

UCLV
Universidad Central
"Marta Abreu" de Las Villas



FH
Facultad de
Humanidades

Departamento de Lengua Inglesa

TRABAJO DE DIPLOMA

Título: Banque des expressions linguistiques pour l'écriture de l'introduction des travaux de recherche en français

Autora : Aniabet Alfonso Valdés

Tutores: MSc. Yuliet González Madariaga y Dr. Amed A. Leiva Mederos

Santa Clara, Junio 2019
Copyright©UCLV

UCLV
Universidad Central
"Marta Abreu" de Las Villas



FH
Facultad de
Humanidades

English Language Department

DIPLOMA THESIS

Title: Banque des expressions linguistiques pour l'écriture de l'introduction des travaux de recherche en français

Author: Aniabet Alfonso Valdés

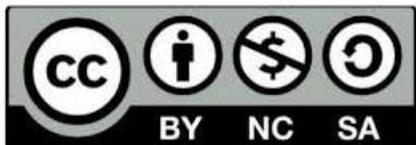
Thesis Directors: MSc. Yuliet González Madariaga y Dr. Amed A. Leiva Mederos

Santa Clara, June, 2019
Copyright©UCLV

Este documento es Propiedad Patrimonial de la Universidad Central “Marta Abreu” de Las Villas, y se encuentra depositado en los fondos de la Biblioteca Universitaria “Chiqui Gómez Lubian” subordinada a la Dirección de Información Científico Técnica de la mencionada casa de altos estudios.

Se autoriza su utilización bajo la licencia siguiente:

Atribución- No Comercial- Compartir Igual



Para cualquier información contacte con:

Dirección de Información Científico Técnica. Universidad Central “Marta Abreu” de Las Villas. Carretera a Camajuaní. Km 5½. Santa Clara. Villa Clara. Cuba. CP. 54 830

Teléfonos.: +53 01 42281503-1419

FACULTÉ DE SCIENCES HUMAINES
DÉPARTEMENT DE LANGUE ANGLAISE



Mémoire présenté en vue de l'obtention du diplôme de Licence ès
Langue Anglaise, option Langue Française

**Banque des expressions linguistiques pour
l'écriture de l'introduction des travaux de
recherche en français**

Auteure : Aniabet Alfonso Valdés

Directrice de recherche: Prof. Yuliet González Madariaga, MA.

Codirecteur: Prof. Amed Abel Leiva Mederos, PhD.

Santa Clara

Juin 2019



Hacemos constar que el presente trabajo fue realizado en la Universidad Central "Marta Abreu" de Las Villas como parte de la culminación de los estudios de la especialidad de Lengua Inglesa, autorizando a que el mismo sea utilizado por la Institución para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos ni publicados sin autorización de la Universidad.

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del Tutor

Firma del Jefe de Departamento

*Tout le talent d'écrire ne consiste après tout
que dans le choix des mots.*

Gustave Flaubert



Dédicace

À mes parents, ma sœur et Alis

À mon fiancé



Remerciements

- ✚ À mes parents pour leur soutien tout au long de ma vie, en particulier pendant ces sept années d'université et pour m'avoir appris à croire en moi. Merci de ne jamais abandonner malgré toutes les souffrances causées par moi, ou de m'avoir laissé abandonner après toutes mes tentatives de le faire. Merci pour votre préoccupation et votre immense amour, tout ce que je suis, je vous le dois. Je remercie infiniment Dieu de m'avoir donné la joie d'être sa fille. Je les aime!
- ✚ À ma Tatica adorée, une femme que j'ai toujours admirée, car elle s'est occupée de moi en tant que ma deuxième mère. Bien que nous ayons eu quelques disputes, je ne pourrai jamais arrêter de t'aimer comme je le fais, ni te remercier pour le soutien inconditionnel que tu m'apportes.
- ✚ Merci à ma princesse Alisita de m'avoir appris à aimer d'une manière totalement nouvelle et de me montrer qu'il faut surmonter tous les obstacles de la vie avec un beau sourire. Merci d'être la raison de mes joies et de m'aimer autant.
- ✚ Merci à mon fiancé Kiusler d'être toujours à mes côtés contre tous les pronostics et de me faire rire même dans les moments les plus difficiles. Merci de m'avoir écouté, de m'avoir encouragé à grandir en tant que professionnel et en tant que personne et de toujours croire en moi. Merci de me montrer que, malgré les problèmes, si nous nous aimons, rien d'autre ne compte, et de faire de moi la femme la plus heureuse du monde.
- ✚ À mes grand-mères, mes tantes et mes oncles pour être en attente et se préoccuper par la progression de cette thèse et me soutenir pendant toute ma vie.
- ✚ À Mirita et Aylé qui, malgré la distance, ont su être présents dans ma vie jour après jour.
- ✚ À la famille de mon fiancé pour m'avoir accueilli comme un autre membre de leur famille.
- ✚ À Yuliet pour ses conseils, ses suggestions et ses efforts inlassables pour mener cette recherche à bien et la perfectionner, malgré la distance et les circonstances. Merci parce que plus que ma directrice de recherche, je la

