



HAL
open science

Enseignement/apprentissage de l'écrit universitaire dans les disciplines scientifiques

Mariam Atoui

► **To cite this version:**

Mariam Atoui. Enseignement/apprentissage de l'écrit universitaire dans les disciplines scientifiques. Linguistique. 2020. dumas-03215656

HAL Id: dumas-03215656

<https://dumas.ccsd.cnrs.fr/dumas-03215656>

Submitted on 3 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Enseignement/apprentissage de l'écrit universitaire dans les disciplines scientifiques

Mariam ATOUI

UFR LLD – Département de Didactique du FLE

Mémoire de Master 2 – 28 ECTS

Master : Didactique des langues, du français langue étrangère et seconde :
métiers de la recherche, de l'enseignement et de l'ingénierie

Sous la direction de Mme : Cristelle CAVALLA

Membres du jury : Mme Catherine Mendonças Dias

Année universitaire 2019-2020

Enseignement/apprentissage de l'écrit universitaire dans les disciplines scientifiques

Mariam ATOUI

UFR LLD – Département de Didactique du FLE

Mémoire de Master 2 – 28 ECTS

Master : Didactique des langues, du français langue étrangère et seconde :
métiers de la recherche, de l'enseignement et de l'ingénierie

Sous la direction de Mme : Cristelle CAVALLA

Membres du jury : Mme Catherine Mendonças Dias

Année universitaire 2019-2020

Résumé

La présente étude porte sur l'amélioration de l'écrit scientifique des étudiants universitaires et se trouve à la croisée de la linguistique de corpus et la phraséologie. Elle est précisément centrée sur la phraséologie transdisciplinaire employée à des fins d'expressions scientifiques.

Notre objectif est d'élaborer une méthodologie d'enseignement-apprentissage basée sur des corpus d'experts, d'enseignants et d'étudiants dans le but d'identifier et de comparer les collocations et les constructions récurrentes associées aux trois notions : "*hypothèse, expérience et résultats*".

Afin de mieux comprendre les difficultés des étudiants en matière de collocations, nous établissons un diagnostic qui comporte d'une part, la comparaison des collocations employées par les experts avec celles trouvées dans les cours d'enseignants et d'autre part, l'analyse des constructions dans des cours d'enseignants et dans des comptes rendus des étudiants. L'analyse des productions écrites a permis de dégager les erreurs d'emploi et les cas de non-emploi de certains phénomènes phraséologiques. Ce diagnostic a guidé également l'élaboration de nos réflexions sur le plan didactique.

Les résultats ont montré que les étudiants, encore novices en matière de rédaction, manifestent des difficultés liées à l'écrit académique notamment en ce qui concerne la phraséologie. Afin de remédier à ces lacunes et répondre aux besoins langagiers, nous avons proposé des démarches didactiques dont l'objectif est de guider les étudiants vers l'autonomie dans leur apprentissage et dans la manipulation des corpus en vue d'une meilleure rédaction d'écrits académiques en français.

MOTS-CLÉS : linguistique de corpus, lexique scientifique transdisciplinaire, lexique scientifique général, collocations, phraséologie, écrits académiques, écriture scientifique.

Abstract

This study focuses on the improvement of university students' scientific writing. The topic of research can be described as a nexus between corpus linguistics and phraseology. We are essentially interested in the interdisciplinary phraseology used for scientific expressions purposes. We aim to develop a corpus-based teaching-learning methodology that concerns collocations and frequent constructions related to notions : “*hypothesis, experiment and results*”. Our goal is to help foreign students enhance their academic writing in french language.

To better understand the difficulties encountered by the students with regards to the found collocations, we establish a diagnosis that includes the comparison between collocations in experts' and professors' corpora as well as the analysis of the constructions used by the professors in their lecture notes and by the students when writing their reports. This analysis enabled us shed light on the errors and the cases of non-use of some phraseology phenomena and was essential to elaborate our reflections from an educational point of view.

The obtained results show that students, who are not experienced yet in writing, face some difficulties related to academic writing especially in what is associated with phraseology. In order to fill this gap and satisfy the linguistic needs of students, we have sketched out some didactic steps so as to guide students towards autonomy in learning phraseology by exploiting corpora.

KEYWORDS : corpus linguistics, interdisciplinary scientific words, general scientific vocabulary, phraseology, collocations, academic writing, scientific writing

Table des matières

Introduction	i
I Concepts clés	1
1 La littératie	3
1.1 Origine et définition	5
1.2 Champs de la littératie	7
1.3 Littératies universitaires	8
1.4 Français sur Objectifs Universitaires (FOU)	12
1.5 Bilan du premier chapitre	14
2 L'écrit scientifique	17
2.1 Écrit et apprentissage	19
2.2 Sciences et discours scientifique	20
2.3 L'écriture académique universitaire	21
2.4 Lexique transdisciplinaire	24
2.5 Bilan	28
II Contextualisation de la recherche	29
3 Contextualisation de la recherche	31
3.1 Terrain de recherche	33
3.2 Public concerné	35

3.3	L'exposition au français à l'UL	39
3.4	Bilan du contexte	41
4	Méthodologie et recueil des observables	43
4.1	Linguistique de corpus	45
4.2	L'usage de corpus informatisés	47
4.3	Constitution des corpus universitaires	50
4.4	Collecte de données	56
4.5	Outils d'extraction des collocations	59
4.6	Bilan de la méthodologie	62
III	Analyse des échantillons	65
5	Analyse des échantillons	67
5.1	Résultats d'analyse des corpus	69
5.2	Observation des CR d'étudiants	76
5.3	Propositions didactiques	81
	Conclusion	87
	Bibliographie	91

Introduction



Tout chercheur qu'il soit débutant ou expert est amené à rédiger des écrits scientifiques pour vulgariser ses connaissances. Ce qui pose un réel défi pour un public universitaire allophone/étranger à qui nous demandons non seulement de maîtriser le contenu scientifique mais aussi le contenu linguistique de ces genres textuels.

L'enseignement supérieur est un espace où se multiplient écrits et écritures. Leurs fonctions sont diverses et leurs contextes le sont tout autant. De façon presque évidente, ces écritures se différencient selon leurs contenus ainsi que leurs références.

Beaucoup de questions se posent autour de l'enseignement/apprentissage de l'écrit universitaire : est-ce que cela se fait au sein de la classe de la discipline ou bien en classe de langue ? Certes les acteurs de ce processus, enseignants et étudiants, écrivent, mais qu'est-ce qu'être enseignant lorsque l'on écrit ? sont-ils des professeurs ou des chercheurs ? La question se décline également pour un étudiant qui écrit d'autres choses que ses notes de cours : s'agit-il d'un futur professionnel ou d'être un disciple acquérant éternellement du savoir ?

À l'université, les experts des écrits universitaires sont les enseignants/chercheurs de la discipline elle-même à qui les étudiants adressent leurs rédactions en vue d'une évaluation. Ces pratiques d'écriture ou des discours propres à chaque discipline devraient de plus en plus faire l'objet d'étude parmi les chercheurs. Les étudiants devraient comprendre les différentes formes et attentes concernant l'écriture pour entrer dans les disciplines. Afin de développer l'expertise des étudiants et combler leur faiblesse en matière d'écriture, nous rejoignons Decambre et Lahani-Reuter (2010) sur la possibilité de "clarifier un certain nombre de dimensions de l'écriture universitaire que les dimensions purement instrumentales ou techniques". Ce qui implique une démarche pour articuler entre recherche et enseignement pour cibler principalement les pratiques universitaires d'écriture et de lecture, en relation avec les contextes disciplinaires concernés.

Certains champs ont déjà exploré les questions des difficultés des étudiants face à l'écriture (et à la lecture) qu'elle soit académique ou de recherche. Mais, il se trouve que la question d'analyse des pratiques d'écriture à l'université au sein des champs de recherche linguistiques et didactiques est un phénomène récent.

Dans ce mémoire, nous proposons de travailler sur l'amélioration de l'écrit scientifique auprès d'un public d'étudiants étrangers qui ont atteint un niveau B1 ou plus et qui poursuivent leurs études supérieures scientifiques. Il s'agit tout d'abord de repérer les difficultés des étudiants à l'université pour ensuite, faciliter l'enseignement et l'appropriation. Nous allons aborder l'aspect linguistique et didactique de ce sujet en se fixant comme objectif de travailler sur la phraséologie d'un rapport ou d'un compte-rendu d'un Travail Pratique (TP).

Contexte de l'étude :

Les étudiants libanais, après des années de scolarisation dans le système éducatif libanais, n'ont souvent pas les acquis correspondant aux objectifs académiques "à la française". Or, ils doivent accomplir les mêmes tâches que leurs collègues français natifs pour réussir leurs études en France, souvent des productions écrites, sous la forme de comptes rendus, de synthèses, de mémoires et de thèses. Nous nous intéressons particulièrement aux étudiants libanais de premier cycle à l'université libanaise et qui ont à rédiger un écrit scientifique (rapport ou compte rendu d'un travail pratique). Même si la plupart des étudiants ont atteint un niveau B1/B2 en français, ils sont souvent confrontés à des difficultés rédactionnelles, en particulier celles qui sont liées à la combinatoire lexicale des mots en contexte. Ces phénomènes linguistiques se révèlent être l'une des principales sources des nombreuses erreurs chez les apprenants de Français Langue Etrangère (FLE).

Objectifs et intérêt du travail :

Afin de poursuivre ses études supérieures, on n'a pas seulement besoin de bien rédiger son CV et sa lettre de motivation mais aussi de rédiger un écrit scientifique, chacun dans son champ d'études. Dans un tel écrit, nous apercevons un chercheur, qu'il soit novice ou expert, qui exprime son esprit scientifique, son sens critique, et qui établit des réalités, compare des idées, argumente son point de vue, prend position, présente des résultats, résume un historique, et conclut son travail. Ces différentes actions rentrent dans l'élaboration du fonctionnement rhétorique et pragmatique de l'écrit scientifique et

ne sont pas triviales pour des étudiants étrangers. À notre sens, il convient de travailler la rédaction scientifique avec les étudiants dès le premier cycle universitaire afin qu'ils soient dotés du bagage nécessaire à la rédaction d'un mémoire ou d'une thèse.

En classe de licence en physique, un compte-rendu écrit est demandé à la suite d'une réalisation d'un dispositif expérimental afin de mettre en forme et communiquer les démarches et les résultats obtenus. Ce type d'écrit comprend des spécificités linguistiques comme tout autre genre discursif rencontré à l'université. C'est à nous, en tant qu'enseignants, d'accompagner les élèves dans la découverte et la maîtrise progressive des codes et formes de l'écrit en usage dans les apprentissages des disciplines.

Dans ce travail, nous nous intéressons particulièrement à la phraséologie transdisciplinaire scientifique, et plus précisément aux collocations relatives aux notions "*hypothèse, expérience et résultats*". Il s'agit des mots situés essentiellement à l'intersection entre la langue courante et les langues spécialisées, qui appartiennent à la langue générale, mais sont utilisés à des fins scientifiques (Yan, 2017). Chaque collocation construite à partir de ces lexies est associée à un sens particulier et obéit à des structures syntagmatiques et à des valeurs sémantiques spécifiques.

Notre motivation pour les collocations de ces notions, faisant partie du lexique scientifique, est liée aux complexités linguistiques de ce phénomène pour les étudiants ainsi qu'aux exigences de l'écrit scientifique. Nous rejoignons Cavalla (2014) qui, selon elle, l'enseignement et l'apprentissage de la phraséologie en général et des collocations en particulier pourrait constituer une aide à la lecture et à l'écriture du texte scientifique. Dans la présente étude, nous souhaitons :

- Tout d'abord, mener une étude à partir des corpus d'experts français et des corpus d'enseignants à l'université libanaise en examinant plus précisément les collocations associées aux notions choisies. Cette comparaison sert non seulement à repérer la manière dont les experts emploient les collocations en question mais aussi à identifier les spécificités de l'emploi des collocations dans les écrits de la discipline auxquels les étudiants sont exposés au fur et à mesure de leurs études universitaires.
- Ensuite, établir un diagnostic des besoins langagiers des étudiants en matière de collocations. Les résultats nous permettront de mieux orienter notre réflexion au plan didactique.

La linguistique de corpus au service de la didactique de l'écrit :

Cette étude est à la croisée de la linguistique de corpus et la phraséologie. La première discipline permet, dans notre travail, d'observer les propriétés syntaxiques et sémantiques des collocations et d'identifier celles qui sont récurrentes dans l'écrit scientifique. Les développements en ce domaine ont contribué au renouvellement de la réflexion sur les contenus d'enseignement et sur la place des connaissances linguistiques dans l'enseignement (Pecman, 2005).

Nous nous posons principalement les questions suivantes : quels corpus de référence permettent l'étude des caractéristiques linguistiques de l'écrit en question ? Quelles données collecter pour identifier les difficultés spécifiques concernant l'emploi des collocations pour les étudiants libanais dans l'écriture académique ?

Les différents corpus utilisés dans notre travail et dont deux sont constitués ont des visées didactiques précises. D'une part, l'utilisation des corpus numériques d'écrits d'experts permet d'étudier la phraséologie recherchée dans des écrits scientifiques de différents genres et faisant partie de plusieurs disciplines. D'autre part, la constitution d'un corpus de cours d'enseignants nous permet d'observer l'emploi des collocations par les "experts" de la discipline dans un contexte précis qui est celui de l'université libanaise ainsi qu'à faire émerger les différences et les points communs avec les écrits d'experts. Le deuxième corpus constitué d'écrits universitaires d'étudiants sert à observer l'utilisation des collocations trouvées dans les corpus d'experts et qui plus est à identifier leurs difficultés en amont de la recherche orientée vers la didactique de l'écrit académique.

Finalement, ces corpus présentent un intérêt pour l'enseignement de la phraséologie. Nous pouvons en servir des données en vue d'enseigner les collocations.

Hypothèses et questions :

Nous pouvons avancer l'hypothèse selon laquelle la phraséologie transdisciplinaire et les collocations plus précisément, présentent des spécificités au niveau sémantique et syntaxique dans l'écrit scientifique. Nous cherchons à répondre aux questions suivantes : Quelle approche à adopter afin d'analyser l'emploi des collocations dans notre étude linguistique ? Quelles données permettent d'observer les phénomènes phraséologiques dans l'écrit scientifique d'une discipline dite non linguistique ? Quelles sont les constructions fréquemment associées aux notions "*hypothèse, résultats et expérience*" ?

L'étude des collocations a pour objectif de mieux outiller l'enseignement de l'écrit scientifique. Nous partons du constat que les étudiants ne sont pas assez exposés à l'ensei-

nement de la phraséologie liée aux écrits scientifiques et se trouvent confrontés à de difficultés à faire passer leurs idées dans la rédaction. La problématique de la recherche porte d'une part, sur le diagnostic des besoins langagiers et d'autre part sur leur transposition didactique. Nous souhaitons alors savoir si les difficultés des étudiants sont seulement liées à leur maîtrise de langue et les outils à mettre en place afin d'envisager l'enseignement/l'apprentissage des collocations.

Plan du mémoire :

Le plan du mémoire s'esquisse de la manière suivante. Le chapitre I est destiné à présenter le cadre théorique et les concepts autour desquels s'articulent notre étude. Il s'agit, dans un premier temps, de discuter l'origine et la définition de la littératie. Nous entrons ensuite dans le champ auquel nous nous intéressons : les littératies universitaires. Nous introduisons, dans un deuxième temps, le Français sur Objectifs Universitaires, une approche qui se rejoint avec les littératies universitaires dans le processus d'enseignement-apprentissage de l'écrit académique à l'université. Les deux domaines permettent de développer les compétences en matière d'écriture scientifique produite dans le cadre d'une activité de recherche.

Dans le chapitre II, nous décrivons d'abord les objets analysés à savoir les genres et discours scientifiques à l'université en mettant en avant l'importance de penser le contexte des écrits produits à l'université comme disciplinaire. Nous entrons ensuite dans des définitions plus précises quant à notre objet d'étude ; en d'autres termes, nous expliquons ce que nous entendons par les collocations et la phraséologie dans l'écrit scientifique avant de délimiter les notions choisies dans notre travail et finalement, souligner l'importance de leur enseignement en vue d'aider à la rédaction universitaire des étudiants.

Le troisième chapitre présente le terrain de notre recherche qui est l'université libanaise ainsi que son public spécifique. Il s'agit d'un public universitaire, bilingue en troisième année de licence en physique qui est amené à rédiger un rapport ou un compte rendu à la suite d'un cours de travaux pratiques. Nous avons par la suite analysé le rapport de ces étudiants au français ainsi que leur exposition à cette langue au sein de l'université en particulier en classe d'une discipline dite non linguistique.

Dans le chapitre IV, nous décrivons la méthodologie adoptée en fonction de nos objectifs de recherche. Nous explicitons ensuite les différents usages de corpus en mettant en avant leurs rôles dans cette recherche. Puis, nous avons expliqué la démarche de construction des corpus universitaires tout en abordant la manière d'extraction des collocations dans

le cadre d'une étude phraséologique.

Le dernier chapitre présente les résultats de l'analyse des corpus. Nous nous appuyons sur les corpus d'écrits universitaires d'étudiants en vue de cerner les erreurs et les spécificités liées à l'utilisation des collocations chez les étudiants libanais. Nous proposons également des pistes didactiques et nos réflexions sur l'enseignement de différentes constructions des notions choisies avec l'utilisation des corpus en classe de langue ou de discipline.

Première partie

Concepts clés

Chapitre 1

La littératie

Contents

1.1	Origine et définition	5
1.2	Champs de la littératie	7
1.3	Littératies universitaires	8
1.3.1	Historique	9
1.3.2	Objets et thèmes de recherche	11
1.4	Français sur Objectifs Universitaires (FOU)	12
1.4.1	Méthodologie du FOU	12
1.4.2	Le FOU en face des LU	13
1.5	Bilan du premier chapitre	14



L'écriture que ce soit en réception ou en production, est une pratique qui se présente dans de différentes institutions, plusieurs contextes, et en formats scripturaux variés. Elle est souvent associée à la notion de la littératie. Cette dernière a connu plusieurs formes, champs et lieux d'application à travers l'histoire. Nous allons nous intéresser, dans le cadre de cette étude, à l'apport de la littératie dans le contexte universitaire et dans un domaine particulier : la didactique du français. Dans ce même contexte, nous allons discuter la relation qu'a la littératie avec l'approche méthodologique qui se focalise sur le développement des compétences langagières des étudiants allophones à l'université : le Français sur Objectifs Universitaires (FOU). Ces deux approches différentes mais qui sont complémentaires l'une à l'autre vont constituer le cadre théorique et méthodologique sur lequel s'appuie notre étude et vont être l'objet autour duquel est axée la discussion du premier chapitre.

1.1 Origine et définition

Les formes sociales et scripturales se sont liées entre elles et évoluées à travers le temps et les époques. La littératie n'a donc jamais eu la même place selon les périodes historiques et s'est toujours adaptées aux contraintes et exigences qu'on lui imposait.

Le terme de la littératie (literacy) apparaît pour la première fois dans le contexte anglo-saxon afin de désigner le contraire de l'illettrisme (illetteracy en anglais). Le mot "literate" était le synonyme de "educated" c.à.d. cultivé. La littératie n'apparaît dans un dictionnaire qu'en 1924 et prend le sens d' "être capable de lire et d'écrire" (Jaffré, 2004). La notion s'est par la suite étendue pour s'associer à d'autres domaines tels que : computer literacy, economic literacy, information literacy, etc.

Les travaux des chercheurs francophones font révéler un emploi emprunté du terme anglais lui-même avec parfois des graphies différentes : littéracie, littératie, litéracie. Comme, notre travail ne fera pas l'objet d'une discussion orthographique, nous référons le lecteur au Jaffré 2004 pour plus de détails sur ce sujet. Nous allons toutefois admettre la coexistence des variantes graphiques dont la graphie "littératie" que nous allons adopter par la suite.

Il s'avère que définir la notion de "littératie" est une tâche difficile à accomplir. Tous les auteurs admettent qu'ils ont trouvé des difficultés dans leurs tentatives de définition voire une impossibilité pour certains (Privat & Kara, 2006). Selon l'Organisation de coopération et de développement économiques (OCDE), la littératie c'est

l'aptitude à comprendre, à utiliser et à lire l'information écrite dans la vie courante, à la maison, au travail et dans la collectivité en vue d'atteindre des buts personnels et d'étendre ses compétences et ses capacités. (OCDE, 1997)

La notion inclut notamment la lecture, l'écriture, le visionnement, l'écoute et la communication orale. Elle permet d'entrer en contact avec le monde et le langage pour pouvoir s'intégrer et mieux comprendre les différents aspects de la vie quotidienne, interagir, communiquer, apprendre, socialiser et effectuer des calculs simples (Lafontaine, Emery-Bruneau et Guay, 2015). De plus, à travers cette notion, on s'intéresse aux différents modes de transmission et d'apprentissage qui peuvent se mettre en place à travers les contextes dans lesquels ils évoluent.

Dans l'ouvrage de Barré-De Miniac et al. (2004), Jaffré propose la définition suivante, fréquemment reprise et citée depuis dans les recherches sur la littératie.

La littératie désigne l'ensemble des activités humaines qui impliquent l'usage de l'écriture, en réception et en production. Elle met un ensemble de compétences de bases linguistiques, graphiques, au service de pratiques, qu'elles soient techniques, cognitives, sociales ou culturelles. Son contexte peut varier d'un pays à l'autre, d'une culture à l'autre, et aussi dans le temps. » (Jaffré J.P., 2004)

Cette définition expose deux points forts : il s'agit d'une part du rapport entre l'homme et la technologie de l'écriture, plus précisément de l'impact que l'utilisation de cette dernière aura sur son développement (cognitif et social) ainsi que sur son fonctionnement (capacités motrices, intellectuelles, comportementales), et d'autre part d'un rapport qui peut changer selon le cadre spatiotemporel. Dans ce contexte, l'écrit n'est plus considéré sous le seul aspect de lecture-écriture, ni dans le cadre de la tradition lettrée, mais selon un ensemble de compétences et pratiques situées traduisant la complexité des relations entre cognition, langue et culture.

Plusieurs dimensions sont ensuite ajoutées à la définition de Jaffré, plus précisément en ce qui concerne les conceptions issues du rapport à l'écrit, le matériel nécessaire à

sa transcription, l'influence des individus et des établissements dans le processus d'enseignement/apprentissage ainsi que la perspective des relations de pouvoir qui y existe (Goody, 2007).

Après avoir discuté la définition de la notion, nous allons voir par la suite les différents champs et contextes de la littératie.

1.2 Champs de la littératie

Le contexte de la littératie peut varier d'un pays à l'autre, d'une culture à l'autre, et aussi dans le temps. Il existe plusieurs champs qui s'intéressent à littératie et qui ont en commun les pratiques l'écriture comme : la linguistique, la psycholinguistique, la psychologie de la cognition, l'illettrisme, l'ethnologie de l'écriture (Jaffré, 2004), la didactique du français, les sciences du langage, la littératie informationnelle, etc.

