmolex)

Les constructions avec complétive ou objet nominal sont également réalisées pour l'ensemble des verbes. Les proportions d'emploi de constructions varient toutefois selon le verbe. Ainsi, la construction 3, intégrant un objet et sujet nominal non-humain, représente moins de 6 % des emplois pour *démontrer* et *montrer* alors qu'elle correspond à presque un tiers des emplois pour *indiquer*. Ce verbe se distingue également par le fait qu'il ne sélectionne que très rarement un sujet humain contrairement à ses deux synonymes.

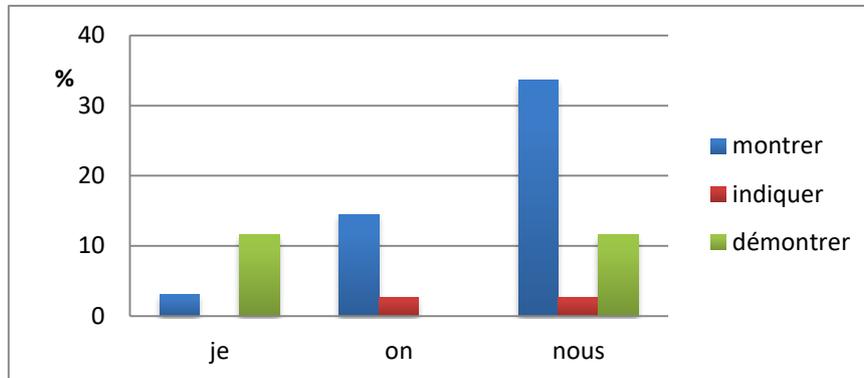
La lecture du tableau confirme ainsi que des verbes sémantiquement proches partagent un grand nombre de constructions syntaxiques (et de compléments sémantiquement proches s'intégrant dans ces patrons), que nous analysons en détail dans la partie suivante.

Nous avons séparé les constructions selon la nature du sujet : humain ou chose. Quand les constructions comprennent un sujet humain, elles prennent le sens 'justifier certains faits scientifiques avec des arguments', tandis qu'avec un sujet chose, le sens peut être décrit comme 'mettre en évidence et révéler certains faits scientifiques'. Pour établir une correspondance entre les constructions et ces deux sens relevés, nous avons alors examiné les type de catégories grammaticales intervenant dans les constructions verbales.

4.2 Construction en complétive que

4.2.1 Pronoms sujets mettant en jeu l'auteur

Si nous prenons en compte l'ensemble des verbes étudiés, la construction la plus fréquente est la complétive *que*. Pour observer les associations, dans cette construction, entre les pronoms sujets renvoyant à l'auteur et les trois verbes analysés, nous avons calculé la proportion des trois pronoms *je*, *on* et *nous* sur le nombre total de pronoms sujets utilisés, incluant également *il*, *ceci*, *cela*.



Graphique 1 : Les pronoms sujets *je*, *on* et *nous* (en % sur le nombre total de pronoms sujets utilisés)

Les résultats, apparaissant dans le graphique 1, sont contrastés. De manière générale, les pronoms sujets renvoyant à l'auteur sont peu fréquemment associés aux verbes analysés, ce qui semble confirmer l'observation de Tutin (2010) qui relevait une faible proportion de ce type de pronoms liés aux verbes de positionnement dans l'écrit scientifique (science de l'éducation, psychologie et linguistique). Nous pouvons en déduire que l'écrit scientifique se caractérise par un effacement énonciatif où la voix de l'auteur n'est pas mise en avant (cf. Tutin 2010 ; Fløttum/Dahl/Kinn 2006). Parmi les trois pronoms, *je* et *on* sont les moins utilisés. Dans le sens de démonstration scientifique, c'est le pronom *nous*, renvoyant à l'auteur « appelé alors 'nous de modestie' » ou aux auteurs, qui est préféré (voir l'exemple 5).