considère mon amie parce qu'elle m'a aidé à surmonter beaucoup de peurs et d'insécurités, et parce que quand vous enseignez avec l'amour que vous faites, il est impossible de ne pas aimer la langue française.

- ✚ À Amed, pour m'avoir ouvert les portes d'un monde totalement nouveau et incroyable, et pour avoir été capable de me guider avec une telle excellence lors du développement de cette recherche. Merci d'avoir travaillé sans relâche et de m'avoir toujours donné espoir et tranquillité.
- ✚ À Lienny car c'est aussi sa recherche, et parce que malgré les obstacles, elle m'a donné la force pour continuer.
- ✚ À la famille King, pour m'avoir donné tant d'amour inconditionnel car, presque sans me connaître, ils m'ont traitée comme une autre de leurs filles et craignent que rien ne me manque.
- ✚ À Didi et à Lissy, car même s'ils sont arrivés à deux moments différents de ma vie, ils ont su gagner un grand espace dans mon cœur, et ils ont toujours été là pour moi.
- ✚ À Miriam et à Jose de m'avoir laissé participer à leurs projets, de leur aide et de m'avoir fait confiance dès le premier instant.
- ✚ À toutes ces camarades de classe qui m'ont offert leur soutien durant ces six années. Janny, Lore, Roxy, Marta, Lenay, je vous remercie et je ne vous oublierai jamais.
- ✚ À tous les professeurs qui ont influencé de quelque manière que ce soit mon apprentissage tout au long de toutes ces années.
- ✚ Enfin, je désire remercier tous ceux qui, de près ou de loin, m'ont aidé à la réalisation de ce mémoire.



Résumé

Résumé

Actuellement, les outils linguistiques pour faciliter la rédaction de différentes sections du travail de recherche sont insuffisants. Cette étude aborde le processus pour l'élaboration d'une banque des expressions linguistiques basée sur des mouvements rhétoriques, à travers du *knowledge graph*, pour aider aux étudiants de la Licence ès Langue Anglaise option Langue Française, dans l'écriture de l'introduction de leurs travaux de recherche en français. Pour ce faire, un corpus de 50 mémoires écrits en français liés aux domaines de la linguistique et de la didactique a été compilé.

On a mis en place une méthodologie mixte avec prédominance quantitative ainsi que diverses techniques et outils d'exploration du texte pour réaliser la sélection des expressions. En outre, des techniques issues de la recherche sociale ont été utilisées pour déterminer le besoin d'une banque.

L'outil proposé dans cette recherche est conforme aux normes internationales requises pour ce type de lexique et facilitera la rédaction de l'introduction du travail de recherche en français.

Mots-clés : banque des expressions linguistiques, travail de recherche, introduction, mouvement rhétorique, *knowledge graph*.



