



HAL
open science

Analyse contrastive des patrons verbaux dans l'écrit scientifique entre scripteurs étudiants et experts

Sylvain Hatier, Rui Yan

► **To cite this version:**

Sylvain Hatier, Rui Yan. Analyse contrastive des patrons verbaux dans l'écrit scientifique entre scripteurs étudiants et experts. CORELA - COgnition, REprésentation, LAngage, 2017, HS-21, 10.13092/lo.78.2952 . hal-01536419

HAL Id: hal-01536419

<https://hal.science/hal-01536419>

Submitted on 11 Jun 2017

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.



Corela

Cognition, représentation, langage

HS-21 | 2017

**Linguistique de corpus : vues sur la constitution,
l'analyse et l'outillage**

Analyse contrastive des patrons verbaux dans l'écrit scientifique entre scripteurs étudiants et experts

Sylvain Hatier et Rui Yan



Éditeur

Cercle linguistique du Centre et de l'Ouest -
CerLICO

Édition électronique

URL : <http://corela.revues.org/4879>

ISSN : 1638-573X

Référence électronique

Sylvain Hatier et Rui Yan, « Analyse contrastive des patrons verbaux dans l'écrit scientifique entre scripteurs étudiants et experts », *Corela* [En ligne], HS-21 | 2017, mis en ligne le 14 mars 2017, consulté le 16 mars 2017. URL : <http://corela.revues.org/4879>

Ce document a été généré automatiquement le 16 mars 2017.



Corela – cognition, représentation, langage est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

Analyse contrastive des patrons verbaux dans l'écrit scientifique entre scripteurs étudiants et experts

Sylvain Hatier et Rui Yan

Cadre et objectifs de l'étude

Notre étude s'inscrit dans le domaine de la linguistique de corpus, et emprunte à celle-ci le « principe d'idiomaticité » (idiom principle) du fonctionnement de la langue (Sinclair, 1991). Nous appliquons ici ce principe à l'étude d'un lexique spécifique, le lexique scientifique transdisciplinaire (désormais LST), ceci afin de rendre compte des constructions polylexicales récurrentes le concernant. Cette approche étendue de la phraséologie permet de prendre en compte des combinaisons lexico-syntaxiques, dont l'intérêt dans l'enseignement/apprentissage d'une langue étrangère nous paraît essentiel, notamment dans le cadre des langues de spécialité ou de l'« écrit académique ».

Dans le présent article, nous nous intéressons aux patrons verbaux en procédant à une analyse contrastive entre un corpus de mémoires d'étudiants chinois apprenants du français, un corpus de mémoires d'étudiants français en master de didactique du Français et un corpus d'articles de recherche en sciences humaines et sociales parus dans des revues à comités de lecture. Le choix des patrons verbaux dans l'écrit scientifique comme objet d'étude est guidé par des objectifs d'ordre linguistique et didactique. D'une part, à l'instar de Hanks (2013), nous considérons que l'analyse du sens lexical ne peut se faire sur des mots isolés. Les mots ont des sens potentiels, pouvant être activés par les patrons dans lesquels ils s'inscrivent. Ainsi, l'étude des patrons permet de restreindre l'ambiguïté du sens du verbe. D'autre part, comme le souligne Paquot (2010 : 118), « les verbes ayant des fonctions rhétoriques ou organisationnelles spécifiques dans l'écrit académique

entrent généralement dans des séquences compositionnelles 'flexibles'¹ ». Or, ce domaine lexico-grammatical est une des difficultés majeures rencontrées par les étudiants dans leurs activités rédactionnelles (Cavalla, 2014).

À l'instar de Tutin & Grossmann (2014), nous prenons pour objet d'étude les verbes du lexique scientifique transdisciplinaire. Ce lexique, au cœur de l'argumentation, permet de décrire l'activité scientifique notamment à travers le raisonnement et le positionnement, exprimés par les verbes d'analyse et d'opinion. Le LST est ainsi à la fois méta-scientifique et méta-discursif.

Nous avons pour objectif didactique premier l'identification des difficultés liées aux patrons verbaux. Nous dégageons alors les difficultés communes aux étudiants chinois et français ainsi que celles propres aux étudiants chinois. Pour ce faire, nous procédons à une analyse contrastive des corpus d'étudiants natifs/non natifs. À travers l'identification de ces difficultés, nous avons pour perspective l'élaboration d'un outil d'aide à la rédaction scientifique. Nous envisageons ainsi d'intégrer les résultats de l'analyse linguistique des patrons dans des exercices se basant sur la méthode inductive et les corpus pour améliorer l'enseignement/apprentissage des patrons verbaux.

Ce travail d'analyse contrastive des patrons verbaux part de plusieurs hypothèses. Premièrement, nous pensons que la variété des patrons verbaux se révélera moindre dans les corpus d'étudiants que dans le corpus d'experts. Au niveau de la comparaison scripteurs natifs/non-natifs, nous supposons que les étudiants français emploient une plus grande diversité de patrons que les étudiants chinois. L'analyse de ces différences dans l'emploi des patrons verbaux nous permettra également d'en analyser les effets sur les dimensions argumentatives et rhétoriques.

Genres des corpus d'analyse

Nous prenons pour corpus d'analyse différents ensembles de textes se rapportant aux genres de l'écrit scientifique et de littéracie avancée. Le genre étant défini ici, à la suite de Biber (1993), comme une catégorie de textes définis situationnellement. Les deux corpus d'étudiants chinois et français peuvent être regroupés dans un même sous-genre dit de littéracie avancée. Le corpus d'experts appartient lui au sous-genre de l'article de recherche. Lors de l'étape d'analyse contrastive, nous tiendrons compte de ces différences de genre dans les corpus. En effet, comme le souligne Rinck (2013), parmi les multiples facteurs de variation de genre, se trouvent notamment la discipline, l'expertise et le style du scripteur. Pour l'ensemble de ses genres d'écrit, nous nous intéressons plus précisément au lexique propre à la démonstration scientifique, le LST.

Lexique scientifique transdisciplinaire

Le LST est le lexique spécifique au genre de l'écrit scientifique. Nous le définissons par des propriétés linguistiques et lexicométriques. Ces propriétés nous ont permis, dans de précédents travaux², de procéder à une extraction semi-automatique du LST et d'élaborer une classification sémantique de ce lexique.

Définition

À la suite de Tutin (2007), nous définissons le LST comme le lexique renvoyant au discours sur les objets et les procédures scientifiques. Il est central dans l'argumentation et la structuration du discours scientifique, comme le notent notamment Drouin (2007) ou Paquot & Bestgen (2009).

Le LST est ainsi défini comme étant :

- transdisciplinaire : statistiquement fréquent dans différents corpus disciplinaires. Ce critère exclut la terminologie, intra-disciplinaire et thématique.
- spécifique au genre de l'écrit scientifique : statistiquement surreprésenté par comparaison à d'autres genres³ (littéraire, journalistique, oral). Ce critère exclut la langue générale, commune aux autres genres.

Dans l'extrait d'article d'anthropologie suivant, les segments soulignés appartiennent au LST :

Du point de vue anthropologique ce déplacement de paradigme met en jeu la définition même de l'identité en renforçant le divorce entre l'identité civile et l'identité personnelle ou sociale⁴

Extraction

À partir d'un corpus d'articles de recherche en sciences humaines et sociales, que nous présenterons en détails dans la partie méthodologie, nous avons procédé à l'extraction d'une liste de mots répondant aux critères de transdisciplinarité et de spécificité. Cette liste a ensuite été validée manuellement par des experts pour aboutir à une ressource du LST composée de 1297 mots pleins (noms, verbes, adjectifs et adverbes).