5) **Nous allons démontrer que** c'est bien le cas et apporter un nouveau regard sur l'interprétation des dimensions fractales. (geographie.xml-s3005)

Si nous nous intéressons aux particularités selon les verbes, nous remarquons que le verbe *indiquer* est en cooccurrence avec des pronoms comme *il(s)*, *ceci*, *cela* ou des noms de chercheurs :

6) À la suite d'une importante recension de la littérature, **Harrison et Martocchio (1998) indiquent que** l'absentéisme de moyen terme, c.-à-d. couvrant un horizon temporel de 3 à 12 mois, s'explique principalement par les attitudes au travail. (psycho_19_XipEmolex)

Ce verbe prend le sens de 'faire savoir' et n'est alors pas strictement synonyme des verbes *démontrer* et *montrer* dans le sens de démonstration. Le cas est différent pour le verbe *démontrer* qui marque une faible proportion dans les associations 'pronoms sujets+verbe', et s'intègre souvent dans des tournures au passif :

7) À ce moment, l'exactitude des hypothèses sur la validité des estimations **est démontrée** en produisant ces mêmes estimations de validité comme un produit de la solution. (sciencespolitiques_46_XipEmolex)

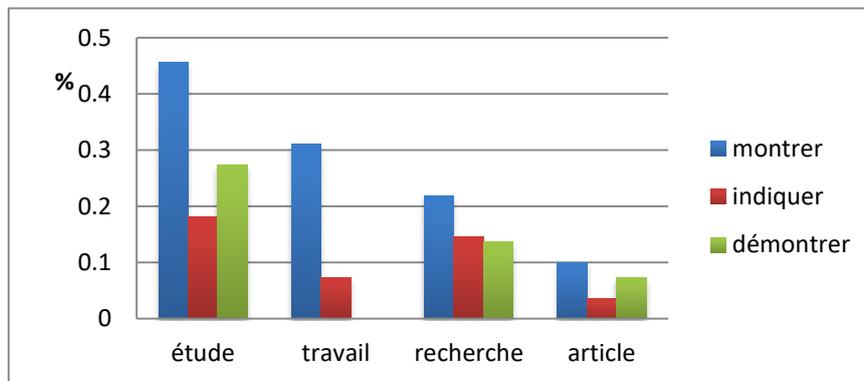
Quant au verbe *montrer*, quand il est associé au pronom *on*, ce dernier peut représenter soit l'auteur unique, soit l'auteur et la communauté scientifique (voir l'exemple 8).

8) Cette orientation est relative au caractère intrinsèquement dialectique des situations argumentatives. Ses mécanismes ont été analysés dans des situations didactiques (Quignard, 2001), où l'**on montre qu'**un débat s'apparente à un jeu gouverné par un ensemble de règles dialectiques... (psycho.xml-s16319)

Dans l'exemple (8), l'emploi de *on* permet d'introduire la prise de position partagée dans la communauté scientifique, ce qui rend le propos plus persuasif.

4.2.2 Sujets métonymiques : cette étude montre que...

Dans l'écrit scientifique, l'auteur peut parfois être mentionné implicitement derrière des sujets métonymiques tels *cette étude*, *ce travail*, etc. Nous avons calculé la proportion de ces sujets métonymiques parmi l'ensemble des sujets nominaux, comme l'illustre le graphique 2.



Graphique 2 : Les sujet métonymiques *étude*, etc. (en % sur le nombre total de noms sujets utilisés)

Le verbe *montrer* est celui qui est le plus fréquemment associé à ces noms métonymiques (*cette étude montre que*=*Nous montrons dans cette étude que*).

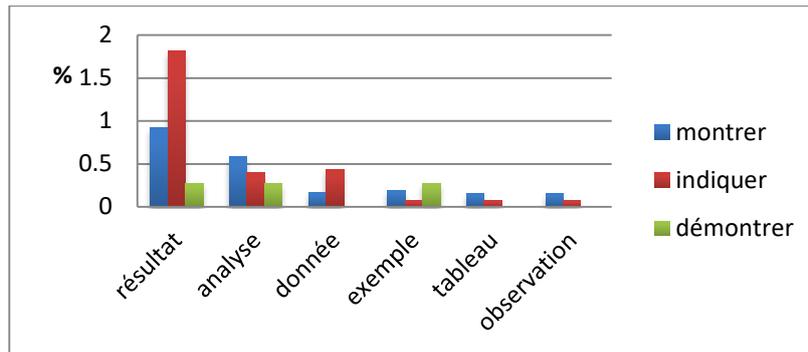
9) **Cette étude montre que** le bois mort reste négativement valorisé dans l'imaginaire collectif. (geographie.xml-s11955)