Table de matières

TABLE DE MATIÈRES

Résumé

Liste des tableaux

Liste des graphiques

Liste des figures

Introduction.....	1
Chapitre I : Cadre théorique de la recherche.....	5
1.1 Écriture académique.....	5
1.1.1 Types de textes académiques.....	6
1.1.1.1 L'article de recherche.....	6
1.1.1.2 Le mémoire.....	6
1.1.1.3 Le travail de recherche.....	7
1.1.2 Marqueurs textuels.....	8
1.2 Les langues de spécialité.....	9
1.2.1 La langue de spécialité et la langue générale.....	10
1.2.2 Classifications des langues de spécialité.....	11
1.3 Genre.....	12
1.3.1 Le travail de recherche en tant que genre.....	14
1.3.1.1 L'introduction en tant que subgenre.....	15
1.3.2 Swales et l'analyse du genre.....	15
1.3.2.1 L'analyse des mouvements rhétoriques.....	16
1.3.2.2 Le modèle CARS de Swales.....	17
1.4 Linguistique de corpus.....	19
1.4.1 Le corpus.....	20
1.4.2 Type de corpus.....	21
1.4.3 Critères pour la construction de corpus.....	22
1.5 Les banques des expressions linguistiques	23
1.5.1 Type des banques des expressions linguistiques.....	24
1.5.2 Importance des banques des expressions linguistiques.....	24

1.5.3	Méthodes pour la construction des banques des expressions linguistiques.....	25
1.5.3.1	Manuel (compilation manuelle)	26
1.5.3.2	Numérique	26
1.6	Ontologies.....	27
1.6.1	<i>Knowledge graphs</i>	28
1.7	Conclusions partielles.....	29
Chapitre II : Références méthodologiques de la recherche.....		30
2.1	Contexte de la recherche.....	30
2.2	Étapes de la recherche.....	30
2.3	Type de recherche.....	31
2.4	Portée de la recherche.....	31
2.5	Opérationnalisation des variables.....	32
2.6	Méthodes de la recherche.....	36
2.7	Outils.....	42
2.8	Population et échantillon.....	44
2.9	Révision bibliographique.....	44
2.10	Limites de notre étude.....	46
2.11	Conclusions partielles.....	46
Chapitre III: Banque des expressions linguistiques: analyse des données et résultats.....		47
3.1	Résultats des techniques de collecte d'informations.....	47
3.2	Étapes pour l'élaboration de la banque des expressions linguistiques.....	49
3.2.1	Sélection du corpus.....	50
3.2.2	Analyse des mouvements rhétoriques.....	51
3.2.3	Extraction des expressions à travers les outils.....	56
3.2.4	Traitement et stockage des données.....	58
3.2.5	Normalisation.....	59
3.2.6	Édition de la banque des expressions linguistiques.....	59
3.2.6.1	Macrostructure.....	60
3.2.6.2	Microstructure.....	60

Conclusions.....	62
Recommandations.....	63
Bibliographie.....	64
Annexe I.....	72
Annexe II.....	73
Annexe III.....	76
Annexe IV.....	78
Annexe V.....	88

LISTE DES TABLEAUX

Tableau 1

Opérationnalisation des variables.....32

Tableau 2

Auteurs principaux.....44

Tableau 3

Années principales.....45

Tableau 4

Types de documents.....45

Tableau 5

Résultats de l'enquête.....47

Tableau 6

Caractéristiques du corpus sélectionné.....51

Tableau 7

Mouvements et étapes identifiés dans la section Introduction des mémoires.....52

Tableau 8

Fréquence des mouvements dans la section de l'introduction.....53

LISTE DES FIGURES

Figure 1

Outil de *Concordance* appliqué à un élément linguistique récurrent.....57

Figure 2

Outil *Word List* appliqué à la section Introduction.....58

Figure 3

Outil *Word Clusters / NGrams* appliqué à la section Introduction.....58

LISTE DES GRAPHIQUES

Graphique 1

Importance accordée à la section d'introduction.....73

Graphique 2

Degré de complexité de l'écriture des introductions.....73

Graphique 3

Connaissance de la structure des introductions des travaux de recherche.....74

Graphique 4

Connaissance des mouvements rhétoriques.....74

Graphique 5

La banque des expressions linguistiques sera-t-elle utile?.....75



Introduction

INTRODUCTION

L'écriture constitue une activité très fréquente dans le domaine universitaire. La plupart de connaissances développées dans ce contexte se reflète dans des documents écrits produits par des étudiants et des professeurs. De toutes les compétences, l'écriture est possiblement l'une des habilités les plus difficiles à acquérir car l'apprenant ne doit pas seulement représenter graphiquement la langue parlée mais aussi la développer et la présenter d'une manière structurée. Quand il s'agit de produire un texte académique, la tâche devient plus compliquée encore.

