



HAL
open science

Routines verbales pour les Français Langue Étrangère : des corpus d'experts aux corpus d'apprenants

Rui Yan, Agnès Tutin, Thi Thu Hoai Tran

► To cite this version:

Rui Yan, Agnès Tutin, Thi Thu Hoai Tran. Routines verbales pour les Français Langue Étrangère : des corpus d'experts aux corpus d'apprenants. *LIDIL - Revue de linguistique et de didactique des langues*, 2018, 58. hal-01927538

HAL Id: hal-01927538

<https://hal.science/hal-01927538>

Submitted on 19 Nov 2018

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.

Copyright



Lidil

Revue de linguistique et de didactique des langues

58 | 2018

**L'enseignement et l'apprentissage de l'écrit
académique à l'aide de corpus numériques**

Routines verbales pour les français langue étrangère : des corpus d'experts aux corpus d'apprenants

Verb Routines in L2 French: From Expert to Learner Corpora

Rui Yan, Agnès Tutin et Thi Thu Hoai Tran



Édition électronique

URL : <http://journals.openedition.org/lidil/5411>

ISSN : 1960-6052

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

ISBN : 978-2-37747-064-8

ISSN : 1146-6480

Référence électronique

Rui Yan, Agnès Tutin et Thi Thu Hoai Tran, « Routines verbales pour les français langue étrangère : des corpus d'experts aux corpus d'apprenants », *Lidil* [En ligne], 58 | 2018, mis en ligne le 02 novembre 2018, consulté le 07 novembre 2018. URL : <http://journals.openedition.org/lidil/5411>

Ce document a été généré automatiquement le 7 novembre 2018.

© Lidil

Routines verbales pour les français langue étrangère : des corpus d'experts aux corpus d'apprenants

Verb Routines in L2 French: From Expert to Learner Corpora

Rui Yan, Agnès Tutin et Thi Thu Hoai Tran

1. Introduction

- 1 Dans cet article, nous nous intéressons aux routines sémantico-rhétoriques, des phénomènes phraséologiques au croisement de plusieurs domaines : l'analyse du discours, la phraséologie et la syntaxe. Ces phénomènes relèvent de la « phraséologie étendue » (Legallois & Tutin, 2013) et renvoient souvent dans l'écrit scientifique et académique au métadiscours et au métatexte (Tutin & Grossmann, 2014). Ces routines sont définies comme « des énoncés récurrents construits autour d'un verbe et d'arguments remplissant différents rôles sémantiques dont une partie est actualisée dans le discours. Chaque routine peut être associée à une fonction discursive/ rhétorique spécifique » (Tutin & Kraif, 2016, p. 121).
- 2 Les fonctions que les routines remplissent sont au cœur de l'énonciation et de l'argumentation dans l'écrit académique : formuler une hypothèse (*on peut supposer que*), établir une relation entre une cause et un effet (*cette différence s'explique par*), exprimer une prise de position (*on s'accorde à/sur*), appuyer un argument par une référence à autrui (*comme le souligne X*), ou justifier le bienfondé de son interprétation (*ce résultat permet d'affirmer que*). La maîtrise de ces routines, dont les réalisations lexicales sont assez diversifiées, permet de faciliter l'intégration des scripteurs dans la « communauté de discours » des chercheurs (Tutin, 2014). L'étude de ces phénomènes linguistiques est intéressante à plusieurs titres. Elle permet, d'une part, de mettre en évidence les stratégies rhétoriques mises en œuvre par les auteurs et de mieux comprendre les démarches scientifiques élaborées (éléments d'argumentation, critères de scientificité). D'autre part, elle apparaît tout à fait pertinente dans une démarche d'enseignement de

l'écrit académique en langue maternelle ou en langue étrangère, ces éléments phraséologiques et leur fonction étant souvent mal maîtrisés par les apprentis scripteurs (González-Rey, 2007).

- 3 Les apprenants du français langue étrangère (ci-après FLE) utilisent-ils à bon escient les routines dans leurs écrits ? Quelles sont les difficultés rencontrées avec ce phénomène ? Ces préoccupations nous ont conduites à nous intéresser à ces questions et à proposer un diagnostic de la maîtrise de ces expressions par les apprenants allophones, dans la perspective d'élaborer des matériaux didactiques adaptés à leurs besoins.
- 4 Nous décrivons en premier lieu l'objet de l'étude, les routines dans l'écrit académique, à travers un ensemble de routines extraites d'un corpus d'écrits scientifiques et modélisées dans le cadre de la thèse de Yan (2017). En second lieu, nous établirons un diagnostic des difficultés en deux temps : d'une part, à travers l'analyse des erreurs d'emploi liées aux routines verbales dans un corpus d'apprenants chinois (corpus Sup-Chinois) ; d'autre part, à travers la comparaison entre le corpus Sup-Chinois, un corpus de novices français (corpus Sup-Natifs) et le corpus d'experts francophones Transdisciplinaire-TermITH, de façon à mettre en évidence les cas de sous-emploi des routines verbales chez les étudiants chinois par rapport aux natifs. Enfin, dans la perspective de remédier aux difficultés observées, nous proposons deux types de ressources pédagogiques destinées à la rédaction académique pour étudiants allophones, une ressource lexicale intégrant des informations sur les routines et des activités d'entraînement.

2. Une ressource de routines verbales pour l'écrit académique

- 5 Dans le cadre de sa thèse, Yan (2017) a constitué une ressource lexicale intégrant les patrons verbaux et les routines verbales, permettant la réalisation d'activités d'apprentissage dans la perspective de l'aide à la rédaction scientifique. À partir du corpus Transdisciplinaire-TermITH (5 millions de mots), environ 50 verbes du lexique scientifique transdisciplinaire (LST) fréquents ont été traités (*montrer, analyser, considérer, expliquer, proposer, supposer, etc.*). Plus de 100 patrons verbaux et routines verbales, spécifiques à l'écrit scientifique, y ont été codés.
- 6 Inscrite dans la linguistique de corpus, l'approche adoptée est fondée sur le modèle *Corpus Pattern Analysis* (CPA, analyse des patrons basée sur les corpus) (Hanks, 2013). Ce modèle visant à analyser les usages réguliers des verbes en contexte, permet d'accéder aux sens du verbe à travers l'analyse des constructions syntaxiques et des co-occurrences qui lui sont associées. Cependant, ce modèle est initialement destiné à la description des verbes anglais de la langue générale. L'étude de Yan se distingue par un objectif plus spécifique, soit l'analyse des emplois usuels du verbe dans le domaine de l'écrit scientifique avec une visée didactique. La méthode d'analyse a combiné des techniques de traitement automatique du langage (TAL) et l'analyse linguistique manuelle. Elle a été appliquée au corpus Transdisciplinaire-TermITH qui se compose de 500 articles représentant 10 disciplines des sciences humaines et sociales (Hatier et coll., 2016). L'extraction et la modélisation des routines verbales ont été appliquées à une liste des verbes établie par Hatier (2016) dans le cadre du projet TermITH¹, selon les étapes suivantes :
 1. Extraction des cadres de sous-catégorisation : pour chaque occurrence verbale, les cadres de sous-catégorisation² sont extraits automatiquement par Hatier à l'aide du corpus analysé en

dépendances syntaxiques et regroupées de façon à faire émerger les constructions syntaxiques les plus fréquentes.