Par ailleurs, nous avons remarqué que lorsque le verbe *indiquer* est associé à un sujet métonymique, celui-ci renvoie à d'autres études/auteurs, comme le montre l'exemple (10). Cette observation semble cohérente avec nos résultats précédents selon lesquels le verbe *indiquer* est rarement lié aux pronoms sujets renvoyant à l'auteur.

10) **Une étude** menée en 2003, confirmant nos investigations, **indique que** 50 % des personnes déplacées souhaitent retourner dans leurs anciens villages (Saybouanlaven, 2004). (geographie.xml-s1485)

4.2.3 Sujets appartenant au LST

L'extraction des constructions verbales nous permet de constater que les sujets chose dans la construction en complétive *que* renvoient souvent aux observables (*résultat*, *donnée*, *exemple*, etc.), aux supports (*tableau*, *figure*, etc.) et aux processus scientifiques (*analyse*, *observation*, etc.) de l'activité scientifique. Le graphique ci-dessous détaille, pour l'ensemble des verbes étudiés, les sujets chose les plus fréquents. Nous avons calculé la proportion de ces noms en tant que sujet par rapport à l'ensemble des sujets nominaux.



Graphique 3 : Les noms sujets *résultat*, etc. (en % sur le nombre total de noms sujets utilisés)

Comme le montre le graphique 3, les noms *résultat* et *analyse* sont les plus fréquemment employés en tant que sujet des trois verbes. Le verbe *démontrer* est celui dont les associations avec ces noms sont les moins fréquentes. Ceci confirme le fait que ce verbe implique un raisonnement rigoureux et est davantage employé dans le sens de la démonstration nécessitant un sujet humain.

11) **Comme nous l'avons démontré** au cours de cette étude, l'armée de l'air a élaboré un dispositif symbolique très sophistiqué... (scinfo.xml-s8901)

4.3 Construction en *comme* et autres constructions syntaxiques fréquentes

Comme nous l'avons déjà constaté, la structure en *comme le _sujet humain ou sujet chose*, bien que peu fréquente, est partagée par les 3 verbes. Ces constructions comportent un verbe au passé composé ou au passif réduit, à valeur d'accompli, ou un verbe au présent : *comme nous l'avons déjà montré*, *comme indiqué ci-dessous*, *comme le démontre l'auteur...* etc.

Ces constructions ont, dans ce cas, une fonction métatextuelle et/ou évidentielle (cf. Grossmann 2014 : 87). D'une part, l'auteur peut faire référence aux travaux précédents, par exemple dans la construction *comme nous l'avons déjà montré*. D'autre part, l'auteur souhaite prouver la véracité de ses propos :

12) La recherche du secret est une obsession constante de la part des conseils de ville, qui y voit l'une des conditions de l'efficacité du gouvernement urbain, **comme le montre notamment le serment imposé** aux conseillers... (histoire_33_XipEmolex)

Nous notons également que le verbe *indiquer* est le plus employé dans ces constructions qui permettent souvent d'introduire une citation :

13) **Comme l'indique Metzger** (2000), les théories du cycle de vie des quartiers, développées dans la lignée des travaux de Hoyt, ont une composante raciale marquée puisque le déclin est lié à l'installation de minorités ethniques. (geo.xml-s9631)

D'autres constructions nous paraissent importantes sur le plan didactique, du fait de leur fréquence dans le corpus d'analyse et de la relative difficulté à les encoder/décoder : *rien n'indique que* (8/906 occurrences), *permettre de montrer que* (80/3114), *tendre à démontrer que* (7/326), *(cela est) démontré* (24/326 occurrences). Ces informations linguistiques doivent être signalées pour aider les apprenants à s'initier à la rédaction des écrits universitaires.