Classification sémantique

À l'aide des ressources lexicographiques de Dubois & Dubois-Charlier (1997, 2010), nous avons par la suite enrichi la ressource du LST en informations sémantiques (glose, classes sémantiques). La classification repose sur une typologie à deux niveaux, classes et sous-classes. Celles-ci sont constituées sur la base de propriétés lexico-syntaxiques, à la manière de Flaux & Van De Velde (2000) et de Dubois & Dubois-Charlier, suite à l'observation et l'analyse des profils combinatoires⁵ des éléments du LST dans le corpus. Nous disposons ainsi d'une ressource lexicale spécifique au genre de l'écrit scientifique, dont un extrait est présenté dans le tableau ci-dessous.

Acception	Catégorie	Classe	Sous-classe	Glose
<i>article</i>	Nom	{communication}	{document}	Écrit
<i>document</i>	Nom	{communication}	{document}	Écrit
<i>ouvrage</i>	Nom	{communication}	{document}	Livre
<i>présenter-1</i>	Verbe	{état}	{inclusion}	comporter
<i>présenter-2</i>	Verbe	{analyse}	{description}	montrer, exposer

Tableau 1 : Exemple d'entrées lexicales du LST

Les entrées lexicales sont des acceptions de mots, avec pour informations, le lemme, la catégorie, la classe, la sous-classe et une courte définition.

Les membres d'une même classe/sous-classe sont définis comme étant co-hyponymes et répondant positivement à la définition et au test d'appartenance de la classe. Par exemple la classe {*support de communication*} a pour définition : élément servant de support de communication écrite ou partie d'un tel élément. Le test lexico-syntaxique permettant de valider l'appartenance d'un élément à cette classe est 'Ce N est consacré à'.

Nous exploitons dans notre étude cette ressource de plusieurs manières. D'une part, les verbes pour lesquels nous analysons les patrons lexico-syntaxiques appartiennent à la ressource du LST. D'autre part, les informations sémantiques de classes et sous-classes sont intégrées dans l'analyse des patrons verbaux. Ainsi, si nous prenons pour exemple le verbe *présenter* dans son sens de *montrer, exposer*⁶. Étant donné la sous-classe {*document*} de la classe {*support de communication*}, un exemple de patron de *présenter* sera :

Sujet (SN-{*document*}) *présenter* Objet (SN)

Ce patron précise la structure lexico-syntaxique dans laquelle s'insère le verbe *présenter*. Il prend ici pour sujet un syntagme nominal dont la tête appartient à la sous-classe {*document*} et prend pour complément d'objet un syntagme nominal dont l'appartenance au LST n'est pas spécifiée.

Ce patron correspond notamment à l'exemple suivant : « *Cet article **présente une plateforme** qui permet d'évaluer l'expérience utilisateur face à des concepts innovants dans des systèmes sociotechniques évolutifs*⁷. »

Nous disposons ainsi d'une ressource sémantique du LST que nous incluons dans l'analyse des patrons verbaux.

Méthodologie

Nous détaillons dans un premier temps la méthodologie commune aux trois corpus analysés, soit les étapes de constitution et d'enrichissement. Dans un second temps, nous présentons le processus d'extraction semi-automatique des patrons verbaux à partir des corpus d'étudiants et d'experts.

Corpus

L'analyse contrastive des patrons repose sur trois corpus :

- un corpus d'écrits d'étudiants chinois, de 600 000 mots, constitué de mémoires de master⁸ de spécialité de français dans une université chinoise. Leur sujet d'étude porte sur la traduction (9), la littérature (12), la linguistique (1) et des thématiques socioculturelles (7) ;
- un corpus d'écrits d'étudiants français, de 460 000 mots, constitué de mémoires de master en spécialité didactique du français. Ce corpus est issu du corpus *Littéracie Avancée*⁹ ;
- un corpus d'experts, de 5 millions de mots, issu du projet Scientext¹⁰, constitué de 500 articles de dix disciplines des sciences humaines et sociales¹¹.

Analyse et enrichissement

L'ensemble des corpus a été analysé syntaxiquement avec XIP¹², analyseur en dépendances. Les corpus sont ainsi enrichis en lemmes, catégories syntaxiques, traits morphosyntaxiques et relations syntaxiques. Les traits de classes et sous-classes sémantiques ont également été projetés dans le corpus afin d'intégrer ces informations sémantiques dans les patrons verbaux extraits.

Extraction des constructions verbales

Nous nous concentrons pour cette étude sur quatre verbes du LST (*montrer, expliquer, décrire, considérer*), de fréquence élevée dans les trois corpus et essentiels du point de vue argumentatif, car mobilisés dans le positionnement et analyse scientifique. En effet, les verbes *montrer* et *considérer* sont mis à contribution pour construire la « posture réflexive » (Rinck, 2010) de scripteur (ex : *nous montrons l'importance, on considère que*). Quant aux verbes *décrire* et *expliquer*, ils sont employés dans les phases de description et d'interprétation des données, permettant ainsi d'établir des relations entre les éléments linguistiques.

Pour chacun de ces verbes, nous procédons à l'extraction de l'ensemble des occurrences. Chaque occurrence verbale est ensuite représentée par un cadre de sous-catégorisation, extrait automatiquement à partir des résultats de l'analyse en dépendances.

Analyse en dépendance et cadre de sous-catégorisation

Comme nous l'avons précisé précédemment, les trois corpus (écrits d'étudiants natifs, d'étudiants non-natifs et d'experts) ont été analysés en dépendances avec XIP. Les corpus sont au format XML-TEI. Pour chaque phrase, sont listés l'ensemble des tokens et l'ensemble des relations de dépendances les impliquant dans la phrase. Considérons la phrase suivante :

Les valeurs sémantiques de prépositions relèvent essentiellement du dictionnaire.

L'analyse syntaxique et l'enrichissement sémantique nous permettent d'aboutir au résultat suivant :

Selon la même approche que Messiant, Gábor, & Poibeau (2010), nous souhaitons acquérir des cadres de sous-catégorisation « sans a priori, pour faire émerger du corpus [ceux] correspondant à l'usage ». Cette démarche « corpus-driven » (cf. Tognini-Bonelli, 2001), permet de faire émerger à partir du corpus des patrons récurrents. Ces patrons sont obtenus après une phase de factorisation des cadres de sous-catégorisation, à la manière de Kupsc (2007). Nous regroupons ainsi certaines relations de dépendance sous une même étiquette. Par exemple, si le verbe est dépendant dans la relation objet avec un semi-modal (*devoir, sembler, aller*), la relation MODAL-OBJ est ajoutée dans le cadre de sous-catégorisation. De plus, les informations sémantiques concernant les cooccurrents du verbe nous permettent de dégager, lorsque cela est possible, des classes ou sous-classes sémantiques préférentiellement associées au verbe pour le cadre en question.

Toute occurrence d'un verbe est alors analysée en termes d'ensemble de relations, tel que l'illustre la figure ci-après :

Le verbe *relever* est dans cet exemple :

– gouverneur dans la relation 'sujet'¹³ avec pour dépendant *valeurs* (nom du LST, dont la

classe et la sous-classe sont renseignées dans les traits) ;

- gouverneur dans la relation 'adverbiale' avec pour dépendant *essentiellement* ;
- gouverneur dans la relation 'complément prépositionnel' avec pour dépendant *dictionnaire*.

Étant donné que nous ne retenons pas les compléments adverbiaux dans cette étude, nous obtenons alors le cadre de sous-catégorisation suivant :

(SUBJ {document}) relever_VERB (COMPLEMENT PREP.).

Pour chaque cadre, nous calculons les mots en cooccurrence statistiquement significative¹⁴ avec le verbe concerné. Le même traitement est effectué au niveau des traits sémantiques afin d'identifier les classes intégrant préférentiellement le cadre. Nous intégrons également le trait +hum lorsque le cooccurrent est un des pronoms suivants : *je, on, nous*.