2. Repérage des acceptions : à partir des cadres extraits, Yan repère les acceptions des verbes du LST qui apparaissent dans le corpus Transdisciplinaire-TermITH en se basant autant que possible sur la ressource lexicographique *Les verbes du français* (ci-après LVF) (Dubois & Dubois-Charlier, 1997).
 3. Classification des verbes et traitement des routines : après avoir relié cadres de sous-catégorisation et acception, Yan propose une classification des verbes du LST en fonction de leurs propriétés sémantiques et syntaxiques et procède à la modélisation des routines verbales par classe de verbes.
- 7 Considérons les exemples (1, 2, 3) du verbe *étudier* dans les phrases suivantes :
- (1)
En revanche avec les bébés âgés de 4 mois, il nous paraît **intéressant d'étudier** non seulement le délai court de 30 s mais également la mémoire à long terme après 2 min de délai. (psychologie)
 - (2)
Pour obtenir un transfert de compétence en sa faveur, la Commission s'empare de ces analyses. L'exemple de la politique vitivinicole **permet ainsi d'étudier le rôle** des chercheurs dans la légitimation des réformes européennes. (sciences politiques)
 - (3)
Mais la connaissance des élites politiques à l'aide d'une approche relationnelle **est peu étudiée** malgré une littérature de qualité sur le thème plus général des élites politiques africaines que l'on peut regrouper en deux modèles explicatifs. (sociologie)
- 8 Le corpus analysé en dépendances syntaxiques nous permet d'obtenir les cadres de sous-catégorisation suivants :
- a. (~ADJMOD)³(~de VIN F)étudier_VERB(OBJ -hum) (10 occurrences)
 - b. (~VerbSup)étudier_VERB(OBJ -hum) (66 occurrences)
 - c. (OBJ -hum)(PASSIF)étudier_VERB (146 occurrences)
- 9 Dans le premier cadre, le verbe *étudier* prend pour objet un nom inanimé et est précédé d'un adjectif et de la préposition *de* qui introduit le verbe à l'infinitif. Dans le deuxième cadre, le code « VerbSup » représente les verbes supports. Ici, le verbe est précédé d'un verbe modal (*permettre, devoir, pouvoir, etc.*), comme le montre l'exemple (2). Quant au troisième cadre, le verbe est utilisé au passif, construction assez fréquente dans le corpus. À l'issue de l'extraction des cadres, les constructions syntaxiques fréquentes associées à chaque verbe du LST ont été identifiées.
- 10 Ensuite, le repérage des « acceptions » constitue l'étape de base de notre modélisation des routines, les verbes transdisciplinaires étant assez polysémiques. Nous mettons ainsi en correspondance un cadre et une acception présente dans notre corpus. Pour qu'une acception soit retenue et considérée comme transdisciplinaire dans le corpus d'analyse, elle doit apparaître au moins 20 fois dans au moins 3 des 10 disciplines. Ce travail est fondé sur la ressource lexicographique LVF⁴ (Dubois & Dubois-Charlier, 1997) — une référence linguistique pour l'étude des verbes français — qui se caractérise par sa large couverture, non seulement en ce qui concerne le nombre de verbes français répertoriés, mais aussi pour l'exhaustivité des propriétés sémantiques et syntaxiques consignées. Par exemple, dans les trois cadres de sous-catégorisation présentés plus haut, ils ont un même sens. Ce sens mobilisé correspond à l'entrée 5 du verbe *étudier* dans le LVF, dont la définition est « analyser, évaluer » (par exemple, *on étudie un projet*).

- 11 À l'issue du repérage manuel des acceptions des verbes du LST, nous proposons une classification sémantique des verbes. Constituée à la manière de Dubois et Dubois-Charlier (1997), cette classification repose sur une typologie à deux niveaux en classes et en sous-classes, qui regroupent des verbes ayant des propriétés syntaxiques et sémantiques similaires. Le verbe *étudier*, par exemple, appartient à la sous-classe « #examen » qui signifie « étudier en détail, faire l'analyse de quelque chose pour en avoir une connaissance approfondie » et apparaît soit dans une construction avec un sujet humain et un objet inanimé, soit dans une construction au passif. Cette sous-classe « #examen » comporte aussi les verbes *examiner*, *analyser* et *explorer*.
- 12 Une fois les classes des verbes établies, nous observons les cooccurrences lexicales les plus significatives de chaque verbe à l'aide de l'outil Lexicoscope⁵ (Kraif & Diwersy, 2012) en utilisant un seuil de 5 occurrences. Ces cooccurrences sont ensuite reliées aux cadres de sous-catégorisations extraits, ce qui permet de faire émerger des routines verbales. Pour la sous-classe « #examen », on relève ainsi une série de routines intéressantes (cf. tableau 1), assez spécifiques au genre de l'écrit scientifique. L'observation du corpus permet de révéler leurs fonctions rhétoriques liées aux contextes caractéristiques d'argumentation ou d'énonciation dans le discours scientifique.

Tableau 1. – Exemples des routines verbales – verbes d'examen repérées dans les écrits scientifiques d'experts (corpus Transdisciplinaire-TermITH).

Routine verbale et nombre d'occurrences dans le corpus Transdisciplinaire-TermITH	Fonctions
1. <nous étudions/examinons/analysons {d'abord/ensuite/enfin}> (21 occurrences)	annoncer le plan d'organisation
2. <si l'on examine/analyse> (19 occurrences)	impliquer le lecteur dans l'activité d'examen pour en faire un témoin
3. <consiste à étudier/examiner/analyser> (12 occurrences)	introduire l'objectif de l'étude
4. <il est {intéressant/important} d'étudier/examiner/analyser> (10 occurrences)	attirer l'attention du lecteur et mettre en évidence un fait saillant
5. <il {convient/importe} d'étudier> (18 occurrences)	attirer l'attention du lecteur et exprimer la nécessité de l'étude
6. <il s'agit d'étudier/examiner/analyser> (22 occurrences)	introduire et mettre en avant l'objet d'étude
7. <{analyse/modèle/exemple} permet d'étudier/examiner/analyser> (35 occurrences)	établir une relation causale en mettant en évidence le bienfondé de l'étude

- 13 Comme on peut le constater, dans la première routine, les verbes d'examen prennent comme sujet le pronom *nous* qui peut renvoyer à l'auteur (singulier ou collectif). Ici, les

routines se rapportent souvent au processus d'écriture, comme l'illustre l'exemple suivant (4) :

(4)

Nous examinerons ensuite comment l'identité métisse est une identité conflictuelle résultant de ces processus contradictoires d'affiliation qui placent les individus à la frontière des groupes ethniques. (anthropologie_15216)

- 14 La deuxième routine (exemple 5) intègre le lecteur par le biais du pronom inclusif *on* et est utilisée pour l'inviter comme témoin dans l'activité d'examen, de façon à rendre le discours plus convaincant.

(5)

Pourtant, si l'on examine attentivement cette table de scolarité située dans l'annexe de l'ouvrage « Population » et l'enseignement [...] et partiellement reproduite ci-dessous (cf. le tableau 1), on note le résultat contraire. (scienceseducation_3044)

- 15 Quant à la troisième routine (exemple 6), comportant des verbes d'examen à l'infinitif introduits par le verbe *consister* et la préposition *à*, elle apparaît souvent dans l'introduction de l'article, pour introduire l'objectif de l'étude.

(6)

Au-delà de la simple description, notre travail consiste, dans le cadre d'une approche ethnographique, à analyser les stratégies par lesquelles un enseignant expérimenté instaure les règles de la vie scolaire et les effets de son action au plan didactique. (scienceseducation_9557)

- 16 Les trois routines (routines 4-6) comportant le pronom impersonnel *il* servent à exprimer l'intérêt ou la nécessité de l'étude. La quatrième routine est introduite par un adjectif de « saillance » (*important, intéressant, utile, etc.*), comme dans l'exemple (7) suivant :

(7)

Pour avoir une chance de démêler les implications religieuses dans les diverses agressions intervenues entre les orthodoxes et les uniates, il est important d'étudier le langage de la violence manifestée lors de ces épisodes. (histoire_18593)

- 17 Enfin, la routine causative (exemple 8) est fréquemment introduite par le verbe modal *permettre*.