L'écriture académique est celle qui se produit dans le domaine universitaire et scientifique. Elle comprend des œuvres produites par des étudiants universitaires ainsi que les textes de l'académie pour la diffusion des connaissances scientifiques. Ainsi, le genre académique est composé de textes spécialisés circulant dans le domaine scientifique et, par conséquent, ces textes-ci ont certaines caractéristiques communes. Particulièrement pour les étudiants, indépendamment de la discipline dans laquelle ils s'insèrent, l'écriture de documents académiques tels qu'articles, mémoires ou thèses devient une activité très stressante car ils sont censés montrer la connaissance par rapport à un thème, en utilisant la structure rhétorique appropriée du document et en ayant un bon contrôle de la langue écrite. Quand les étudiants doivent utiliser l'écriture académique en seconde langue, la situation est encore plus difficile pour eux car il est très important de maîtriser des structures rhétoriques et linguistiques qui, parfois, sont très éloignées de celles de leur propre langue.

À l'heure actuelle, l'approche de l'analyse de genre s'avère être une méthodologie très utile pour contribuer à l'apprentissage de l'écriture (Paltridge, 1996 ; Dudley-Evans et Kay, 1997 ; Biber, 2007 ; Hyland, 2008). Selon le linguiste John Swales (1990), la notion de genre peut être définie comme un groupe d'événements communicatifs dans lesquels, des formes similaires ou le contenu du langage sont utilisés comme une représentation significative des objectifs communicatifs partagés par les membres de la même communauté discursive. En se basant sur ce critère, Swales envisage l'analyse de mouvements rhétoriques (*moves*) comme un élément essentiel pour la définition du genre. Cet auteur

considère le mouvement rhétorique comme une unité sémantique, définie par son intention de communication, et fondée sur la détermination du genre auquel appartient le texte.

Le mouvement rhétorique constitue l'une des principales sous-catégories pour organiser une banque des expressions linguistiques (Morley, 2014), un outil lexical qui contribue à améliorer l'écriture de divers genres, dont l'académique. Les connaissances liées à la banque des expressions linguistiques ont été attribuées de manière empirique à la technologie informatique dans l'identification de modèles phraséologiques récurrents dans de très grands corpus à l'aide de logiciels spécialisés.

L'intégration de graphes est l'une des procédures les plus efficaces pour résoudre le besoin de construire des systèmes linguistiques supportés automatiquement par RDF (*Resource Description Framework*) en tant que banques des expressions linguistiques (Leiva-Mederos, 2018).

Situation Problématique

Les étudiants de la Licence ès Langue Anglaise, option Langue Française doivent faire face à l'écriture d'un nouveau genre discursif pour en deuxième année de l'année scolaire: la rédaction du travail de recherche, en particulier, de la section *Introduction*. Ils peuvent choisir de le faire soit en anglais soit en français. Dans les années ultérieures, les étudiants doivent continuer à élaborer d'autres sections telles que le cadre théorique, méthodologique et des résultats (autrement dit, le *Développement* et les *Conclusions*). Malheureusement, pour ces étudiants qui choisissent le français comme langue de leur travail de recherche, l'écriture devient extrêmement complexe parce qu'ils ne possèdent pas d'outils linguistiques et lexicaux suffisants qui facilitent la rédaction de chaque section de ce type de texte académique, dans ce cas, de l'*Introduction*. En outre, dans le pays il n'y a non plus de bibliographie suffisante reliant les mouvements rhétoriques aux *knowledge graphs*, ce qui pourrait aider les étudiants de la Licence ès Langue Anglaise, option Langue Française dans l'élaboration de leur travail de recherche en français.

La situation précédemment exposée nous conduit à la **question de recherche suivante**: comment peut-on contribuer à améliorer la rédaction en

français de l'introduction du travail de recherche des étudiants de Licence ès Langue Anglaise, option Langue Française de l'Université Centrale « Marta Abreu » de Las Villas?

Hypothèse: L'élaboration d'une banque des expressions linguistiques à travers du *knowledge graph* contribuera à améliorer la rédaction en français de l'introduction du travail de recherche des étudiants de la Licence ès Langue Anglaise, option Langue Française de l'Université Centrale « Marta Abreu » de Las Villas.

C'est pour cela que nous menons à bien une recherche ayant pour **objectif général** : Élaborer une banque des expressions linguistiques basées sur des mouvements rhétoriques pour l'écriture de l'introduction du travail de recherche en français des étudiants de la Licence ès Langue Anglaise, option Langue Française de l'Université Centrale « Marta Abreu » de Las Villas.