La phase de factorisation des cadres permet de regrouper au maximum les occurrences et ainsi d'analyser les structures les plus fréquentes. Nous obtenons pour le verbe *montrer*, dans le corpus d'experts, les cadres présentés ci-dessous.

Il nous est alors possible d'identifier le cadre le plus fréquent (ici 762 occurrences) pour le verbe *montrer*, comprenant un sujet non humain et une complétive introduite par *que*. Sont également renseignés, pour chaque cadre, des exemples, ainsi que les mots et classes les plus fréquents. Les résultats de cette phase d'extraction et de factorisation des cadres de sous-catégorisation intègrent ainsi des informations d'ordres syntaxique, lexical et sémantique. Ces informations sont alors utilisées dans la seconde phase de regroupement manuel présentée dans la partie suivante.

Patrons lexico-syntaxiques et modèle Corpus Pattern Analysis

Le modèle Corpus Pattern Analysis (CPA), du domaine de l'analyse lexicale, envisage le sens en contexte, inscrit dans les patterns. Cette approche est dans la continuité de la linguistique de corpus de Sinclair dans la mesure où les patterns ne sont repérables que par le biais d'analyses de fréquence inductives (corpus driven approach). Elle se distingue des travaux précédents sur les patterns, notamment l'approche de Pattern Grammar (Hunston & Francis, 2000) en accordant un rôle clé aux collocations lexicales pour déterminer la notion de pattern :

A pattern is a statement of the clause structure (valency) associated with a meaning of a verb, together with typical semantic values of each argument, realized by salient collocates [...] Different semantic values of arguments activate different meanings of each verb¹⁵. (Hanks, 2013)

Ce modèle est d'abord utilisé dans la construction du *Pattern Dictionary of English Verbs* (PDEV¹⁶) pour l'anglais langue générale (Hanks, 2013). Nous illustrons ci-dessous un exemple de pattern dans le PDEV :

Dans cet exemple du verbe *to execute*, nous pouvons voir que le pattern définit pour chacun de ses arguments des « types sémantiques » ou concepts cognitifs tels que « *human* », « *institution* », « *action* » etc. Les types sémantiques peuvent être complétés par des « rôles sémantiques » en vue d'affiner la définition du pattern. Ces rôles représentent les propriétés extrinsèques des mots au niveau de chaque argument, en fonction du contexte du pattern. Dans l'exemple ci-dessus, la séquence « *human 1 =Ruler or Military* » signifie que le type sémantique « *human* » s'instancie dans le rôle du souverain ou du militaire dans ce contexte. L'implicature renvoie ainsi à une paraphrase du pattern.

Plusieurs projets en cours montrent que le modèle CPA est adapté à des objectifs divers :

- étude lexicale dans un domaine spécifique comme l'écrit académique - l'*E-Advanced Learner's Dictionary of Verbs in Science* (DicSci) (Williams, Million & Alonso, 2012) ;
- étude quantitative des structures verbes arguments en espagnol (Nazar & Renau, à paraître) ;
- étude des verbes espagnols à visée pédagogique pour le *Diccionario de aprendizaje del español como lengua extranjera* (DAELE¹⁷) (Renau & Battaner, 2011).

Concernant notre étude, nous appliquons le modèle CPA à l'écrit scientifique en français dans une perspective didactique. L'identification des difficultés sur les constructions verbales dans les écrits scientifiques constitue la première étape dans cet objectif. En effet, ce diagnostic nous amène à poser des hypothèses explicatives qui nous guident dans notre réflexion didactique afin d'aider les étudiants (non-natifs) dans la rédaction scientifique.

Les cadres de sous-catégorisation sont ensuite analysés manuellement et regroupés dans des patrons proches du modèle CPA, incluant des informations lexicales, syntaxiques et sémantiques. Ainsi, si nous prenons pour exemple le verbe *expliquer*, considérons un de ses cadres de sous-catégorisation :

- (SUBJ-hum) (SE se_PRON) expliquer_VERB par (92 occurrences).
- Ex 1 : *La différence s'explique sans doute par la proximité de ces élus avec leurs administrés, souvent plus favorables à Bonn.* (corpus experts : géographie)

Ce cadre décrit un emploi pronominal du verbe *expliquer*. Dans ce cadre, le trait -hum du sujet renvoie à un nom non humain dont les plus fréquents sont *différence*, *situation*, *résultat*, *fait*. Les types sémantiques utilisés préférentiellement comme sujets dans ce cadre sont la classe {*relation*}, avec le nom *différence*, et la sous-classe {*situation*} de la classe {*objet scientifique*}, avec les noms *situation* et *fait* comme illustré dans l'exemple 2 :

- Ex 2 : *Le fait que les corrélations avec l'absentéisme soient sensiblement plus élevées au niveau des unités de soins qu'au niveau individuel peut s'expliquer par le fait que la variable absentéisme reflète la seule variance entre les unités.*

À ces classes s'ajoutent des noms renvoyant à la classe {*processus cognitifs*}, tel *choix*. Ces occurrences correspondent aux cas dans lesquels l'auteur se justifie, se positionne, comme l'illustre l'exemple 3 :

- Ex 3 : *Ce choix s'explique par le fait qu'elles sont particulièrement appropriées aux échantillons composés de personnes qui occupent un emploi commercial comme dans le cas de notre échantillon.* (corpus experts : psycho)

En identifiant ainsi les classes sémantiques en cooccurrence significative avec le verbe, nous pouvons intégrer cette dimension sémantique dans les patrons. Cependant, le recouvrement entre classes sémantiques et noms rentrant effectivement en cooccurrence avec le verbe est évidemment partiel. Une classe sémantique peut être fréquemment en relation avec le verbe à travers une minorité de ses éléments tandis que ses autres éléments n'entrent jamais en cooccurrence avec le même verbe. L'objectif n'est pas de relever l'intégralité des combinaisons lexicales mais d'intégrer des informations sémantiques afin de décrire le plus finement possible les patrons.

Si nous reprenons l'exemple du cadre pronominal du verbe *expliquer*, nous voyons qu'il est suivi de la préposition *par*, qui introduit l'objet de la cause. Cet élément est fréquemment réalisé par un nom de la classe {*qualité*} (*nécessité*, *difficulté*, *propriété*) et de la

classe {*objet scientifique*} (*phénomène, méthode*). Nous pouvons donc présenter le patron lexico-syntaxique comme suit :

Patron : [[relation|objet scientifique|processus cognitif]] s'explique par [[qualité|objet scientifique]]

Implicature : [[qualité|objet scientifique]] constitue la raison de [[relation|objet scientifique|processus cognitif]] et le rend compréhensible.

Lexet [[relation|objet scientifique|processus cognitif]] : *différence, fait, phénomène, choix, situation, résultat, problème*

À la suite de cette étape d'analyse des cadres de sous-catégorisation et de création des patrons, nous pouvons procéder à l'étude contrastive des constructions verbales entre les trois corpus.

Étude contrastive des constructions verbales

Le corpus d'experts nous servira ici de référence pour identifier les patrons les plus fréquents dans l'écrit scientifique. Nous nous appuyerons sur cette base pour confronter les deux corpus d'étudiants, comparables au niveau de la taille et du contexte d'écriture (mémoire de fin d'études).

Fréquences et analyse quantitative

Avant de nous intéresser plus spécifiquement à quatre verbes du LST (*décrire, expliquer, montrer, considérer*) et à leurs constructions, nous présentons dans cette partie les points communs et spécificités des corpus pour l'ensemble des verbes du LST. Ce lexique, constitué à partir du corpus d'experts, est composé de 342 verbes, dont 334 sont présents dans le corpus d'étudiants chinois et 336 dans le corpus d'étudiants français. Cela rejoint notre remarque précédente selon laquelle ce lexique est central dans le genre de l'écrit scientifique et essentiel à maîtriser pour les étudiants.