(8)

Ces données permettront d'examiner la propension des systèmes éducatifs européens à regrouper dans les mêmes écoles des élèves faibles ou défavorisés. (scienceseducation_57)

- 18 L'extraction automatique des cadres de sous-catégorisation, associée à l'observation des contextes textuels et à l'analyse linguistique, permet ainsi de modéliser les routines verbales et de mettre en évidence les fonctions rhétoriques spécifiques au genre. Dans une perspective didactique, nous examinerons dans la partie suivante l'emploi des routines verbales dans les écrits des étudiants allophones.

3. Les routines verbales dans les écrits des étudiants allophones : vers un diagnostic des difficultés

- 19 Dans le domaine de l'écrit académique, de nombreuses études ont montré que les apprenants d'une langue étrangère peinent à manier les éléments phraséologiques d'une manière fluide et experte (Granger & Paquot, 2009). Différents types d'éléments phraséologiques ont fait l'objet de ces études, en particulier les collocations (Nesselhauf,

2005 ; Cavalla, 2015) et les séquences récurrentes (Lake, 2004 ; Paquot, 2010). L'apparition des corpus d'apprenants (Granger et coll., 2015), en particulier pour l'anglais, a facilité les recherches sur l'écrit académique en permettant, d'une part de relever les erreurs dans les productions des apprenants, et d'autre part de « cibler des éléments de langue sous- ou surexploités (comparés aux productions de natifs ou d'autres apprenants) » (Boulton et coll., 2013, p. 24). Néanmoins, faute de réels corpus d'apprenants en FLE, les études sur l'utilisation des phénomènes phraséologiques des apprenants du français sont beaucoup moins nombreuses, hormis quelques études récentes effectuées autour du projet Scientext : erreurs liées à l'utilisation des collocations verbales (ex. : *mettre comme hypothèse*) (Cavalla, 2015), sous-utilisation des marqueurs discursifs (ex. : *en d'autres termes*) (Tran, 2014) et des constructions verbales (ex. : *ceci s'explique par*) (Hatier & Yan, 2017). À notre connaissance, l'analyse des difficultés des étudiants allophones en matière de routines verbales n'est pas encore abordée.

- 20 Dans la suite de cet article, nous présenterons les corpus utilisés dans le diagnostic puis le diagnostic des difficultés en deux volets : le repérage des erreurs et les comparaisons quantitative et qualitative entre les productions des locuteurs natifs et non natifs.

3.1. Les corpus utilisés

- 21 Pour établir un diagnostic lié à l'utilisation des routines verbales chez les apprenants chinois par rapport aux scripteurs natifs (novices français et experts), trois corpus sont utilisés, comme indiqué dans le tableau 2 :

Tableau 2. – Les trois corpus utilisés dans la comparaison.

Corpus	Taille	Annotation	Genre	Public	Discipline
Transdisciplinaire-TermITH	5 000 000 mots	annotation morphosyntaxique, annotation syntaxique en dépendances	500 articles de recherche en sciences humaines	experts francophones	10 disciplines sciences humaines et sociales
Sup-Chinois	600 000 mots	annotation morphosyntaxique	29 mémoires de recherche en formation en M1	apprenants chinois	FLE
Sup-Natifs	460 000 mots	annotation morphosyntaxique	41 mémoires de recherche en formation en M1 et M2	étudiants français	didactique du FLE

- 22 Dans le cadre de cette étude, nous avons exploité un corpus d'apprenants, appelé Sup-Chinois, qui se compose de 28 mémoires de masterisants chinois en spécialité de français (environ 600 000 mots), dont le sujet d'étude porte sur la traduction (9), la littérature (12), la linguistique (1) ou des thématiques socioculturelles (6). Nous appuyant sur ce corpus, nous souhaitons relever dans un premier temps les difficultés liées à l'emploi des routines verbales. Le corpus d'experts, corpus Transdisciplinaire-TermITH (Hatier et coll., 2016), d'une taille plus conséquente, permet de relever les routines verbales les plus significatives, comme vu dans la section précédente. Ce corpus constitue en quelque sorte un modèle d'experts que les scripteurs novices souhaitent atteindre. Ensuite, à l'instar de Gilquin et Paquot (2008), nous pensons que, par rapport aux corpus d'experts, les écrits des étudiants natifs fournissent des données mieux adaptées pour cibler les difficultés des apprenants non natifs dans la mesure où ils permettent d'effectuer une comparaison d'apprenants de niveau d'étude équivalent. Nous avons donc aussi utilisé un corpus d'étudiants natifs, le corpus Sup-Natifs⁶, composé de 41 mémoires d'étudiants de master (1 et 2) en spécialité didactique du français. Étant plus proche du corpus Sup-Chinois, notamment aux niveaux de la taille, du genre et du type de texte, il servira donc de référence pour la comparaison.
- 23 À l'aide de ces corpus, nous avons repéré les routines récurrentes chez les experts, et observé par la suite leur fréquence et leur diversité chez les apprenants chinois et les novices français. Nous passons maintenant au repérage des erreurs dans le corpus Sup-Chinois.

3.2. Repérage des erreurs liées aux routines verbales

- 24 Le repérage des erreurs s'est fait manuellement par Yan (2017). Ce travail a été réalisé sur 10 verbes du LST (*choisir, considérer, comparer, décrire, étudier, expliquer, montrer, noter, proposer, supposer*) qui représentent au total 1 593 occurrences dans le corpus Sup-Chinois. Les verbes retenus sont fréquents et présentent des constructions/fonctions importantes et variées dans l'écrit scientifique. Dans sa thèse, Yan a relevé 221 erreurs sur 1 593 occurrences dans l'emploi de ces 10 verbes (les erreurs au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la syntaxe, au niveau du registre). Parmi les erreurs relevées, nous nous concentrons ici sur les erreurs liées aux routines verbales que l'on peut regrouper en deux grands types : les erreurs au niveau de la syntaxe et du lexique, les erreurs au niveau du registre.

3.2.1. Les erreurs au niveau de la syntaxe et du lexique

- 25 Parmi les 221 cas d'erreurs recensés par Yan (2017), 75 erreurs relèvent des structures syntaxiques erronées, soit 34 % des cas. Il s'agit de l'omission d'un élément dans la phrase (les exemples 9 et 10), de structures déformées ou inexistantes, ou encore de formulations maladroites.

(9)

Comme explique l'auteur : « l'élément psychologique, comme l'élément pictural, se libère insensiblement de l'objet avec lequel il faisait corps... » (littérature_5)
(Correction proposée : comme l'explique l'auteur)

(10)

Le traducteur transmet l'intention de l'auteur original (**Il est à noter ici**, il ne

comprend que l'intention reçue par le traducteur après la communication avec l'auteur original...) (traduction_8) (structure erronée : il est à noter qu'ici)

- 26 Les cas d'omission peuvent être attribuables soit à la négligence des apprenants soit à une maîtrise insuffisante des structures syntaxiques. Dans l'exemple 9, il manque le pronom *le* devant le verbe *expliquer*, qui renvoie à la citation entre guillemets. Dans l'exemple 10, l'apprenant aurait dû ajouter la conjonction de subordination *que* pour introduire la proposition complétive.
- 27 Dans l'exemple 11, le sujet est absent. La construction *pousser à étudier et analyser* est rarement utilisée dans l'écrit scientifique pour établir une relation de causalité. Il apparaît qu'ici l'apprenant manque de moyens linguistiques pour justifier le choix de son sujet portant sur l'étude du passé simple et de l'imparfait.

(11)

Après avoir lit la lecture de quelques livres linguistiques de Gosselin, Touratier, Bres, etc., **nous pousse à étudier et analyser** le passé simple et l'imparfait. (linguistique_1) (Correction proposée : La lecture des ouvrages de [...] nous amène à nous intéresser à l'étude et à l'analyse du passé simple et de l'imparfait.)