On assiste depuis quelques années à l'émergence du mot en pluriel « litératies » faisant référence à la diversification de domaines associées (didactique, linguistique, psycholinguistique, sociolinguistique, ethnologique, économie, etc.) ou aux divers mondes scientifiques (anglo-saxons, québécois, français, etc.). Plusieurs chercheurs revendiquent la polyvalence de la notion de littératie pour s'écarter d'une conception de la "literacy" comme résultat d'une activité cognitive individuelle et pour désigner également l'application de cette connaissance dans des contextes spécifiques et à des objectifs différents (Jaffré, 2004). Ils se rejoignent sur le fait que la notion renvoie à la diversité des pratiques de lecture et écriture, des compétences et des fonctionnements cognitifs et sociaux qu'elles mobilisent et des situations où elles s'exercent.

Dans ce qui suit, nous allons présenter brièvement quelques facettes de la littératie en lien avec la didactique du français langue étrangère (FLE) :

- La littératie *vs* illettrisme : L'entrée de la notion de la littératie dans le domaine français est, pour sa part, relativement tardive. L'ouvrage *La littéracie* (Barré-De Miniac et al.) définit ce terme comme étant "le versant positif de ce que celui d'illettrisme désigne en négatif ; le premier serait l'apprentissage de l'écrit au lieu du "désapprentissage" que stigmatise le terme d'illettrisme". Apparu dans les années 70, l'illettrisme touche le public qui a été peu instruit et dans l'incapacité aujour-

d'hui de lire ou d'écrire un texte simple. Il s'agit de l'élaboration d'une nouvelle compétence chez un individu. Dans l'appropriation de la littératie, la compétence est réorientée, réorganisée en fonction de l'émergence de nouveaux besoins en lien avec les usages du nouveau contexte auquel l'individu est exposé.

Un lien entre les deux notions n'a été établi que dans les années 80. L'OCDE a rapproché les définitions de la littératie et de l'alphabétisation fonctionnelle afin de prouver que l'un peut venir au secours de l'autre. C'est à partir de ce moment-là que le terme de littératie a trouvé sa place en France. Le recours à la littératie comme notion a ainsi permis de rendre compte des fonctionnements cognitifs et sociaux des pratiques de lecture-écriture, ces derniers étant alors considérés dans leurs contextes d'usage.

- La littératie numérique (Gerbaut, 2012) : aujourd'hui, la notion de littératie s'est élargie avec l'arrivée massive de moyens numériques. Nous communiquons et interagissons désormais par écran virtuel et message écrit tapé sur un clavier que l'on peut effacer, remanier comme bon nous semble. Cela montre bien que les bases de la littératie sont avant tout sociales et culturelles et qu'elles varient d'une époque à une autre.
- La littératie universitaire (LU) (Garnier, S., Rinck, F., Sitri, F. et De Vogüé, S., 2015) : Ce champ a pour objet de décrire les écrits scientifiques des étudiants en contexte universitaire en représentant "les genres et les modes de discours universitaires ainsi que les difficultés rencontrées chez les étudiants dans leur mise en pratique" (Lafontaine, Emerey-Bruneau & Guay, 2015)

Quelle que soit l'orthographe choisie de cette notion et les adjectifs qui lui sont associés, elle a connu un succès international depuis son émergence ainsi que des développements variés en fonction des contextes dans lesquels elle s'inscrit. Nous allons par la suite décrire plus précisément le champ qui nous intéresse dans le cadre de ce travail : les littératies universitaires.

1.3 Littératies universitaires

De leur côté, les LU se sont également appuyées sur un historique particulier puisque ces recherches trouvent leurs racines dans leur pendant anglo-saxon et américain.

1.3.1 Historique

Au courant des années 1970, les États-Unis connaissent un nombre croissant d'institutions avec une augmentation importante de la population étudiante et de la diversité des publics académiques (Nystrand, 2006). L'arrivée de publics issus de milieux défavorisés et la massification des effectifs à l'entrée à l'université ont entraîné une réflexion sur l'écriture. Cela a remis en question les habiletés en écriture requises à l'université et à une réflexion sur le format de l'enseignement à prodiguer. Ainsi, les enseignants ayant été encouragés par les universités américaines à s'intéresser au domaine de la composition et à l'écriture de leurs étudiants et les besoins du public ont été considérés sous l'angle de la mise à niveau et de l'aide (Russel, 2012).

De courants tels que "Composition Studies", "Writing Across the Curriculum (WAC)" et "Writing in the Disciplines (WID)" ont émergés. Ils s'inscrivent dans une stratégie développée par les États-Unis qui visait à développer l'enseignement des particularités des genres discursifs à l'université et au sein des disciplines afin d'accompagner l'écriture et l'élaboration de contenus conceptuels dans le Supérieur ainsi que les intégrer comme des composantes de la formation universitaire.

Parallèlement à cela, nous voyons se déployer au Royaume Uni, dans les années 1980, ce qu'on appelle "Academic Literacies (AcLit)". Il s'occupait essentiellement de la mise à niveau linguistique des étudiants allophones ainsi que d'analyses des genres et conventions textuelles. À la différence des mouvements américains, cela s'inscrit directement dans une vision de l'écriture comme étant une action sociale qui construit l'individu en relation avec les cultures institutionnelles et disciplinaires. Les pratiques de lecture-écriture sont alors envisagées de manière bien plus enchevêtrée dans la dimension sociolinguistique du contexte. Nous voyons bien que le -s- du pluriel est bien présent : ceci afin de considérer non seulement la diversité des pratiques d'écriture mais aussi pour marquer que ces pratiques sont sociales et culturelles, liées aux contextes et ne sont pas individuelles.

Dans le monde francophone, la constitution du champ des LU est plus récente. Elle s'inscrit dans des contextes scientifiques, sociaux et intellectuels différents de ceux de l'univers anglo-saxon. Le concept a commencé à faire son apparition en France à la fin des années 1990 et s'est bien établi en tant que champ de recherches au courant des an-

nées 2000. En effet, du côté des Sciences du Langage et de la didactique du français, des travaux sur l'écriture et ses relations aux savoirs, pratiques, modèles et genres prennent essor à partir des années 1980 (Barré-De Miniac, 1995).

Les modèles d'enseignement de l'écriture qui existaient dans les années 1970, reposaient sur une conception instrumentale de la langue et sur des compétences méthodologiques ; ils proposaient ainsi des remises à niveau centrées sur la maîtrise de la langue et sont soupçonnés de technicisme. Les travaux en LU récusent ces approches réductrices des problématiques liées à l'écrit universitaire. Ils s'inscrivent ainsi dans un domaine propre établi comme "nécessairement transdisciplinaire" (Delcambre et Lahanier-Reuter, 2010) comme il emprunte aux champs de recherche voisins en s'appuyant en partie sur leurs concepts et méthodologies comme par exemple : les genres, analyse de discours et les pratiques d'écriture.

Malgré l'existence des différences entre les AcLit et les LU (Delcambre et Lahanier-Reuter, 2010)), ce nouveau courant épistémologique emprunte des aspects fondamentaux aux AcLit. le premier, comme l'indiquent Delcambre et Lahanier-Reuter, la notion des LU considère que « les apprentissages de l'écrit se déroulent dans un continuum depuis les premiers contacts avec l'écrit avant les apprentissages scolaires jusqu'aux usages épistémiques de l'écrit ». Le deuxième est représenté par le -s- du pluriel associé à la notion qui considère les pratiques comme contextualisées socialement et culturellement.

L'idée est de penser l'écriture (et la lecture) à l'université comme des pratiques sociales (liées aux situations, aux contenus, aux communautés, aux enjeux, aux usages, etc.), et comme un mode d'entrée vers les savoirs qui structurent les contenus par des genres, des formes d'écrit et des pratiques. Il ne s'agit plus de rejeter les problèmes au milieu scolaire, ni sur une maîtrise linguistique jugée insuffisante, mais de les penser comme issus de la socialisation universitaire et de l'acculturation aux codes académiques. En conséquence, l'objectif des LU est de théoriser la spécificité des pratiques d'écriture académique, de la formation, de l'enseignement et de la recherche. Elles cherchent à décrire les usages discursifs et langagiers en fonction du contexte, et à situer les besoins des étudiants en relation avec l'enseignement supérieur et les situations de communication qui y prennent place.

La section suivante présente l'objet principal qui fait l'objet de recherche en LU : les genres académiques ainsi que les thèmes de recherche qui s'inscrivent dans ce domaine.

1.3.2 Objets et thèmes de recherche

Les LU s'attachent, dans une démarche analytique et descriptive, à décrire les genres du discours, notions sur lesquelles nous revenons dans le deuxième chapitre, et leur diversité dans le but de mieux cerner les difficultés rencontrées par les étudiants, de clarifier les dimensions, autres que seulement techniques et instrumentales, liées à l'écrit universitaire (compétences rédactionnelles générales, maîtrise des genres académiques, posture réflexive) pour finalement développer l'expertise des étudiants en les formant aux pratiques d'écriture instituées dans les disciplines. Les problèmes que les étudiants rencontrent sont dès lors pensées en relation au contexte en termes de difficultés éprouvés à se conformer aux genres universitaires et disciplinaires.

Penser les objets de recherche didactiques en terme de littératie, amène à envisager des pratiques situées dans les contextes précis, à articuler entre plusieurs pratiques langagières (écrit et oral par exemple), à analyser les difficultés des élèves en fonction des conditions sociales et académiques et non plus seulement aux seules compétences des apprenants, ainsi qu'à élargir l'observation des pratiques de l'écrit à différentes disciplines, au-delà du français.

Les discours et genres universitaires écrits constituent l'objet principal des recherches menées dans le cadre des LU vus sous l'angle transdisciplinaire ou disciplinaire. Les recherches françaises en ce domaine s'appuient aussi bien sur des recueils de données importants quantitativement tels que des corpus de textes ou des questionnaires que sur des analyses qualitatives de textes représentant d'une certaine situation.

De très nombreuses études ont été menées dans le cadre des LU dans le but d'améliorer et de didactiser les pratiques écrites des étudiants. Les besoins et les objectifs des étudiants restent la clé de voûte des contenus. Parmi ces travaux menés sous l'égide des LU, nous repérons des sujets portant sur la pratique du copier-coller (Rinck F., 2019), sur l'influence des disciplines et du niveau d'étude dans les pratiques de l'écrit (Delcambre et Lahanier-Reuter, 2010). D'ailleurs, plusieurs auteurs mettent en avant la dimension linguistique dans la formation à l'écrit. Parmi les catégories qui font l'objet de la dimension linguistique on repère : l'accord orthographique, la manière d'étayer un avis critique, les collocations transdisciplinaires et la phraséologie scientifique, les anaphores résomptives, les pratiques énonciatives et citationnelle, etc.

Pour s'éloigner d'une conception instrumentale de la langue, la dimension linguistique devrait être pensée dans le cadre de la littératie qui prône une prise en compte globale des compétences liées à la lecture et à l'écriture envisagées comme des compétences cognitivo-discursives : reformuler, critiquer, problématiser.

Il se trouve cependant que les LU ne traitent presque pas certains phénomènes liés aux problèmes qu'elles cherchent à résoudre comme par exemple, les pratiques langagières des enseignants. Cela semble être l'objet d'une autre approche s'inscrivant dans le domaine du Français Langue Étrangère (FLE) et spécifique au contexte universitaire : le Français sur Objectifs Universitaires (FOU). Dans la section suivante, nous allons présenter cette démarche.

1.4 Français sur Objectifs Universitaires (FOU)

Développée dans de nombreuses universités françaises ainsi que dans plusieurs établissements à l'étranger, le FOU est issu du Français sur Objectifs Spécifiques (FOS) dont la méthodologie consiste à construire les programmes de formation linguistique au plus près des situations ciblées. Il a pour objectif de développer les programmes nécessaires à un public étranger voulant poursuivre ses études supérieures en français dans le monde universitaire francophone ou français. Nous trouvons au coeur de ces interrogations les besoins relatifs à la maîtrise des formes langagières du monde universitaire, académique et de spécialité, une compétence essentielle à la réussite des études universitaires.

Il s'agit de s'occuper des besoins des étudiants qui devraient s'acculturer aux comportements et aux modalités de travail d'un contexte institutionnel qu'il s'agisse de l'intégration administrative, de la transmission des connaissances ou de la validation des apprentissages (Mangiante et Parpette, 2012).

1.4.1 Méthodologie du FOU

L'élaboration d'un programme du FOU passe par l'étape centrale du FOS qui est la collecte de données dont l'analyse permet de déterminer les compétences langagières à développer chez les candidats à l'intégration universitaire, et d'en tirer les documents qui serviront de supports de formation.

Nombreux sont les informations qui structurent le déroulement des études et devraient être connues pour un étudiant universitaire. Ce dernier se trouve en contact avec le monde des savoirs disciplinaires, l'univers didactique (les consignes, les explications, les définitions, etc.), les tâches à effectuer (prise de notes par exemple), le système universitaire, etc. Le FOU s'attache à mettre ces réalités toutes en ligne de compte quasiment simultanément.

Les préoccupations pédagogiques essentielles du FOU ciblent le développement des compétences de réception et de productions écrites et orales des étudiants dans le but d'une meilleure maîtrise des discours universitaires et de la réalisation des tâches d'apprentissage attendues. En s'inscrivant dans une approche actionnelle, elle se centre sur l'apprenant et la réalisation de tâches dans le contexte de l'enseignement supérieur en français. La réception orale du cours doit aboutir à la production d'écrits qui répondent à des règles d'écriture précises souvent implicites. Ces règles déterminent des genres textuels universitaires qui constituent la majeure part de l'évaluation. Afin d'aider les étudiants à réussir leurs études, le FOU vise non seulement le développement des compétences transversales mais aussi la maîtrise des genres académiques. En effet, comme l'indiquent Mangiante et Parpette (2012), la démarche s'inscrit ainsi largement dans une perspective d'une acquisition de compétences linguistiques combinée à une acquisition de savoir-faire en situation, en l'occurrence de savoir-faire universitaires.

L'approche adoptée par le FOU vise à travers sa méthodologie un développement des compétences des étudiants allophones afin qu'ils maîtrisent mieux les modalités des discours écrits et oraux dans les universités francophones. Il s'avère alors qu'elle a des points de rencontre avec les LU en ce que les deux domaines s'intéressent à chercher comment appréhender l'écriture. Dans la section suivante, nous allons tenter de ressortir les points communs et les différences entre les LU et le FOU.

1.4.2 Le FOU en face des LU

Bien que les LU et le FOU ne s'adressent pas au même public (natifs vs étrangers), ces champs partagent des points communs. En effet, ils soulèvent des interrogations semblables autour de l'écrit universitaire et l'acculturation des étudiants à ses conventions ainsi qu'ils accordent une importance à la question des genres de discours universitaires et disciplinaires. De plus, leurs démarches sont similaires : ils adoptent une approche

pragmatique du contexte d'enseignement et de la situation d'apprentissage.

À la lumière de ce qui précède, la formation des étudiants est dirigée vers la préparation aux différents discours employés à l'université et plus particulièrement dans les disciplines. Les compétences langagières ne sont pas ciblées en tant que telles mais considérées comme des prérequis permettant la construction de ces discours.

D'un autre côté, au moment où les LU s'intéressent à la description de certaines dimensions liées à l'écriture académique dans le but de les clarifier aux étudiants et de permettre aux enseignants de sensibiliser les apprenants à mieux les appréhender, le FOU opte pour sa part pour une démarche plus pédagogique qui porte prioritairement sur les types de discours et la structure discursive qui en est dépendante. Nous voyons dans la première une approche théorique (description du savoir et de ses formes) et dans la deuxième une qui est moins théorique mais plus technique, applicative et interventionniste permettant de proposer les outils nécessaires aux étudiants. Ce sont sur des extrémités du triangle didactique (Lang, 2020).

Allier les deux champs des LU et du FOU nous permettrait de penser la production écrite des étudiants d'un point de vue centré sur la dynamique d'appropriation de l'étudiant dans le cadre des pratiques scripturales universitaires.

1.5 Bilan du premier chapitre

Ce chapitre commence par introduire la notion de la littératie à travers sa définition et son histoire. Nous avons présenté la diversité des champs d'application et les domaines qui y correspondent. Ensuite, nous nous sommes arrêtée sur la notion dans un contexte précis : à l'université. Le domaine des littératies universitaires (LU) se focalise principalement sur la description épistémologique de genres de discours en lien avec leurs contextes disciplinaires c.à.d. relation entre écriture, disciplines et constructions du savoir.

Le domaine des LU se rejoint avec celui du Français sur Objectifs Universitaires (FOU) notamment dans le processus d'enseignement-apprentissage de l'acculturation des étudiants universitaires à l'écrit ainsi que dans la recherche des difficultés rencontrées par les étudiants dans leur mise en pratique. Nous avons introduit, dans un deuxième temps, en quoi consiste le FOU qui s'intéresse au développement des habiletés langagières des étudiants en vue d'une meilleure insertion dans le monde universitaire. Cette démarche prend en considération les pratiques langagières de tous les acteurs à l'université (ensei-

gnant, institutionnel, etc.). À travers l'acquisition des méthodologies, le FOU cherche à développer les compétences discursives en "discours scientifique", entendu ici au sens de discours produit dans le cadre de l'activité de recherche à des fins de construction et de diffusion du savoir.

Le chapitre suivant va être l'occasion d'aborder plus profondément les discours scientifiques et les genres de discours à l'université. Nous allons ainsi parler de l'objet principal de notre étude et qui caractérise tout écrit scientifique : la phraséologie.

Chapitre 2

L'écrit scientifique

Contents

2.1	Écrit et apprentissage	19
2.2	Sciences et discours scientifique	20
2.3	L'écriture académique universitaire	21
2.3.1	Genres universitaires	21
2.3.2	Discipline et écrit	23
2.4	Lexique transdisciplinaire	24
2.4.1	La phraséologie	26
2.5	Bilan	28



Dans le cadre des littératies universitaires, nombre de dispositifs de formations sont proposées dans le but d'améliorer les écrits scientifiques des étudiants. Nous pouvons citer, à titre d'exemple, les formations aux genres universitaires, à l'écrit académique, à l'écrit professionnel, et à une maîtrise élémentaire de l'écriture. Jusque là, les approches adoptées dans de telles formations se trouvent normatives, basées sur la maîtrise du système linguistique et sont soupçonnées de technicisme. Elles investissent des activités résumantes telles que les plans, les résumés, les synthèses et les prises de notes. L'enjeu est comment aller au-delà de ces approches pour envisager l'écrit en lien avec l'appropriation des savoirs disciplinaires et le raisonnement.

Dans ce chapitre, nous allons présenter ce qu'est l'écrit académique et quelles en sont les genres et les caractéristiques. Ensuite, nous allons nous arrêter sur l'une de ses particularités : la phraséologie.

2.1 Écrit et apprentissage

Dans l'univers d'apprentissage disciplinaire (enseignement, recherche), l'écrit occupe une place primordiale. L'enseignant vise à ce que les étudiants soient capables de reproduire les connaissances transmises par lui, signe qu'ils les ont intégrées. Cela nous appelle d'être particulièrement attentifs pour tenir compte de l'usage du code écrit pour plusieurs raisons :

- L'écrit permet de conserver une trace stable d'évènement de pensée, par exemple : éléments d'observations, mesures diverses, résultats obtenus au terme d'une activité expérimentale.
- Le travail d'analyse et de réflexion a besoin d'un support écrit pour se développer et exige un retour permanent sur les notations déjà effectuées (tableaux statistiques, courbes graphiques, schémas, etc.). La forme écrite est le point de départ et le point d'aboutissement des activités des étudiants de même que le chercheur, quand il entame un travail de recherche, il reprend ce que l'on appelle la "littérature scientifique", la trace écrite, sur le sujet (publications antérieures, références méthodologiques, etc.)
- L'élaboration d'un objet scientifique à travers l'observation ne peut se faire sans

le recours au langage qui s'efforce de bannir, autant que faire se peut, toute appréciation personnelle liée à la sensibilité de l'observateur.

La formation à l'écriture et la formation par l'écriture (Rinck, 2011) sont les deux facettes du "faire écrire" au sein de l'université. Il existe une relation fondamentale entre la pratique de l'écrit à l'université et la construction des savoirs disciplinaires. L'écriture permet en effet de construire des savoirs dans toutes les disciplines. Certains auteurs ont montré que dans le contexte universitaire les aptitudes en production écrite constituent un facteur déterminant dans la réussite des études. (Rinck, 2011).

2.2 Sciences et discours scientifique

Un discours désigne un "ensemble de messages, textes et documents écrits, oraux, iconiques, sonores ou autres, qui s'échangent dans un domaine particulier" (Duverger, 2011). Nous abordons le terme "science", dans les *discours scientifiques*, dans un sens large incluant les sciences humaines et sociales (SHS) ainsi que les sciences exactes¹.

Tout domaine disciplinaire a ses discours qui possèdent des caractéristiques repérables : lexicale, structures récurrentes : temps, lieu, cause/conséquence, opposition, condition/hypothèse, variété discursive (narratif, argumentatif, injonctif, descriptif), reformulation. Tout discours scientifique est constitué d'un langage composite qui peut comprendre des éléments verbaux, numériques, algébriques, graphiques, cartographiques, schématiques, etc. Il est élaboré par des spécialistes d'un domaine - des experts - traitant une problématique clairement posée à laquelle ils tentent d'apporter des réponses. L'écrit scientifique remplit des critères qui renvoient à des pratiques intellectuelles spécifiques imposant des genres spécifiques codifiés fondés sur le raisonnement scientifique, les opérations d'analyse et le positionnement. Ces genres permettent de distinguer l'écrit scientifique des écrits journalistiques et littéraires qui se caractérisent par un langage descriptif, informatif, poétique ou métaphorique (Tran, 2014). Le langage formel de l'écrit scientifique se compose de traits textuels, pragmatiques, syntaxiques et lexicaux, ce qui pose de grandes difficultés pour les non spécialistes du domaine en question (Yan, 2017).

1. Les sciences qui sont purement formelles comme les mathématiques, ou sciences plus expérimentales comme les sciences physiques, les sciences naturelles, les sciences biologiques sont dites dures ou exactes à l'opposé des sciences humaines et sociales faisant partie des sciences molles

2.3 L'écriture académique universitaire

L'écrit académique est un genre de discours élaboré dans un contexte académique en milieu universitaire et qui relève d'un sous-domaine des écrits scientifiques. Il nécessite un certain apprentissage spécifique vu le nombre de traits caractéristiques sur les plans linguistique, rhétorique et argumentatif. L'importance de la maîtrise de ces écrits est liée au fait qu'ils restent le moyen d'évaluation privilégié des études supérieures et constituent donc un défi pour les étudiants étrangers en vue de valider leurs formations et obtenir un diplôme. Il nous semble que la production de textes écrits est une activité spécifique qui ne devrait pas se réduire à une forme de l'évaluation des connaissances. L'apprentissage de l'écriture scientifique est l'un des objets majeurs des enseignements disciplinaires. C'est dans ce sens que nous orientons notre étude.