Si nous nous intéressons aux verbes les plus fréquents, nous constatons que parmi les 50 verbes les plus présents dans chaque corpus d'étudiants, 28 sont en commun : *devenir, exister, utiliser, exprimer, présenter, considérer, constituer, montrer, décrire, développer, analyser, concerner, sembler, apparaître, permettre, produire, choisir, expliquer, représenter,, réaliser, établir, proposer, étudier, paraître, apporter.*

Les corpus d'étudiants diffèrent cependant sur la fréquence de certains verbes. Par exemple, le verbe semi-modal *permettre* est le plus fréquent chez les étudiants français (fréquence par million de mots de 2 982 contre 294 pour le corpus d'étudiants chinois). Il est ainsi largement utilisé dans des constructions où un {*processus_cognitif*} ou un {*objet scientifique*} précède '*permettre de*', une construction essentielle dans l'argumentation scientifique (voir exemples 4 et 5).

– Ex 4 : Les **exemples** ci-dessus nous **permettent de** tirer la conclusion : la qualité de traduction littéraire est qu'elle conserve le plus possible l'image [...] (corpus d'étudiants chinois)

– Ex 5 : Cette **prise de conscience** du sens de l'écriture par la production d'écrit, **permet d'**impliquer d'avantage les élèves dans leur apprentissage [...] (corpus d'étudiants français)

À l'inverse, parmi les verbes fréquents dans le corpus d'étudiants chinois mais absents ou peu fréquents dans le corpus d'étudiants français, nous identifions les verbes suivants :

traduire, obtenir, former, signifier, influencer, posséder, indiquer, composer, manifester, mentionner, publier, correspondre, marquer, atteindre, refléter, conduire, construire, exercer.

Le verbe traduire (fréquence par million de mots de 1 083 contre 53 pour le corpus d'étudiants français) est le plus fréquent chez les étudiants chinois. Ceci n'est cependant pas dû à l'emploi de l'acception scientifique ('être l'expression d'un phénomène'), mais à son acception appartenant à la langue générale 'formuler dans une autre langue étrangère', liée à la thématique du corpus d'étudiants chinois, la traduction. Nous faisons le même constat pour d'autres verbes (obtenir place/permission, manifester sentiment) pour lesquels l'acception dans le corpus ne correspond à aucune entrée lexicale de notre ressource du LST. Ainsi, certains de ces verbes ne sont pas représentatifs de ce lexique spécifiquement mobilisé dans les écrits scientifiques. De plus, d'autres verbes, bien qu'éléments du LST, sont employés dans un sens ne s'inscrivant pas dans notre définition du LST.

Par exemple, si nous comparons les exemples suivants :

- Ex 6 : *Pas mal de chercheurs **traduisent** leurs propres essais académiques en une autre langue pour les publier dans les magazines ou journaux étrangers [...].* (corpus d'étudiants chinois)

- Ex 7 : *L'inacceptabilité de contre dans ces contextes **traduit** là encore le fait que l'individu visé par ces sentiments n'est supposé exercer aucune contre-force.* (corpus experts : linguistique)

Nous constatons dans l'exemple 6 un emploi non-LST du verbe traduire dont la haute fréquence dans le corpus d'étudiants chinois est bien causée par la thématique des écrits. De la même manière, la thématique du corpus d'étudiants français étant la didactique, nous y trouvons parmi les verbes les plus fréquents *écrire* et *lire*. L'analyse qualitative des concordances est ici nécessaire pour savoir si les verbes sont effectivement utilisés dans leur sens scientifique ou si les étudiants (français et/ou chinois) ne maîtrisent pas l'emploi de ce sens spécialisé.

Concernant les verbes surreprésentés dans le corpus d'étudiants français, en comparaison du corpus d'étudiants chinois, nous identifions : *observer, impliquer, effectuer, intéresser, adapter, favoriser, constater, souligner, définir, correspondre, amener, aborder, participer, interroger, noter, préciser, reconnaître, faciliter, évaluer, progresser, identifier, appuyer*.

Ces verbes sont également fréquemment employés dans la description de l'activité scientifique dans le corpus d'experts. Ils appartiennent notamment aux classes sémantiques suivantes : {constat} (*observer, constater, noter*), {opinion} (*souligner*), {analyse} (*définir, préciser*).

Nous notons ainsi une certaine convergence concernant les verbes du LST employés par les étudiants chinois et français. Cependant, les différences de fréquence sont un indice de la variabilité de l'emploi de ces verbes, et notamment de leur usage spécifique dans l'écrit scientifique. Afin d'étudier ces possibles divergences, nous procédons à une analyse qualitative des patrons présents dans les différents corpus, et ce, pour quatre verbes du LST.

Analyse qualitative avec l'outil Anatext

Nous précisons ici que notre méthodologie d'extraction semi-automatique des patrons a dans un premier temps été élaborée pour le corpus d'experts dont la taille est dix fois supérieure à celle des corpus d'étudiants. Les résultats de l'étape de factorisation des cadres (reposant sur des effectifs nombreux) sont alors insuffisants pour une analyse

basée uniquement sur les calculs de fréquence. Ce changement d'échelle explique le recours à un outil supplémentaire pour procéder à l'analyse qualitative des patrons, et notamment à l'étude des classes sémantiques s'inscrivant dans les patrons. Nous utilisons ainsi, pour l'étape d'observation du corpus, l'outil de textométrie Anatext¹⁸, conçu par Olivier Kraif.

Les quatre verbes du LST (*montrer*, *considérer*, *décrire* et *expliquer*) que nous étudions spécifiquement sont parmi les plus fréquents pour chaque corpus analysé, ce qui permet d'extraire et d'observer le maximum d'occurrences. Le tableau suivant présente leur fréquence par million de mots (et fréquence absolue entre parenthèses) dans chaque corpus :

	Experts	Étudiants français	Étudiants chinois
<i>montrer</i>	649 (3 114)	780 (363)	449 (271)
<i>considérer</i>	533 (2 259)	367 (171)	469 (283)
<i>décrire</i>	236 (1 134)	221 (103)	384 (232)
<i>expliquer</i>	341 (1 637)	896 (417)	265 (160)

Tableau 2 : Fréquences des quatre verbes dans les trois corpus

Comme nous pouvons le constater, les verbes *montrer* et *considérer* sont les plus fréquemment utilisés dans les trois corpus. Le verbe *décrire* est surreprésenté dans le corpus d'étudiants chinois, le verbe *expliquer* dans le corpus d'étudiants français. Nous posons l'hypothèse que ces variations ont deux facteurs principaux : d'une part, la thématique des mémoires (didactique de l'écrit pour les étudiants français et traduction/littérature pour les étudiants chinois), et d'autre part la compétence discursive des deux groupes de scripteurs>

Nous présentons dans la partie suivante l'analyse contrastive des patrons grammaticaux, en introduisant tout d'abord les patrons lexico-syntaxiques répertoriés en se basant sur le corpus d'experts. Les patrons grammaticaux relèvent de l'approche « *Pattern Grammar* » et n'incluent pas les collocations. Dans un premier temps, afin de faciliter la comparaison et pour avoir une vision générale des constructions, l'analyse des patrons dans les corpus d'étudiants se fera au niveau des patrons grammaticaux. Ces patrons grammaticaux résultent de l'extraction des cadres de sous-catégorisation. Les patrons lexico-syntaxiques se situent à un niveau d'analyse supérieure, nécessitant un traitement manuel des arguments verbaux récurrents, uniquement effectué sur le corpus d'expert. La prise en compte des phénomènes collocatifs sera effective pour l'étape d'analyse suivante, qui prend notamment appui sur les calculs de cooccurrence.

Verbe *montrer*

Afin de ne pas devoir traiter un trop grand nombre de cadres de sous-catégorisation différents, nous fixons le seuil de fréquence minimale par cadre à 15 pour le corpus d'experts et à 5 pour les corpus d'étudiants.