- 28 Par ailleurs, au-delà des problèmes d'ordre syntaxique, la maîtrise insuffisante des routines verbales dans l'écrit scientifique conduit souvent l'apprenant à construire des formulations maladroites, comme le montrent les exemples 12 et 13 :

(12)

Se pose alors la question de savoir, **comme le considère Gosselin**, si le caractère essentiel de l'imparfait est bien d'être un temps grammatical du passé ou au contraire... (linguistique_1) (structure erronée : comme s'interroge Gosselin)

(13)

Il est hors de doute que Meschonnic considère la poétique comme une théorie plus profonde et plus complète que la linguistique. (Correction proposée : Meschonnic considère...) (traduction_1)

- 29 Ces derniers exemples peuvent être considérés comme des maladroites puisqu'ils ne sont pas strictement erronés du point de vue linguistique. Ils montrent que les apprenants ne disposent pas de moyens linguistiques suffisants pour exprimer adéquatement certaines fonctions rhétoriques, qu'il s'agisse de faire référence à un chercheur (exemple 12) ou d'introduire un point de vue personnel moins tranché dans un discours à visée objectivante (Boch, 2013) (exemple 13). Nous croyons que ces erreurs ou plutôt maladroites sont attribuables au fait que les apprenants sont peu familiarisés avec le genre de l'écrit académique, en particulier aux routines spécifiques ou fréquentes dans ce genre d'écriture.

3.2.2. Les erreurs au niveau du registre

- 30 Les erreurs au niveau du registre représentent 3,6 % des cas recensés dans Yan (2017) (8 erreurs sur 221 erreurs recensées). Ces erreurs se rapportent au cas où l'apprenant emploie une expression ou une structure appartenant à un registre inapproprié, en mélangeant notamment des usages oraux et écrits. Il est important de souligner que ces erreurs sont parmi les moins fréquentes, mais nous pensons qu'elles sont plus problématiques, car elles ne respectent ni la norme linguistique, ni les codes linguistiques reconnus par la communauté de discours, comme l'illustrent les deux exemples suivants (14, 15) :

(14)

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...
(littérature_12) (Correction proposée : il donne les explications suivantes)
(15)

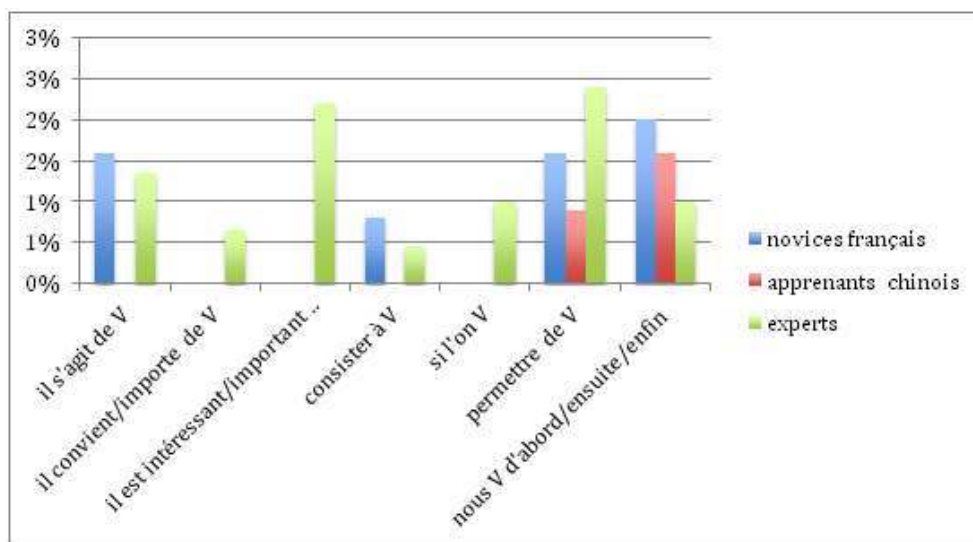
Ce qu'on doit étudier, c'est que dans un modèle ou une circonstance si centralisée, comment on a un pluralisme d'enseignement, le système des différents niveaux du grade. (socioculturel_2) (Correction proposée : Dans un modèle [...] il nous semble important d'étudier comment on peut avoir un enseignement pluraliste)

- 31 Ces énoncés ne respectent pas la norme linguistique de la communauté de discours. L'expression *comme ça* et la structure *ce que... c'est que*, très orales, ne sont pas acceptables dans un écrit académique. Ce type d'erreurs peut être attribuable au fait que l'apprenant n'est pas suffisamment conscient de la différence entre les usages oraux et écrits de la langue. Nous pensons que le fait de sensibiliser l'apprenant aux constructions et aux routines verbales fréquentes et à la notion d'usage dans l'écrit scientifique permet de réduire ce type d'erreur.
- 32 D'une manière générale, on voit bien que les erreurs liées à l'utilisation des routines verbales sont bien présentes pour ce public. Les sources possibles des erreurs relevées sont assez diverses : l'influence de la langue maternelle, le manque de savoir sur les propriétés sémantique et syntaxique des verbes français, le manque de connaissances sur les codes linguistiques dans l'écrit scientifique. Parmi ces causes, nous pensons que les deux dernières paraissent primordiales. En effet, à notre connaissance, le genre de l'écrit scientifique ne fait pas l'objet d'un enseignement spécifique pour ce public. Les apprenants semblent peu familiarisés avec le genre de l'écrit académique, ce qui nous amène à supposer que les lacunes liées à cette dimension empêchent l'apprenant d'employer les routines appropriées au genre. C'est pourquoi nous souhaitons aborder dans la partie suivante l'analyse des cas de sous-emploi afin de mettre en évidence ces lacunes.

3.3. Analyse des cas de sous-emploi

- 33 En dehors des erreurs, la langue de l'apprenant se distingue de celle des natifs par des cas de sous-emploi de mots, d'expressions et de structures syntaxiques (autrement dit « des spécificités »), lesquels « associés aux erreurs pures et simples, trahissent l'origine allophone des apprenants » (Granger, 2007, p.132). En effet, en construisant leur interlangue, les apprenants peuvent commettre des erreurs grammaticales, élaborer des traductions spécifiques et sous-employer des mots ou des expressions. Ces cas se cumulent et se rencontrent fréquemment dans un même paragraphe. Dans le domaine de l'*English for Academic Purposes*, ces spécificités sont étudiées de manière assez systématique, notamment avec la méthode d'analyse *Contrastive Interlanguage Analysis* (Granger, 1996) qui consiste à effectuer des comparaisons quantitative et qualitative entre les productions des locuteurs natifs et non natifs ou celles entre les différents groupes de locuteurs non natifs. Notre analyse comparative porte sur les trois corpus présentés plus haut et se rapporte aux routines présentées dans le tableau 1, illustrées dans le corpus d'experts Transdisciplinaire-TermITH. À titre d'exemple, nous avons calculé la fréquence relative de ces routines et observé dans un deuxième temps la fréquence et la diversité de ces routines dans les corpus Sup-Natifs et Sup-Chinois. Les résultats de comparaison apparaissent dans la figure 1.

Figure 1. – La proportion des routines sous-employées ou absentes chez les apprenants chinois (en % sur le nombre total de verbes d'examen).