Nous avons aussi établi comme **objectifs spécifiques** :

1. Identifier des fondements théoriques et méthodologiques de l'écriture académique, le genre et les mouvements rhétoriques, la linguistique de corpus et les banques des expressions linguistiques.
2. Diagnostiquer l'emploi des banques des expressions pour la construction des textes académiques.
3. Élaborer une banque des expressions linguistiques basée sur des mouvements rhétoriques pour l'écriture de l'introduction du travail de recherche.

La banque des expressions linguistiques basée sur des mouvements rhétoriques pour l'écriture de l'introduction du travail de recherche sera très utile pour les étudiants ainsi que pour les professeurs de la section Français. Premièrement, cette étude contribuera à faciliter l'acquisition de la compétence d'écriture, plus spécifiquement elle facilitera l'utilisation de l'écriture académique. Deuxièmement, les étudiants disposeront d'un outil lexical nécessaire pour comprendre et défier l'écriture académique. La recherche est destinée à faire prendre conscience aux apprenants des structures et objectifs des textes appartenant au discours académique. En outre, cette étude fournira une description détaillée des mouvements rhétoriques devant figurer dans l'introduction

des documents de recherche et des expressions linguistiques qui accompagnent ces unités rhétoriques. Enfin, cette étude fournira une banque des expressions linguistiques, un outil qui facilitera la rédaction de l'introduction du travail de recherche, ce qui sera très pratique pour les étudiants.

Afin de développer cette étude, un corpus monolingue spécialisé a été compilé. Il comprend 50 mémoires écrits en français. Les mémoires collectés sont liés aux domaines de la linguistique et la didactique de la langue française et ont été écrits par des francophones. En plus, ils ont été publiés entre les années 1998 et 2018. Cette période a été choisie pour la sélection des mémoires en raison de la mise à jour continue des connaissances dans ces deux domaines.

Structure du travail

La dissertation a été organisée de la manière suivante:

Dans le chapitre 1, les définitions les plus générales et les concepts théoriques en relation avec la recherche linguistique sont examinés. Les définitions comprennent le concept d'écriture académique, genre, mouvements rhétoriques, linguistique de corpus, les banques des expressions linguistiques, et les *knowledge graphs*.

Le deuxième chapitre concerne la méthodologie utilisée pour cette étude. Cette partie de la recherche est basée sur une méthode de recherche européenne et vise à décrire le processus de recherche, en tenant compte du contexte, des étapes, des types et de la portée de la recherche, ainsi que de l'opérationnalisation des variables, des méthodes de la recherche, des outils utilisés, de la population, de l'échantillon, et finalement, de la révision bibliographique.

Le troisième chapitre vise à décrire les résultats de l'étude. La méthodologie employée dans la recherche pour élaborer la banque des expressions linguistiques est explorée. Les étapes principales suivies sont également exposées pendant ce processus jusqu'à la présentation finale de la banque.

Cette thèse est complétée par une section de conclusion, suivie des recommandations, des références et des annexes correspondantes.



Premier

Chapitre

CHAPITRE I : CADRE THÉORIQUE DE LA RECHERCHE

Dans ce chapitre, les concepts d'écriture académique, genre, langues de spécialité, linguistique de corpus, corpus, banques des expressions linguistiques, ontologie et *knowledge graphs* vont être abordés. On fera allusion aux différents types de textes académiques, aux classifications des langues de spécialité, au travail de recherche et à l'introduction respectivement en tant que genre et sous-genre, aux types de corpus, et, aux méthodes pour la construction des banques des expressions linguistiques. On expliquera aussi l'analyse des mouvements rhétoriques et le modèle CARS de Swales pour faire cette analyse.

1.1 L'écriture académique

L'écriture académique est celle qui se produit dans les domaines universitaire et scientifique. Elle comprend le travail réalisé par les étudiants universitaires (examens et textes de divers genres tels que des travaux pratiques, évaluations, etc.), ainsi que les textes de l'Académie pour la diffusion des connaissances scientifiques. De cette façon, le genre académique est composé de textes spécialisés circulant dans le domaine scientifique et, par conséquent, qui ont certaines caractéristiques communes (Ramírez, 2007).