2.3.1 Genres universitaires

De différents genres discursifs existent à l'université. Parmi ceux pratiqués par les étudiants on trouve : note de lecture, compte rendu, mémoire, thèse, rapport, examen écrit, etc. Ces écrits, à part leur "nom de genre", se distinguent par certains traits linguistiques particuliers, une finalité explicite et une certaine structure compositionnelle. À côté de genres à peu près stabilisés existent ce que l'on peut appeler à la suite des "écrits de transition" caractérisés par une consigne complexe : note de lecture associant résumé et commentaire, compte rendu et élaboration d'une problématique par exemple. Travailler sur ces textes permet de mieux comprendre les stratégies mobilisées dans des écrits longs de recherche en formation tel qu'un mémoire.

Ces différents écrits, bien qu'ils aient des enjeux différents, partagent la propriété commune qu'ils sont produits à destination d'enseignants dans un but d'évaluation. Par exemple, on rédige une thèse dans le but d'évaluer les résultats d'un travail de recherches mené en guise d'obtention d'un doctorat. De même, dans un examen écrit, les étudiants cherchent à exhiber un savoir présenté avec une certitude absolue dans le but d'une évaluation de connaissances acquises. Comme il est précisé dans le numéro rédigé par Garnier et al. (2015) autour de la formation à l'écrit, ces écrits possèdent des caractéristiques générales communes telles que : la posture de l'auteur (prise de position vs effacement critique), la conduite textuelle (élaboration d'un texte cohérent et convaincant), l'adaptation au lectorat (Garnier et al., 2015).

Les genres et les discours universitaires ainsi que les difficultés rencontrées chez les étu-

dians dans leur mise en pratique font l'objet d'études des LU. Ce champ avec celui du FOU tentent d'ailleurs d'articuler l'analyse linguistique de productions écrites et l'analyse des discours des acteurs, étudiants et enseignants. Il s'agit de voir les situations dans lesquelles apparaît l'écrit et quelles sont les pratiques et quelle place occupent celles-ci dans l'évaluation des acquis des étudiants. Le travail universitaire s'articule majoritairement sur la relation entre la compréhension orale des cours magistraux ou encore travaux dirigés et pratiques et la production d'écrits qui constituent l'objet de l'évaluation des étudiants.

Au vu de la place importante de l'écrit dans la formation universitaire, la maîtrise de la production écrite reste toujours au centre des discussions et des études. Dans le contexte universitaire français, les différents écrits produits répondent aux exigences méthodologiques, à une codification d'écriture, à des règles de composition qui génèrent de véritables genres textuels. Face à ces genres académiques, les étudiants étrangers sont face à plusieurs défis. Ils doivent rendre compréhensible leurs écrits, respecter les normes de langue écrite et gérer les stratégies rédactionnelles, ce qui implique un degré de maîtrise de la langue qui n'est pas la leur. Or, un texte bien écrit doit non seulement répondre aux exigences grammaticales et orthographiques, mais aussi à diverses prescriptions textuelles et socioculturelles. En effet, comme le mentionne Cavalla (2007), les étudiants doivent manipuler simultanément deux savoirs : le français en tant que langue étrangère et le savoir scientifique dont ils ne sont pas encore experts. Leurs besoins se situent sur quatre dimensions, à savoir les dimensions scientifique, méthodologique, terminologique et linguistique. À la dimension scientifique correspond le savoir disciplinaire, la dimension méthodologique aborde la structure des écrits universitaires, la dimension terminologique traite du lexique savant ; enfin, la dimension linguistique étudie les structures qui contribuent au sens.

Il est vrai que les besoins linguistiques sont les plus manifestants en SHS qui ont été davantage investiguées, mais cela n'empêche pas d'aborder les écrits dans des disciplines de sciences dures tel le cas de notre étude. En effet, à l'opposé de ce que Tran (2014) pense, il nous semble que la rédaction d'un écrit académique dans le domaine des sciences dures n'est pas plus aisée que dans le domaine des SHS. Bien qu'en SHS il y a intérêt à une manipulation plus exigeante de la langue, la structuration de l'écrit, la mise en relation de termes utilisés et le bon emploi des unités phraséologiques faisant partie de la langue qui véhicule la science restent tout de même une étape importante à franchir

particulièrement par les étudiants non natifs qui rédigent en langue étrangère.

Les différences disciplinaires sont aujourd'hui pointées comme essentielles, de même que la diversité des genres, qui renvoie à la diversité des activités dans le monde de la recherche et notamment à la question de savoir comment elles se situent par rapport à la logique de la découverte et la logique de l'exposition (Rinck, 2010).

2.3.2 Discipline et écrit

Chaque spécialité ou chaque discipline comprend ses propres spécificités et marqueurs linguistiques qu'il faudrait maîtriser dans le but non seulement de respecter les normes spécifiques d'un certain genre du discours mais aussi afin de véhiculer en cohérence la présentation des idées ou en d'autres termes, établir une bonne structuration sémantique. Nous rejoignons Delcambre et Lahanier-Reuter (2010) sur l'importance de penser le contexte des écrits produits à l'université comme disciplinaire. Les discours disciplinaires constituent un cadre d'élaboration et d'acquisition de savoirs spécialisés. Plusieurs recherches sur l'écriture à l'université ont mis en évidence la relation entre écriture et discipline. Par exemple, I. Delcambre (1998) montre comment l'écriture descriptive, des classes et des élèves, dans les mémoires professionnels des enseignants en formation est orientée selon la discipline : la description est objectivante dans les mémoires scientifiques avec des outils de type statistique, tandis que la description est réaliste du type portrait dans les mémoires de Lettres. Un autre exemple cité par Delcambre et Lahanier-Reuter (2010) explore les spécificités des articles de recherche dans les champs des Sciences du langage et de la Littérature (F. Rinck, 2005) écrits par des chercheurs experts ou débutants. Rinck montre comment les différences disciplinaires l'emportent largement sur les différences liées au statut du scripteur en mettant en évidence les différences entre les deux champs dans l'organisation structurelle des textes et dans l'emploi du lexique scientifique (termes référant à l'analyse, aux arguments, aux concepts, aux relations logiques d'effet et de conséquence). Les différences apparaissent également à l'intérieur du même champ disciplinaire. D'après une analyse des pratiques d'écriture en statistique dans deux contextes d'enseignement différents (scientifique ou en sciences humaines), Dominique Lahanier-Reuter (2003 a, 2003b) montre les variations importantes quant aux normes d'écriture, aux pratiques d'écritures entre ces deux contextes. Ces études diverses permettent de réfléchir au rôle important que jouent les contextes et les disciplines dans l'élaboration des écrits. Elles nous poussent ainsi à orienter l'analyse des difficultés des

étudiants vers la prise en compte de l'inscription disciplinaire des genres écrits auxquels ils sont confrontés.

Dans ce cadre, nous allons étudier les genres académiques qui apparaissent en physique, une des disciplines scientifiques qui fait partie du domaine des sciences naturelles se consacrant à étudier les lois fondamentales régissant la nature qui nous entoure. En particulier, nous nous intéressons à l'écriture des rapports ou comptes rendus d'un Travail Pratique (TP). L'écriture des étudiants sera étudiée dans le contexte qui se lie à la production écrite des enseignants à travers les supports de cours rédigés. Nous allons revenir en détails sur la description de ces genres ciblés dans le chapitre 4. À notre connaissance, la plupart des études portant sur l'écriture scientifique ont été menées en SHS et rares sont les études qui ont pris en compte les disciplines de sciences exactes.

Les écrits académiques dépendent de la discipline étudiée comme nous avons déjà expliqué et par suite comprennent un certain éventail du lexique de spécialité. Bien que les caractéristiques les plus saillantes du discours scientifique apparaissent sur le plan lexical, dans notre étude, nous nous focalisons sur le lexique des écrits scientifiques dans une perspective large au niveau de la construction de la phrase. La section suivante introduit les différents types du lexique qui existent dans un écrit et délimite notre choix.

2.4 Lexique transdisciplinaire

Le discours académique répond à des contraintes de productions discursives spécifiques (faire une introduction, exposer un dispositif de recherche, citer ses sources...). D'où l'importance de travailler la compétence discursive en FOU. Mais si cette compétence transphrastique (au niveau du texte) est un préalable nécessaire à la rédaction de discours académiques, elle ne suffit pas à la production de l'écrit au niveau micro de la phrase. Plusieurs auteurs s'accordent sur l'existence d'une Langue Scientifique Générale. Nous citons dans ce cadre Pecman qui s'est intéressée dans sa thèse publiée en (2004) aux unités phraséologiques de ce qu'on appelle "*la langue transdisciplinaire générale (LSG)*" définie comme "*Pratique langagière spécifique à une communauté de discours composée de chercheurs en sciences exactes dont les objectifs communicatifs poursuivis émanent des préoccupations partagées par des scientifiques à travers le monde et indépendamment*

de leurs spécificités disciplinaires”. Elle a fait son étude sur la LSG dans une perspective contrastive anglais-français qu’elle décrit à partir d’un corpus en physique, chimie et en biologie.

D’autres travaux se fondent sur des corpus venant de disciplines SHS tels que ceux menés par Tutin (2007a). Selon elle, l’écrit est composé de cinq strates de lexique :

1. Lexique propre aux écrits scientifiques
2. Lexique abstrait général
3. Lexique méthodologique disciplinaire lié à la spécialité
4. Lexique terminologique
5. Lexique de la langue générale

Donc, à côté du lexique de spécialité, on voit apparaître un lexique transversal largement partagé par les disciplines et qui correspond à la scientificité de l’écrit. Nous allons particulièrement nous intéresser à ce type de lexique. Ce choix se justifie dans notre objectif didactique. Il s’agit du lexique qui décrit les activités scientifiques : de l’exposition du cadre de la recherche aux résultats en passant par la méthodologie et l’argumentation. Les unités lexicales de ce lexique ne désignent pas véritablement des éléments disciplinaires, mais elles renvoient à la démarche et aux activités scientifiques, au raisonnement et à l’écriture scientifiques.

Donc, aux genres de discours scientifiques, on associe un lexique appelé «lexique scientifique transdisciplinaire » (LST) (Tutin, 2007b) et une phraséologie scientifique transdisciplinaire (Cavalla, 2009 ; Tutin et Grossmann, 2014). Les éléments du lexique sont “transdisciplinaires” en ce qu’ils traversent les disciplines et correspondent au lexique stable du genre textuel, par opposition au lexique disciplinaire propre à une sphère scientifique particulière. Nous retrouverons ainsi dans la plupart des disciplines — des sciences humaines aux sciences les plus “dures” - des lexèmes comme *hypothèse*, *analyser*, *résultats*, mais aussi des expressions comme dans un premier temps, *pour conclure*, qui sont étroitement liées à la démarche scientifique et à l’écriture scientifique. À côté des mots simples, il convient d’intégrer dans ce lexique la phraséologie, qu’il s’agisse d’expressions complètement figées comme “*point de vue*” ou “*prendre en compte*”, mais aussi les collocations comme *résultats encourageants*, *analyse approfondie* ou *jouer un rôle*, qui correspondent à des associations lexicales binaires privilégiées.

Nous allons présenter dans la section suivante en quoi consiste plus précisément le terme de phraséologie.

2.4.1 La phraséologie

La phraséologie, “placée à mi-chemin entre le lexique et la syntaxe [...], s’occupe des signes polylexicaux composés d’au moins deux mots, stables, répétés et souvent figurés” (González-Rey, 2010). Elle correspond donc à l’étude des segments répétés, des concordances, collocations entre plusieurs unités lexicales. Ces segments figés, ou chunks, constitueraient près de la moitié du discours d’un locuteur natif (Cavalla, 2009). Ce sont ces “prêts-à-parler”, si difficiles à acquérir, notamment pour un locuteur non natif. D’où il est nécessaire de mettre l’accent sur l’importance d’un enseignement/apprentissage plus systématique de ces unités phraséologiques, essentiellement des séquences récurrentes.

Les éléments phraséologiques représentent l’ensemble des unités de type figé et comprennent une variété d’éléments tels que : les collocations (Hausmann, 1979) et les routines langagières (Firth, 1957; Sinclair, 1991; Tutin, 2010). Elles partagent certaines caractéristiques et diffèrent par d’autres. Dans ce qui suit, nous allons discuter ces deux éléments.

2.4.1.1 Collocations vs routines langagières

Les collocations et les routines langagières sont largement employés dans les écrits et suscitent l’intérêt des chercheurs en linguistique vu le pan non négligeable qu’ils représentent dans le lexique.

Les routines langagières représentent l’ensemble des unités lexicales dont le sens est conservé et qui se trouvent d’une façon fréquente dans un écrit telles que par exemple : *des résultats intéressants*. On peut trouver une compositionnalité sémantique, des insertions lexicales entre les éléments, par exemple : *des résultats très intéressants*, et des substitutions quasi-synonymiques comme : *des résultats importants* (Cavalla, 2018a).

Par contre, une collocation, est une association lexicale figée de deux éléments formée des éléments identifiés selon la combinaison : *base+collocatif*, c’est un ensemble de lexies en bloc qui, selon Cavalla, “ne permet pas la compositionnalité des deux éléments associés”. La “*base*” est l’élément porteur du sens de l’unité phraséologique, le “*collocatif*” est celui dont le sens est modifié. Dans l’exemple “une peur bleue”, cité dans (Cavalla, 2018a), la base est le mot “peur” qui conserve le sens habituel connu par les locuteurs non natifs, tandis que “bleue” ne renvoie pas dans cette expression à la couleur, mais signifie que la peur est “très intense”, un sens qui est non prévisible par un apprenant de français L2

mais connu par les locuteurs natifs. C'est ainsi qu'on attribue à ce dernier élément qui est modifié sémantiquement, la position d'un *collocatif*.

Les collocations se divisent en plusieurs catégories (Cavalla, 2007) :

- Les collocations générales qui sont aussi présentes dans la langue générale que dans les articles scientifiques telles que : *remettre en question, prendre en compte, mener à bien, prendre du recul, prendre contact, etc.*
- Les collocations spécifiques que nous trouvons notamment dans des écrits scientifiques : *émettre/avancer une hypothèse, avancer un postulat, réfuter une thèse, passer en revue, remettre en question, mener l'expérience, etc.*

Les étudiants francophones ayant une langue maternelle différente du français ou en d'autres termes, les locuteurs non natifs (LNN) manifestent certaines difficultés face aux écrits universitaires telles que celle d'ordre linguistique liée à la non maîtrise des formes discursives, souvent d'ordre phraséologique (Cavalla, 2018a). De telles difficultés ne peuvent pas être comblées par les dictionnaires électroniques, outils disponibles à leurs portées, proposant des traductions de mots isolés sans prendre en compte le contexte. D'où provient la nécessité d'aborder l'enseignement de la phraséologie avec les étudiants universitaires afin d'améliorer leurs écrits scientifiques. Nous allons nous intéresser par la suite à la phraséologie spécifique aux écrits universitaires en Physique et qui contribue à la structuration linguistique des écrits en question.

2.4.1.2 Importance et choix de la phraséologie

La phraséologie, récurrente dans différentes disciplines, joue des rôles clés en ce qu'elle permet tout d'abord à l'auteur/chercheur de renforcer son appartenance à une communauté de discours, avec l'emploi d'un langage "codé" (nous écartons l'hypothèse, etc.), et elle facilite, par son caractère préconstruit, l'écriture (Tutin, 2014). Dans le cadre didactique, la spécificité des écrits scientifiques, à travers la phraséologie, a été étudiée dans la perspective de l'enseignement universitaire comme par exemple, les travaux du Français sur Objectifs Universitaires (FOU) en direction des étudiants étrangers devant s'acculturer à la langue universitaire française (Cavalla, 2010).

Dans cette perspective, nous allons étudier la phraséologie qui peut être employée dans écrits académiques du type compte rendu d'un travail pratique (TP) en physique. Nous allons viser plus particulièrement la phraséologie relative à la notion repérée dans la LSG

“*expérience* (Pecman, 2004), et celles qui font partie du LST “*hypothèse*” et “*résultats*” (Tutin 2007a, Tutin 2007b, Tutin 2014, Cavalla 2007, Cavalla 2010, Cavalla 2018a) afin d’étudier leurs emplois et usages variés dans les disciplines de sciences dures et les disciplines des sciences humaines et sociales.

Nous pensons que le fait de travailler la phraséologie à un niveau d’étude tel que la Licence sensibilise les étudiants à l’emploi des expressions convenables dans la discussion de leurs futurs travaux de recherche dans des écrits plus présentatifs tels que le mémoire demandé en Master. L’enseignement de ces moyens lexicaux est susceptible d’aider les apprenants à répondre aux “normes stylistiques d’un genre particulier” qui consiste non seulement en “les choix grammaticaux ou lexicaux, mais aussi la sélection des unités polylexicales d’une manière appropriée” (Tran, 2014).

2.5 Bilan

Dans ce chapitre, nous avons présenté l’écrit académique, l’objet principal vers lequel nous orientons notre étude. Dans notre cas, il s’agit de comptes rendus des travaux pratiques rédigés dans une discipline venant des sciences dures : la physique.

Plus particulièrement, nous nous focalisons sur la phraséologie employée dans ce type d’écrit. Nous avons également délimité les notions lexicales qui nous intéressent et qui font partie du LST et/ou de la LSG.. Notre étude se focalise alors autour des collocations et des routines langagières construites à partir de ces notions. En raison des spécificités de ce lexique, les apprenants rencontrent souvent des erreurs d’ordre syntaxique, sémantique et organisationnel. Les raisons sont multiples, mais la raison principale est relative au manque d’enseignement systématique de ces éléments phraséologiques malgré leur importance dans le discours. Un enseignement de ces unités sera bénéfique pour les étudiants afin de leur donner plus d’aisance dans l’écriture et afin de leur permettre de mieux se positionner et se démarquer au sein de la communauté scientifique.

Nous allons par la suite, présenter le contexte de notre étude à savoir l’université et le public concernés, ainsi que l’exposition au français au sein de la classe.

Deuxième partie

Contextualisation de la recherche

Chapitre 3

Contextualisation de la recherche

Contents

3.1	Terrain de recherche	33
3.2	Public concerné	35
3.2.1	Compétences langagières	35
3.2.2	Autour du niveau B2	36
3.2.3	Culture éducative des étudiants	38
3.3	L'exposition au français à l'UL	39
3.3.1	DdNL en enseignement bilingue	39
3.3.2	L'écrit et l'oral en classe	40
3.4	Bilan du contexte	41



Le contexte visé à travers cette étude comprend des étudiants libanais à l'université. Selon les chiffres publiés, la France se positionne aujourd'hui au premier rang des destinations des étudiants libanais en mobilité¹, pour plusieurs raisons : l'intérêt culturel du pays, l'apprentissage d'une langue en immersion ainsi que la qualité de la formation universitaire. En 2017/2018, on dénombre 5665 étudiants libanais en France, soit 27% de la totalité des étudiants libanais à l'étranger.

Le système éducatif au Liban repose sur un enseignement bilingue et est caractérisé par la liberté de choix de la langue d'enseignement. Nous entendons par l'enseignement bilingue un enseignement où les deux langues de travail pour les élèves sont d'une part la langue maternelle (L1) qui est la langue officielle du pays (l'arabe dans notre cas), et d'autre part une langue étrangère (L2). Les L2 d'enseignement au Liban sont le français et l'anglais. Le volume horaire d'enseignement scolaire en L2 varie selon le secteur d'enseignement, le cycle et la discipline. Par exemple, dans un contexte scolaire, le français ou l'anglais restent les langues véhiculaires des disciplines scientifiques (Maths, Sciences de la vie et de la terre, Chimie, Physique et Informatique). La géographie, l'histoire et l'éducation civique sont assurées en arabe.

Le français reste très largement une langue de scolarisation et il s'avère que le niveau des élèves en français évolue avec l'âge mais reste relativement bas en particulier dans les établissements publics compte tenu du degré d'exposition à la langue hors du cadre scolaire (dans un cadre social et/ou familial par exemple).

Dans ce chapitre, nous allons présenter le terrain à exploiter dans le cadre de notre recherche à savoir l'université libanaise, nous allons ainsi décrire le public concerné en revenant sur ses compétences langagières, et sur l'exposition au français au sein de l'établissement universitaire en question tout en s'arrêtant sur la façon dont s'y déroule l'enseignement disciplinaire.

3.1 Terrain de recherche

Parmi les établissements d'enseignement supérieur au Liban, on trouve l'Université libanaise (UL), la seule université publique jusqu'à nos jours. Elle est fondée en 1951 dans le but d'assurer la formation des enseignants et instituteurs dans son école normale

1. <https://lb.ambafrance.org/Tout-savoir-sur-la-hausse-des-frais-d-universite-en-France>

supérieure. Durant les années 1950, elle s'est développée et plusieurs facultés ont vu le jour (en 1959 : faculté des sciences sociales, faculté de droit et faculté des sciences ; en 1965 : institut des beaux-arts). À l'époque, le Liban représentait un centre intellectuel qui absorbait des flux significatifs d'étudiants de la région, les pays arabes voisins et ceux du Golfe n'ayant pas encore instauré leurs propres institutions d'enseignement supérieur (Nahas, 2009).

L'UL possède l'effectif d'étudiant le plus élevé, malgré une baisse significative due à l'émergence de petites universités privées (67 309 étudiants ou 34,6 % du total des étudiants pour l'année 2015, contre 49,5 % en 2004 et 60 % en 2000) (Gharib et Hamdan Saadé, 2017). Le même système éducatif scolaire se perpétue dans cet établissement dans le sens où les matières scientifiques qui ont été enseignées à l'école en langue étrangère -en français notamment- seront également enseignées à l'université dans la même langue. Pratiquement, la faculté des sciences de l'UL offre la possibilité aux étudiants de poursuivre leurs études selon leur L2 au secondaire. Les mêmes cours sont dispensés en français et en anglais, soit par deux professeurs soit par le même professeur.

Le présent travail se déroule à la faculté des Sciences de l'université Libanaise de Beyrouth². Cette faculté regroupe des classes de Licence et de Master de toutes les filières de sciences naturelles qu'elles soient des sciences exactes ou expérimentales : Les disciplines des mathématiques, informatique et statistiques dépendent du département des mathématiques pures et appliquées. La Biologie, la chimie et la biochimie sont liées au département des sciences de la vie et de la Terre. Quant à la physique et l'électronique, elles dépendent du département intitulé "Physique et électronique". Plus particulièrement, au niveau microsocial, nous ciblons les séances d'un Travail Pratique (TP) où les étudiants ont à réaliser et à observer des travaux expérimentaux en Physique nucléaire et atomique et où à l'issue de ce travail, un écrit du type compte-rendu rédigé en français est attendu.

Dans la section suivante, nous allons présenter ce qui caractérise le public en question et quelles sont ses compétences langagières.

2. L'université libanaise comprend huit facultés des sciences répartis sur l'ensemble du territoire libanais. Notre travail a eu lieu à Beyrouth, la première faculté établie.

3.2 Public concerné

Le public auquel nous nous intéressons dans la présente étude, est constitué d'un ensemble d'étudiants en troisième année de Licence en Physique générale. Ils ont suivi un enseignement bilingue dès la maternelle avec le français comme première langue étrangère (L2) et l'arabe comme langue maternelle (L1). Pour la majorité, le français apparaissait comme une simple discipline scolaire. Qui plus est, les étudiants de l'UL, issus généralement des milieux sociaux moyens et défavorisés, ont suivi un enseignement dans des institutions publiques où l'enseignement du français est loin de donner entière satisfaction. La représentation qu'ils en avaient est plutôt négative dans la mesure où, à son inutilité apparente et à leur difficulté d'apprentissage, s'ajoute qu'il constitue assez souvent une cause d'échec scolaire.