- 34 De manière générale, l'emploi des routines chez les experts se distingue par une plus grande diversité. Dans la figure 1, on voit bien que plusieurs routines sont absentes chez les apprenants chinois : *il s'agit de V*, *consister à V*, *il convient/importe de V*, *si l'on V*, *il est intéressant/important de V*. Notons par ailleurs que les trois dernières routines sont également absentes chez les novices français. Ces omissions peuvent être liées à des différences au niveau du genre et de la discipline, mais semblent surtout révéler que les étudiants français comme les apprenants chinois ont une connaissance limitée de l'emploi de ces routines verbales. On peut avancer que ces deux publics sont confrontés à fois à la diversité et à la complexité de ces routines verbales en tant que novices dans ce genre discursif. Autrement dit, ils n'ont pas à leur disposition les moyens linguistiques permettant de souligner l'intérêt ou la nécessité de l'étude, mettre en évidence un fait saillant, etc.
- 35 En outre, il est intéressant de constater que les deux routines *permettre de V* et *nous V d'abord/ensuite/enfin* sont bien présentes chez les apprenants chinois. Certaines fonctions rhétoriques comme « annoncer le plan d'organisation » et « établir une relation causale » posent moins de problèmes pour les apprenants chinois. En revanche, la fonction de topicalisation (*il convient/importe de V*, *il est intéressant/important de V*, etc.) visant à expliciter ou à mettre en relief le topique ou le focus (Creissels, 2004) et la fonction de dialogisme (Grossmann & Tutin, 2010) consistant à impliquer le lecteur (*si l'on V*) semblent ignorées, non seulement par les apprenants chinois, mais aussi par les novices français. Ces lacunes peuvent donc s'expliquer par le fait que ces publics n'ont pas encore construit une représentation précise du fonctionnement du discours scientifique, étant peu familiarisés avec certaines fonctions spécifiques au genre.
- 36 Si l'on observe la proportion des routines *il s'agit de V*, *consister à V*, on constate que l'emploi des routines des novices français se rapproche davantage de celui des experts, tant en termes de diversité que de fréquence. En fait, ce type de routines qui renvoie davantage à des procédures est moins complexe du point de vue rhétorique. En outre, on peut supposer qu'une meilleure maîtrise de la langue facilite sans conteste la rédaction scientifique des étudiants français. Ce n'est pas le cas des étudiants chinois qui, en outre,

n'ont pas appris à argumenter dans le système éducatif chinois et sont moins sensibilisés au genre textuel de l'écrit scientifique (Bi, 2016). L'exemple (16) suivant semble confirmer notre hypothèse :

(16)

Il nous vaut mieux de considérer le paysage de la littérature comme une combinaison de la littérature et le paysage, au lieu de considérer la littérature comme un miroir solitaire qui ne reflète ou déforme que le monde extérieur. (littérature_7) (Correction proposée : il nous semble préférable de considérer...)

- 37 Dans l'exemple (16), il semble que l'étudiant souhaite émettre un jugement. La construction *il nous vaut mieux de considérer* est incorrect. Il est donc possible de penser que l'étudiant ne dispose pas de moyens linguistiques pour exprimer ce sens.

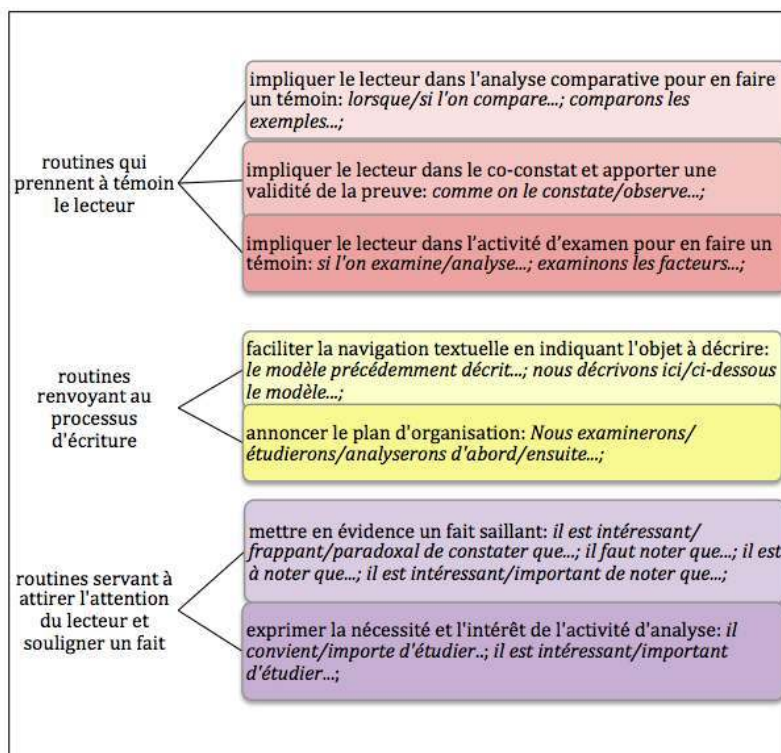
3.4. Synthèse des difficultés

- 38 L'analyse des erreurs associée au repérage des cas de sous-emploi nous a permis d'identifier les principales difficultés liées à l'utilisation des routines chez les étudiants chinois. Ces difficultés peuvent relever de deux niveaux : structures syntaxiques et fonctionnement discursif des routines. D'abord, les structures syntaxiques des routines verbales restent une source d'erreurs principales (l'omission d'un élément, structures inexistantes ou maladroites). Ensuite, il apparaît que les étudiants français comme les apprenants chinois ne maîtrisent pas la diversité des routines rhétoriques récurrentes chez les experts. En effet, les routines verbales sont des éléments phraséologiques complexes du fait de leurs variations syntagmatiques et de la spécificité des associations lexicales. L'absence ou le sous-emploi de certaines routines verbales/ fonctions rhétoriques chez les apprenants chinois nous amène à penser que d'une part, ils ont une connaissance limitée des routines pour exprimer leurs idées, par exemple, les routines comportant le pronom impersonnel *il* (*il convient/importe de V*) semblent non maîtrisées par les apprenants chinois ; et d'autre part, qu'ils ne maîtrisent pas l'emploi de certaines fonctions rhétoriques complexes comme celles qui mettent en jeu le dialogisme interlocutif et la topicalisation.
- 39 Si l'on compare les novices français aux apprenants chinois, on observe de réelles différences. Comme Gilquin et Paquot (2008), nous pensons que les écrits académiques des étudiants natifs occupent une place intermédiaire entre ceux des apprenants chinois et ceux des experts. Le statut de novice en matière de rédaction scientifique confronte de manière comparable les étudiants natifs et non natifs aux difficultés des routines rhétoriques, en particulier celles qui sont spécifiques au discours scientifique. Les étudiants natifs connaissent peut-être mieux la langue, mais ne maîtrisent pas encore les normes du genre. Les difficultés sont plus grandes pour les apprenants chinois, du fait de leur statut d'apprenants du français, qu'il s'agisse des confusions de registre, des erreurs syntaxiques ou des emplois maladroits.
- 40 Tenant compte des difficultés relevées, nous souhaitons proposer maintenant des pistes didactiques pour enseigner les routines verbales et aider les étudiants allophones dans la rédaction académique.