Cependant, ces textes ne sont pas seulement produits au sein de la communauté scientifique, mais également le genre académique intègre des textes diffusés dans le domaine universitaire et qui n'ont pas des origines académiques. Autrement dit, l'expéditeur du texte est un spécialiste dans un domaine spécifique et ses destinataires sont généralement aussi des spécialistes ou semi-spécialistes dans la même discipline que l'auteur (Ramírez, 2007). L'objectif principal du discours académique est de communiquer une découverte scientifique. Cependant, ce n'est pas le seul but, il doit aussi convaincre de la légitimité et de l'importance scientifique de cette conclusion (Ramirez, 2007). Ainsi, les fonctions du texte académique sont à la fois d'informer et de convaincre. Par conséquent les trames textuelles principales sont l'explicative et l'argumentative (Fernández & Bressia, 2009):

Les formats ou schémas organisationnels, qui aident à l'efficacité communicationnelle, sont les caractéristiques déterminantes des textes

académiques et scientifiques. Le degré élevé de conventionalité répond fondamentalement à deux faits: premièrement, les locuteurs exposent les formes textuelles avec certains contenus et, dans une procédure inverse, ils attribuent les sujets à certains genres; deuxièmement, l'organisation en sections met en évidence les étapes, les moments, les contextes et les procédures utilisés dans le processus de recherche (Puiatti, 2005: 26 dans Fernández et Bressia, 2009).

1.1.1 Types de textes académiques

Il existe une grande variété de textes académiques tels que le résumé, l'article de recherche, le mémoire, l'essai, la monographie, le compte-rendu, et le travail de recherche. Nous ne ferons référence qu'à ceux que nous considérons les plus pertinents pour notre recherche et nous soulignerons le type de texte académique sur lequel notre étude se concentrera.

1.1.1.1 L'article de recherche

L'article de recherche consiste à présenter les résultats d'une recherche scientifique. Pour cette raison, « ... le sujet de l'article de recherche est toujours très spécifique. Telle spécificité a pour conséquence que la formulation explicite du sujet consiste généralement en une construction nominale relativement étendue et complexe » (Adelstein et Kuguel, 2005: 56).

En tant que communication entre pairs et présentant les résultats d'une recherche, ce type d'article est constitué d'un texte court dont l'objectif est sa publication (environ 5 000 mots, chiffre qui peut varier), il utilise une terminologie spécifique, suit une structure canonique par sections, et représente une contribution originale à la communauté discursive scientifique (toujours liée au sujet de l'article) à laquelle elle est adressée (Swales et Feak, 2004).

L'article de recherche aspire à informer et à convaincre. Dans le cas de la deuxième de ses fonctions, le lecteur doit être convaincu de l'importance et de la validité des résultats de la recherche (Aldestein & Kuguel, 2005). Ainsi, les informations présentées dans l'article sont au service de la validité scientifique de l'hypothèse et des résultats présentés (Fernández & Bressia, 2009).

1.1.1.2 Le mémoire

Selon Fernández et Bressia (2009), un mémoire est une recherche effectuée par un étudiant, généralement intercédée par un enseignant ou un tuteur, en vue d'obtenir un diplôme universitaire. Outre le tuteur, il est également fréquent de confier à l'étudiant une commission de professeurs, qui viennent constituer avec le tuteur le comité d'évaluation du mémoire. Celui-ci est composé de certaines parties obligatoires et d'autres à caractère facultatif.

Fernández et Bressia (2009), établissent que, dans ce qui fait référence au sujet, celui-ci doit être suffisamment large pour permettre une recherche relativement longue, mais pas suffisamment pour le rendre non viable. De plus, il est nécessaire d'offrir une amélioration ou un nouveau point de vue par rapport à la réalité qui a existé jusqu'à présent.

L'objectif du mémoire est d'appliquer les connaissances acquises au cours de toutes ses années d'études, à un problème particulier. En outre, il a comme but de communiquer de manière organisée la méthode et les techniques de recherche, ainsi que les résultats ou conclusions obtenus (Fernández & Bressia, 2009).

1.1.1.3 Le travail de recherche

L'article de recherche et le mémoire sont deux types de textes largement utilisés dans le contexte académique. Tous deux doivent présenter une pertinence et une nouveauté d'investigation, et avoir pour destinataire une communauté discursive connaisseuse du sujet dont traitent ces textes.