4.4 Association avec des adverbes

Pour toutes les constructions relevées, nous avons constaté une forte association entre les verbes étudiés et les adverbes discursifs : *en effet*, *cependant*, *également*, *enfin*, *par ailleurs*, *clairement* pour citer les plus fréquents. Ces adverbes ont ainsi tendance à apparaître dans des constructions verbales récurrentes. Par exemple, pour la construction en *comme* avec des verbes au passé composé/passif réduit, nous notons la présence d'adverbes (*précédemment*, *déjà*) qui renvoient au texte précédent, assurant alors une fonction métatextuelle.

14) Comme on l'a déjà indiqué **précédemment**, les niveaux d'équilibre du capital mondial et de la consommation mondiale sont parfaitement déterminés et ne dépendent que des paramètres du modèle. (eco.xml-s17997)

Dans la construction en complétive *que*, le verbe s'associe alors préférentiellement à des adverbes à fonction argumentative introduisant une cause, une conséquence ou un argument supplémentaire pour soutenir l'idée défendue.

15) L'analyse de l'évolution socio-spatiale des territoires de Ciudad Sandino dans leurs rapports avec ceux de Managua, à partir d'une approche temporelle, démontre **en effet** qu'il n'y a pas forcément coexistence permanente entre marginalité spatiale et marginalité sociale. (geographie.xml-s2347)

5 Intégration dans Dicorpus, dictionnaire basé sur corpus

Comme nous l'avons mentionné plus haut, un dictionnaire spécialisé d'apprentissage, intégré dans un outil d'aide à la rédaction scientifique, doit permettre à l'apprenant d'accéder facilement à l'encodage des informations où les collocations et les patrons occupent une place importante. L'outil Dicorpus²⁰ semble approprié à cet objectif didactique. Il a en effet été conçu pour une utilisation pédagogique et permet de combiner les résultats de requêtes sur le corpus Scientext et les descriptions linguistiques effectuées précédemment. Les exemples sont présélectionnés manuellement d'après le corpus, ce qui évite à l'apprenant de se perdre dans des exemples trop fournis ou trop complexes. Comme le soulignent Cavalla et Loiseau (2014), l'enseignement des éléments phraséologiques à l'aide du corpus doit guider l'apprenant vers l'autonomie de l'apprentissage. En suivant le même objectif didactique, nous adoptons l'approche « teaching to exploit » (le corpus pour enseigner et apprendre la langue où l'apprenant s'approprie l'outil d'exploitation du corpus) de Fligelstone (1993).

L'interface de Dicorpus propose deux choix de consultation :

1. accès par approche sémasiologique (accès direct aux verbes/patrons verbaux)
2. accès par approche onomasiologique (accès au sens, par exemple, 'pour justifier un fait scientifique')

Au niveau de la microstructure, nous mettons à la disposition de l'apprenant les résultats de la modélisation des patrons et les présentons d'une façon structurée et compréhensible. Les informations suivantes sont indiquées :

²⁰ Outil, conçu par Achille Falaise, faisant office d'assistant phraséologique pour la rédaction scientifique en français.

1. la fréquence absolue des verbes dans le corpus : très fréquent, moyennement fréquent, peu fréquent²¹ ;
2. les composants grammaticaux (sujet, objet, etc.), ex : sujet (*on, nous, je*) ;
3. l'association avec les adverbes, ex : *en effet, cependant, etc.* ;
4. les constructions verbales, ex : *le modèle que nous avons montré* ;
5. le sens du patron décrit par une définition ;
6. les exemples extraits et validés manuellement du corpus Scientext

Dans la figure 2 ci-dessous, l'interface de Dicorpus est présentée dans son entrée sémasiologique. Il s'agit des verbes *montrer* et *démontrer* dans leur sens de démonstration.