4. Pistes didactiques

- 41 Les difficultés des apprenants chinois relèvent non seulement du lexique, mais aussi du discours. À l'instar de Cavalla (2009), nous pensons que l'acquisition des routines verbales ne saurait se réduire à l'acquisition d'un savoir sur les formes et leur sens, mais qu'il faut, en outre, introduire une composante discursive. Par exemple, dans quelle partie du texte peut-on utiliser les routines ? À quoi servent les routines ? Nos propositions didactiques visent, d'une part, à aider les étudiants allophones à bien repérer l'emploi des routines verbales récurrentes dans l'écrit scientifique et, d'autre part, à les faire réfléchir sur ces éléments dans l'objectif de prendre en compte les spécificités liées au genre du discours.
- 42 Notre approche didactique s'inscrit dans les travaux en *data-driven learning* (DDL) initiés par Johns (1991). Le DDL vise à donner aux apprenants « un accès direct aux données linguistiques » et à les placer au cœur de l'apprentissage, ce qui leur permet d'endosser un rôle de « détective » et de construire par conséquent leurs propres savoirs de façon inductive (Johns, 1991, p. 30). Selon McEnery et Xiao (2010), le corpus donne accès à une « approche opérationnelle des trois “i” », c'est-à-dire, « Illustration – Interaction – Induction » :
- Par illustration on entend le fait de présenter à la classe d'étudiants en langues étrangères des données réelles ; l'interaction concerne la deuxième phase du travail des étudiants qui échangent leurs opinions et leurs observations ; enfin l'induction permet à l'étudiant de se créer sa propre « règle » (qui sera affinée au fur et à mesure que le corpus de données s'élargit) sur le phénomène étudié. (Di Vito, 2013, p. 160)
- 43 L'utilisation des corpus en classe de langue n'est pas nouvelle. De nombreuses études empiriques soulignent que les corpus se révèlent efficaces par rapport à d'autres formes de pratiques, et ceci, dans différents contextes d'enseignement (Cobb & Boulton, 2015), dans la mesure où ils facilitent l'accès aux données et favorisent une activité d'observation et des compétences d'induction. Notre objectif est donc d'enseigner les routines verbales à l'aide de corpus en classe de FLE. Plus concrètement, nous avons adopté deux modes d'enseignement à l'aide des corpus en classe proposés par Fligelstone (1993), à savoir « exploiter les corpus pour enseigner » (*exploiting to teach*) ou « apprendre à exploiter les corpus » (*teaching to exploit*). La première peut se rapporter aux fins d'utilisation didactique, consistant donc à « exploiter les corpus et l'interprétation des résultats pour enseigner une langue » (Cavalla & Loiseau, 2014, p. 165) ; la deuxième vise à former l'apprenant à travailler sur les données des corpus ou à exploiter les corpus, par exemple, pour apprendre une langue étrangère dans le cas qui nous intéresse ici.
- 44 En ce qui concerne l'approche « exploiter les corpus pour enseigner », nous avons montré dans la section 2 comment l'exploitation du corpus Transdisciplinaire-TermITH permettait de construire une ressource lexicale sur les routines verbales, qui servira de matériau pédagogique pour aider les apprenants dans la rédaction scientifique. Cette ressource regroupe les données d'environ 50 verbes et propose un accès onomasiologique par fonction rhétorique (Yan, 2017). À l'instar de Pecman et Kübler (2011), nous considérons que la proposition d'un accès onomasiologique par fonction discursive est pertinente dans la conception d'un outil d'aide à la rédaction scientifique. La figure 2 illustre des exemples de routines classées par fonctions rhétoriques.

Figure 2. – Exemples de routines verbales accompagnées des fonctions rhétoriques.



45 Quant à l'approche « apprendre à exploiter les corpus », il s'agit d'amener l'apprenant à travailler sur des activités conçues à partir de concordances présélectionnées et de le guider vers l'exploitation autonome des corpus. Dans cet article, nous souhaitons proposer des activités à l'aide du corpus Transdisciplinaire-TermITH afin de montrer comment on peut aborder la question de l'enseignement/apprentissage des routines. Ce corpus spécialisé a déjà été utilisé pour mettre en place plusieurs types de séquences didactiques autour des phénomènes phraséologiques (Cavalla, 2009 ; Tran, 2014 ; Yan, 2017). Ces travaux ont montré que l'utilisation du corpus spécialisé selon le DDL est bénéfique à l'enseignement/apprentissage de l'écrit scientifique. Dans la même ligne que ces travaux, nous proposons ici deux types d'activités d'enseignement au croisement de la phraséologie (ici les routines verbales) et de l'écrit académique :

1. L'exploration des routines : de quel type d'expressions s'agit-il et à quoi servent-elles ? (Cf. activité 1)
2. L'observation des fonctions associées aux routines : quelle place dans le texte pour quelle fonction ? (Cf. activité 2)

46 Comme le soulignent Cavalla et Loiseau (2014), le travail sur les concordances permet aux étudiants de mener une réflexion métalinguistique, mettant en jeu une activité d'observation et des compétences d'induction. L'activité réflexive des apprenants est essentielle pour les sensibiliser au lexique.

En se posant tout haut les questions qu'il se pose tout bas, c'est-à-dire en laissant apparaître des traces de ses activités de réflexion sur le nouveau système, l'apprenant se donne une chance d'aboutir à des réponses, c'est-à-dire de développer plus efficacement son nouveau système linguistique. (Arditty & Vasseur, 1996, p. 76-77)

- 47 L'activité 1 est entièrement métalinguistique. Nous proposons de faire observer aux apprenants la routine <il est ADJ d'étudier/examiner/analyser> pour qu'ils comprennent le fonctionnement de la structure syntaxique (cf. tableau 3).

Tableau 3. – Concordancier autour de la routine <il est ADJ d'étudier/examiner/analyser>.

Consigne. – En binôme, observez les phrases et répondez aux questions suivantes.		
Contexte gauche	Occurrences	Contexte droit
Au-delà des questions de niveau technologique et de savoir-faire industriel que nous avons analysées dans la section précédente,	il est également important d'analyser	les secteurs qui sont porteurs d'enjeux stratégiques, aussi bien en termes politiques qu'économiques.
Dans cette perspective,	il est intéressant d'examiner	le modèle de construction du sens proposé par Weick pour un contexte organisationnel, et d'observer si des rapprochements sont possibles avec le modèle de Dervin.
Le pays 1 est plus impatient que l'autre, il consomme toute sa richesse puis disparaît tandis que l'autre pays devient l'économie mondiale.	il est nécessaire d'étudier	la répartition du capital avant de pouvoir poser une condition nécessaire et suffisante sur les paramètres qui assure que la richesse du pays 2 est initialement positive.

- Quels adjectifs se trouvent avant le verbe ? Connaissez-vous d'autres adjectifs qui peuvent apparaître dans ces phrases ? Lesquels ?
- Est-il possible de remplacer le verbe par des synonymes ? Lesquels ?
- Quel est le sens de l'expression *il est nécessaire d'étudier* ? Elle sert à exprimer une nécessité ou une évaluation ?

- 48 Les questions posées doivent guider l'apprenant à repérer la structure. Ensuite, il faut faire découvrir que cette forme préfabriquée est liée à une liste limitée d'adjectifs et de verbes. Après avoir repéré la forme, l'enseignant peut demander à l'apprenant de cerner le sens ou la fonction liée à cette routine.
- 49 Une autre activité que nous souhaitons mettre en avant consiste à classer les routines verbales selon leurs fonctions rhétoriques. Nous supposons que l'apprenant les a déjà travaillées en amont. Cette activité permet à l'apprenant de mieux comprendre à quoi servent les routines étudiées et de produire sa propre liste lexicale. Dans la figure 3, nous proposons de classer les formules suivantes dans trois cases différentes selon leurs fonctions rhétoriques. L'enseignant pourrait également guider les apprenants à chercher d'autres routines ayant les mêmes fonctions à partir des classes de verbes.

Figure 3. – Exercice de classement sémantique.

Consigne : En binôme, classez les formules suivantes dans le tableau ci-après. Ajoutez-y les synonymes récoltés.