Or, notre étude est centrée sur le *travail de recherche* parce que c'est le type de texte académique le plus employé par les étudiants de notre université. Il permet d'approfondir leurs connaissances sur un sujet donné, de clarifier leurs idées et de les communiquer de façon logique et rigoureuse. Il a pour but d'observer, d'expliquer, d'interpréter, de découvrir de nouvelles relations entre les faits et, après vérification, de reconstituer une réalité afin de donner une portée universelle aux faits étudiés. Il implique généralement de développer une thèse et de proposer des arguments qui la soutiennent. Le travail de recherche doit donc être très documenté (Malo, 2006).

Il exige sa défense devant un jury lors de l'examen oral correspondant, selon les procédures établies dans les règles de fonctionnement du programme et dans

le manuel des procédures relatives aux modalités d'obtention du diplôme préparé et approuvé par le comité académique (Swales et Feak, 2004).

Ce type de travail se divise en trois parties distinctes : l'introduction, le développement et la conclusion. Le corps du travail est structuré en chapitres, épigraphes, et sous-épigraphes.

D'après Fernández & Bressia (2009), l'introduction cherche à expliquer les contextes dans lesquels la recherche est menée. Dans celle-ci, l'hypothèse est présentée, le problème est justifié, l'intérêt scientifique du sujet est exposé, les objectifs de la recherche et les déclarations sur l'état actuel des connaissances sont indiqués. Les références bibliographiques doivent être analysées, développées; il ne s'agit pas d'exposer des concepts ou des problèmes, alors que les antécédents doivent donner un cadre théorique et méthodologique au travail. Dans certains cas, la méthodologie utilisée est présentée. Les informations incluses dans l'introduction sont réparties dans des sous-catégories préétablies:

- Établissement du domaine ou du territoire scientifique de la recherche.
- Établissement de la niche ou du sujet non encore étudié.
- Établissement du sujet de recherche (Puiatti, 2005).

1.1.2 Marqueurs textuels

Dans la section précédente, nous analysons certains textes académiques et nous nous concentrons sur celui qui sera analysé plus en profondeur. Des textes académiques ou des textes de toute autre nature doivent présenter des connexions indispensables pour maintenir un sens et une cohésion, ces connexions sont établies à travers de marqueurs textuels (également appelés connecteurs ou connectives). Ce ne sont que des unités superficielles qui guident les relations entre les propositions dans leur ensemble, c'est-à-dire la communication. Les unités peuvent être connectées sans cohérence visible, la connexion n'est donc pas une condition suffisante pour l'acceptabilité du discours. Selon Van Dijk (1998), une série de conditions doit être remplie pour que le lien sémantique entre phrases et propositions se produise: les faits dénotés doivent appartenir à des mondes liés; ils doivent être des faits compatibles; les concepts en jeu ne doivent pas être trop distants les uns des autres et, enfin, il doit y avoir

un sujet de conversation, c'est-à-dire un thème qui formule les propositions qui composent le discours (Gil-Leiva, 2008).

De nombreux auteurs ont étudié les connecteurs et chacun d'entre eux a proposé une nouvelle classification, mais l'une des plus complètes est, à notre avis, celle proposée par Narcisse Kouassi (2018).

D'après cet auteur, les marqueurs discursifs sont repartis en cinq grands groupes qui sont à leur tour composés de différents sous-groupes à savoir: Les structurateurs de l'information intégré par les sous-groupes des ordonnateurs, des commentateurs, des marqueurs de digression; les connecteurs formés par des connecteurs d'addition, des connecteurs consécutifs, des connecteurs contre-argumentatifs; les reformulateurs constitués des reformulateurs explicatifs, des reformulateurs rectificatifs, des reformulateurs de distanciation, et des reformulateurs récapitulatifs; les opérateurs argumentatifs divisés en des opérateurs de renforcement argumentatif, des opérateurs de concrétion; et les opérateurs conversationnels (Kouassi, 2018).