Les verbes dans les écrits scientifiques | démonstration

MODE D'ACCÈS

- ▶ Accès par sens
- ▶ Accès par expression

EXPRESSION

montrer

()

montrer (** très fréquent)

démontrer (** fréquent moyen)

TYPE DE VERBES
verbes de démonstration

COMPOSANTS GRAMMATICaux
sujet : nous | on | je | étude | travail | recherche | article

ASSOCIATION AVEC ADVERBES
en effet | cependant | également | enfin | par ailleurs

CONSTRUCTIONS
on montre que | on montre comment | on montre qch. d'abstrait (importance) | permettre de montrer que | comme le montrer (qn.) | pour montrer que | il est montré que | en montrant que

SIGNIFICATION
justifier certains faits scientifiques avec des arguments

Exemples validés sur le corpus Scientext

164 occurrences trouvées

▼ N°	▽ Contexte gauche: <input type="text" value="10"/> mots	Occurrence:	▽ Contexte droit: <input type="text" value="10"/> mots
5	On peut le	démontrer	pragmatiquement par un raisonnement de combinatoire , en écrivant toutes les
8	corrélats linguistiques de divers paramètres de ces situations a été	démontrée	par les travaux de D. Biber (cf .
9	Halliday , 1991) . En	démontrant	l' impact du contexte linguistique et extra linguistique sur la
10	Par conséquent , afin de	démontrer	le succès d' une recherche à visée théorique , il
15	hypothèses , un des objectifs de cette étude est de	démontrer	que les expressions temporelles voient leur capacité à signaler une
32	par Mayer (ibid .) , a été	démontrée	au travers de nombreuses expériences relatives à la compréhension de phénomènes scientifiques
34	argumentation développée par L. Cherchi Cherchi L. tend à	démontrer	que l' essence de l' ellipse est son rôle dans
36	Ceci a magistralement été	démontré	par l' étude de Z. Harris Harris Z. S. sur
37	Pour	démontrer	ceci , l' auteur s' est , dans cet article
40	tous les cas de figure et alors , comme le	démontrer	avec force J.P. (1992) , on génère énormément de
41	des méthodes employées , nous tendrons ainsi à en	démontrer	l' assise scientifique .
42	L' étude de l' alignement a	démontré	un fait important : que la relation traductionnelle entre les
43	Vis-à-vis de la pratique de la traduction , nous avons	démontré	la non- pertinence d' une telle conception .
44	Ce qui tendrait à	démontrer	qu' il y a une certaine récurrence du contexte traduit

Figure 2 : Base de données des verbes de démonstration dans Dicorpus – Entrée sémasiologique

²¹ « Très fréquent » correspond à au moins 10 occurrences dans le corpus, « moyennement fréquent » entre 5 et 10 occurrences, et « peu fréquent » moins de 5 occurrences.

La figure 3 illustre l'interface de Dicorpus avec l'entrée onomasiologique. L'entrée se fait par le sens, ici 'révéler et mettre en évidence un fait scientifique'.

Les verbes dans les écrits scientifiques | révélation

MODE D'ACCÈS

- ▶ Accès par sens
- ▶ Accès par expression

POUR EXPRIMER...

...

... révéler et mettre en évidence un fait scientifique

EXPRESSION

montrer (** très fréquent)

démontrer (** fréquent moyen)

indiquer (** fréquent moyen)

TYPE DE VERBES
verbes de révélation

COMPOSANTS GRAMMATICAUX
sujet : résultat | analyse | donnée | exemple | tableau | observation

ASSOCIATION AVEC ADVERBES
en effet | cependant | également | enfin | par ailleurs

CONSTRUCTIONS
résultat nous montre que | résultat montre qch d'abstrait (effet) | comme le montrer l'analyse

SIGNIFICATION
laisser apparaître clairement la vérité d'un fait ou les caractéristiques d'une entité abstraite