Désormais, le regard négatif envers la langue française se change. À l'université, les étudiants ont un regard positif envers l'apprentissage de la langue française. Elle est considérée comme un moyen indispensable d'apprentissage des ses cours universitaires. La maîtrise du français va leur permettre non seulement de réussir son parcours mais aussi d'être ancrés dans la réalité de la science dans le monde francophone et de se lancer dans une recherche scientifique dans le futur.

3.2.1 Compétences langagières

Les cours de langues s'invitent à la faculté des sciences de l'UL aux deuxième et troisième années de la Licence. Les étudiants passent un test d'aptitude à la fin de leur première année. En se basant sur un classement à trois niveaux (Level 1, Level 2 et Level 3), le test permet l'affectation du niveau de maîtrise du français et détermine le cours de langue auquel les étudiants devraient s'inscrire. Les étudiants qui n'ont pas eu la note qui correspond au niveau minimal (Level 1) suivent, pendant l'été, un cours intensif de remise à niveau avant qu'ils intègrent leur deuxième année universitaire. Par la suite, tous les étudiants seront obligés de suivre, parallèlement au cursus universitaire, le cours de langue qui leur est proposé à raison d'une séance de deux heures par semaine. Au cours de chaque semestre, ce cours est évalué dans un examen partiel et un autre final en évaluant les compétences à l'écrit ainsi qu'à l'oral. Il est obligatoire d'y assister d'autant plus que l'obtention du diplôme est conditionnée par la validation du cours de langue. En troisième année de la Licence, les étudiants sont censés être au niveau le plus haut

selon l'échelle de classement adopté à l'UL c.à.d. le niveau Level 3. Nous pourrions dire en d'autres termes que ces étudiants ont un niveau B1 en vue d'obtention du B2 selon la description du Cadre Européen Commun de Références des Langues (CECRL). Cela leur permettrait d'intégrer les universités françaises dont la majorité exige un niveau B2 pour l'admission aux études scientifiques³.

Pour cette étude, nous allons considérer que les étudiants de la troisième année de licence ont le niveau B2, niveau ciblé en fin de compte par l'université. Nous allons explorer brièvement ce que ce niveau exige comme compétences en productions écrites.

3.2.2 Autour du niveau B2

Le niveau B2 correspond à un niveau intermédiaire. Il vise selon le CECRL à rendre compte des spécifications de l'utilisateur indépendant et il exige une argumentation efficace : rendre compte de ses opinions de façon claire en apportant des explications appropriées, des arguments et des commentaires, développer un point de vue sur un sujet en soutenant tour à tour les avantages et les inconvénients des différentes options, construire une argumentation logique, évaluer les choix possibles, faire des hypothèses et y répondre.

En termes d'aptitudes pour la production écrite, le niveau B2 exige des compétences à la fois méthodologiques et discursives faisant preuve d'une capacité à rédiger un texte *clair et détaillé*, sur la base d'une *synthèse d'informations et d'arguments* provenant de sources diverses :

Peut écrire des textes clairs et détaillés sur une gamme étendue de sujets relatifs à son domaine d'intérêt en faisant la synthèse et l'évaluation d'informations et d'arguments empruntés à des sources diverses. (CECRL, Conseil de l'Europe, 2001)

Le CECRL propose une seconde échelle pour les types d'écrits attendus au niveau B2 tels que les rapports et les essais, genres discursifs rencontrés à l'université. Au premier palier, nous trouvons des compétences plus générales où l'accent est mis sur la façon méthodologique de l'argumentation qui devrait être étayée par les points importants. Les compétences plus ciblées présentées au second palier demandent un passage au-delà

3. En France, le niveau C1 en français est obligatoirement requis pour toutes les formations en master relevant des Unités de Formation et de Recherche (UFR) de lettres, arts, communication et sciences du langage.

d'une simple argumentation : justifier, préciser son point de vue et son opinion, expliquer et synthétiser.

En complément aux deux échelles déjà présentées, le CECRL donne également les attentes en termes de compétence discursive. Au sein de cette dernière, ils se trouvent "*une connaissance de l'organisation des phrases et de leurs composantes et une capacité à les maîtriser*" (enchaînements, thème/rhème, etc.), "*une capacité de gérer et de structurer le discours*" (organisation thématique, cohérence et cohésion, style et registre, etc.), et "*une capacité à structurer le plan du texte*".

En Europe, les enseignants, les établissements d'enseignement supérieur, les employeurs et organisations gouvernementales s'appuient régulièrement sur le CECRL pour prendre des décisions. Pour intégrer une université, avoir atteint le niveau B2 constitue la norme. À leur arrivée en France, les étudiants allophones se trouvent face à des professeurs qui parlent la langue du pays et ils doivent suivre leurs cours en cette langue qui n'est pas leur langue maternelle. Les tests utilisés pour déterminer un niveau en langue, évaluent un français général. Mangiante et Parpette pointent les taux de réussite des étudiants allophones inférieur de 40% à ceux des natifs. Les recherches menées dans ce sens (Coulon et Paivandi, 2003) expliquent ce taux d'échec par la grande difficulté rencontrée à l'université qui est l'utilisation de la langue française. Il s'avère que le taux d'échec des étudiants allophones est moins important lorsqu'ils ont le niveau B2 ou plus. D'ailleurs, plusieurs types d'écrits sont attendus des étudiants chacun selon sa discipline et selon son année universitaire. C'est le cas également des étudiants de l'UL.

Le français ciblé dans cet établissement est plutôt un français général qui permet aux étudiants d'obtenir le niveau B2 et d'intégrer, s'ils le veulent, des universités francophones. Il ne vise pas particulièrement à développer des compétences dans l'écrit académique. La faculté des sciences offre aux étudiants une formation qui est essentielle en langue française mais il reste encore du travail à accomplir.

La réussite à des tests standardisés de type DELF utilisés pour déterminer un niveau en langue, et qui évaluent un français général, ne sont pas garants de la réussite universitaire, et cela est expliqué par le fait qu'il existe différents types de français pour différents objectifs. Des compétences en matière de réception (écrite et orale) comme en matière de production (écrite et orale) entrent dans les conditions de réussite universitaire. La maîtrise du français reste donc une condition nécessaire mais non suffisante à la compréhension ou à la réussite universitaire, et l'élément majeur à prendre en compte

est celui du contexte.

Comme le précise Yan (2014), le taux d'échec des étudiants étrangers inscrits en SHS est deux fois supérieur à celui des Français, ce qui est essentiellement imputable aux compétences langagières des étudiants (Mangiante & Parpette, 2011).

Il nous semble alors pertinent d'intervenir, tout en alliant les principes des LU et du FOU, afin de combler les difficultés des étudiants liées à l'écrit académique, notamment en matière de ce qui nous intéresse dans cette étude, la phraséologie.

Après avoir présenté les compétences langagières ciblées à travers les cours de langue à la faculté des sciences de l'université libanaise, nous allons procéder pour parler de la culture éducative des étudiants, leur emploi de temps et leur attitude vis-à-vis l'apprentissage.

3.2.3 Culture éducative des étudiants

Pendant le deuxième semestre, les étudiants que nous ciblons ont à valider six cours. Selon eux, ils ont un emploi du temps assez chargé et par la suite, ne travaillent pas leurs cours de façon quotidienne. Même lorsque les travaux dirigés leur sont distribués une semaine avant la résolution, la plupart des étudiants admet qu'ils n'ont pas trouvé le temps pour réfléchir au sujet des exercices ni à la façon de les résoudre.

Mis à part les plus motivés, ils attendent la période des examens pour réviser le cours. Pour eux, l'essentiel est de réussir et avoir la moyenne. Le manque de temps n'est pas la seule raison mais aussi parce que les connaissances acquises sont soumises à deux examens écrits (partiel et final) et sont les seules formes d'attribution de notes. Les professeurs n'ont pas à évaluer ni la participation ni les rendus des travaux dirigés (TD).

Concernant leur apprentissage, les étudiants sont plongés dans les équations et les formules mathématiques, que certains professeurs demandent de retenir par cœur, de telle façon que lorsqu'une définition ou un historique leur sont présentés, ils posent directement la question : "est-ce que cela est à retenir par cœur ou à comprendre?". Notons que le fait de comprendre une idée permet de l'exprimer sans difficultés particulières. Mais, c'est l'expression en français qui s'avère dans certains cas compliquée ne laissant aux étudiants que le recours à la mémorisation pour plusieurs raisons : soit qu'ils craignent que le professeur n'accepte pas une autre formulation que la sienne soit à cause d'un manque du lexique correspondant.

Dans les sections suivantes, nous allons discuter le degré d'exposition des étudiants au français dans une classe de discipline au sein de l'UL. Nous allons également décrire l'alternance entre les langues maternelle et seconde, à l'écrit et à l'oral par les différents acteurs : étudiants et enseignants.

3.3 L'exposition au français à l'UL

Dans un tel contexte d'apprentissage hétéroglotte, les apprenants n'ont pratiquement pas d'occasions de pratiquer le français en dehors de la classe à l'université dans des situations réelles de communication. Par exemple, ils utilisent l'arabe lorsqu'ils s'adressent à leurs professeurs, sollicitent des clarifications, demandent un renseignement à l'administration, s'échangent entre eux, etc. Dans le contexte de l'université libanaise, seuls la classe de langue et les cours des disciplines constituent des situations où la L2 est plus ou moins employée par les étudiants. Mais qu'en est-il des professeurs enseignants de la discipline ?

3.3.1 DdNL en enseignement bilingue

Les disciplines enseignées à la faculté des sciences relevant des sciences (plus ou moins) exactes et expérimentales peuvent être appelés "disciplines non linguistiques". Cette appellation est d'après certains inappropriée (Duverger, 2011). Selon eux, il est plus convenable de parler de discipline "dites" non linguistiques (DdNL). Nous utiliserons dans ce travail cette dernière dénomination comme toutes les disciplines utilisent des langues pour être enseignées et, qui plus est, des variantes spécifiques et originales de ces langues. Le professeur bilingue de la DdNL est un nouveau type de professeur. Il enseigne sa discipline dans une langue qui n'est ni sa langue maternelle ni celle des élèves.

L'enseignement en L2 permet d'améliorer le rendement de l'apprentissage de cette L2 chez les étudiants dans la mesure où il permet d'utiliser la L2 de manière naturelle ici et maintenant au travers des cours. Il doit permettre également d'améliorer les savoirs disciplinaires enseignés en cette langue.

Les textes officiels relatifs à l'organisation de l'UL stipulent que l'arabe est la langue d'enseignement et laissent la latitude de recourir à d'autres langues que l'arabe en cas de besoin. "Dès lors, le choix de la langue d'enseignement peut, dans une certaine mesure, être déterminée par les *convictions des responsables* des facultés ou des sections d'une

part et par les impératifs liés à la disponibilité d'enseignants de formation francophone, d'autre part" (Hoyek, 2004). Il est à noter que la majorité des professeurs à l'UL ont fait leurs études supérieures y compris le doctorat en France et sont titulaires de diplômes français.

Nous allons voir par la suite comment s'utilisent les langues et à quel degré dans une classe de DdNL à l'UL.

3.3.2 L'écrit et l'oral en classe

Le professeur parle seul plus ou moins longtemps ou même durant tout le cours. Puisque la langue maternelle des libanais est l'arabe, les professeurs sont amenés à employer oralement l'arabe à plusieurs reprises en salle de classe. Ils font fonctionner le français (la L2 des étudiants et des professeurs) en évoquant des termes et notions scientifiques. Les langues mobilisées ne sont pas identiques selon que le professeur renvoie ses élèves au manuel, qu'il donne des instructions ou qu'il commente des résultats. A priori, rien n'empêche l'alternance des langues et l'enseignement se fait partiellement en L2.

L'organisation de la classe en DdNL ne prend pas la forme strictement orale entre le professeur et ses élèves mais elle est constituée d'une combinaison d'éléments écrits et oraux. L'écrit est là : titre, plan et notes écrites par le professeur au tableau, des activités, des données, résumés lus, énoncés d'exercices, documents de manuels, etc. La prise de notes par les apprenants est affaire personnelle, puisque celle-ci correspond à des stratégies propres à chacun. Selon la variation du support et les situations d'activités, les formes de la langue pourront connaître d'importantes variations mais, le français demeure la seule langue employée à l'écrit académique que les étudiants doivent rendre aux professeurs. Les compétences du professeur de la DdNL en matière de compréhension écrite doivent être excellentes (il devrait comprendre parfaitement des manuels rédigés en L2), ses compétences peuvent être en revanche moindres à l'oral. Le professeur, selon ses compétences, met en place des formes de communication appropriées.

Dans une discipline telle que la physique, les étudiants devraient être capables de verbaliser un langage mathématique qui nécessite d'articuler et de formuler les différentes étapes d'un raisonnement, de communiquer le résultat et de commenter à propos de la solution. Qui plus est, relevant des sciences expérimentales et exactes, le langage verbal sert à mettre en place des situations d'observation, des situations expérimentales, à

commenter des données numériques, ainsi qu'à présenter et commenter des résultats.

3.4 Bilan du contexte

Ce chapitre présente le contexte de notre recherche à savoir l'université libanaise et le public spécifique. Nous nous intéressons particulièrement aux étudiants libanais de premier cycle qui ont à rédiger un écrit scientifique rapport ou compte-rendu de TP et qui, dans le futur, sont amenés à présenter des écrits du type rapport de recherche, mémoire, thèse ou article de recherche. La plupart des étudiants en Licence 3 sont censés atteindre un niveau B1/B1+ en français général mais il s'avère que ce n'est pas suffisant pour la poursuite des études supérieures dans des universités francophones. En effet, les étudiants allophones sont souvent confrontés à des difficultés en français académique notamment d'ordre rédactionnel. Afin de mieux cerner les réflexions à mener par la suite, nous avons analysé la culture éducative des étudiants, l'emploi des langues au sein de la classe que ce soit par l'enseignant ou par les étudiants et le rapport qu'entretient chacun d'eux avec la langue.

Nous allons passer maintenant au chapitre 4 où nous présentons les données utilisées dans notre étude et où nous détaillons les genres académiques qui vont être l'objet de l'analyse à mener.

Chapitre 4

Méthodologie et recueil des observables

Contents

4.1	Linguistique de corpus	45
4.2	L’usage de corpus informatisés	47
4.2.1	Corpus Scientext	48
4.2.2	Base de données LST	49
4.3	Constitution des corpus universitaires	50
4.3.1	Choix de la discipline	51
4.3.2	Choix des genres	51
4.3.3	Choix de sources	54
4.4	Collecte de données	56
4.4.1	Les cours retenus	57
4.4.2	Recueil des CR	58
4.5	Outils d’extraction des collocations	59
4.5.1	ScienQuest	60
4.5.2	Extractions dans l’interface LST	61
4.5.3	Extraction dans les corpus universitaires	62
4.6	Bilan de la méthodologie	62



Si la stratégie d'internationalisation des universités, par la création de parcours en partie ou totalement en anglais, semble au premier abord menacer le français comme langue des sciences et techniques, le développement du FOU pourrait contre-balancer ce phénomène, d'autant plus si le FOU, initialement axé sur la méthodologie universitaire dans une perspective généraliste, s'adapte aux besoins langagiers de chaque discipline, grâce à l'approche basée sur les données disciplinaires. Le passage du FOU au Français Langue Académique, autrement dit de la méthodologie universitaire à l'écriture académique, permet d'affiner progressivement les compétences langagières des (futurs) chercheurs.

Dans un nouveau jeu d'enchâssement, la phraséologie scientifique se spécifie dans chaque discipline, le discours académique se déclinant en autant de genres que de filières, chacune ayant sa propre terminologie (médecine, sociologie, lettres, géographie, etc.), voire en autant de sujets de recherche. C'est à ce niveau-là que l'approche par corpus d'experts peut répondre de manière très fine aux besoins spécifiques des (futurs) chercheurs.

Afin de se familiariser avec les discours spécialisés et les genres, nous allons examiner plusieurs corpus académiques rédigés en français, relevant de genre et de disciplines différents. Le but n'est pas tant pour lire leur contenu que d'y observer le fonctionnement de la phraséologie.

Dans ce chapitre, nous allons présenter et décrire la construction d'une méthodologie adaptée à nos objectifs de recherche, à commencer par justifier le choix de l'approche sur corpus adoptée pour l'étude de la phraséologie. Ensuite, nous allons présenter le choix des corpus sur lesquels nous nous sommes appuyée dans cette étude. Enfin, et après avoir expliqué la construction des corpus universitaires, nous abordons le choix des unités phraséologiques (UP) et la question de leur exploitation .

4.1 Linguistique de corpus

Notre approche relève de la linguistique de corpus qui facilite le développement de la compétence discursive et plus encore phraséologique, occupant une place de premier plan dans le discours scientifique, tout en accompagnant de ce fait l'émergence d'un nouvel objet d'enseignement-apprentissage : le français scientifique (trans)disciplinaire.

Les corpus sont devenus un élément essentiel de la recherche en linguistique. Fabre (2001)

les définit en mettant l'accent sur leur finalité, qui est un paramètre clé dans la constitution de corpus : "Les corpus sont des collections de données textuelles rassemblées dans un objectif précis." Ils fournissent le matériel authentique pour découvrir la langue en particulier, des corpus académiques écrits tels que le recueil de textes scientifiques, d'extraits de productions universitaires, mémoires, thèses, etc. permettant de décrire les phénomènes linguistiques tels que la phraséologie de manière plus précise et plus complète. L'exploitation de ces corpus académiques à des fins didactiques a fait l'objet de plusieurs études relevant de l'analyse des discours universitaires et de leur exploitation pédagogique, sous forme d'applications de la linguistique de corpus par des spécialistes et leur réception par les utilisateurs (Cavalla et Hartwell, 2018), ou sous forme d'analyse des énoncés récurrents produits par les experts et les étudiants dans leurs écrits académiques (Tutin, Yan et Tran, 2018).

Cette méthodologie appliquée à l'enseignement des langues étrangères a déjà presque cinquante ans d'existence : théorisée par Tim Johns à partir d'un corpus de textes scientifiques authentiques, pour aider des étudiants allophones dans la rédaction universitaire en anglais, l'analyse sur corpus est apparue en France avec l'élaboration du Corpus d'Orléans, base de matériel pédagogique (Biggs et Dalwood, 1976).

Pour les langues bien dotées en ressources linguistiques, tel le français, des corpus spécialisés couvrent de nombreux genres textuels, et chacun comporte souvent plusieurs millions de mots. Les linguistes utilisent ces ressources dans une démarche scientifique afin de confronter les théories à l'usage concret de la langue en général, mais aussi en faisant ressortir des contrastes entre types d'usages de la langue. Les corpus permettent de montrer, par exemple, qu'au sein de ce type d'écrit, la situation n'est pas homogène : les habilitations à diriger des recherches, écrites par des chercheurs confirmés, recourent ainsi beaucoup plus à la première personne que les thèses de doctorat, écrites par des chercheurs débutants. Par ailleurs, l'écrit scientifique se caractérise par une phraséologie spécifique que seule l'expérience peut nous apprendre avec quel sens et dans quelles structures nous pouvons effectivement employer.

Plusieurs chercheurs ont utilisé ce corpus afin d'élaborer des séquences didactiques pour aider la rédaction académique en français (Tran, 2014 ; Cavalla, 2015). Comme ces travaux, notre intérêt porte sur l'étude des phénomènes phraséologiques dans une perspective didactique. Mais, ce qui est différent dans cette étude est que nous cherchons à orienter les objets d'enseignement à partir des difficultés observés dans les corpus d'étudiants. Il existe de nombreuses études consacrées à la description des phénomènes

phraséologiques de la langue, mais peu d'entre elles se centrent sur les difficultés des étudiants en français comme langue étrangère (Yan, 2017).

Notre travail vise une certaine phraséologie employée dans les écrits académiques à l'université. Afin d'analyser et faire ressortir, le cas échéant, les similarités et les différences d'usage de la phraséologie dans différentes disciplines et dans plusieurs genres académiques, nous allons prendre en compte deux ensembles de données se rapportant aux genres de l'écrit scientifique et académique. D'abord, le corpus spécialisé du discours académique et de textes scientifiques d'experts : Scientext (Tutin et Grossman, 2012), publié à partir de 2008, utilisé pour relever des concordances pré-sélectionnées et, la ressource (LST) du lexique scientifique transdisciplinaire développée par Hatier et Tutin (2016). Ensuite, le corpus que nous avons élaboré à partir des supports de cours de Physique donnés par des enseignants à l'université libanaise. Il est constitué dans le but non seulement de repérer la phraséologie ciblée mais aussi afin d'analyser le degré de son utilisation à l'université par les différents acteurs. Enfin, le corpus de CR des étudiants, constitué afin de relever les erreurs ainsi que les cas de sous-emploi et de suremploi liés à l'utilisation de la phraséologie en question.

Dans ce qui suit, nous allons décrire deux interfaces différentes desquelles nous pouvons extraire les unités phraséologiques et les collocations dites transdisciplinaires puisqu'on a affaire à de plusieurs disciplines.

4.2 L'usage de corpus informatisés

D'une manière générale, l'avantage de toute recherche linguistique fondée sur un corpus réside dans la possibilité de travailler sur un échantillon de la langue telle qu'elle se présente dans la réalité. Dans le cadre de la phraséologie, le corpus joue un rôle tout à fait particulier. Par exemple, lui seul permet en effet de confirmer le statut phraséologique de certaines combinaisons de mot. Comme Habert et al. (1997) le remarquent :

[...] étant donné une expression jugée « contrainte » quant à ses possibilités de transformation, les corpus permettent de chercher si ses réalisations effectives confirment ce jugement.

En outre, la prolifération de corpus textuels informatisés se trouve à l'origine du développement des études en phraséologie. Les corpus informatisés ont ce mérite d'avoir

rendu possible le repérage des séquences de langue récurrentes :

The increasing use of computer-held text corpora containing many millions of words has allowed linguists to establish lexico-grammatical patterns in language that were previously unavailable to observers. (Gledhill, 2000)

Nous allons voir dans la section suivante, le premier corpus informatisé qu'on a utilisé dans ce travail.

4.2.1 Corpus Scientext

Parmi les corpus académiques les plus connus se trouve le corpus Scientext¹. Il comprend plusieurs corpus élaborés dont deux d'écrits scientifiques en français et en anglais. Dans notre travail, nous allons exploiter le corpus du français qui est un corpus public élaboré par le Laboratoire de Linguistique et de Didactique du Français Langue Étrangère et Maternelle (LIDILEM). Il compte 4.8 millions de mots issus des différents genres et disciplines : les sciences humaines (linguistique, psychologie, sciences de l'éducation, traitement automatique du langage), les sciences expérimentales (biologie et médecine) et les sciences de l'ingénieur (électronique et mécanique). Pour des raisons de droits, il n'était pas possible d'y intégrer des corpus de Physique. Ce corpus constitué de textes authentiques issus de discours spécifiques au domaine visé permet de mettre en évidence le lexique propre au champ disciplinaire. En effet, le chercheur doit maîtriser, en plus du discours académique, une terminologie spécialisée. Concernant les genres académiques, ils intègrent des articles de recherche, des communications écrites, des thèses de doctorat et des mémoires d'habilitation à diriger les recherches (HDR). Chaque texte est en outre découpé en parties textuelles (développement, introduction, conclusion, résumé, notes, titres, remerciements et annexes).