1. Consister à étudier (ex : *L'analyse des résultats consiste à étudier la nature des erreurs des participants.*)
2. Si l'on examine (ex : *Si l'on examine les exemples suivants...*)
3. Il est nécessaire d'analyser (ex : *Il est nécessaire d'analyser les secteurs qui sont porteurs d'enjeux stratégiques.*)
4. Il importe d'étudier (ex : *Il importe d'en étudier les effets sur l'équité des systèmes éducatifs.*)
5. Avoir pour objectif d'étudier (ex : *Cet article a pour objectif d'étudier la motivation des salariés.*)
6. Examinons d'abord (ex : *Examinons d'abord la terminologie utilisée pour définir le champ disciplinaire.*)

Exprimer un objectif	Inclure le lecteur	Exprimer la nécessité

50 Nous avons montré que le corpus spécialisé Transdisciplinaire-TermITH permet de créer du matériel pédagogique sur les routines verbales. Nous avons insisté tout particulièrement sur le lien entre l'approche inductive et l'introduction du corpus en classe de FLE en proposant notamment un apprentissage fondé sur l'observation des phénomènes langagiers pour développer le métalangage des apprenants. Néanmoins, l'utilisation du corpus selon le DDL en classe de langue représente un enjeu majeur. La difficulté principale est liée à l'accessibilité cognitive des corpus (Ciekanski, 2014). En effet, la complexité des outils à maîtriser et la présentation de masses importantes de données peuvent constituer une surcharge cognitive pour les apprenants. Du côté de l'apprenant, la question du développement de l'autonomie se pose. À l'instar de Ciekanske (2014), nous pensons que l'un des principaux freins à l'utilisation des corpus provient de l'enseignant. Il est donc nécessaire de faire comprendre aux enseignants l'utilité des corpus et de les former à la méthodologie de linguistique de corpus tant sur les plans théorique, que méthodologique et technique (Kübler, 2014) afin de mieux guider les apprenants dans leur apprentissage. Les propositions didactiques présentées ci-dessus pourraient donner quelques pistes et pourront être intégrées dans une formation d'initiation aux normes universitaires destinée aux étudiants allophones.

5. Conclusion

51 Dans cet article, nos objectifs étaient à la fois linguistiques et didactiques. Nous avons présenté quelques routines verbales abordées dans la thèse de Yan (2017), mises en évidence avec une méthode combinant les techniques de TAL et l'analyse linguistique manuelle. Il a été montré que les routines se caractérisent non seulement par la diversité des réalisations lexicales, mais aussi par des fonctions rhétoriques spécifiques du genre.

52 L'analyse des productions des apprenants chinois a révélé que les difficultés liées à l'utilisation des routines verbales relèvent non seulement des structures syntaxiques,

mais aussi des fonctions rhétoriques associées. Certaines routines et fonctions se montrent plus complexes que d'autres à maîtriser (le dialogisme et la topicalisation), même pour les novices français, et doivent faire l'objet d'un enseignement spécifique de l'écrit académique. À partir des difficultés relevées, nous avons proposé des pistes didactiques pour : 1) favoriser une réflexion métalinguistique sur les routines ; 2) associer les formes lexicales aux fonctions rhétoriques.

- 53 Sur le plan didactique, nous avons montré l'intérêt des corpus pour plusieurs tâches : 1) pour la construction des ressources lexicales à partir d'un corpus d'experts ; 2) pour le repérage des difficultés chez les étudiants allophones dans le corpus d'apprenants ; 3) pour enseigner l'écrit académique — repérage du lexique, réflexion métalinguistique, utilisation de concordances sélectionnées.

BIBLIOGRAPHIE

- ARDITTY, Joseph & VASSEUR, Marie-Thérèse. (1996). Les activités réflexives en situation de communication exolingue. *AILE*, 8, 57-88.
- Bi, Xiao. (2016). *Rhétorique de la dissertation : étude contrastive des conventions d'écriture académique en français et en chinois* (Thèse de doctorat en didactique des langues et des cultures publiée). Université Sorbonne Paris Cité.
- BOCH, Françoise. (2013). Former les doctorants à l'écriture de la thèse en exploitant les études descriptives de l'écrit scientifique. Dans F. Komesu & L. Tenani (dir.), *Écriture et discours, Revista Linguagem em (Dis)curso*, 3(13), 543-568.
- BOULTON, Alex, CANUT, Emmanuelle, GUERIN, Emmanuelle, PARISSÉ, Christophe & TYNE, Henry. (2013). Corpus et appropriation de L1 et L2. *Linx. Revue des linguistes de l'Université Paris Nanterre*, 68-69, 9-32.
- CAVALLA, Cristelle. (2009). La phraséologie en classe de FLE. *Les Langues modernes*, 2009(1). Disponible en ligne sur <<http://www.aplv-languesmodernes.org/...article2292>> (consulté le 12 novembre 2017).
- CAVALLA, Cristelle. (2015). Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. Dans F. Sitri & F. Rinck (dir.), *Former à l'écrit universitaire* (p. 95-110). Paris : Linx.
- CAVALLA, Cristelle & LOISEAU, Mathieu. (2014). Scientext comme corpus pour l'enseignement. Dans F. Grossmann & A. Tutin (dir.), *L'écrit scientifique : du lexique au discours. Autour de Scientext* (p. 163-180). Rennes : Presses universitaires de Rennes.
- CIEKANSKI, Maude. (2014). Les corpus : de nouvelles perspectives pour l'apprentissage des langues en autonomie ? *Recherches en didactique des langues et des cultures*, 11(1). Disponible en ligne sur <<http://journals.openedition.org/rdlc/1710>> (consulté le 17 février 2018).
- COBB, Tom & BOULTON, Alex. (2015). Classroom Applications of Corpus Analysis. Dans D. Biber & R. Reppen (dir.), *Cambridge Handbook of English Corpus Linguistics* (p. 478-497). Cambridge : Cambridge University Press.

- CREISSELS, Denis. (2004). *Cours de syntaxe générale. Topicalisation et focalisation* (chapitre 17). Disponible en ligne sur <<https://fr.scribd.com/doc/280633674/Creissels-COURS-DE-SYNTAXE-pdf>> (consulté le 5 novembre 2017).
- DI VITO, Sonia. (2013). L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE. *Linx*, 68-69, 159-176.
- DUBOIS, Jean & DUBOIS-CHARLIER, Françoise. (1997). *Les verbes français*. Paris : Larousse-Bordas.
- FLIGELSTONE, Steven (1993). Some Reflections on the Question of Teaching, from a Corpus Linguistics Perspective. *ICAME Journal*, 17, 97-109.
- GONZÁLEZ-REY, Isabel. (2007). *La didactique du français idiomatique*. Fernelmont : E.M.E.
- GILQUIN, Gaëtanelle & PAQUOT, Magali. (2008). Too Chatty: Learner Academic Writing and Register Variation. *English Text Construction*, 1(1), 41-61.
- GRANGER, Sylviane. (1996). From CA to CIA and Back : An Integrated Approach to Computerized Bilingual and Learner Corpora. Dans K. Aijmer, B. Altenberg & M. Johansson (dir.), *Languages in Contrast. Text-Based Cross-Linguistic Studies* (p. 37-51). Lund University Press : Lund.
- GRANGER, Sylviane. (2007). Corpus d'apprenants, annotation d'erreurs et ALAO : une synergie prometteuse. *Cahiers de lexicologie*, 91(2), 117-132.
- GRANGER, Sylviane & PAQUOT, Magali. (2009). Lexical Verbs in Academic Discourse: A Corpus-Driven Study of Learner Use. Dans M. Charles, D. Pecorari & S. Hunston (dir.), *Academic Writing: At the Interface of Corpus and Discourse* (p. 193-214). Londres : Continuum.
- GRANGER, Sylviane, GILQUIN, Gaëtanelle & MEUNIER, Fanny (dir.). (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge : Cambridge University Press.
- GROSSMANN, Francis & TUTIN, Agnès. (2010). Les marqueurs verbaux de constat : un lieu de dialogisme dans l'écrit scientifique. Dans *Actes du colloque « Dialogisme : langue, discours »*. Disponible en ligne sur <www.praxiling.fr/dialogisme-langue-discours.html> (consulté le 22 décembre 2017).
- HANKS, Patrick. (2013). *Lexical Analysis: Norms and Exploitations*. MIT Press.
- HATIER, Sylvain. (2016). *Identification et analyse linguistique du lexique scientifique transdisciplinaire. Approche outillée sur un corpus d'articles de recherche en SHS* (Thèse de doctorat en sciences du langage). Université Grenoble Alpes, Grenoble.
- HATIER, Sylvain, AUGUSTYN, Magdalena, JACQUES, Marie-Paule, TRAN, Thi Thu Hoai, TUTIN, Agnès & YAN, Rui. (2016). French Cross-Disciplinary Scientific Lexicon: Extraction and Linguistic Analysis. Dans T. Margalidze & G. Meladze (dir.), *Lexicography and Linguistic Diversity: Proceedings of the XVII EURALEX International Congress* (p. 355-366). Tbilissi, Géorgie : Ivane Javakhishvili Tbilisi University Press. Disponible en ligne sur <<https://euralex2016.tsu.ge/publication2016.pdf>>.
- HATIER, Sylvain & YAN, Rui. (2017). Analyse contrastive des patrons verbaux dans l'écrit scientifique entre scripteurs étudiants et experts. Dans I. Eshkol-Taravella & A. Lefevre-Halftermeyer (dir.), *Linguistique de corpus : vues sur la constitution, l'analyse et l'outillage*, CORELA, numéro spécial HS-21. Disponible en ligne sur <<http://journals.openedition.org/corela/4879>> (consulté le 20 décembre 2017).
- JACQUES, Marie-Paule & RINCK, Fanny. (2017). Un « corpus de littéracie avancée » : résultat et point de départ. Dans J. David, C. Doquet & S. Fleury (dir.), *Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement*, CORPUS, 16, 217-237.