Exemples validés sur le corpus Scientext

159 occurrences trouvées

▼ N°	▽ Contexte gauche: 10 mots	Occurrence:	▽ Contexte droit: 10 mots
4	L' évaluation sur corpus de certains systèmes de prédiction	montre	qu' ils ont le potentiel d' éviter la saisie d'
15	Les arguments ci-dessus	montrent	l' importance et la difficulté de l' évaluation dans le
16	Cette analyse	montre	que les principaux champs de recherche du TAL se laissent
17	demeure une question difficile à étudier objectivement , comme le	montre	l' exemple de l' analyseur syntaxique cité à la section
23	autres exemples (e .g . les relations anaphoriques)	montrent	qu' il existe des phénomènes discursifs non représentables avec des
27	La figure 2	montre	un exemple de vue de type arborescent dont le contenu
30	Le tableau 3	montre	les opérateurs définis pour tester des conditions sur le rapport
48	Néanmoins , la section 5	montrera	que la capacité générative forte de RST est moins grande
49	: Dans le tableau 1 , la première ligne	montre	les quatre arbres RST avec une relation nucleus- satellite enchâssée
50	La seconde ligne	montre	les DAG de dépendances équivalents , à savoir (Ib)(IVb
51	La troisième ligne du tableau 1	montre	les graphes SDRT équivalents .
52	: Dans le tableau 2 , la première ligne	montre	les quatre arbres RST avec une relation multinucléaire enchâssée ,
53	La seconde ligne	montre	les DAG de dépendances équivalents , à savoir (Ib)(IVb
54	La troisième ligne du tableau 2	montre	les graphes SDRT équivalents .
69	Les analyses	montrent	une forte convergence entre ceux -ci .
76	Les tableaux 6 et 7	montrent	aussi la complémentarité de l' utilisation d' une approche endogène
78	Par ailleurs , ces évaluations ont également	montré	qu' approche endogène et approche exogène sont complémentaires et que
79	Néanmoins , l' évaluation de la section 7.3	montre	que cette taille n' est pas nécessairement importante puisque les

Figure 3 : Base de données des verbes de révélation dans Dicorpus – Entrée onomasiologique

6 Conclusion et perspectives

En combinant description linguistique et TAL, nous avons extrait de façon semi-automatique les constructions verbales les plus fréquentes pour les trois verbes étudiés. Nous proposons ainsi, pour ces verbes, une modélisation des patrons lexico-syntaxiques présents dans l'écrit scientifique. La modélisation nous permet de distinguer ces trois verbes sur le plan syntaxique et sémantique dans le contexte scientifique. Bien que les verbes *montrer*, *démontrer* et *indiquer* aient des propriétés sémantiques et syntaxiques comparables, ils ne sont pas synonymes dans tous les contextes. L'analyse quantitative et qualitative des patrons nous montre que les verbes *montrer* et *démontrer* sont employés dans le sens de 'prouver' lorsqu'ils sélectionnent un sujet humain ou métonymique. D'un autre côté, nous notons que

les verbes *montrer*, *démontrer* et *indiquer* sont synonymes, au sens de ‘mettre en évidence’ lorsqu’ils sélectionnent un sujet chose. La construction de la base de données des verbes dans Dicorpus nous a permis de proposer un outil d’aide à la rédaction universitaire qui combine la description linguistique et les données du corpus. Ce modèle nous semble ainsi pertinent dans le cadre de l’enseignement/apprentissage des éléments phraséologiques à partir du corpus.

Nous projetons d’intégrer dans les patrons des informations morpho-syntaxiques sur les réalisations les plus fréquentes, par exemple pour certaines constructions qui prennent pour sujet un syntagme nominal déterminé préférentiellement par un démonstratif, ou se réalisent avec un verbe conjugué à un certain temps/mode. À partir de notre classification du LST, nous prévoyons d’intégrer dans les patrons des informations sur les classes et sous-classes sémantiques des arguments, ceci afin d’aboutir à une description plus fine sémantiquement.