L'utilisateur de Scientext peut sélectionner son corpus d'après différents critères (comme les disciplines, les parties textuelles, les types textuels) ou bien écarter certains textes jugés non pertinents. Plusieurs modes de recherche sont proposés :

- (a) mode guidé et simple
- (b) mode sémantique d'après un thème
- (c) mode avancé requérant un langage de requête assez complexe mais qui permet de formuler des requêtes plus spécifiques

1. Site du projet Scientext : <https://scientext.hypotheses.org>

Les recherches peuvent porter sur les formes, les lemmes, les catégories syntaxiques et les relations syntaxiques entre les mots. Les résultats sont visibles sous forme de concordances et sous forme de statistiques dans l'onglet **Résultats**.

Ce corpus nous permettra de décrire les usages pratiqués par les scripteurs natifs et de fournir des extractions permettant une description du fonctionnement de la langue en contexte. Il constitue un modèle d'experts que les scripteurs novices souhaitent atteindre, car il fournit les éléments linguistiques qui nous intéressent en termes de fréquences. Deux principaux types d'informations seraient pris en compte dans notre recherche : les collocations et leurs fréquences c.à.d. le nombre de fois d'apparition.

En plus de Scientext, nous avons consulté une base de données contenant un lexique scientifique transdisciplinaire (LST) que nous allons décrire brièvement dans la section suivante.

4.2.2 Base de données LST

Dans le cadre d'un travail de deux thèses de doctorat (Tran, 2014), puis (Hatier, 2016) relevant du projet TermITH du LIDILEM, un corpus transdisciplinaire des écrits de sciences humaines et sociales a été constitué afin de faciliter l'indexation des textes de sciences humaines. Ce corpus comporte 500 articles répartis dans 10 disciplines de sciences humaines, tirés de revues de référence. Par la suite, une liste des mots simples et locutions figées du LST (verbes, noms, adjectifs et adverbes), accompagnés d'informations sémantiques, a été extraite (Hatier et Tutin, 2016). En même temps, un ensemble de collocations reliées aux mots simples et locutions a été extrait et traité. La base de collocations contient un peu plus de 1500 collocations.

Une plateforme web² a été développée afin de permettre l'exploration et l'utilisation des ressources du LST. Il autorise la consultation des mots simples du LST ainsi que des collocations. La base est présentée à travers quatre pages principales : la consultation des entrées lexicales, celle des collocations, la présentation de la classification sémantique et textes annotés. Le développement de cette interface a pour objectif de permettre une consultation plus pratique de la ressource, et une compréhension plus avancée de sa structuration. Des exemples d'utilisation en contexte sont associées à chaque entrée lexicale, des statistiques sur les occurrences ainsi que des informations sur les propriétés

2. <http://lst.demarre-shs.fr>

sémantiques de l'élément en question. Pour ce qui concerne les collocations, l'utilisateur peut avoir accès à des informations à partir d'une liste alphabétique, ou à partir d'une recherche libre d'un mot de la collocation.

L'usage de cette ressource dans le cadre de notre étude sera un complément de l'exploitation de la phraséologie que nous effectuerons à partir de Scientext. Cela nous permet d'accéder directement à la liste de collocations proposées et de comparer les résultats de recherche et les fréquences d'emploi des unités repérées avec les résultats obtenus dans Scientext.

Après avoir précisé les usages des interfaces Scientext et LST dans le cadre de notre étude, nous présentons dans la section suivante les corpus que nous avons constitué à partir de productions écrites d'enseignants et d'étudiants.

4.3 Constitution des corpus universitaires

Dans notre travail, en dehors du corpus d'articles scientifiques, nous utiliserons deux autres corpus pour répondre à nos objectifs de recherche, à savoir le corpus de cours d'enseignants et un corpus de CR d'étudiants. L'objectif du corpus de cours étant de vérifier s'il y a le même usage de collocations dans le champ de la physique que celles repérées dans les corpus présentés précédemment. Le corpus de cours est utilisé comme un corpus de contraste, par rapport à celui de Scientext, qui permet de ressortir les spécificités d'utilisation de la phraséologie en Physique.

Le deuxième corpus est un corpus de comptes rendus de TP en français. Constitué pour notre étude didactique, il nous permet de comprendre et de repérer les principaux problèmes dans l'utilisation de la phraséologie (trans)disciplinaire dans les productions écrites des étudiants. Ce type de corpus d'étudiants peut donner accès aux erreurs et aux difficultés des apprenants liées à l'utilisation de la phraséologie.

Il est important de souligner que le corpus d'experts Scientext n'est pas comparable avec le corpus de CR des apprenants tant au niveau scientifique qu'au niveau des objectifs de communication. En effet, les textes du Scientext rédigés par les experts de disciplines visent à diffuser les connaissances scientifiques tandis les CR des étudiants sont des écrits de recherche en formation requis pour valider le cours de TP dans le cas que nous

études. Par rapport aux corpus d'experts, les cours des enseignants fournissent des données mieux adaptées, une référence permettant non seulement d'étudier la phraséologie dans une discipline de sciences dures mais aussi de mettre en évidence les difficultés propres aux étudiants de l'UL. En effet, la comparaison des corpus d'apprenants avec les corpus des enseignants permet de révéler les utilisations spécifiques liées aux étudiants et d'aider à cibler des éléments linguistiques précis qui sont absents des discours d'apprenants. Ainsi, ces corpus répondent aux objectifs d'enseignement puisque les résultats de diagnostic peuvent être intégrés dans les méthodes d'aide à la rédaction proposées (Yan, 2017). Il nous semble important de prendre en compte à la fois les erreurs ainsi que les cas de non-emploi de la phraséologie.

Nous pensons que ces deux corpus de petite taille ne peuvent présenter que des situations d'enseignement précises. En dépit de leur valeur représentative moins importante, ils nous fournissent "une plus grande facilité d'analyse, de prise en main, d'interprétation et de constitution". (Aston, 2002 ; cité par Cavalla & Loiseau, 2014) permettant d'établir un diagnostic plus ciblé et de créer des matériaux didactiques plus spécifiques.

Dans ce qui suit, nous allons décrire la discipline et les genres dont ils relèvent non deux corpus.

4.3.1 Choix de la discipline

Afin de pouvoir travailler sur la phraséologie, il était essentiel d'axer notre collecte et notre analyse des associations collocatives sur le domaine scientifique qui nous intéresse et dont je suis spécialiste : la physique et non seulement sur les disciplines que Scientext comprend. La diversité des domaines est indispensable. À cette fin, un corpus recueillant des textes du domaine de la physique a été ainsi mis en place. Les autres domaines des sciences dites dures, les sciences spéculatives (telles que les mathématiques) pourraient être intégrées au cours d'études futures.

4.3.2 Choix des genres

Parmi les genres d'écrits scientifiques souvent évoqués à l'université, à savoir le mémoire, le compte rendu, l'article, la thèse, nous nous focalisons tout particulièrement sur les comptes rendus CR de TP des étudiants ainsi que sur les notes des cours magistraux CM

des enseignants pour nos différents objectifs linguistique, méthodologique et didactique. À l'université, le cours magistral est une des formes les plus courantes de transmission des savoirs. Les cours magistraux prennent en général un volume important dans l'ensemble des enseignements dispensés à l'université (dans une discipline, 60% des cours sont des cours magistraux (CM) contre 40% de TD). Il a fait l'objet principal des études menées en FOU (Bouchard & Parpette, 2012 ; Mangiante & Parpette, 2012).

4.3.2.1 Caractéristiques du CM

Le CM est un genre académique nouveau en comparaison du secondaire souvent associé à une nouvelle discipline. Pour les étudiants, c'est un discours complexe, contextualisé et à fort implicite culturel. Il comporte également des caractéristiques discursives et linguistiques transversales. Selon Bouchard et Parpette (2012)

Ce genre académique est une manifestation de la littéracie, consubstantielle à la notion d'université comme à celle d'école en général : l'écrit y est un outil et d'enseignement et d'apprentissage.

On a souvent du CM une représentation qui tend à en faire un discours, dense, de transmission disciplinaire. Ce qui est effectivement sa fonction fondamentale. Mais cette transmission s'effectue dans une perspective pédagogique, avec un enjeu non seulement de formation individuelle, mais de validation officielle par le biais de l'obtention d'un diplôme, dans une institution particulière qui est l'université.

Le cours magistral n'est pas une simple lecture de notes écrites, mais il comprend une double activité : celle de production écrite et orale de l'enseignant et de la production écrite de la part des étudiants par le biais de prises de notes. Ces deux activités doivent rester synchrones tout au long de l'évènement de la parole. Notre étude ne se focalise pas sur la partie oralographique ou bien sur la production orale de l'enseignant dans un CM tel est le cas étudié par Bouchard et Parpette (2012), mais nous allons nous servir des supports écrits que l'enseignant utilise afin d'assurer son CM.

La salle de classe, avec son tableau, montre le rôle important que joue l'écriture dans la transmission du savoir. L'utilisation du tableau s'impose plus lorsque le CM porte sur un contenu qui graphiquement va utiliser un autre codage que la langue naturelle tel que le cas en physique, en mathématiques, en chimie où les démonstrations et explications sont quasi intégralement reconstruites au tableau. Dans ce cas, les supports du cours soit sous la forme de photocopiés soit sous forme de diapositives du type PowerPoint offrent

une meilleure possibilité de découvrir la langue écrite en particulier, la phraséologie que l'enseignant peut, même s'il l'a dit à l'oral, ne pas la noter au tableau pour de raisons de simplifications d'inscriptions.

Dans certaines universités au Liban, les cours sont élaborés à l'avance par les instances officielles avec des thématiques et des contenus originaux. À l'université libanaise, par contre, on donne le programme du cours et c'est le professeur qui élabore le contenu en suivant par exemple, une référence en lien avec le cours à enseigner, ou bien il se réfère à des cours déjà élaborés dans des universités étrangères telles que France ou ailleurs, il fait des ajouts, des incises, des ouvertures, etc. Bref, il adapte selon les besoins.

Dans notre travail, puisque nous ciblons l'activité d'écriture, nous allons donc considérer ce qui relève de la production écrite à savoir les photocopiés et les documents PPT des enseignants tels qu'ils sont rédigés. Les cours des enseignants consultés dans ce cadre se basent sur plusieurs références rédigées en anglais et en français. Ils constituent un support intéressant révélant d'une certaine réflexion méthodologique et scientifique structurée.

Si ces derniers permettent d'étudier la phraséologie telle qu'elle est employée par les enseignants, elles ne permettent pas de vérifier l'utilisation de la phraséologie par les étudiants. Pour ce faire, Nous allons utiliser des CR écrites dans un cours des TP . Dans la section suivante, nous allons voir les caractéristiques d'un tel écrit.

4.3.2.2 Les CR de TP

Les comptes rendus scientifiques ou les rapports de TP tels qu'ils sont nommés à l'université libanaise partagent une structure générale qui vise à décrire, expliquer et justifier une certaine démarche expérimentale suivie ainsi qu'à mettre en valeur et analyser son travail. La présentation des résultats obtenus lors du travail expérimental doit être accompagnée d'une interprétation et d'une explication. Un CR a la caractéristique d'un genre structuré qui peut se diviser en sections identifiables. Dans le corps d'un compte rendu de TP, on retrouve différentes sections comme l'introduction (le contexte de l'expérience, les hypothèses de départ), les objectifs, la démarche expérimentale, les résultats et la conclusion. Nous allons nous focaliser sur les deux dernières parties où il y aurait une présentation et une interprétation des résultats et où, dans la conclusion, il s'agit

d'exposer la réponse au problème et d'indiquer si le but du TP est réellement atteint, en d'autres termes, si les résultats viennent de confirmer ou de réfuter les hypothèses. Ce genre académique sert d'un exercice important aux étudiants novices qui, dans le futur, seront amenés à faire des études de recherche plus rigoureuses ainsi qu'à rédiger des genres plus développés tels que des mémoires, des thèses et des articles scientifiques. Le corpus de CR d'étudiants et celui des enseignants appartiennent à de genres différents qui ne sont comparables en tant que tels. Mais, le fait d'étudier l'emploi de la phraséologie dans les deux nous permettra de faire un diagnostic plus ciblé de l'utilisation de la phraséologie chez des étudiants de niveau avancé pour en révéler les lacunes et les erreurs, ce qui permettra ensuite de proposer des activités qui sont susceptibles de répondre au mieux aux besoins spécifiques de ces étudiants dans la rédaction de leur travail.

Après avoir repéré la discipline et les genres académiques ciblés par notre recherche, nous allons décrire les critères selon lesquels nous avons choisi les textes constituant nos corpus.

4.3.3 Choix de sources

Le choix des sources est une étape importante de la constitution de nos corpus. L'objectif était de construire un corpus qui serait le plus représentatif possible de la phraséologie dans un contexte universitaire francophone à l'étranger. Nous avons mis en place les six critères évoqués par Pecman (2004) pour la sélection des textes en vue de la création du corpus. Ils assurent l'homogénéité du corpus, seul garant d'une analyse linguistique rigoureuse et fiable. Voici les critères qui nous ont guidée dans notre choix des sources textuelles :

1. **Période de rédaction** : Les textes des cours choisis sont tous donnés pendant l'année courante (en 2020). Ils sont rédigés entre 2017 et 2019, mais cette période reste relativement récente. Les comptes rendus disponibles ont été écrits l'année dernière, en 2019 et non pas cette année. Mais, ceci ne gêne pas à l'étude. En effet, nous avons l'intention de faire le recueil des CR que des étudiants en Licence 3 auraient rédigés suite à une classe de TP pendant cette année d'étude, mais, suite aux circonstances de confinement et de fermeture de l'université, les classes de TP ont été reportées jusqu'à nouvel ordre. Nous avons alors contacté des étudiants de Master et leur avons demandé la permission d'exploiter les CR qu'ils ont rédigés

en Licence 3 (c.à.d. en 2019). Nous n'avons pas pris en compte les TP rédigés en Licence 2. La troisième année de Licence est particulièrement ciblée puisque les étudiants sont censés avoir un niveau plus avancé en français. Ce qui correspond à notre étude.

2. **Nature des textes** : Les textes exploités sont des textes scientifiques du type : cours rédigés par les enseignants de l'UL et comptes-rendus scientifiques. Seuls les corps de textes rédigés en langage naturel ont été pris en compte. Les graphiques, les figures et les tables, les équations, consistant souvent en un langage certes typique du discours scientifique, mais symbolique, ont été écartés. Ce processus de nettoyage est nécessaire pour le traitement de ces corpus. En effet, nous sommes intéressés à la façon dont les étudiants et les enseignants structurent leur écrit en parlant des "résultats" et de la "conclusion".
3. **Fiabilité des textes** : Pour s'assurer une certaine fiabilité des textes, il était essentiel de veiller à leurs origines. Tous les textes de cours et de CR exploités sont respectivement la propriété d'enseignants et celle d'étudiants, rédigés dans un cadre institutionnel et officiel. Ce statut justifie qu'on les considère comme faisant autorité en matière de langage.
4. **Langue des textes** : Un des critères principaux de la sélection des textes était leur disponibilité dans la version française. Les notes de l'enseignant, les références à disposition des étudiants, les documents distribués en classe, ainsi que les TD sont rédigés en français. Nous avons voulu rendre possible l'étude de la langue scientifique dans son usage à l'université libanaise. Le professeur, dont le français n'est pas la langue maternelle, écrit directement son cours en français. Par ailleurs, pour un cours donné en français, l'enseignant peut utiliser une référence rédigée en anglais dont il traduit en partie le contenu ou bien, il peut distribuer aux étudiants des documents rédigés par des professeurs français. Ces textes sont également inclus dans notre étude. Le fait de travailler sur des productions écrites d'auteurs d'origines linguistiques diverses, va en fin de compte dans le sens de nos préoccupations.
5. **Format des textes** Une partie des comptes rendus était disponible en version électronique doc. L'autre partie était au format papier scanné/numérisé. De même pour les textes des cours : une partie était disponible en version pdf et l'autre au format papier. La conversion de textes au format électronique (doc par exemple) aurait été plus avantageuse notamment pour les besoins des analyses automatiques,

et pour une meilleure gestion des données. Mais pour l'étude actuelle, nous avons décidé de travailler avec le format que nous avons. Dans le cas de textes au format papier, l'exploitation a été alors faite manuellement page par page. La disponibilité d'autres textes en version électronique nous a permis de faire une recherche sur ordinateur de la phraséologie ciblée.

6. **Droits d'exploitation et de reproduction des textes** : Les textes de nos corpus sont issus des cours d'enseignants et des comptes rendus des étudiants qui nous ont accordé leur autorisation d'exploitation dans le cadre de la recherche. Les cours exploités m'appartiennent à l'except d'un seul cours dont l'enseignante a fait elle-même la recherche demandée. Par ailleurs, nous avons expliqué aux étudiants le cadre didactique de notre recherche et pour ceux qui ont envoyé leurs productions écrites, ils nous accordé la permission d'exploitation.

Après avoir décrit les critères qu'ils nous ont servi pour le choix de ressources textuelles et ainsi construire nos corpus, nous allons décrire précisément les observables qui ont été finalement retenus et recueillis pour réaliser notre étude.

4.4 Collecte de données

L'objet d'apprentissage visé se situe à la croisée de FOU (méthodologie et discours) et du français langue académique FLA (travail de l'écrit académique à travers la phraséologie). La démarche utilisée dans ce travail est alors celle de FOU dans le sens où elle consiste à explorer des situations universitaires pour analyser des besoins et d'en tirer des données qu'il faut ensuite analyser et largement recomposer pour tenter de créer un espace d'apprentissage permettant aux étudiants de préparer au mieux leur contact avec l'écrit dans le contexte universitaire. L'élaboration d'un programme de FOU passe par l'étape centrale du FOS qui est la collecte des données dont l'analyse permet de déterminer les compétences langagières à développer chez les étudiants, et d'en tirer éventuellement les documents qui serviront de supports de formation.

Nous avons collecté nos corpus d'enseignants et d'étudiants à la faculté des sciences de l'UL au Liban. En particulier, nos données relèvent de la discipline Physique. Le choix de l'établissement peut s'expliquer par plusieurs raisons : tout d'abord, j'avais suivi mon cursus de licence et de master en physique à cette université et j'y suis actuellement enseignante. J'ai donc un contact personnel avec le département de physique où

je donne mes cours. J'ai constaté d'après mon expérience avec les étudiants que les pratiques rédactionnelles des étudiants sont encore loin d'être satisfaisantes. Ils se trouvent confrontés à des difficultés langagières en raison du manque d'outils appropriés et de formation spécifique à l'écrit académique.

Nous allons voir dans ce qui suit les cours et les CR retenus et qui constituent les données sur lesquelles seraient centrées notre réflexion et analyse.

4.4.1 Les cours retenus

Pour une meilleure généralisation des résultats d'analyse, nous avons essayé de diversifier les sous-disciplines ou bien les spécialités en physique ainsi que les niveaux d'études des étudiants dont ces cours font partie. Les contenus diversifiés permettent de nous donner un point de vue d'ensemble sur l'aspect de la phraséologie, en particulier celle que nous ciblons, dans différents domaines de la physique. Il se peut que les étudiants y sont exposés dans un cours plus qu'un autre selon qu'ils soient dans des cycles différents ou bien selon qu'ils soient dans le même cycle. Par conséquent, le rapport de ces étudiants à la langue n'est pas le même.

Nous allons partir de l'hypothèse que les cours en Licence sont beaucoup plus développés que les cours donnés en M1. La majorité des cours que nous avons retenus font partie des cours donnés en Licence. Cette hypothèse de départ se justifie par le besoin d'une meilleure explication aux étudiants du premier cycle qui sont des novices par rapport aux étudiants du deuxième cycle qui ont plus d'acquis. Donc, du point de vue linguistique, cela va dans le sens de nos préoccupations en termes d'extraction de collocations puisque la probabilité d'en trouver est plus grande.

En outre, nous voulions augmenter le nombre de textes de cours constituant notre corpus. Cependant, le confinement sanitaire imposé dans le monde nous a empêché de faire l'intervention directe à l'université et ainsi de demander à d'autres professeurs la permission d'exploiter leurs cours. À noter que généralement, à l'université libanaise dans une filière telle que la physique, les cours sont basés majoritairement sur des démarches de calcul et on ne note que rarement des phrases complètes soit dans le support du cours soit au tableau. Même des séquences polylexicales telles que : "*on en déduit, ce qui implique*" sont parfois remplacées par des symboles tels que : " \Rightarrow ".

Nous présentons ici la liste de supports de cours qui ont été retenus pour la construction de notre corpus.

- Cours de “*physique des ondes*” constitué de deux chapitres et de 40 pages environ. Il est destiné aux étudiants de la Licence 2.
- Le livre “*Physique exercices incontournables*” (Beury, 2014) de 300 pages environ et dont certains extraits sont mis à la disposition des étudiants de la Licence 2.
- Cours de “*physique biomédicale*”, destiné aux étudiants en Licence 3. Ce cours est composé de quatre parties.
- Fiches descriptives, d’une vingtaine de pages, distribuées aux étudiants en Licence 3 afin de leur montrer l’aspect théorique et les étapes de la réalisation des expériences.
- Cours “*introduction à la théorie électrofaible*” dédié aux étudiants de Master. Ce cours est composé de cinq chapitres et il est de 80 pages environ.

Parallèlement aux cours d’enseignants, nous avons recueilli des productions écrites des étudiants dont nous allons décrire les modalités de recueil dans la section suivante.

4.4.2 Recueil des CR

Les données recueillies sont composées de six CR en spécialité de physique générale. Chaque CR est rédigé par un groupe de quatre personnes. Leur sujet d’étude porte soit sur la physique atomique soit sur la physique nucléaire. Lors de la rédaction de CR, les étudiants sont en troisième année de Licence en physique et ont pris, pendant deux ans de leur cursus, des cours de français à l’université. Il est à noter que ces CR rédigés en français sont corrigés par des enseignants de Physique libanais sachant que les corrections portent sur le respect de différentes parties du CR, les résultats obtenus après avoir fait l’expérience et plus rarement sur les erreurs de langue (grammaire, orthographe, lexique ou syntaxe).

Pour recueillir les productions écrites, nous avons contacté les étudiants en envoyant un message collectif sur le groupe Whatsapp dédié au cours de Master 1 que j’assure. Nous leur avons expliqué que le recueil c’est dans le cadre d’une étude qui vise à améliorer l’écrit académique à l’université et que leurs productions ne seront en aucun des cas jugées. Certains étudiants n’ont manifesté aucune réaction et d’autres nous ont envoyé

leurs CR. Au final, nous avons reçu des CR de la part de six étudiants qui ont accepté de participer à cette étude. chacun d'eux a envoyé de deux à six CR dont nous allons prendre en compte un seul pour cette étude.