Les patrons lexico-syntaxiques extraits du corpus d'experts sont filtrés selon le critère de transdisciplinarité. Ce corpus étant composé d'articles de 10 disciplines, nous ne retenons que les cadres non spécifiques à une seule discipline. Ce critère de transversalité est également appliqué pour les compléments nominaux appartenant au LST intégrant ces cadres lexico-syntaxiques.

Nous présentons ci-dessous un exemple des patrons lexico-syntaxiques du verbe *montrer*. Le formalisme, intégrant informations syntaxiques, lexicales et sémantiques, est inspiré du modèle CPA.

patron 1.1 : [[humain =auteur]] montre [complétive-que/comment/quel]
patron 1.2 : [[humain =auteur]] montre [[{qualité}{relation}]]
patron 1.3 : {Comme} {le} montre [[humain =auteur]], [proposition]
patron 1.4 : [[{événement}{support}{objet scientifique}]] (datif nous) permet de montrer [complétive-que/comment/quel]
Ex : on montre que/on montre la différence/comme on le montre/le résultat nous permet de montrer que
patron 2.1 : [[{événement}{support}{objet scientifique}]] montre [complétive-que/comment/quel]
patron 2.2 : [[{événement}{support}{objet scientifique}]] montre [[état_qualité relation]]
patron 2.3 : (Comme) (le) montre [[{événement}{support}{objet scientifique}]], [proposition]
Ex : l'analyse montre que/l'exemple montre la différence/comme le montre le tableau

Tableau 3 : Patrons lexico-syntaxiques du verbe *montrer* dans le corpus d'experts

Dans les patrons pour lesquels le verbe sélectionne un sujet humain, *montrer* prend le sens de 'donner des arguments pour faire savoir et justifier certains faits scientifiques'.

Lorsque *montrer* sélectionne un sujet de type chose, le verbe prend le sens 'mettre en évidence'.

La modélisation des patrons permet ainsi d'associer des constructions lexico-syntaxiques à des acceptions identifiées.

Nous confrontons maintenant ces patrons, issus du corpus d'experts, aux patrons grammaticaux du verbe *montrer* dans les deux corpus d'étudiants. Le tableau ci-dessous présente les rangs de fréquence pour les patrons du verbe *montrer* dans l'ensemble des corpus. Un rang 1 signifie que le patron est le plus utilisé pour le verbe en question dans le corpus. La valeur '-' signifie que le patron est absent du corpus, ou que sa fréquence est inférieure au seuil de 5 occurrences.

Patrons grammaticaux	Experts	Français	Chinois
(SUBJ chose) montrer que	1	2	3
(SUBJ humain) montrer que	2	1	3
(SUBJ humain) montrer (OBJ chose)	3	3	2

(SUBJ chose) montrer (OBJ chose)	3	3	1
Comme le montrer (SUBJ humain)	4	-	-
Comme le montrer (SUBJ chose)	4	-	-
(SUBJ chose) permet de montrer que	4	4	-

Tableau 4 : Ordre des fréquences des patrons du verbe montrer dans les trois corpus

Ce tableau nous révèle que les étudiants chinois ont plus de tendance à utiliser des patrons intégrant un complément d'objet nominal chose tandis que les experts et les étudiants français emploient plus fréquemment la structure en complétive *que*. Parmi les patrons avec complétive, celui prenant un sujet humain est fréquemment employé par les experts et étudiants français. Cependant, l'observation des concordances fait apparaître une divergence dans l'emploi de ce patron. Ce dernier est utilisé notamment pour marquer le positionnement, associé aux pronoms personnels *on* ou *nous* dans le corpus d'experts. Cependant, ce positionnement n'est pas directement « assumé » par les scripteurs étudiants, qui renvoient alors souvent à un autre chercheur du domaine pour soutenir l'argumentation. Le recours au discours rapporté ne marque alors pas une opinion explicite et le point de vue de l'auteur est dissimulé en faisant dire par autrui ce qu'il souhaiterait exprimer. L'argumentation se construit ainsi comme dans l'exemple suivant :

- Ex 8 : Nadeau et Fisher (2011), **montrent** que les connaissances explicites sont corrélées au degré de certitude des apprenants concernant leurs réponses. (corpus d'étudiants français).

Le même constat est fait pour les étudiants chinois qui emploient très peu les pronoms *on* ou *nous* mais plutôt des renvois :

- Ex 9 : De plus, la conclusion de XU Minglong **montre** que la « sinomanie » est un phénomène important dans l'histoire des échanges culturels entre la Chine et l'Europe. (corpus d'étudiants chinois).

Cependant, les patrons les plus fréquemment employés par les étudiants chinois activent le sens 'faire voir' du verbe *montrer*, comme dans l'exemple 10 :

- Ex 10 : Le tableau en bas peut **montrer** les diplômes différents dans les étapes différentes. (corpus d'étudiants chinois).

Si nous considérons les autres patrons présents dans le corpus d'experts, nous constatons que les étudiants emploient peu de constructions infinitives. Ainsi, les patrons incluant les verbes *suggérer*, *nécessiter*, *permettre* et les constructions en *comme* sont peu utilisées par les étudiants. Pourtant, il nous semble que ces constructions sont essentielles dans l'argumentation (*ces résultats permettent de montrer*) et dans la structure du texte (*comme on l'a montré*). Les étudiants semblent ainsi avoir des difficultés à exprimer leurs propres opinions et à faire des renvois textuels.

Verbe *considérer*

L'analyse qualitative des patrons du verbe *considérer* fait émerger des différences comparables à l'exemple du verbe *montrer*.

Nous présentons ci-dessous les patrons modélisés à la suite des traitements automatiques et de l'observation du corpus d'experts :

patron 1 : [[humain =auteur]] considère [[{objet scientifique}{relation}]] comme
Ex : on considère la relation comme
patron 2 : [[humain =auteur]] considère [complétive-que]
Ex : on considère que
Patron 3 : Si/Lorsque [[humain =auteur]] considère [[{ objet scientifique}{qualité} =sujet], [proposition]
Ex : si l'on considère l'exemple

Tableau 5 : Patrons lexico-syntaxiques du verbe considérer dans le corpus d'experts

Ce verbe correspond à trois entrées lexicales, ou acceptions, dans notre ressource du LST : 'donner un jugement', 'exprimer une opinion' et 'prendre en considération'.

Le patron 3, recouvrant le dernier sens, est notamment utilisé lorsque l'auteur introduit un nouvel argument, la [proposition]. Cet argument est alors en corrélation avec le nom objet de *considérer*, représenté par : [[{objet scientifique}{qualité} =sujet]]. Ainsi, dans l'exemple suivant, le nom *pertinence*, de la classe du LST {qualité} permet d'introduire la [proposition] qui suit :

– Ex 11 : Si l'on **considère**, maintenant, les pertinences évoquées par l'orateur, elles sont de deux ordres : la pertinence coranique, tout d'abord, puisqu'il affirme que le projet de loi a été examiné par une assemblée d'experts religieux... (corpus experts : sociologie)

L'emploi de ce patron est ici justifié par sa fonction méta-discursive.

Le tableau ci-après présente les fréquences des patrons du verbe *considérer* dans les différents corpus,

Patrons grammaticaux	Experts	Français	Chinois
(SUBJ humain) considérer que	2	2	2
(SUBJ humain) considérer (OBJ chose) comme	3	1	1
(Si) considérer (OBJ chose)	1	-	-

Tableau 6 : Ordre des fréquences des patrons du verbe considérer dans les trois corpus

Nous voyons que les étudiants chinois et français ont un emploi comparable du verbe *considérer* en termes de fréquences de patrons. Ils privilégient les deux premiers patrons au détriment du dernier, pourtant le plus employé par les experts. Pour ce verbe *considérer*, nous avons observé une grande variété de constructions chez les experts, en comparaison du peu de patrons employés par les étudiants. Ces constructions, assurant

diverses fonctions discursives (par exemple la fonction de topicalisation, en abordant un nouvel argument), se réalisent notamment ainsi : *considérons les deux exemples, le cas considéré, si l'on considère la relation*. Ces constructions sont ainsi essentielles dans l'écrit scientifique et leur absence dans les corpus d'étudiants doit être prise en compte dans le cadre d'une formation d'aide à la rédaction.