Disponible en ligne sur <<https://journals.openedition.org/corpus/2806>> (consulté le 20 décembre 2017).

JOHNS, Tim. (1991). Should You Be Persuaded: Two Examples of Data-Driven Learning. Dans T. Johns & P. King (dir.), *Classroom Concordancing, English Language Research Journal*, 4, 1-16.

KÜBLER, Natalie. (2014). Mettre en œuvre la linguistique de corpus à l'université : vers une compétence utile pour l'enseignement/apprentissage des langues ? *Recherches en didactique des langues et des cultures. Les Cahiers de l'Acedle*, 11(1), 37-77.

KRAIF, Olivier & DIWERSY, Sascha. (2012). Le Lexicoscope : un outil pour l'étude de profils combinatoires et l'extraction de constructions lexico-syntaxiques. Dans G. Antoniadis, H. Blanchon & G. Sérasset (dir.), *Actes de la conférence conjointe JEP-TALN-RECITAL 2012* (vol. 2, p. 399-406). Grenoble : ATALA & AFCP.

LAKE, John. (2004). Using "On the Contrary": The Conceptual Problems for EAP Students. *ELT Journal*, 58(2), 137-144.

LEGALLOIS, Dominique & TUTIN, Agnès. (2013). Présentation – Vers une extension du domaine de la phraséologie. *Langages*, 189(1), 3-25.

MCENERY, Tony & XIAO, Richard. (2010). What Corpora Can Offer in Language Teaching and Learning. Dans E. Hinkel (dir.), *Handbook of Research in Second Language Teaching and Learning* (p. 364-380). Londres : Routledge.

NESSSELHAUF, Nadja. (2005). *Collocations in a Learner Corpus*. Amsterdam : John Benjamins.

PAQUOT, Magali. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis*. Londres : Continuum International Publishing Group.

PECMAN, Mojca & KÜBLER, Natalie. (2011). ARTES: An Online Lexical Database for Research and Teaching in Specialized Translation and Communication. Dans *Proceedings from International Workshop on Lexical Resources (WoLeR)* (p. 87-93). Ljubljana, Slovénie.

TRAN, Thi Thu Hoai. (2014). *Développement d'une aide à l'écrit scientifique. Description de la phraséologie scientifique et réflexion didactique pour l'enseignement à des étudiants non natifs* (Thèse de doctorat en sciences du langage). Université Grenoble Alpes, Grenoble.

TUTIN, Agnès. (2014). La phraséologie transdisciplinaire des écrits scientifiques : des collocations aux routines sémantico-rhétoriques. Dans A. Tutin & F. Grossmann (dir.), *L'écrit scientifique : du lexique au discours. Autour de Scientext* (p. 27-44). Rennes : Presses universitaires de Rennes.

TUTIN, Agnès & GROSSMANN, Francis. (2014). *L'écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes : Presses universitaires de Rennes.

TUTIN, Agnès & KRAIF, Olivier. (2016). Routines sémantico-rhétoriques dans l'écrit scientifique de sciences humaines : l'apport des arbres lexico-syntaxiques récurrents. *Lidil*, 53, 119-141.

YAN, Rui. (2017). *Étude des constructions verbales scientifiques dans une perspective didactique : utilisation des corpus dans le diagnostic des besoins langagiers du FLE à l'aide des techniques de TAL* (Thèse de doctorat en sciences du langage). Université Grenoble Alpes, Grenoble.

NOTES

1. TermITH (Terminologie et Indexation de Textes en sciences Humaines) : ANR-12-CORD-0029 CONTINT. ATILF, INIST, LIDILEM, LINA, INRIA NGE et Saclay : <www.atilf.fr/ressources/termith/> (consulté le 20 décembre 2017).
 2. Les cadres de sous-catégorisation sont un ensemble des relations de dépendances impliquant le verbe étudié. Le travail de l'extraction des cadres a été effectué par Sylvain Hatier (2016).
 3. ~ signifie que le pivot (soit le verbe *étudier* dans l'exemple) est en position de dépendant. ADJMOD représente un modifieur adjectival.
 4. Il existe en deux versions (papier et électronique) : <<http://rali.iro.umontreal.ca/rali/?q=fr/LVF>> (consulté le 2 décembre 2017).
 5. Lexicoscope permet d'extraire à la fois des concordances et des lexicogrammes, c'est-à-dire des tables de cooccurrences ; la technique est basée sur la mesure d'association – le *loglike*. Il est accessible en ligne sur <<http://phraseotext.u-grenoble3.fr/lexicoscope/index.php>>.
 6. Il s'agit d'un sous-corpus du corpus *Littéracie avancée* (Jacques & Rinck, 2017), constitué pour analyser les compétences rédactionnelles en français langue maternelle à un niveau avancé et développer, par la suite, des ressources didactiques.
-

RÉSUMÉS

Dans cet article, qui porte sur l'utilisation des corpus à des fins didactiques, nous abordons les routines langagières qui sont particulièrement présentes dans les corpus d'écrits scientifiques et académiques en français et nous présentons les difficultés liées à l'utilisation de ce phénomène chez les étudiants allophones. L'étude de diagnostic nous permet ensuite de proposer des exemples d'activités développées à l'aide du corpus Transdisciplinaire-TermITH auprès d'étudiants allophones.

In this article, which focuses on the use of corpora for didactic purposes, we will discuss the linguistic routines that are particularly present in scientific and academic writings in French. We will present the difficulties related to this lexicon amongst non-native speaking students. This initial diagnostic work then allows us to propose examples of activities for foreign language students developed with the Transdisciplinaire-TermITH corpus.

INDEX

Mots-clés : linguistique de corpus, routines, fonctions rhétoriques

Keywords : corpus linguistics, routines, rhetorical functions

AUTEURS

RUI YAN

LIDILEM, Université Grenoble Alpes
rui.yan@univ-grenoble-alpes.fr

AGNÈS TUTIN

LIDILEM, Université Grenoble Alpes
agnes.tutin@univ-grenoble-alpes.fr

THI THU HOAI TRAN

GRAMMATICA, Université d'Artois / LIDILEM, Université Grenoble Alpes
tthoi.tran@univ-artois.fr