Afin de faciliter l'analyse, les CR ont été regroupés dans des dossiers nommés par les noms des étudiants. Puisque les étudiants de M1 à la faculté des sciences à Beyrouth sont issus de diverses régions et alors de différentes branches de l'UL, nous avons ajouté au nom du dossier la branche de laquelle l'étudiant est issu.

Après avoir décrit le processus de constitution de nos corpus de cours et de comptes rendus, nous abordons l'étape suivante du traitement des phénomènes collocatifs, et la localisation des unités phraséologiques dans de tels écrits.

4.5 Outils d'extraction des collocations

Rappelons que, notre étude est centrée sur l'observation de la phraséologie liée aux notions fortement présentes dans les écrits académiques et les CR en physique telles que "expérience", "hypothèse" et "résultats". Les notions choisies ont été déjà mentionnées dans des travaux précédents sur le lexique scientifique transdisciplinaire (LST) ou général (LSG). Certaines avaient pour but de réfléchir au fonctionnement de la phraséologie et du lexique dans des écrits scientifiques et à en constituer un inventaire de base (Pecman, 2007; Tutin, 2014), d'autres se basent sur ces dernières afin de développer des pistes didactiques en termes d'écriture en direction des étudiants allophones (Cavalla 2015, 2019). Par ailleurs, il existe dans les genres textuels que nous traitons d'autres notions telles que celles qui sont utilisées pour conclure un travail, mais elles ne seraient pas incluses dans le cadre de ce travail.

Notre objectif ne constitue pas à extraire tous les mots spécifiques aux écrits scientifiques de la physique pour en faire un inventaire complet mais, de se servir de quelques-unes pour faire une étude représentative de leur utilisation par les étudiants à l'université.

La constitution du corpus d'étudiants nous sert à effectuer une étude d'ordre quantitatif et qualitatif. Nous visons à relever la fréquence des UP en question dans leurs écrits ainsi qu'à analyser comment les étudiants les utilisent. Ce qui va dans le sens de nos préoccupations en direction des locuteurs non natifs entreprenant des études scientifiques et devant s'acculturer aux écrits académiques en français.

Nous allons discuter, dans les sections qui suivent, les outils qui nous ont guidée dans la

tâche d'extraction d'unités phraséologiques.

4.5.1 ScienQuest

Un sous-ensemble du corpus Scientext est consultable en intranet avec l'outil ScienQuest initialement développé pour les sous-corpus du projet Scientext. Sur l'interface ScienQuest, pour faire des recherches, la première partie consiste à sélectionner le corpus d'analyse. Comme il est montré par la figure, l'interface propose trois possibilités de sélections par disciplines, par genre et par parties textuelles. Dans notre étude, nous nous intéressons, au premier temps, aux unités phraséologiques dans toutes les disciplines, tous les genres et toutes les parties textuelles.

The screenshot shows the ScienQuest interface with the following sections:

- Header:** ScienQuest – Corpus Écrits scientifiques en français. Langue de l'interface: [fr] [en]
- Navigation:** Corpus, Textes, Recherche, Outils, Connexion, À propos, Aide.
- Left Panel:**
 - Choisissez les textes:** Par défaut, tous le corpus est sélectionné. Si vous le souhaitez, vous pouvez retirer certains textes (ou parties de textes) de votre sélection, soit de manière groupée (onglet Critères), soit texte par texte (onglet Liste).
 - Votre sélection:** Actuellement votre sélection comporte 5 596 291 mots (205 textes), sur les 5 596 291 mots (205 textes) du corpus.
- Discipline:**
 - Sciences expérimentales
 - Biologie (744 534 mots, 28 textes)
 - Médecine (97 139 mots, 8 textes)
 - Sciences humaines
 - Linguistique (1 633 579 mots, 67 textes)
 - Psychologie (528 554 mots, 17 textes)
 - Sciences de l'éducation (1 280 515 mots, 60 textes)
 - TAL (753 025 mots, 17 textes)
 - Sciences appliquées
 - Mécanique (139 912 mots, 3 textes)
 - Électronique (419 234 mots, 5 textes)
- Genre:**
 - HDR (517 025 mots, 7 textes)
 - Article (348 168 mots, 45 textes)
 - Communication (532 333 mots, 112 textes)
 - Thèse (4 198 966 mots, 41 textes)
- Parties textuelles:**
 - Parties principales
 - Développement (4 310 279 mots)
 - Introduction (154 732 mots)
 - Conclusion (202 371 mots)
 - Autres parties
 - Résumé (42 330 mots)

FIGURE 4.1 – Les critères de recherche et la sélection des corpus dans ScienQuest

Ensuite, concernant les modes de recherche nous allons adopter le mode de recherche simple et guidé en raison de sa simplicité d'accès par rapport à la recherche avancée qui est davantage réservée aux linguistes familiarisés avec le traitement automatique du langage. Il nous permet de sélectionner des formes, des lemmes, des catégories et des relations syntaxiques d'une façon simple. Nous avons également utilisé le mode sémantique renvoyant à des fonctions sémantiques spécifiques préétablis (ex : formulation d'une

hypothèse) pour rechercher les collocations ou les combinaisons fréquentes de mots qui nous intéressent.

Une fois les requêtes terminées, les statistiques sur les occurrences trouvées peuvent être obtenues en fonction de différents critères : forme, lemme, catégorie, traits morphosyntaxiques, partie textuelle, genre et discipline.

En dehors de l'outil ScienQuest, nous avons recours à l'interface LST pour compléter notre recherche. Cet outil a l'avantage d'accéder directement aux collocations liées aux notions que nous ciblons.

4.5.2 Extractions dans l'interface LST

La requête libre dans la page collocations permet de repérer facilement tous les co-occurents de la notion. Elle permet également d'avoir les collocations relatives à une entrée donnée sous forme d'une liste. La page donne le choix d'ordonner cette dernière selon l'ordre alphabétique, la fréquence et la position syntaxique du mot. Dans notre étude, nous allons nous intéresser par la fréquence d'usage de certaines collocations et par la position syntaxique de la notion recherchée.

The screenshot shows the 'Collocations du LST' interface. At the top, there is a navigation bar with 'LST', 'Entrées', 'Classes Sémantiques', 'Collocations', and 'Textes annotés'. On the right, there are links for 'Inscription' and 'Connexion'. Below the navigation bar, the title 'Collocations du LST' is displayed. There are three buttons: 'Fréquence', 'Log-Like', and 'Patron'. A search bar labeled 'Rechercher' is present. The main content is a grid of 20 collocations arranged in 4 rows and 5 columns. At the bottom, there are logos for UGA, Huma-Num, and TermITH, along with an 'NR' logo and an upward arrow icon.

à la suite des travaux	constituer un facteur	favoriser la construction	moyen nécessaire	remplir une fonction
à partir d'indices	constituer un indicateur	favoriser la diffusion	nature de l'objet	rencontrer des difficultés
à partir de l'analyse	constituer un indice	favoriser le développement	nature différente	renforcer la capacité
à partir de l'enquête	constituer un instrument	figure centrale	nier l'existence	renforcer un caractère
à partir de la caractéristique	constituer un moyen	figure de référence	niveau d'analyse	renvoyer à l'ouvrage
à partir des cas	constituer un obstacle	figure géométrique	niveau de connaissance	renvoyer à la notion
à partir des documents	constituer un outil	fonction centrale	niveau de traitement	renvoyer à une problématique
à partir des données	constituer un support	fonction essentielle	nombre croissant	renvoyer directement
à partir des éléments	constituer un terrain	fonction principale	nombre d'éléments	renvoyer l'image
	constituer une dimension	fonctionnement du	nombre élevé	

FIGURE 4.2 – Page des collocations sur l'interface LST.

4.5.3 Extraction dans les corpus universitaires

La localisation des UP consiste en un dépouillement du corpus en vue de la collecte des collocations relatives à notre étude et qui sont liées aux mots : “hypothèse”, “résultats” et “expérience”. L’extraction est une étape de travail essentielle qui sert à dégager le degré de manifestation des UP en question dans le contexte universitaire.

Les deux corpus universitaires que nous avons créés sont de tailles assez suffisantes pour pouvoir localiser la phraséologie ciblée le plus de fois possible. Mais les deux corpus restent à l’échelle d’une extraction manuelle des unités phraséologiques à partir des textes. Aucun logiciel n’a été alors utilisé pour répondre à nos besoins. Une collecte manuelle a été employée pour l’extraction de la phraséologie à partir des textes qui existent en format papier. Un dépouillement automatique à l’aide de l’outil recherche sur ordinateur a été effectué avec les documents disponibles en versions pdf ou doc.

Qu’elle soit manuelle ou par ordinateur, la procédure de collecte des UP est une procédure menée avec le plus grand soin. Elle a été organisée de façon à explorer chacun des cours et des CR dans un ordre préétabli. Les premiers documents exploités étaient les cours de Physique et ensuite les CR des étudiants. Nous avons commencé par le cours de “physique des ondes”, celui qui contient plus d’explications et plus de mots rédigés vu le niveau d’étude auquel il s’adresse. Nous précisons que l’ordre d’exploitation des cours n’est pas important en soi mais il est préférable que le dépouillement du corpus s’effectue selon une procédure ordonnée.

4.6 Bilan de la méthodologie

Ce chapitre a permis de présenter une méthodologie d’extraction des collocations dans le cadre d’une étude phraséologique de la langue scientifique employée dans des écrits académiques du type CR à l’université libanaise. D’abord, nous avons introduit l’approche de la linguistique de corpus qui constitue de nos jours une méthodologie de recherche linguistique notamment dans le domaine de l’écrit scientifique et académique. Les corpus présentent le double avantage de comprendre le fonctionnement et l’usage de la langue et de favoriser son apprentissage.

À ce titre, nous avons présenté les corpus qui font l’objet de notre analyse linguistique avec comme objectif didactique l’aide à la rédaction académique dans des domaines de sciences tels que la physique. Après avoir parlé des interfaces Scientext et LST dont les

corpus sont issus d'écrits d'experts de différentes disciplines, nous nous sommes concentrés sur la présentation et la constitution des corpus universitaires d'enseignants et d'étudiants qui permettent d'effectuer une étude de diagnostic des besoins langagiers dans un contexte universitaire relevant de sciences. Cela constitue un nouvel élément dans le domaine du FLE où il existe peu de corpus d'apprenants inscrits dans un genre académique précis.

Finalement, nous avons expliqué quels sont les différents outils d'exploitation (Scien-Quest, LST et autres) utilisés dans notre analyse linguistique sur corpus.

Après avoir expliqué la méthode adoptée dans la présente étude, nous procédons à l'analyse des corpus et aux résultats obtenus dans le chapitre suivant.

Troisième partie

Analyse des échantillons

Chapitre 5

Analyse des échantillons

Contents

5.1	Résultats d'analyse des corpus	69
5.1.1	La phraséologie dans Scientext	69
5.1.2	Collocations du LST	72
5.1.3	Phraséologie du corpus de cours	73
5.1.4	Retour sur la synonymie dans Scientext	74
5.1.5	Bilan de l'analyse du corpus d'experts <i>vs</i> corpus de cours . . .	75
5.2	Observation des CR d'étudiants	76
5.2.1	Fréquence de l'emploi des notions	76
5.2.2	L'emploi des collocations	77
5.2.3	Les étudiants face aux collocations	80
5.2.4	Récapitulatif du diagnostic	80
5.3	Propositions didactiques	81



Dans le chapitre précédent, nous avons établi et proposé une méthodologie d'étude de la phraséologie dans des genres académiques à l'université. Avant de mener une réflexion didactique à ce sujet, il nous semble pertinent de s'arrêter sur l'analyse de besoins des étudiants liés à l'utilisation de la phraséologie afin de mieux cibler leurs difficultés et de proposer par la suite des orientations pédagogiques appropriées aux besoins.

Nous allons exposer ici les résultats tirés de l'analyse de corpus informatisés ainsi que ceux relatifs aux discours collectés dans le milieu universitaire. Le diagnostic comporte deux volets : le premier s'agit de la comparaison entre les écrits scientifiques des experts et les notes de cours des professeurs. Le deuxième consiste à confronter les écrits des étudiants à ceux des professeurs afin de mettre en évidence et étudier les cas d'emploi et du non-emploi de la phraséologie ciblée.

Nous essaierons de répondre aux questions suivantes : Les étudiants à l'université libanaise maîtrisent-ils la phraséologie des écrits scientifiques ? Quelles sont leurs lacunes, le cas échéant, en matière de phraséologie ?

5.1 Résultats d'analyse des corpus

5.1.1 La phraséologie dans Scientext

Après avoir sélectionné le mot simple, par exemple "*résultats*", nous avons recherché tous les cooccurents privilégiés fréquents qui entretiennent une relation syntaxique avec ce mot. Nous avons repéré plusieurs emplois et plusieurs patrons syntaxiques dont, par exemple, celles utilisées pour :

- Évaluer les résultats obtenus (Nom+Adjectif) : résultats préliminaires/généraux/très satisfaisants/significatifs, les meilleurs/de très bons résultats, résultats comparables, etc.
- Parler des apports des résultats (Nom+adjectif+Verbe) : les résultats (trouvés/obtenus/établis/issus de) (donnent/contribuent à/laissent envisager/confortent les choix faits/ montrent que/suggèrent que/valident l'hypothèse)
- Présenter les résultats : énoncer des résultats sur, la figure illustre/résume les résultats, etc.
- Autres usages : vérifier les résultats, l'interprétation des résultats/interpréter les

résultats dans une optique, etc.

De plus, la requête sémantique permet de dégager les verbes suivants de résultats et apports scientifiques : mettre en évidence, on en déduit, nous montrons que/comment, nous concluons que, nous soulignons, les relations que nous dégagions, nous démontrons, etc...

ScienQuest – Corpus Écrits scientifiques en français Langue de l'interface: [fr] [en]

Corpus | Textes | Recherche | Résultats | Outils | Connexion | À propos | Aide

Sémantique | Libre | Avancée

Recherche
Vous pouvez effectuer des recherches de plusieurs manières: en mode "libre" (avec un assistant), ou en mode "avancé" (avec un langage de requête).

Options de recherche
Max 1 000 000 résultats
Contexte 10 mots (max 100)
 Couper les contextes sur les limites de phrases.

Recherche libre
En mode libre, vous pouvez chercher un ou plusieurs mots (cliquez sur le bouton "+" à droite de "Mot 1" pour ajouter un mot). Ces mots peuvent être à la suite (par défaut), ou bien reliés par des relations syntaxiques (dans ce cas, l'ordre des mots est ignoré).
Pour chaque mot, vous pouvez entrer une forme, un lemme, ou une catégorie. Il n'est pas

Mot 1
Forme: résultats
Lemme:
Catégorie: Nom
Traits:
Mot 2
Forme:
Lemme:
Catégorie: Verbe
Traits:
+

Relations syntaxiques
+

Ordre des mots: Aucune relation syntaxique n'est définie. Les mots seront recherchés exactement dans cet ordre.

Chercher 1772 occurrences

#	Contexte gauche	Occurrences	Contexte droit	Texte
1	dans la troisième séance après que le professeur ait rappelé les	résultats trouvés	[23] pour Pn : Pn & gt ;	#1 - HDR - Sciences de l'éducation - Développement
2	renvoyer des déclarations ou des questions de validité relatives aux	résultats formulés	(dimension de la connaissance théorique) . Cette dernière	#1 - HDR - Sciences de l'éducation - Développement
3	de connaissances , et l' application par l' élève des	résultats établis	. Cependant cette alternative s' avère dans certains cas peu	#1 - HDR - Sciences de l'éducation - Développement
...	.	résultats	sont bien meilleurs que ceux du modèle de	#2 - HDR - Mécanique -

FIGURE 5.1 – Représentation de la recherche libre du mot “résultats” via Scientext

Du point de vue statistique, les formes “résultats obtenus” et “résultats montrent” sont les plus fréquents. D’ailleurs, les résultats montrent que la phraséologie recherchée est présente majoritairement dans un écrit du type “thèse de doctorat” (voir figure 5.2) et notamment dans la partie textuelle “résumé”. Concernant l’emploi de ce mot dans chaque discipline (figure 5.3), il s’avère que les disciplines de sciences de l’ingénieur (électronique+mécanique) passent en première place avec un pourcentage de 33%.

Nous voyons bien que l’écrit scientifique ne constitue pas un genre textuel homogène : le lexique et les formes discursives varient sensiblement suivant les disciplines (même en mettant de côté le vocabulaire technique propre à chaque discipline), le type de texte, et la position de l’énoncé dans le texte (et donc dans le raisonnement).

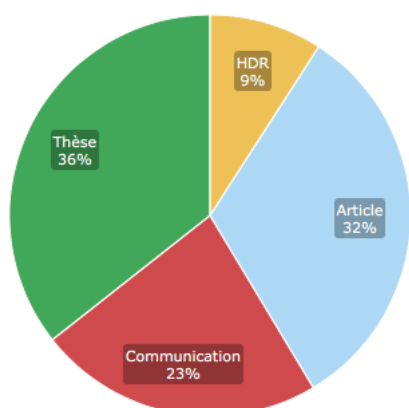


FIGURE 5.2 – Diagramme représentant la distribution statistique du mot "résultats" en fonction du genre

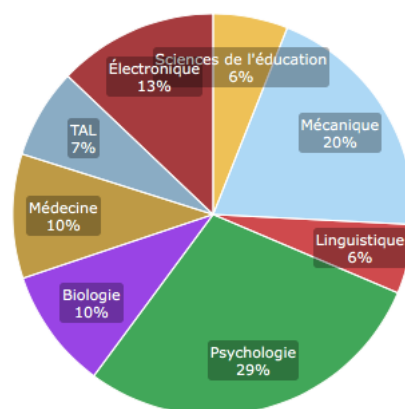


FIGURE 5.3 – Diagramme représentant la distribution statistique du mot "résultats" selon la discipline

Par ailleurs, en ce qui concerne la notion "*expérience*", parmi les verbes utilisés avec ce mot, on trouve le verbe "*mener*" qui d'ailleurs admet plusieurs associations telles que "*mener une analyse, mener une enquête*". Il se trouve aussi que "*mener*" n'est pas le seul verbe qui se trouve associé au mot "*expérience*". Dans ScienText, en cherchant le lemme *expérience*, nous remarquons qu'il est associé aux verbes suivants : *faire, réaliser, illustrer, concevoir, mener, mettre en place, recommencer, répéter, présenter, effectuer, décrire, résumer, confronter/comparer/recueillir les expériences*.

Le dernier ensemble de collocations recherché est lié à la notion *hypothèse*. Ce mot a été déjà traité dans des travaux de recherche portant sur les collocations et la phraséologie transdisciplinaire (Cavalla & Grossmann, 2005 ; Tutin, 2014 et Cavalla, 2015). Les collocations relatives à ce mot sont : *confirmer/étayer/vérifier/valider l'hypothèse, faire/formuler une hypothèse, écarter/réfuter l'hypothèse, les résultats corroborent l'hypothèse*. On peut également ajouter à ces constructions actives, les constructions passives (ex : l'hypothèse est formulée) et passives réduites (ex : l'hypothèse formulée) (Tutin, 2014). Nous allons prendre en considération les différentes formes active, passive et passive réduite lors de l'analyse des constructions associées aux notions choisies dans les écrits des enseignants et des étudiants.

5.1.2 Collocations du LST

Après avoir consulté la ressource du LST et accédé à la page dédiée aux collocations¹, les résultats de recherche, illustrés dans la figure 5.4, nous montrent que la collocation la plus fréquente entretenant une relation syntaxique avec le mot “résultats” est “obtenir des résultats”. En deuxième et troisième positions viennent respectivement les collocations “les résultats montrent” et “présenter les résultats”.

La collocation la plus fréquente associée au mot “hypothèse” est “faire une hypothèse”. Nous apercevons par la suite que ce mot, selon l’ordre en fréquence, est associé aux verbes “tester, rejeter, confirmer, avancer, formuler, émettre”. Les résultats de collocations relatives aux “résultats” et “hypothèse” sont identiques à celles trouvées dans Scientext. En ce qui concerne l’emploi d’ “expérience”, deux collocations sont seulement trouvées. La première associée au verbe “montrer” : “l’expérience montre” est déjà repérée dans Scientext et la deuxième : “décrire l’expérience”, constitue une nouveauté par rapport à ce qui a été trouvé.



FIGURE 5.4 – Résultats des collocations contenant le mot “résultats” via la plateforme LST. À côté de chaque collocation on trouve sa fréquence en gris.

L’exploitation de deux différentes bases de données permet de confirmer la similitude qui existe dans l’emploi des collocations par les experts. Nous allons, par la suite, procéder à

1. <http://lst.demarre-shs.fr>

l'analyse du corpus d'enseignants en prenant comme objet de contraste les collocations trouvées dans les parties précédentes.

5.1.3 Phraséologie du corpus de cours

La consultation des corpus de physique permet d'identifier le mot “*résultats*” dans différentes situations. Il est en effet utilisé pour :

- Comparer les valeurs obtenues par différentes méthodes : retrouver les mêmes résultats que/qu'avec la méthode X, avoir des résultats compatibles/qui divergent, avoir les mêmes résultats que ceux établis précédemment.
- Évaluer les résultats obtenus : les résultats sont satisfaisants.
- Parler des apports des résultats : ce qui signifie que le résultat de X est, on peut appliquer les résultats de.
- Ainsi que pour d'autres usages : interpréter les résultats (dans une consigne), on va démontrer le résultat général suivant, exploitation de résultats, on consigne les résultats dans le tableau suivant, les résultats mis dans ces tableaux, résultats de mesures.

En outre, il se trouve que pour évoquer des résultats numériques, le mot “valeurs” s'emploie dans plusieurs endroits. Dans un tel contexte, il désigne ce que vaut un certain observable ou une variable que nous pourrions mesurer. La notion de “valeurs” se substitue alors à celle de “résultats” puisque les résultats en physique sont indiqués en grande partie sous forme de valeurs numériques, par exemple : *On note les valeurs de la tension U et de la longueur d'onde λ .*

Concernant le mot “*expérience*”, les expressions suivantes y sont associés :

mener/réaliser/faire/recommencer l'expérience sous formes active ou passive.

D'un autre côté, nous remarquons la présence de l'expression : “*montage expérimental*” pour désigner une expérience. Cette combinaison de mots est utilisée dans des collocations telles que : *réaliser le montage expérimental.*

Quant à la notion “*hypothèse*”, nous avons extrait : *on fait des hypothèses, vérifier les hypothèses, vérifications de(s) l'hypothèse(s), ce qui justifie l'hypothèse faite précédemment, bien connaître les hypothèses.* D'ailleurs, de point de vue sémantique, afin d'avancer une

hypothèse en physique, nous constatons l'emploi fréquent du verbe “supposer” avec une construction en complétive “que”, par exemple : “supposons que les particules sont discernables”, “supposons que l'évolution du système est isentropique”, “nous supposons que les constantes sont égales”, “on peut supposer que”, etc.