Bibliographie

- Aït-Mokhtar, Salah/Chanod, Jean-Pierre/Roux, Claude (2002): “Robustness beyond shallowness : incremental deep parsing”. *Natural Language Engineering* 8/2–3: 121–144.
- Campos, Araceli Alonso/Araque, Irene Renau (2013): “Corpus Pattern Analysis in determining specialised uses of verbal lexical units”. *Terminàlia* 7: 26–33.
- Cavalla, Cristelle (2008): « Les collocations dans les écrits universitaires : un français spécifique pour les apprenants étrangers ». In: Bertrand, Olivier/Schaffner, Isabelle (eds.): *Apprendre une langue de spécialité : enjeux culturels et linguistiques*. Paris, éditions École Polytechnique: 93–104.
- Cavalla, Cristelle (2014): « Collocations transdisciplinaires d’apprentissage : réflexions pour l’enseignement ». In: González-Rey, Isabel (ed.): *Outils et méthodes d’apprentissage en phraséodidactique. Editions Modulaires Européennes*: 151–169. <https://hal-ens.archives-ouvertes.fr/hal-01216841> [30.06.2015].
- Cavalla, Cristelle/Loiseau, Mathieu (2014): « Scientext comme corpus pour l’enseignement ». In : Tutin, Agnès/Grossmann, Francis (eds.): *L’écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes, Presse universitaire de Rennes : 63–180.
- Condamines, Anne/Rebeyrolle, Josette (1996): « Point de vue en langue spécialisée ». *Meta* 42/1: 174–184.
- Condette, Marie-Hélène/Marin, Rafael/Merlo, Aurélie (2012): « La structure argumentale des noms déverbaux : du corpus au lexique et du lexique au corpus ». *SHS Web of Conferences* 1: 845–858.
- Coxhead, Averil (2002): “The academic word list: A corpus-based word list for academic purposes”. *Language and Computers* 42/1: 73–89.
- Drouin, Patrick (2007): « Identification automatique du lexique scientifique transdisciplinaire ». *Revue française de linguistique appliquée* 12/2: 45–64.
- Dubois, Jean/Dubois-Charlier, Françoise (1997): *Les verbes français*. Paris: Larousse-Bordas.
- Dubois, Jean/Dubois-Charlier, Françoise (2010): « La combinatoire lexico-syntaxique dans le Dictionnaire électronique des mots. Les termes du domaine de la musique à titre d’illustration ». *Langages* 3: 31–56.
- Fligelstone, Steve (1993): “Some Reflections on the Question of Teaching, from a Corpus Linguistics Perspective”. *ICAME Journal* 17: 97–109.

- Fløttum, Kjersti/Dahl, Trine/Kinn, Torodd (2006): *Academic voices: Across languages and disciplines*. Amsterdam: Benjamins.
- Granger, Sylviane/Paquot, Magali (2009): "Lexical Verbs in Academic Discourse : A Corpus-driven Study of Learner Use". In: Charles, Maggie/Hunston, Susan/Pecorari, Diane (eds.): *Academic Writing: At the Interface of Corpus and Discourse*. London, Continuum: 193–214.
- Granger, Sylviane/Paquot, Magali (2010): "Customising a general EAP dictionary to meet learner needs". In: Granger, Sylviane/Paquot, Magali (eds.): *eLexicography in the 21st century: New challenges, new applications. Proceedings of elex2009*. Louvain-la-Neuve, Presses universitaires de Louvain: 87–96.
- Grossmann, Francis (2014): « Les verbes de constat dans l'écrit scientifique ». In : Tutin, Agnès/Grossmann, Francis (eds.): *L'écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes, Presses Universitaires de Rennes: 85–100.
- Hanks, Patrick (2013): *Lexical Analysis : Norms and Exploitations*. Cambridge, MA: MIT.
- Hatier, Sylvain (2013): « Extraction des mots simples du lexique scientifique transdisciplinaire dans les écrits de sciences humaines : une première expérimentation ». In: Morin, Emmanuel/ Estève, Yannick (eds.): *Actes de la conférence TALN-RÉCITAL 1*. Les Sables d'Olonne, ATALA: 85–100.
- Hatier, Sylvain/Yan Rui (2015): « Comparaison de constructions verbales entre un corpus d'apprenants et un corpus d'articles de recherche ». *Conférence donnée à l'occasion des 8es Journées Internationales de Linguistique de Corpus*. Orléans, 2-4 September 2015.
- Hyland, Ken (2008): "Academic clusters: text patterning in published and postgraduate writing". *International Journal of Applied Linguistics* 18/1: 41–62.
- Jezek, Elisabetta/Frontini, Francesca (2010): "From Pattern Dictionary to Patternban". In: de Schryver, Gilles-Maurice (ed.): *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala, Menha: 215–239.
- Kraif, Olivier/Diwersy, Sascha. (2012): « Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques ». In: Antoniadis, Georges/Blanchon, Hervé/Sérass, Gilles (eds.): *Actes de la conférence conjointe JEP-TALN-RECITAL*. Vol. 2. TALN: 399–406.
- Kupsc, Anna (2007): « Extraction automatique de cadres de sous-catégorisation verbale pour le français à partir d'un corpus arboré ». Toulouse, *TALN*. <https://halshs.archives-ouvertes.fr/halshs-00751131> [30.06.2016].
- Legallois, Dominique/Tutin, Agnès (2013): « Présentation : Vers une extension du domaine de la phraséologie ». *Langages* 189: 3–25.
- Messiant, Cédric/Gábor, Kata/Poibeau, Thierry (2010): « Acquisition de connaissances lexicales à partir de corpus : la sous-catégorisation verbale en français ». *Traitement automatique des langues* 51/1: 65–96.
- Nesselhauf, Nadja (ed.) (2005): *Collocation in a Learner Corpus*. Amsterdam/Philadelphia: Benjamins.
- Paquot, Magali (2010): *Academic vocabulary in learner writing : From extraction to analysis*. London/New York: Bloomsbury.
- Pecman, Mojca (2005): « Les apports possibles de la phraséologie à la didactique des langues étrangères ». *Apprentissage des langues et systèmes d'information et de communication* 8/1: 109–122.

- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam/Philadelphia: Benjamins.
- Tran, Thi Thu Hoai (2014): *Description de la phraséologie transdisciplinaire scientifique et réflexions didactiques pour l'enseignement à des étudiants non-natifs. Application aux marqueurs discursifs*. Thèse sous la direction d'Agnès Tutin et de Cristelle Cavalla, Université Stendhal.
- Tutin, Agnès (2007): « Autour du lexique et de la phraséologie des écrits scientifiques ». *Revue française de linguistique appliquée* XII/2: 5–14.
- Tutin, Agnès (2010): « Dans cet article, nous souhaitons montrer que... Lexique verbal et positionnement de l'auteur dans les articles en sciences humaines ». *Lidil. Revue de linguistique et de didactique des langues, Énonciation et rhétorique dans l'écrit scientifique* 4: 15–40.
- Tutin, Agnès/Falaise, Achille (2014): « Expressions polylexicales dans le discours scientifique : une base de données lexicales basée sur corpus ». *Conférence donnée à l'occasion du Colloque Europhras 2014 : La Phrase Ologie : Ressources, Descriptions et Traitements Informatiques*. 10–12 Septembre 2014, Paris.
- Tutin, Agnès/Grossmann, Francis. (eds.) (2014): *L'écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes: Presse universitaire de Rennes.
- Verlinde, Serge/Selva, Thierry/Binon, Jean (2009): « La base lexicale du français : de la lexicographie d'apprentissage à l'environnement d'apprentissage ». In: Heinz, Michaela (ed.): *Le dictionnaire maître de langue. Lexicographie et didactique*. Berlin, Franck/Timme: 289–306.
- Williams, Geoffrey (2008): “Verbs of Science and the Learner’s Dictionary”. In: Bernal, Elisenda/DeCesaris, Janet (eds.): *Proceedings of the XIII EURALEX International Congress*. 15–19 July 2008. Barcelona, Universitat Pompeu Fabra: 797–806.
- Williams, Geoffrey/Million, Chrystel/Alonso, Araceli (2012): “Growing naturally: The DicSci Organic E-Advanced Learner’s Dictionary of Verbs in Science”. In: Fjeld, Ruth Vatvedt/Torjusen, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress*. 7–11 August 2012. Oslo, University of Oslo: 1008–1013.