Nous observons également une divergence entre scripteurs experts et étudiants au niveau des pronoms sujets dans les patrons. Le verbe *considérer* faisant partie de la classe {*opinion*} dans notre ressource du LST, il est souvent associé aux pronoms *on* et *nous*, représentant l'auteur, dans le corpus d'experts. De la même manière que pour *montrer*, les étudiants emploient peu ces constructions, spécifiquement les étudiants chinois. La variabilité s'explique ici tant par la différence de type d'écrits (articles de recherche et mémoire) que par la différence entre scripteurs (chercheurs « publiants » et étudiants).

Verbe *décrire*

L'analyse des patrons et l'observation des concordances dans le corpus d'experts nous ont permis d'identifier un seul sens pour le verbe *décrire* : 'représenter en détail'. Ce verbe est régulièrement utilisé au passif comme dans l'exemple suivant :

– Ex 12 : *Les indicateurs de ségrégation ont été calculés selon une méthode **décrite** par Stephen Gorard et Chris Taylor (2002)*. (corpus experts : sciences de l'éducation)

patron 1 : [[humain =auteur]{processus}] décrit [[{processus}]{objet scientifique}] (comme)
Ex : <i>nous décrivons les effets</i> souvent au passif : <i>effets décrits</i>

Tableau 7 : Patron lexico-syntaxique du verbe *décrire* dans le corpus d'experts

Ce patron sélectionne un sujet humain ou un sujet métonymique du type *étude, travaux*, renvoyant à un humain :

– Ex 13 : *Des travaux en sciences de l'éducation, étudiant les pratiques enseignantes [...] **décrivent**, de façon très précise, le comportement de l'enseignant*. (corpus experts : sciences de l'éducation)

Nous observons cependant que cet emploi métonymique est peu fréquent chez les étudiants, en particulier chez les étudiants chinois.

L'analyse des concordances nous révèle également que les experts privilégient la construction au passif réduit, à l'inverse des étudiants, comme on peut le constater dans le tableau suivant :

Patrons grammaticaux	Experts	Français	Chinois
(OBJ chose) être décrit (passif)	1	2	2
(SUBJ humain) décrire (OBJ chose)	2	1	1

Tableau 8 : ordre des fréquences des patrons du verbe décrire dans les trois corpus

Les deux exemples suivants illustrent cet emploi du verbe décrire dans les corpus d'étudiants :

- Ex 14 : Et SENORE (2011 ; p60) **décrit** « la tâche » de cet enseignement comme conduisant « les élèves vers le pouvoir d'écrire ». (corpus d'étudiants français)
- Ex 15 : Dans le livre écrit par l'Abée¹⁹ Gainet que nous avons mentionné dessus, il a **décrit** les conditions de vie des Chinois ordinaires vues et racontées par un missionnaire français en Chine. (corpus d'étudiants chinois)

Notons par ailleurs que la construction il décrit + objet chose est assez récurrente chez les étudiants chinois. Les noms objets de décrire sont variés et appartiennent rarement au LST : image, scène, condition. Le verbe décrire est ainsi peu utilisé dans la description des données chez les étudiants.

Verbe *expliquer*

Concernant le verbe *expliquer*, l'analyse des cadres et la modélisation des patrons fait ressortir cinq constructions fréquentes dans le corpus d'experts. Ces patrons correspondent aux deux sens 'faire comprendre' et 'justifier et être la cause de'. Ces deux sens se distinguent par la nature de leur sujet, humain ou chose, comme illustré dans le tableau ci-après.

patron 1.1 : [humain =auteur]] explique [[{événement}]]
patron 1.2 : [humain =auteur]] explique [complétive-que/pourquoi/comment]
Ex : on explique le fait
Patron 2.1 : [[{qualité}]{relation} =cause]] explique [[{événement} =conséquence]]
patron 2.2 : [[{qualité}]{relation} =cause]] explique [complétive-que/pourquoi/comment]
patron 2.3 : [[{événement} =conséquence]] s'explique par [[{qualité}]{relation} =cause]]
Ex : ce facteur explique l'augmentation

Tableau 9 : Patrons lexico-syntaxiques du verbe expliquer dans le corpus d'experts

L'analyse contrastive des trois corpus pour les patrons de ce verbe révèle des usages bien distincts, résumés dans le tableau ci-dessous.

Patrons grammaticaux	Experts	Français	Chinois
(SUBJ chose) expliquer (OBJ chose)	1	3	3
(OBJ chose) s'expliquer par	2	-	-
(OBJ chose) expliquer que	3	-	-
(SUBJ humain) expliquer que	-	1	2

(SUBJ humain) expliquer (OBJ chose)	-	2	1
-------------------------------------	---	---	---

Tableau 10 : ordre des fréquences des patrons du verbe expliquer dans les trois corpus

Ce tableau nous montre que les experts emploient d'avantage deux types d'alternances syntaxiques : le passif et le pronominal de sens passif. De la même manière que pour les patrons du verbe *montrer*, les étudiants français favorisent les patrons avec complétive lorsque les étudiants chinois privilégient un complément d'objet nominal.

Au niveau de sujets, les étudiants emploient peu de sujets choses, le sens 'justifier et être la cause de' est ainsi très peu présent dans les corpus d'écrits d'étudiants, ce qui peut marquer, au niveau discursif, une difficulté à argumenter ou à exprimer des relations entre les éléments textuels.

Dans le cas des patrons à sujet humain les plus fréquemment utilisés par les étudiants, le sujet renvoie généralement à un autre chercheur, afin de s'appuyer sur celui-ci :

– Ex 17 : *De plus Dominique Bucheton **explique** que les élèves se rendent souvent compte que leurs ébauches ne sont pas forcément toutes justes.* (corpus d'étudiants français)

Pour identifier les différences entre étudiants chinois et français, il faut ici s'intéresser aux compléments d'objets nominaux et complétives présents dans les patrons. Ainsi, les étudiants chinois emploient des constructions incluant des noms génériques, peu spécifiques, et des tournures peu conformes au genre de l'écrit scientifique et de littéracie avancée. Les exemples suivants attestent ainsi de l'intérêt de l'élaboration d'un outil d'aide à la rédaction scientifique intégrant des informations sur les patrons selon le verbe et le sens mobilisé.

– Ex 18 : *Il faut ajouter des annotations pour leur **expliquer** ce que c'est ou ce que signifie.* (corpus d'étudiants chinois).

– Ex 19 : *Quand Simon parle du sentiment de la création de La Route des Flandres, il **explique** comme ça : toutes les choses surgissent ensemble dans mon esprit.* (corpus d'étudiants chinois).

À travers l'analyse contrastive des patrons de quatre verbes, nous avons pu identifier plusieurs différences entre les écrits d'étudiants et les écrits d'experts. Ainsi, les constructions verbales présentes dans les corpus d'étudiants se caractérisent par une sous-utilisation :

- de certaines alternances syntaxiques : passif, passif pronominal ;
- d'éléments lexicaux : mots du LST, pronom sujet *on* et *nous* ;
- de certaines constructions spécifiques : *comme nous* verbe, *si nous* verbe.