Nous tenons à noter que la fréquence de l'emploi de ces notions et des expressions qui y sont associées varie d'un cours à un autre. Par exemple, dans le cours de physique biomédicale, nous remarquons une absence de mots “expérience” et “hypothèse” puisque le contenu ne s'articule pas autour des travaux expérimentaux mais sur la présentation des outils techniques associés à l'imagerie médicale. Par ailleurs, dans un cours de physique théorique basé sur des calculs mathématiques tel que celui de la “théorie électrofaible” donné en Master, les deux points sont souvent utilisés dans la construction des phrases telles que : “l'une de ces expériences : Wu et al.” au lieu d'une construction verbale comme “l'une de ces expériences est celle menée/réalisée par Wu et al.”.

Dans ce qui suit, nous allons nous intéresser aux collocations que les étudiants emploient et de plus, nous allons chercher à savoir s'ils utilisent les mots alternatifs que nous avons observés dans le corpus des enseignants. Avant d'entamer cette analyse, nous allons réexploiter les corpus Scientext et LST afin de vérifier si les synonymes “valeurs”, “montage expérimental” et “supposer que” apparaissent dans les écrits d'experts avec les sens attendus.

5.1.4 Retour sur la synonymie dans Scientext

Dans la base LST, nous n'avons pu repérer aucune collocation construite avec les mots et expressions alternatifs observés. Nous nous focalisons alors sur l'outil ScienQuest pour rechercher l'emploi des mots. Tout d'abord, il apparaît que “valeurs” est plus fréquemment employé dans des sous-disciplines de la physique : électronique 41%, et mécanique 19%. Si nous faisons ensuite la recherche dans les écrits appartenant à ces deux disciplines, en particulier dans la partie “conclusion” où en général on parle des résultats, nous obtenons des emplois tels que : “valeurs expérimentales”, “valeurs/moyennes de X”, “valeurs associées à ”, “valeurs mesurées pour X”, “valeurs importantes/significatives de”, etc. Le corpus Scientext a ainsi révélé que le mot “valeurs” est utilisé pour parler des résultats numériques des observables physiques dans une construction qui précise toujours la nature de ces valeurs et à quoi renvoient-elles.

De même, nous trouvons des constructions avec le verbe “supposer que” qui sont utilisées afin d’avancer une hypothèse dans différentes disciplines telles que la linguistique : “*on peut supposer que pendant cette phase d’attente, l’enfant préactive la variante lexicale de ce mot.*” ainsi qu’en électronique où le mot est le plus fréquent (21%) : “*Cette hypothèse revient en fait à supposer que les 3 fréquences de coupure sont relativement peu différentes au sein d’un filtre*”. Notons que dans cette phrase, le lexie hypothèse est suivi du verbe “supposer que”. Ce qui valide la relation sémantique que nous avons observée et établie entre les deux formes.

Concernant l’expression “montage expérimental”, sept occurrences sont seulement trouvées. L’expression évoque un certain dispositif à réaliser dans le but de mener l’expérience. Elle est plus fréquente en électronique où le pourcentage d’emploi est de 82%.

Ces résultats confirment ce que nous avons trouvé à propos des mots “synonymes” dans les corpus des enseignants de physique tout en montrant l’importance de la consultation des corpus d’enseignants révélant la présence des caractéristiques spécifiques à la discipline. Cette partie nous indique également que nous pouvons exploiter, dans le futur, les écrits de l’électronique dans Scientext si nous nous intéressons à l’écrit scientifique relevant du champ de la physique.

5.1.5 Bilan de l’analyse du corpus d’experts *vs* corpus de cours

Le travail de recherche et la comparaison qu’on a menés à travers les plateformes des corpus d’experts et les cours écrits des enseignants permettent de dire qu’il existe une grande similarité dans les collocations et les routines langagières utilisées dans différentes disciplines. Ce qui confirme le point de vue de Tutin selon lequel elle affirme qu’un écrit scientifique est illustré à travers un lexique transdisciplinaire “*différent de la terminologie des domaines scientifiques concernées*” (Tutin, 2014). C’est un lexique “*partagé par la communauté scientifique mis en œuvre dans la description et la présentation de l’activité scientifique [...] qui ne renvoie pas aux objets scientifiques des domaines de spécialité, mais au discours sur les objets et les procédures scientifiques*” (Tutin, 2007a).

D’un autre côté, ce diagnostic a révélé la présence d’autres notions alternatives fréquemment employées par les enseignants. Elles indiquent la spécificité du domaine de la physique comme : “montage expérimental” pour décrire l’“expérience”, “valeurs” pour

représenter les “résultats”, “supposons que” pour désigner ou faire une “hypothèse”.

Après avoir analysé les cours des enseignants de physique en les confrontant aux écrits d’experts dans différents domaines de sciences et après avoir tiré les résultats correspondants, l’étape suivante consiste à analyser les productions écrites des étudiants afin de faire un diagnostic ciblé de la phraséologie liée à l’ensemble de notions choisies.

5.2 Observation des CR d’étudiants

Dans cette section, nous explorons les comptes rendus des étudiants. Dans un premier temps, nous allons analyser la présence des notions retenues dans les écrits tout en comparant leurs fréquences, puis, dans un deuxième temps, nous allons examiner les cas d’emploi et de non-emploi de la phraséologie associée à ces notions. Ces données seront confrontées à l’analyse de corpus d’enseignants pour mettre en lumière les différences. Nous émettons l’hypothèse que la phraséologie recherchée dans les écrits des étudiants est moins variée que celle des experts et des enseignants.

5.2.1 Fréquence de l’emploi des notions

Dans un premier temps, pour s’assurer de l’homogénéité des échantillons à analyser, nous avons choisi des comptes rendus portant sur la même thématique et présentant la même expérience (ou travail pratique) à mener. Notre étude s’articule autour de la phraséologie, nous n’allons pas alors nous arrêter sur l’organisation textuelle des écrits qui diffère d’un étudiant à un autre ni sur les erreurs de grammaire, et d’orthographe présents comme ces points ne constituent l’objet de notre étude.

Si l’on examine la présence des trois notions choisies pour l’étude (hypothèse, expérience, résultats), nous remarquons que la fréquence de leurs emplois varie entre 0 et 2 selon l’étudiant. En particulier, le mot “hypothèse” n’est présent dans aucun de ces écrits. Nous avons observé l’absence de l’ensemble des notions dans trois CR.

Par ailleurs, le mot “expérience” est employé par deux étudiants dans le titre d’une section. Un de ses étudiants utilise ce mot dans le corps de son écrit.

Les mêmes étudiants utilisent le mot “résultats” dans le titre d’une autre section mais également dans le corps de leurs comptes rendus. D’autre part, un étudiant utilise le mot “résultats” dans le corps de son CR sans le mettre dans le titre d’une section.

Le tableau 5.1 récapitule l'emploi des notions et l'endroit de leurs emplois dans le texte.

Notion	Fréquence	
	Titre d'une section/Légende	Corps du texte
HYPOTHÈSE	0	0
EXPÉRIENCE	3 occurrences	3 occurrences
RÉSULTATS	2 occurrences	5 occurrences

TABLE 5.1 – *La fréquence d'emploi des notions par les étudiants*

Les résultats obtenus permettent de dire que les étudiants connaissent les notions “résultats” et “expérience” et les utilisent à plusieurs reprises et à de différents endroits dans leurs écrits. Mais le fait que ces notions sont présentes dans leurs écrits ne veut pas pour autant dire qu'ils en maîtrisent les constructions. L'absence du nom “hypothèse” et la faible fréquence des notions dans le corpus des étudiants peuvent être renvoyées à la taille du corpus mais aussi au fait qu'ils ne savent pas à quel moment les utiliser. Il y a donc lieu de penser que ce public n'est pas conscient de la valeur sémantique de ces notions.

Dans ce qui suit, nous allons considérer la phraséologie liée à l'ensemble des mots choisis et qui est en particulier employée dans le corps de textes.

5.2.2 L'emploi des collocations

Cette section reprend les trois notions qui nous intéressent afin d'examiner la présence et la construction de collocations liées. Comme le mot “hypothèse” n'est pas employé dans les écrits des étudiants, nous allons nous focaliser sur la phraséologie qui correspond aux mots “résultats” et “expérience”.

Phraséologie associée au “résultats” :

Nous avons relevé plusieurs patrons syntaxiques associés au mot “résultat(s)” avec des constructions au passif telles que : *les résultats obtenus*. Dans les écrits des étudiants, la collocation est soit suivie de deux points comme : *Les résultats obtenus .:*, soit d'un verbe : *Les résultats obtenus...sont les suivants .:*

Nous trouvons également la forme active avec “résultat” au singulier : *“Le résultat expérimental est de la forme :”* et *“L’expérience a vérifié les résultats théoriques”*.

Dans ces deux exemples, nous trouvons que le lemme “résultat” se combine avec les adjectifs : expérimental et théorique. Dans la première phrase, il a pour fonction sujet et dans la deuxième c’est un complément d’objet.

Les extractions trouvées présentées ci-dessus indiquent que les étudiants maîtrisent la collocation “*les résultats obtenus*”, une des collocations associées au mot “résultat” sous forme passive réduite. Les formes passives dans l’écrit scientifique ont pour fonctions principales d’ “assurer une progression à thème constant”, d’ “effacer le sujet, alléger la structure phrastique” et de contribuer à “l’impersonnalité du texte technoscientifique” (Yan, 2017). Toutefois, il convient de noter que cette collocation est soit suivie d’un tableau soit d’une courbe qui représentent les valeurs numériques des observables mesurés. Les étudiants n’utilisent pas des constructions verbales servant à rajouter une évaluation de résultats ou bien pour parler de leurs apports comme c’est le cas des collocations relevées dans les écrits d’experts et celles des enseignants. En effet, les adjectifs “théoriques” et “expérimental” qui s’associent au mot définissent la nature de résultats mais ne contribuent pas à leurs évaluations. C’est, par ailleurs, une caractéristique des écrits novices où les scripteurs débutants ne savent pas trop comment se démarquer au sein de la communauté scientifique.

Sur le plan de la discipline, nous avons remarqué que, pour parler des résultats, les étudiants se réfèrent, tout comme les enseignants de la discipline, au mot “valeurs” qu’ils privilégient pour, par exemple, comparer les résultats obtenus à travers deux méthodes : *“Les valeurs théoriques sont en accord avec les valeurs expérimentales”*, ou bien pour présenter le résultat sous forme d’une valeur : *“La valeur obtenue par cette expérience est :”*. Notons que pour parler d’une valeur, il convient de préciser à quoi correspond cette valeur c.à.d. l’observable dont on calcule/mesure la valeur. Nous pouvons considérer que cette absence de complément de nom dans les écrits des étudiants est une erreur par rapport à ce qui est employé par les enseignants où on précise l’observable mesuré.

Phraséologie associée à “expérience” :

Concernant ce mot, une collocation déjà relevée à partir des écrits d’experts a été trou-

vée dans le corpus d'un étudiant qui l'a employée dans une forme nominale :

“Description de l'expérience”. Elle correspond au titre d'une section. Aucune collocation verbale construite avec le mot “expérience” n'est trouvée dans les CR des étudiants.

Bien que nous pourrions renvoyer cet emploi à la collocation trouvée dans la base LST *“décrire l'expérience”*, l'écrit de l'étudiant témoigne de deux constructions erronées avec le mot “expérience”. La première construction se trouve dans : *La valeur obtenue par cette expérience*, le deuxième emploi de la notion est repéré dans la phrase : *L'expérience a vérifié les résultats théoriques*.

Si nous prenons le premier exemple, l'étudiant souhaite dire : *“la valeur de X obtenue après avoir mené cette expérience”* ou bien *“Le résultat obtenu dans cette expérience”*. L'étudiant a fait une construction au passif où le mot “expérience” est un complément d'agent, ce qui ne peut pas s'appliquer ici. En effet, si on transforme la phrase à la forme active, on voit bien que “cette expérience” ne peut pas “obtenir la valeur”. La collocation “l'expérience obtient” n'existe pas.

En outre, l'emploi du verbe vérifier, dans le deuxième exemple, n'est pas adéquat. L'emploi erroné du verbe “vérifier” nous permet de voir que l'étudiant semble ne pas bien maîtriser le sens de ce verbe utilisé pour signifier : *“examiner l'exactitude, l'authenticité, la conformité d'une chose. Prouver, confirmer.”* plutôt que *“démontrer, mettre en évidence”* tel qu'il est employé dans l'écrit. Ici, l'étudiant exprime que l'expérience montre que les résultats obtenus sont compatibles avec les prédictions théoriques, donc il vaut mieux indiquer : *“les résultats de l'expérience ont vérifié/confirmé les prédictions de la théorie”* ou bien en utilisant la collocation *“l'expérience montre que”* : *“l'expérience montre que les résultats obtenus sont compatibles avec les résultats théoriques”*. Le choix du collocatif “vérifier” ne semble pas correspondre à une collocation typique de la base “expérience”.

Nous avons relevé des difficultés pour ce public dans la maîtrise des collocations associées au mot “expérience”. Nous estimons que l'état d'interlangue imparfaite, associé à la non-maîtrise du genre de l'écrit scientifique et au déficit de la compétence discursive, constitue un handicap qui gêne considérablement les étudiants dans la rédaction académique.

Avant de dresser un bilan de l'analyse menée, il nous semble important de nous arrêter sur l'exposition des étudiants aux collocations en salle de classe.

5.2.3 Les étudiants face aux collocations

En salle de cours, les étudiants de physique ne sont presque jamais exposés aux collocations que les experts et les enseignants emploient dans leurs rédactions scientifiques. En effet, rares sont les professeurs qui notent des phrases complètes lors de l'explication du cours. Ils se contentent d'écrire au tableau des phrases du type nominales ou bien le noyau d'une phrase c.à.d. le nom pour noter ce qu'ils étaient entrain d'expliquer. Pareil pour ceux qui utilisent un diaporama pour expliquer le cours. Par exemple : au lieu d'utiliser explicitement l'expression "*faire/avancer/émettre l'hypothèse suivante*", ils notent le mot "*hypothèse :*" pour alléger l'écriture au tableau. D'ailleurs, à l'oral, lorsque le professeur emploie l'expression complète en classe c'est en arabe qu'il exprime le verbe et il préserve l'emploi français au mot "*hypothèse*" qui est plus technique. La majorité des professeurs ne voient l'utilité de l'emploi complet de ces phraséologies que lorsqu'ils rédigent un article de recherche ou lorsqu'ils s'occupent de la correction d'un travail rendu par les élèves du niveau supérieur tel que le mémoire d'un stage de Master 2.

5.2.4 Récapitulatif du diagnostic

Notre diagnostic a mis en évidence un écart entre les pratiques des écrits des enseignants et experts d'un côté, et celles des étudiants d'un autre côté. À l'issue de l'analyse, nous constatons que les collocations sont très peu fréquentes dans les écrits des étudiants. Cette absence de l'utilisation de collocations liées aux lexies : expérience et hypothèse s'explique, selon nous, par le manque d'outils à leur disposition et par le manque de présentation dans les méthodes de français.

Les erreurs présentées ci-dessus nous conduisent à supposer que les étudiants manquent de moyens pour trouver les collocatifs ou les cooccurrents appropriés. Les dictionnaires et les outils de traduction ne leur proposent pas ces informations de combinatoire lexicale ; ils choisissent par conséquent soit de créer une forme inexistante, soit d'employer un terme qui ne traduit pas le sens visé.

Ce qui est encourageant, c'est que certains étudiants s'inspirent de la manière dont les enseignants rédigent leurs cours. Ils présentent des emplois similaires à ceux identifiés dans le corpus de cours comme par exemple, l'utilisation de mot "valeurs" pour parler des résultats. Cependant, ils ne le font pas d'une manière appropriée. Leurs écrits sont légers et manquent de développement phrastique. D'un autre côté, nous trouvons que les étudiants n'utilisent pas les collocations employées par les experts. Seule la colloca-

tion “les résultats obtenus” a été repérée dans le corps du CR mais encore sans aucun développement supplémentaire. En effet, si on considère la construction dans laquelle la collocation “résultats obtenus” est suivie de deux points, nous pouvons affirmer que, de manière générale, les étudiants ne maîtrisent pas les collocations verbales puisqu’ils s’arrêtent avec deux points. Il apparaît qu’ils ne sont pas familiarisés avec les sens du verbe appartenant à l’écrit scientifique, ignorant ainsi les rôles de ces verbes dans le contexte scientifique, car cette dimension ne leur est pas enseignée. Autrement dit, ils n’ont pas à leur disposition les (ou pas suffisamment de) moyens linguistiques permettant de se positionner et de parler des apports de résultats, de faire un commentaire personnel sur les résultats, etc. Ces lacunes peuvent également être liées au fait qu’ils n’ont pas encore construit une représentation précise du fonctionnement de la langue scientifique.

D’où il nous semble important de les sensibiliser au phénomène de collocations/co-occurrences. Rappelons que cette analyse des spécificités des CR des étudiants a pour objectif de mieux outiller l’Enseignement/apprentissage de la phraséologie. Nous proposerons dans la section suivante, des pistes didactiques.

5.3 Propositions didactiques

Les apprenants ignorent le rôle de ces collocations dans la structuration du discours, soit parce qu’ils ne se sentent pas à l’aise en les utilisant, soit parce qu’ils ne savent pas comment les utiliser. Nous pouvons donc affirmer l’importance de prendre en considération ces collocations comme objet d’enseignement en tant que tel afin d’aider les apprenants à comprendre leur fonctionnement pour ensuite les réinvestir dans leurs écrits. Des entretiens plus ciblés avec les apprenants et les enseignants nous semblent intéressants à envisager. Ils serviraient d’un côté à évaluer les habitudes d’utilisation de la phraséologie scientifique chez les apprenants et à révéler les raisons derrière leur faible emploi.

D’un autre côté, nous pourrions adapter l’enseignement afin de développer de différentes stratégies d’apprentissage chez les apprenants. Il manque, nous semble-t-il, un enseignement systématique des collocations et routines langagières liées à l’écrit scientifique de la part des enseignants de français et de la discipline. Si le français assuré à l’université libanaise au sein de la faculté des sciences est d’ordre général, il est alors essentiel de développer l’enseignement du français académique lié à la discipline, soit pendant les cours de Français dispensés auprès des étudiants à l’université, soit à travers une forma-

tion dédiée à l'écrit universitaire, soit par les enseignants de la discipline lorsque le cours le permet tel qu'il est le cas des cours de TP. À notre sens, il est pertinent de profiter des cours où un type d'écrit est à rendre afin de mieux travailler l'écrit académique des scripteurs novices à un stade d'étude relativement précoce. Dans ce cas, nous suggérons la mise en place d'une évaluation comprenant en partie la rédaction scientifique.

Afin de remédier aux lacunes relevées dans la rédaction de CR, nous proposons d'expérimenter auprès des étudiants des exercices reposant sur des concordances présélectionnées liées aux notions choisies. Les exercices peuvent varier entre des exercices lacunaires (collocations synonymes), de reformulation et d'association jusqu'à ce qu'ils terminent par une tâche finale de production écrite d'un court paragraphe dans laquelle les étudiants pourraient réemployer les collocations apprises. Il est préférable que l'activité rédactionnelle soit dans le contexte d'étude. Elle pourrait s'articuler, par exemple, autour de la présentation et de l'interprétation de certains résultats donnés auparavant aux étudiants (si nous ciblons en particulier la notion "résultats") ou bien la rédaction d'une conclusion si nous ciblons les différentes notions choisies.

Avant de procéder aux exercices, il est important que les étudiants passent par une phase d'initiation au cours de laquelle ils découvrent le genre de l'écrit scientifique, apprennent à repérer les différentes parties d'un article scientifique et découvrent la structure organisationnelle de l'écrit scientifique (Yan, 2017).

Par ailleurs, il nous semble important qu'il y aurait des exercices traitant à la fois des extraits de cours relevant de la discipline et des extraits des comptes rendus des étudiants afin de les confronter les uns aux autres. L'apprenant est amené à découvrir comment interroger le texte scientifique pour ensuite être capable de chercher seul des éléments linguistiques. Cette mise en contraste pousse les étudiants à travailler sur certaines concordances présélectionnées et les aide à appréhender leurs erreurs d'emploi de collocations. D'un côté, les erreurs commises par les étudiants constituent des exemples qui pourraient être travaillés durant les activités afin qu'ils prennent conscience de l'emploi correct des collocations. Par exemple : *l'expérience vérifie* vs *l'expérience montre*. D'un autre côté, la collocation "*les résultats obtenus*" suivie de deux points, peut être travaillée davantage pour étoffer la phrase en employant les collocations verbales tirées des corpus d'experts et d'enseignants. Il nous semble essentiel aussi d'introduire la notion d'"*hypothèse*" aux étudiants pour qu'ils puissent utiliser les collocations correspondantes dans leurs écrits scientifiques.

En outre, il faut relier davantage la phraséologie et l'utilisation du corpus dans l'enseignement/apprentissage de la langue. L'objectif des corpus étant d'amener non seulement les formateurs à des fins d'utilisation didactique pour enseigner une langue mais aussi l'étudiant à "s'appropriier l'outil d'exploitation du corpus afin de le manipuler à sa guise" (Cavalla & Loiseau, 2014). Former les étudiants à exploiter les corpus pour maîtriser les phénomènes phraséologiques et apprendre l'écriture académique permet à chacun de travailler à sa manière sur les éléments qui lui sont directement utiles. Pour ce faire, nous proposons d'adopter la démarche mise en place par Cavalla (2019) dans l'enseignement de la phraséologie transdisciplinaire.

L'approche commence par une phase de découverte de l'outil corpus Scientext, les possibilités de requêtes, son appropriation libre puis avec des contraintes linguistiques. Afin de se familiariser avec l'outil, nous proposons qu'ils cherchent dans un premier temps ce qui est relatif aux notions d'hypothèse, de l'expérience et des résultats à travers l'entrée lexicale (chercher les éléments associés au mot "hypothèse) ainsi qu'à travers une entrée sémantique (chercher comment parler des résultats et apports dans un écrit académique). Finalement, une tâche finale écrite d'un court paragraphe leur permettrait de restituer les éléments phraséologiques déjà trouvés pour écrire un paragraphe du type présentation et interprétation de résultats.

Cavalla a appliqué sa démarche didactique auprès des étudiants de différents niveaux en langue et de différentes disciplines et niveaux d'études. Nous suggérons de travailler, avec les étudiants en physique, sur des textes relevant de la physique ou bien une discipline proche telle que l'électronique et la mécanique disponibles sur Scientext. Il revient alors d'adapter la démarche et les outils à son public en fonction des besoins spécifiques de ce dernier.

À travers cette approche, l'étudiant exploite et découvre le corpus notamment les collocations liées à l'écrit d'un compte rendu de TP pour ensuite être capable de chercher et trouver seul les éléments linguistiques dans le corpus. Cette approche leur permet d'apprendre à apprendre, de découvrir par eux-mêmes les régularités. La formulation des règles par eux-mêmes développe chez eux une prise de conscience linguistique ce qui facilite une bonne mémorisation.

D'autres outils peuvent être également exploités par les apprenants et les enseignants

afin de travailler l'écrit académique. À ce sujet, nous présentons Dicorpus² conçu par Tran et Falaise (2016) et déjà utilisé dans des travaux didactiques autour de la phraséologie transdisciplinaire (Yan, 2017; Cavalla 2018b). Cet outil fonctionne comme un dictionnaire mettant à disposition des utilisateurs des caractéristiques syntaxiques et sémantiques du LST ainsi que des exemples des concordances extraites. Par exemple, pour exprimer un enchaînement thématique et ajouter une nouvelle information, nous utilisons l'entrée onomasiologique (accès par sens), c.à.d. nous nous situons d'abord dans un contexte puis consultons les entrées lexicales possibles. Nous trouvons alors les expressions : “*de plus, en outre, de surcroît, entre autres, par ailleurs, du reste, d'ailleurs*”. En cliquant sur ces expressions, des exemples d'emplois sont directement fournis. La figure 5.5 illustre l'interface de Dicorpus selon l'entrée onomasiologique.

Séquences lexicalisées à fonction discursive dans les écrits scientifiques			
MODE D'ACCÈS	POUR EXPRIMER...	SOUS-CATÉGORIE	EXPRESSION
<ul style="list-style-type: none"> ▶ Accès par sens ▶ Accès par expression 	<ul style="list-style-type: none"> une exemplification ... un enchaînement thématique ... une énumération ... une reformulation ... une comparaison 	<ul style="list-style-type: none"> ... insister sur une idée ... ajouter une nouvelle information ... mettre en relief deux aspects 	<ul style="list-style-type: none"> de plus (****) en outre (****) de surcroît (*) entre autres (***) par ailleurs (****) du reste (*) d'ailleurs (****)
<p>CATÉGORIE adverbe</p> <p>POSITION DANS LA PHRASE en position postverbale</p> <p>PORTÉE - Locale (syntagme nominal) - Globale (proposition)</p> <p>DÉFINITION par exemple, sert à présenter un élément / un exemple qu'on veut distinguer d'un ensemble</p> <p>PONCTUATION Précédé d'une virgule</p> <p>FRÉQUENCE ***</p>			

FIGURE 5.5 – Entrée onomasiologique de l'outil Dicorpus “ajouter une nouvelle information”.

Nous soulignons qu'après avoir réalisé l'expérimentation, les résultats peuvent se varier d'un étudiant à l'autre et que l'efficacité de l'approche à employer va fortement dépendre des styles d'apprentissages. L'enseignement devrait non seulement tenir compte du niveau et des besoins des apprenants, mais également de leur style d'apprentissage. Nous pensons que le fait de travailler l'approche basée sur le corpus avec des étudiants de Master pourrait leur être plus bénéfique puisqu'ils sont censés avoir un niveau plus

2. <https://dicorpus.aiakide.net>

avancé en langue et qui plus est, une motivation plus grande pour apprendre les outils de la rédaction scientifique. Les étudiants de la licence sont encore les plus novices en matière de rédaction scientifique. Il faudrait leur proposer un apprentissage progressif et cadré et ne pas fournir des exemples trop complexes.

Pour conclure, les différentes propositions d'activités constituent un exemple illustrant notre réflexion sur l'enseignement de collocations au moyen de corpus (relatif à la discipline ou autre). Cette approche accorde une place importante à la réflexion métalinguistique du fait que le travail sur les concordances met en jeu une activité d'observation et des compétences d'induction. Une expérimentation reste à faire pour étudier les apports ainsi que les limites des réflexions établies.

Conclusion

La présente recherche met en jeu un processus d'extraction de données phraséologiques spécifiques aux écrits en physique. La méthodologie élaborée semble fournir des résultats importants. Elle apparaît comme une réponse adéquate aux besoins des étudiants en ce qui concerne les phénomènes collocatifs et plus précisément ceux associés aux noms "*résultats, hypothèse et expérience*" que nous pourrions trouver dans les rédactions académiques de la discipline en question. Le propos principal de ce travail de recherche a été de relever et comparer les collocations employées par les experts, les enseignants d'un côté et par les étudiants d'un autre côté. En mettant en relief les rôles que les collocations jouent dans l'écrit scientifique et en ciblant les besoins langagiers des étudiants à l'université libanaise, nous avons souligné leur importance dans l'enseignement et l'apprentissage de l'écrit académique. Nous allons retracer les résultats essentiels de notre travail avant de prendre du recul et de proposer des perspectives pour de futures études.

D'abord, du point de vue méthodologique, l'utilisation des corpus informatisés d'experts a permis d'extraire les collocations récurrentes liées aux lexies choisies. Ensuite, l'étude du corpus des cours a mis en évidence la similarité entre l'emploi de la phraséologie par les enseignants dans leurs cours et celle utilisée par des experts dans leurs travaux de recherche. En examinant les notes de cours, nous avons fait émerger d'autres mots récurrents, propres à la discipline et qui s'utilisent pour désigner ou évoquer les mêmes notions choisies. Les résultats d'analyse obtenus offrent ainsi de nouvelles données concernant la phraséologie employée dans la discipline ciblée.

Il convient de souligner un apport principal de cette recherche qui relève du diagnostic

des besoins langagiers des étudiants libanais à partir d'un corpus constitué de leurs productions écrites ainsi qu'un autre comprenant les cours des enseignants. Nous pensons avoir contribué à montrer l'utilité des corpus d'enseignants et d'étudiants dans l'enseignement et l'apprentissage de l'écrit académique. En établissant un diagnostic centré à la fois sur les erreurs et les cas d'emploi et de non-emploi des collocations, nous avons montré que les difficultés des étudiants sont principalement liées aux lacunes discursives et non seulement lexicales. Les difficultés que rencontrent les novices en matière de rédaction scientifique les empêchent d'assumer le rôle de chercheur, d'étayer leur position par rapport aux résultats obtenus et de construire la cohérence textuelle. Les lacunes lexicales et grammaticales peuvent être comblées au fur et à mesure de l'apprentissage de la langue, mais les difficultés d'ordre discursif, nécessitent un accompagnement pédagogique spécifique et présentent un enjeu didactique.

L'analyse de données phraséologiques dans les CR des étudiants a permis d'observer les phénomènes collocatifs pratiqués et d'en proposer d'éventuelles améliorations. Pour répondre aux besoins langagiers, nous avons proposé des démarches d'enseignement des collocations en intégrant des corpus d'écrits scientifiques en classe de langue ou en classe de la discipline dite non linguistique lorsqu'un écrit est attendu. De telles approches présentent le grand intérêt de sensibiliser les étudiants aux formules discursives, aux processus d'analyse ou au positionnement scientifique.

Finalement, cette recherche inscrite dans la linguistique de corpus a confirmé l'intérêt des corpus pour le repérage des difficultés chez les apprenants et le repérage d'exemples d'utilisation de concordances sélectionnées.

Bien que cette étude a conduit à mieux comprendre les spécificités des écrits scientifiques en physique et de découvrir les besoins langagiers des étudiants, elle connaît plusieurs limites. La première limite concerne le petit nombre de notions traitées. La perspective ouverte de cette recherche consiste à inclure d'autres notions utilisées pour par exemple : conclure et reformuler. La deuxième limite est liée à la taille du corpus d'étudiants. Afin de tester si les résultats de diagnostic sont également valables pour d'autres étudiants, il est indispensable de constituer un corpus d'étudiants plus important et inclure des mémoires ainsi que des thèses rédigés en français par des étudiants de l'UL. Il nous semble d'ailleurs important de comparer d'un côté, des comptes rendus de TP des français natifs et des comptes rendus des francophones non natifs. La comparaison sert à vérifier si

le statut novice en matière de rédaction scientifique confronte de manière identique les deux groupes aux difficultés de l'usage des collocations.

La troisième limite est associée au fait que nous n'avons pas pu tester les propositions didactiques. L'expérimentation étant une étape principale pour dégager les limites des approches et faire des ajustements si besoin. Concernant l'utilisation du corpus, la question suivante se pose : quel extrait de texte choisir pour l'enseignement ? Il se peut que les exemples tirés de corpus spécialisé Scientext soient difficiles pour notre public d'étudiants en Licence. Il est possible d'envisager une expérimentation dans laquelle on pourrait donner des exemples venant d'un corpus s'inscrivant dans le domaine d'études des étudiants tel qu'un support de cours.

Puisque les corpus spécialisés à notre portée ne contiennent des textes français en physique, inclure des cours de la discipline rédigés par des enseignants français nous permet d'étudier les pratiques langagières ainsi que de préciser s'il est possible de considérer les cours des enseignants comme étant des écrits des experts.

La mise en place d'un corpus d'écrits d'experts en physique nous paraît indispensable afin de sensibiliser les étudiants aux régularités langagières relevant de la discipline et par la suite, développer l'écriture académique en français. Il s'agit d'un apprentissage guidé par l'enseignant et qui vise l'autonomie. En complément des outils classiques tels que les dictionnaires, le corpus constitue une aide supplémentaire à la rédaction académique.

Dans une étape ultérieure, nous cherchons à travailler l'écriture académique en anglais pour un public d'étudiants libanais francophones, en particulier les collocations les plus fréquentes repérées dans des articles de recherche en physique. La question est comment s'appuyer sur les connaissances des étudiants en termes de phénomènes collocatifs en français pour travailler leurs écrits dans une troisième langue qui est l'anglais. Ce serait intéressant par exemple de travailler la rédaction d'un résumé sans avoir recours à la traduction littérale des mots mais en adoptant une approche onomasiologique c.à.d à travers le sens.

Bibliographie

Aston, G. (1997). Small and large corpora in language learning. *PALC 97 : Paractical applications in language corpora*, 51 ? 62.

Barré-De Miniac, Ch. (1995). La didactique de l'écriture : nouveaux éclairages pluridisciplinaires et état de la recherche. *Revue française de pédagogie*, 113 (1), 93-133.

Barré-de Miniac, C., Brissaud, C., et Rispaïl, M. (dir.) (2004). *La littératie. Conceptions théoriques et pratiques d'enseignement de la lecture-écriture*. Paris, L'Harmattan.

Beillet, M. (2016). Conception d'un test d'évaluation des compétences en langue des étudiants allophones à l'entrée dans l'enseignement supérieur en France. (Thèse de Doctorat inédite). Université de Mons.

Beury, 2014) Beury, J.-N. (2014). *Physique, exercices incontournables*. Dunod. Paris.

Biggs, P. et Dalwood, M. (1976). *Les Orléanais ont la parole : Teaching Guide and Tapescript* (Livre du maître). Londres : Longman.

Bouchard, R. & Parpette, Ch. (2012). Littéracie universitaire et oralographisme : le cours magistral, entre écrit et oral. *Pratiques* 153-154.

Cavalla, C. (2018a). Lexique transdisciplinaire et enseignement aux étudiants allophones. Dans Tutin, A. & Jacques, M.-P (dir.). *Lexique transversal et formules*

discursives des sciences humaines, ISTE Editions, 191-214.

Cavalla, C. (2018b). Exemple d'enseignement de la phraséologie transdisciplinaire à l'aide de corpus numériques en FLE. *La lettre de l'AIRDF*, Association internationale de recherche en didactique du français, La didactique du lexique, 43-47.

Cavalla, C. et Hartwell, L. (2018). L'enseignement et l'apprentissage de l'écrit académique à l'aide de corpus numériques. *LIDIL*, 58. OpenEdition Journals.

Cavalla, C. (2015). Collocations transdisciplinaires dans les écrits de doctorants FLS/FLE. *Linx*, Presses universitaires de Paris Nanterre, 95-110.

Cavalla, C., & Loiseau, M. (2014). Scientext comme corpus pour l'enseignement. In F. Grossmann & A. Tutin (Éd.), *L'Écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes : Presse universitaire de Rennes, 163-180.

Cavalla, C. (2010). Propositions didactiques pour l'enseignement d'éléments phraséologiques en FLE. Dans O. Galatanu, M. Pierrard, D. V. Raemdonk et M.-E. Damar (dir.), *Enseigner les structures langagières en FLE*. Peter Lang, 147-158.

Cavalla, C. (2007). Réflexion pour l'aide à l'écrit universitaire auprès des étudiants étrangers entrant en Master et Doctorat. Dans Jan Goes et Jean-Marc Mangiante (dir.), *L'accueil des étudiants étrangers dans les universités francophones*, Artois Presses Université, 37- 48.

Cavalla, C., & Grossmann, F. (2005). Caractéristiques sémantiques de quelques "Noms scientifiques" dans l'article de recherche en français. *Akademisk prosa*, Skrifter fra KIAP Romansk institutt, Universitetet i Bergen, 47-59.

Delcambre, I., & Lahanier-Reuter, D. (2010). Les littéracies universitaires. Influence des disciplines et du niveau d'étude dans les pratiques de l'écrit. *Diptyque*, 18, 11-42. Repéré à <http://www.forumlecture.ch/sysModules/obxLeseforum/Artikel/431/>

Les-litteracies-universitaires.pdf

Delcambre, I., (1998). Parler d'une classe (la décrire ?) dans un mémoire professionnel. *Pratiques, La description*, 99, 89-103.

Duverger, J. (dir.) (2011). *Enseignement bilingue : le professeur de « Discipline Non Linguistique » : statut fonctions pratiques pédagogiques*. ADEB, Paris.

Fabre, C. (2001). Traitement automatique de textes : techniques linguistiques. *Techniques de l'ingénieur*, vol. HA, réf. H7258 (<http://www.techniques-ingenieur.fr/affichage/noeud.asp?nID=109407>)

Firth, J. R. (1957). A Synopsis of Linguistic Theory 1930–1955. Dans J. R. Firth (Dir.), *Studies in Linguistic Analysis* (Vol. Special Volume of the Philosophical Society, 1-32. Oxford : Basil Blackwell.

Garnier, S., Rinck, F., Sitri, F. et De Vogüé, S. (2015). Former à l'écrit universitaire, un terrain pour la linguistique ?, *Linx*, Presses universitaires de Paris Nanterre, 72.

Gerbault, J. (2012). Littératie numérique. Dans Molinié, M., et Moore, D. (dir.), *les littératies : une Notion en Questions en didactique des langues (NeQ)*, Recherches en didactique des langues et des cultures, les cahiers de l'Acedle, 9(2). Repéré à <https://journals.openedition.org/rdlc/3960>, DOI : <https://doi.org/10.4000/rdlc.3960>

Gharib, Y., et Hamdan Saadé, N. (2017). L'enseignement supérieur au Liban : un paysage complexe *Revue internationale d'éducation de Sèvres*, Centre international d'études pédagogique, 76.

Gledhill, C. (2000). *Collocations in science writing*. Tübingen : Gunter Narr Verlag Tübingen.

González-Rey, M.-I. (2010), La phraséodidactique en action : les expressions figées comme objet d'enseignement, *La Clé des Langues* [en ligne], Lyon, ENS de LYON/DGESCO (ISSN 2107-7029). Repéré à <http://cle.ens-lyon.fr/espagnol/langue/traduction/la-phraseodidactique-en-action-les-expressions-figees-comme-objet-d-enseignement>.

Goody, J. (2007). Pouvoirs et savoirs de l'écrit. Dans J.-M. Privat (Dir.) *La Dispute*. (traduit par M. Claires). Paris

Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Paris : Masson & Armand Colin.

Hatier, S., & Tutin, A., (2019). Lexique et phraséologie scientifiques transdisciplinaires en sciences humaines : de la modélisation à la création d'une ressource lexicale. Dans Cavalla, C., Tutin, A., & Burrows, A. (Dir.), *Quelle place pour le français scientifique dans un contexte universitaire ?*, Francophonie et innovation à l'université, n°1, 139 - 145.

Hatier, S. (2016). *Identification et analyse linguistique du lexique scientifique transdisciplinaire*. Thèse de doctorat, Université Grenoble Alpes.

Hoyek, S. (2004). Le français dans l'enseignement scolaire et universitaire au Liban. Cahiers de l'Association internationale des études françaises, n°56, 49-56.

Jaffré, J.-P. (2004). La littéracie : histoire d'un mot, effets d'une notion. La littéracie. Dans : Barré-de Miniac, C., Brissaud, C., & Rispaïl, M. (dir.). *Conceptions théoriques et pratiques d'enseignement de la lecture-écriture*. Paris, L'Harmattan, France.

Lahanier-Reuter D. (2003a). Tableaux apparaissant dans des textes à visée formatrice. *Spirale, L'organisation visuelle des tableaux*, 32, 33- 46.

Lahanier-Reuter D. (2003b). Différents types de tableaux dans l'enseignement des statistiques. *Spirale, L'organisation visuelle des tableaux*, 32, 143-154.

Lang, E. (2019). *L'écrit(ure) universitaire, une tâche située et complexe : approche holiste du processus d'adaptation de la compétence scripturale chez les apprenants avancés en FLE.* (Thèse de Doctorat inédite). Université de Strasbourg.

Mangiante, J.-M. et Parpette, Ch. (2012). Le Français sur Objectif Universitaire : de la maîtrise linguistique aux compétences universitaires. *Synergies Algérie*. 15, 147-166.

Nahas, C. (2009). *Financing and Political Economy of Higher Education in Lebanon*, Beirut : Economic Research Forum.

OCDE (1997). *Littératie et sociétés du savoir*. OCDE. Paris.

Pacman, M. (2005). Les apports possibles de la phraséologie à la didactique des langues étrangères. *Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication*, 8(1), 109-122.

Pecman, M. (2004). *Phraséologie contrastive anglais-français : analyse et traitement en vue de l'aide à la rédaction scientifique*. Thèse de doctorat. Université de Nice-Sophia Antipolis.

Privat, J.-M. & Kara, M. (2006). La littératie. Autour de Jack Goody. *Pratiques*, 131-132, 3-6.

Rinck, F. (2005). Images of scientific activity through the research article : a comparison between linguistics and literary studies, in K. Fløttum and O. Korsnes dirs., *Academisk Prosa*, Department of Romance Studies, University of Bergen, 3, 75-86.

Russel, D.-R. (2012). Écrits universitaires/écrits professionnalisants/écrits professionnels : est-ce qu'“écrire pour apprendre” est plus qu'un slogan ? *Pratiques*, 153-154, 21-34.

Tran, T. T. H., & Falaise, A. (2016). Ergonomie des dictionnaires pour l'aide à la rédaction, état de l'art et propositions. Présenté à *Journée "Work in progress" : autour des discours scientifiques*, Grenoble. Repéré à <https://pro.aiakide.net/publis/2016WIP-tran-falaise-slides.pdf>

Tran, T. T. H (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, Application aux marqueurs discursifs*. Thèse de doctorat, Université Grenoble Alpes.

Tutin, A., Rui, Y., Tran, T.-T.- H. (2018). *Routines verbales pour le Français Langue Étrangère : des corpus d'experts aux corpus d'apprenants*, dans Cavalla, C. et Hartwell, L. (dir.), *L'enseignement et l'apprentissage de l'écrit académique à l'aide de corpus numériques*, *LIDIL*, 58, OpenEdition Journals.

Tutin, A. (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*. 27-43. Rennes : Presses Universitaires de Rennes.

Tutin, A. et Grossmann, F. (dir). (2012). *Autour du corpus Scientext : de la constitution d'un corpus d'écrits scientifiques à l'étude des marques du positionnement et du raisonnement*. Presses Universitaires de Rennes.

Tutin, A. (2010). Evaluative adjectives in academic writing in the humanities and social sciences. Dans Lores-Sanz, R., Mur-Duenas, P. et Lafuente-Millan, E. (Dir.), *Constructing Interpersonality : Multiple Perspectives on Written Academic Genres*, Cambridge : Cambridge Scholars Publishing, 219-240.

Tutin, A., 2007a, Collocations du lexique transdisciplinaire des écrits scientifiques : annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique, *Cahiers de l'Institut de linguistique de Louvain*, 31(2-4), 247-262.

- Tutin, A., (Dir.) 2007b, Lexique et écrits scientifiques, *Revue française de linguistique appliquée*, XII-2.
- Nystrand, M. (2006). The social and historical context for writing research. Dans C. A. MacArthur, S. Graham, & J. Fitzgerald (dir.), *Handbook of writing research*, 11-27. New-York : The Guilford Press.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.
- Tutin, A. & Grossmann, F. (dirs.) (2014). *L'écrit scientifique : du lexique au discours. Autour de Scientext*. Rennes : Presse universitaire de Rennes.
- Yan, R. (2017). *Étude des constructions verbales scientifiques dans une perspective didactique : utilisation des corpus dans le diagnostic des besoins langagiers en FLE à l'aide des techniques de TAL..* Thèse de doctorat, Université Grenoble Alpes.

Résumé

La présente étude porte sur l'amélioration de l'écrit scientifique des étudiants universitaires et se trouve à la croisée de la linguistique de corpus et la phraséologie. Elle est précisément centrée sur la phraséologie transdisciplinaire employée à des fins d'expressions scientifiques. Notre objectif est d'élaborer une méthodologie d'enseignement-apprentissage basée sur des corpus d'experts, d'enseignants et d'étudiants dans le but d'identifier et de comparer les collocations et les constructions récurrentes associées aux trois notions : "*hypothèse, expérience et résultats*". Afin de mieux comprendre les difficultés des étudiants en matière de collocations, nous établissons un diagnostic qui comporte d'une part, la comparaison des collocations employées par les experts avec celles trouvées dans les cours d'enseignants et d'autre part, l'analyse des constructions dans des cours d'enseignants et dans des comptes rendus des étudiants. L'analyse des productions écrites a permis de dégager les erreurs d'emploi et les cas de non-emploi de certains phénomènes phraséologiques. Ce diagnostic a guidé également l'élaboration de nos réflexions sur le plan didactique. Les résultats ont montré que les étudiants, encore novices en matière de rédaction, manifestent des difficultés liées à l'écrit académique notamment en ce qui concerne la phraséologie. Afin de remédier à ces lacunes et répondre aux besoins langagiers, nous avons proposé des démarches didactiques dont l'objectif est de guider les étudiants vers l'autonomie dans leur apprentissage et dans la manipulation des corpus en vue d'une meilleure rédaction d'écrits académiques en français.

MOTS-CLÉS : linguistique de corpus, lexique scientifique transdisciplinaire, lexique scientifique général, collocations, phraséologie, écrits académiques, écriture scientifique.

Abstract

This study focuses on the improvement of university students' scientific writing. The topic of research can be described as a nexus between corpus linguistics and phraseology. We are essentially interested in the interdisciplinary phraseology used for scientific expressions purposes. We aim to develop a corpus-based teaching-learning methodology that concerns collocations and frequent constructions related to notions : "*hypothesis, experiment and results*". Our goal is to help foreign students enhance their academic writing in french language. To better understand the difficulties encountered by the students with regards to the found collocations, we establish a diagnosis that includes the comparison between collocations in experts' and professors' corpora as well as the analysis of the constructions used by the professors in their lecture notes and by the students when writing their reports. This analysis enabled us shed light on the errors and the cases of non-use of some phraseology phenomena and was essential to elaborate our reflections from an educational point of view. The obtained results show that students, who are not experienced yet in writing, face some difficulties related to academic writing especially in what is associated with phraseology. In order to fill this gap and satisfy the linguistic needs of students, we have sketched out some didactic steps so as to guide students towards autonomy in learning phraseology by exploiting corpora.

KEYWORDS : corpus linguistics, interdisciplinary scientific words, general scientific vocabulary, phraseology, collocations, academic writing, scientific writing