À l'inverse, les écrits étudiants, en comparaison des écrits d'experts, sont caractérisés par une sur-utilisation du pronom *il* et de noms appartenant à la langue générale. Ainsi, les étudiants, français et chinois, ont des difficultés à maîtriser certaines constructions spécifiques à l'écrit scientifique, permettant de développer l'argumentation, d'organiser le texte. Ceci peut alors se traduire par des difficultés à établir une structure discursive cohérente, construire des opinions au sujet des savoirs d'une discipline. Ces difficultés sont encore plus grandes pour les étudiants chinois, du fait de leur statut d'apprenant du français, et aboutissent à des confusions de registre, erreurs de collocations, expressions/ tournures maladroites.

Réflexions pour l'enseignement/apprentissage des patrons verbaux en classe du FLE

Afin de remédier aux difficultés des étudiants relevées par l'étude contrastive, nous proposons un enseignement des patrons verbaux, en classe de FLE, basé sur corpus et privilégiant l'approche inductive.

L'utilisation du corpus dans l'apprentissage répond aux principes du Data-Driven Learning (DDL) (Johns, 1991) et à l'instar de Chambers et al. (2011) et de Cavalla & Loiseau (2014), nous pensons que l'enseignement/apprentissage du lexique peut être facilité par la mise en contexte et l'analyse des concordances, tant dans la phase de la découverte des patrons que dans la phase de production écrite.

L'approche inductive est largement utilisée dans l'enseignement de la grammaire (Chartrand, 1996), du lexique (Cavalla, 2014) et dans l'acquisition langagière (Colletta, 2004). Cette approche vise à construire l'autonomie et la « posture réflexive » pour l'apprenant (Rinck, 2010 ; Boch & Buson, 2012).

L'étape de transposition didactique des patrons verbaux doit suivre une progression allant de la découverte à la réutilisation en passant par le renforcement et l'approfondissement des connaissances pour l'apprenant. De plus, l'utilisation du corpus doit être guidée par l'enseignant pour éviter de confronter les apprenants à des exemples trop nombreux et/ou complexes.

Nous présentons dans le tableau 11 un exemple de concordances (extraites du corpus d'experts) du verbe *considérer*. Ces concordances permettent d'appuyer l'apprentissage sur des emplois attestés, en contexte.

1. Dans la préface à l'édition des lettres de Jean Hus de 1537, Luther	<i>considère</i>	<i>le réformateur de la Bohême comme un « vrai chrétien », qui a produit de bons fruits .</i>
2. Puisque la nature ordinaire regroupe les milieux et espèces les plus communs et abondants , on peut	<i>considérer</i>	<i>que c' est bien celle-ci qui fournit le plus de ressources à l' humanité .</i>
3. Il convient néanmoins de	<i>considérer</i>	<i>le rôle didactique du théâtre et le contexte mouvant des années soixante avant de conclure à une provocation outrancière .</i>
4. Deux points sont à	<i>considérer</i>	<i>: d'une part , le sens qu' exprime une préposition sous les différentes formes qu' elle est capable de prendre , d'autre part...</i>
5. Et de l' autre , un échange tellement indirect qu' on peine à voir comment le principe de réciprocité est respecté.	<i>Considérons</i>	<i>tout d'abord la première forme d' échange , qui équivaut au mariage avec la cousine croisée patrilatérale ,</i>

Tableau 11 : concordances du verbe *considérer*

L'apprenant peut ainsi observer les différents aspects des constructions, guidé par les exemples validés par l'enseignant : associations lexicales, alternances syntaxiques (passif,

impératif), acception mobilisée, fonction discursive (topicalisation, discours rapporté), constructions spécifiques (*il convient de considérer, on peut considérer*), etc.

Les exemples facilitent ainsi la compréhension et l'appropriation des patrons associés à un sens donné du verbe, et autorisent par exemple la mise en place d'exercices de reformulation, de mise en correspondance de verbes synonymes, pour guider l'apprentissage des constructions verbales.

En associant l'observation des concordances à l'étude des patrons, l'enseignant peut ainsi orienter l'apprenant dans la compréhension et l'appropriation de ces constructions complexes.

Perspectives et conclusion

L'extraction semi-automatique des patrons lexico-syntaxiques des verbes du LST nous a permis d'identifier les constructions les plus fréquentes pour plusieurs types de scripteurs. L'analyse contrastive des patrons a mis en évidence les spécificités des étudiants par rapport aux experts et les points de convergence et de divergence entre écrits d'étudiants français et chinois. Notre hypothèse d'une variété des constructions supérieure dans le corpus d'experts a pu être confirmée par les résultats. Certaines différences doivent cependant être confrontées à l'hétérogénéité des corpus. Nous avons en effet pu constater que les sous-genres de l'article de recherche et du mémoire ne sont pas entièrement comparables, notamment au niveau de la structure textuelle. De plus, la thématique des écrits (spécifiquement pour les mémoires) a été identifiée comme facteur de variabilité important dans l'emploi des patrons verbaux et dans les acceptions mobilisées pour chaque verbe étudié. Ainsi, le patron « l'enseignant explique que » est surreprésenté dans les mémoires de didactique de l'écrit lorsque les constructions du type « l'écrivain décrit la scène » sont surreprésentées dans les mémoires en littérature/traduction.

L'analyse contrastive montre également que les difficultés liées aux fonctions discursives sont largement partagées par les étudiants natifs et non-natifs, contrairement à notre intuition de départ. La variété des constructions syntaxiques est ainsi moindre pour les étudiants par rapport au corpus d'experts, notamment pour les constructions syntaxiques typiques de l'écrit scientifique, ayant une fonction argumentative ou d'organisation textuelle. Une différence importante entre scripteurs étudiants et experts se situe au niveau lexical des patrons. Ainsi, les compléments sujets des patrons d'écrits étudiants sont peu marqués par la présence de l'auteur, ce qui peut dénoter une dimension énonciative peu présente et un rejet de positionnement. Nous avons également pu observer certaines variations entre le corpus d'étudiants chinois et français, avec par exemple la différence marquée entre l'emploi de compléments d'objet (généralement des noms de la langue générale pour les premiers versus des complétives pour les seconds).

La non-maîtrise de certains patrons dans les corpus d'étudiants est combinée à une surreprésentation du sens concret du verbe. Pourtant, d'après l'observation du corpus d'experts, le sens par extension/abstrait est le plus employé dans l'écrit scientifique. Ceci rejoint notre constat sur l'importance de la maîtrise du LST et de l'emploi des patrons verbaux de ce lexique. Nous remarquons également que certaines lacunes (notamment lexicales) propres aux étudiants non-natifs leur rendent la maîtrise des patrons encore plus complexe.

Sur le plan didactique, nous proposons des pistes de réflexions pour enseigner les patrons verbaux, à travers l'approche inductive et en se basant sur les corpus, afin d'améliorer l'enseignement/apprentissage de ces constructions essentielles dans l'écrit scientifique.

BIBLIOGRAPHIE

- Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational linguistics*, 19(2), 219-241.
- Blumenthal, P. (2008). Combinatoire des prépositions: approche quantitative. *Langue française*, 157(1), 37. <http://doi.org/10.3917/lf.157.0037> [25/02/2016]
- Boch, F., Buson, L. 2012. « Orthographe & grammaire à l'université. Quels besoins ? Quelles démarches pédagogiques ? ». *Scripta*, vol. 16, n°30, p. 31-51.
- Cavalla, C. (2014). Collocations transdisciplinaires d'apprentissage : réflexions pour l'enseignement. In : González-Rey Isabel (éd.) : *Outils et méthodes d'apprentissage en phraséodidactique* : 151-169. Éditions Modulaires Européennes.
- Chartrand, S.-G. 1996. « Apprendre la grammaire par la démarche active de découverte ». In : *Pour un nouvel enseignement de la grammaire*, n° 2, p. 19-225.
- Colletta, J.-M. 2004. *Le développement de la parole chez l'enfant âgé de 6 et 11 ans : corps, langage et cognition*. Sprimont: Editions Mardaga.
- Drouin, P. (2007). Identification automatique du lexique scientifique transdisciplinaire. *Revue française de linguistique appliquée*, vol. XII(2), 4564.
- Dubois, J., & Dubois-Charlier, F. (1997). *Les verbes français*. Larousse.
- Dubois, J., & Dubois-Charlier, F. (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.
- Flaux, N., & Van de Velde, D. (2000). *Les noms en français : esquisse de classement*. Éditions Ophrys.
- Hanks, P. (2013). *Lexical Analysis : Norms and Exploitations*. Mit Press.
- Kupsc, A. (2007). Extraction automatique de cadres de sous-catégorisation verbale pour le français à partir d'un corpus arboré. *TALN 2007*.
- Messiant, C., Gábor, K., & Poibeau, T. (2010). Acquisition de connaissances lexicales à partir de corpus : la sous-catégorisation verbale en français. *Traitement automatique des langues*, 51(1), 65-96.
- Nazar, R., & Renau, I. (à paraître). A quantitative analysis of the semantics of verb-argument structures.
- Paquot, M. (2010). *Academic vocabulary in learner writing : From extraction to analysis*. Bloomsbury Publishing.
- Paquot, M., & Bestgen, Y. (2009). Distinctive words in academic writing : A comparison of three statistical tests for keyword extraction. *Language and Computers*, 68(1), 247269.

- Renau, I., & Paz, B. (2011). The Spanish Learner's Dictionary DAELE on the Panorama of the Spanish E-lexicography. In : I., Kosem & K., Kosem (eds.) : *Electronic lexicography in the 21st century : new applications for new users : proceedings of eLex 2011*, 10-12 November 2011, Bled, Slovenia, Ljubljana : Trojina, Institute for Applied Slovene Studies, 221-226.
- Rinck, F. 2010. Réflexivité et écrits de recherche. Propositions pour une formation universitaire à et par la littéracie. Dans *De la maîtrise du français aux littéracies dans l'enseignement supérieur (Actes des journées d'études du Centre de Méthodologie Universitaire " CMU 2010 ")* (p. 79-91). Belgique.
- Rinck, F. (2013). L'article de recherche en sciences humaines: approche en corpus. In *Séminaire des Aspects concrets de la thèse, EHESS*. Paris, France. Consulté à l'adresse <http://halshs.archives-ouvertes.fr/halshs-00998105> [25/02/2016]
- Sinclair, J. (1991). *Corpus, concordance, collocation* (Vol. 1). Oxford University Press Oxford.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing.
- Tutin, A. (2007). Autour du lexique et de la phraséologie des écrits scientifiques. *Revue française de linguistique appliquée*, vol. XII(2), 55.
- Tutin, A., & Grossmann, F. (Éd.). (2014). *L'Écrit scientifique : du lexique au discours. Autour de Scientext*. Presses universitaires de Rennes.
- Williams, G., & Million, C. & Alonso, A. (2012). Growing naturally : The DicSci Organic E-Advanced Learner's Dictionary of Verbs in Science. *Proceedings of the 15th EURALEX International Congress*. University of Oslo.

NOTES

1. « verbs that serve specific rhetorical or organizational functions in academic prose generally enter compositional and flexible sequences. »
2. Hatier, S., Augustyn, M., Tran, T.T.H., Yan, R., Tutin, A., & Jacques, M-P. (2015). French cross-disciplinary scientific lexicon : extraction and linguistic analysis – article soumis
3. Nous avons pour une étude précédente constitué un corpus de grande taille du français (120 millions de mots) nous servant de corpus de contraste pour le calcul de spécificité
4. Dubey Gérard, « Nouvelles techniques d'identification, nouveaux pouvoirs. Le cas de la biométrie », *Cahiers internationaux de sociologie* 2/2008 (n° 125), p. 263-279
5. Les profils combinatoires sont, selon Blumenthal (2008), « l'image que donne du comportement d'un mot de base l'ensemble de ses collocatifs ».
6. Correspondant à l'acception 6 selon le *Dictionnaire Électronique des Mots* (Dubois & Dubois-Charlier, 2010), consulté le 11/01/2016 à l'adresse : <http://rali.iro.umontreal.ca/LVF+1/alphabetique/P.html#presenter>
7. Wolff, M. (2003). 7. Apports de l'analyse géométrique des données pour la modélisation de l'activité. *Le Travail humain*, 1, 195-227.
8. Mémoires de 60 pages environ, hors annexes.
9. Corpus constitué par Rinck, F. Boch, F., & Jacques, M-P, disponible à l'adresse suivante : <http://lidilem.u-grenoble3.fr/ressources/corpus-du-labo/article/corpus-litteracie-avancee> (consulté le 11/01/2016).
10. Dont un descriptif est consultable à l'adresse suivante : <http://scientext.msh-alpes.fr/>
11. Anthropologie, économie, géographie, histoire, linguistique, psychologie, sciences de l'éducation, sciences politiques, sciences de l'information, sociologie
12. Ait-Mokthar, S., & Chanod, J. P. (2002). Robustness beyond Shallowness : Incremental Dependency Parsing. *Special issue of NLE Journal*.

13. [u3_advmod], [u3_de_vmod] et [subj] réfèrent aux étiquettes de relations de dépendances issues de l'analyseur et de nos grammaires locales.
 14. À l'aide du calcul de rapport de vraisemblance.
 15. Un patron est une déclaration de la structure de proposition (valence) associée à un sens d'un verbe, avec les valeurs sémantiques typiques de chaque argument qui se réalisent par des cooccurrences saillantes [...] Les différentes valeurs sémantiques des arguments activent les différents sens de chaque verbe.
 16. <http://www.pdev.org.uk/#browse?q=f=C>. Ce dictionnaire est encore en cours de construction, 1 348 verbes en anglais y sont répertoriés et l'objectif est d'atteindre 5 396 verbes.
 17. <http://www.iula.upf.edu/rec/daele/>
 18. <http://olivier.kraif.u-grenoble3.fr/anaText/index.php>
 19. Nous conservons dans les exemples les graphies d'origine présentes dans les corpus
-

RÉSUMÉS

Nous présentons dans cet article une étude des patrons verbaux employés par des scripteurs étudiants (apprenants du français et natifs) et par des scripteurs experts (chercheurs). Nous nous intéressons ici à l'emploi de verbes fréquents dans l'écrit scientifique (*montrer, considérer, décrire, expliquer*) par les scripteurs étudiants. Nous analysons, pour ces derniers, les constructions verbales utilisées préférentiellement ainsi que celles qui leur posent le plus de problèmes. Pour ce faire, en nous basant sur des corpus analysés en dépendances, nous extrayons semi-automatiquement les patrons verbaux les plus fréquents et analysons les points de convergence et de divergence entre les différents types de scripteurs.

This study focuses on the comparison between verbal patterns used in academic writing by students (native and French learners) and expert writers (researchers). We aim at identifying how students use verbs that frequently appear in academic writing (*montrer / show, considérer / consider, décrire / describe, expliquer / explain*). Our objective is to pinpoint verbal constructions preferentially used by students and verbal constructions they do not master. Using syntactically parsed corpora of scientific papers and dissertations, we collect and modelize lexico-syntactic patterns for investigated verbs. Thus we analyze shared and specific verbal patterns between student and expert writers.

INDEX

Mots-clés : écrit scientifique, sous-catégorisation verbale, linguistique de corpus, patron lexico-syntaxique

Keywords : academic writing, verbal subcategorization, corpus linguistics, lexico-syntactic pattern

AUTEURS

SYLVAIN HATIER

LIDILEM, Univ. Grenoble Alpes
Sylvain.Hatier@u-grenoble3.fr

RUI YAN

LIDILEM, Univ. Grenoble Alpes
Rui.Yan@u-grenoble3.fr