1.2 Les langues de spécialité

La notion de langue de spécialité fait référence à la terminologie propre, à une situation de communication spécifique dans un domaine particulier (Lavagnino, 2012). Malgré les nombreuses études sur cette langue, il est encore difficile de bien la définir, les chercheurs ne sont pas toujours d'accord sur l'extension à donner à ce concept. Ci-dessous, certains de ces concepts sont présentés:

1. Langues utilisées dans les situations de communication (orale ou écrite) qui impliquent la transmission d'une information relevant d'un champ d'expérience particulier (Galisson et Coste, 1976: 511).
2. Moyens de communication linguistique nécessaires pour transmettre des informations sur des sujets particuliers à des spécialistes du même sujet (Sager et al., 1980: 21, 182).
3. Le concept proposé par l'ISO (2000: 2) suit le même sens: «Langue utilisée dans un domaine et caractérisée par l'utilisation de moyens d'expression linguistiques particuliers. [...] Les moyens d'expression linguistiques spécifiques

incluent toujours une terminologie et une phraséologie spécifiques du domaine et peuvent également présenter des traits stylistiques ou syntaxiques.»

Malgré leurs divergences, ces définitions partagent plusieurs points communs. Par conséquent, *grosso modo*, la langue de spécialité est une variété de la langue générale. Elle présente un caractère linguistique global mais limité par la spécialité. En plus, elle a une fonction linguistique de communication, indique qu'elle appartient à une spécialité et met l'accent sur l'information et la connaissance.

1.2.1 La langue de spécialité et la langue générale

Comme on l'a déjà mentionné dans cette épigraphe, la langue de spécialité est une variété, un style de la langue générale. Bien qu'elle présente des caractéristiques particulières, elle partage des qualités communes avec le langage ordinaire utilisé quotidiennement. Elle diffère, entre autres choses, par le fait qu'elle est utilisée et comprise par un groupe de spécialistes qui l'utilisent pour communiquer entre eux.

Selon Auger (2008), la langue de spécialité met l'accent sur l'écriture. Son objectif est l'intellectualisation, c'est-à-dire la précision sémantique, la systématisation conceptuelle, la neutralité émotionnelle, l'économie formelle et sémantique. Il a aussi affirmé que dans cette langue, les unités lexicales ont tendance à être définies; à éviter la polysémie et l'homonymie; à supprimer la synonymie; à simplifier les moyens syntaxiques; à neutraliser les émotions et la subjectivité. Le lexique de la langue de spécialité est généralement très large. En fait, c'est le phénomène cognitif et linguistique le plus distinctif de cette langue.

Bien sûr, la langue de spécialité présente d'autres caractéristiques linguistiques spécifiques, telles que l'utilisation de termes et d'unités phraséologiques typiques de la spécialité, qui contribuent grandement à la concision sémantique et à l'utilisation d'un vocabulaire abstrait reflétant l'impersonnalité du texte scientifique.

À ces deux caractéristiques, il faut ajouter l'utilisation de coordinateurs et d'autres unités de connexion conjonctives et adverbiales permettant l'expression d'une pensée technique et scientifique par le biais de la classification, du

dénombrément et de la description, de l'induction et de la déduction, ainsi que l'argumentation (Auger, 2008).

En fait, dans la plupart des cas, la terminologie d'un texte est considérée comme l'élément le plus important qui différencie non seulement la langue spécialisée de la générale, mais aussi celle d'autres langues spécialisées (Cabré, 1993).

Dans ces textes, le présent indicatif prédomine en tant que temps verbal et la déclaration en tant que type de proposition. Les constructions impersonnelles et les phrases passives sont également utilisées comme les pronoms *nous* et *on* de modestie et le pronom invariable *il*, pour atteindre l'impersonnalité de la déclaration technoscientifique.

Les phrases sont aussi syntaxiquement complexes. En fait, la longueur des phrases, l'utilisation de constructions avec des participes et des constructions incisées, ainsi que la diversité des coordinateurs et des subordonnés contribuent à la complexité. D'autre part, des processus tels que la nominalisation et l'adjectivisation, sont utilisés pour limiter le nombre de verbes (Auger, 2008).

Bref, on parle souvent de la langue naturelle pour l'opposer soit à la langue artificielle, soit à la langue technique prise dans son acception la plus générale et recouvrant tous les langages « spéciaux » utilisés pour parler tant des sciences que des techniques. Selon les termes de Jean-Pierre Sérís (2000 : 238), « la langue naturelle est la langue qui n'a pas été inventée et imposée par la technique, la langue qui est transmise et assure sa pérennité ». En plus, les principales différences entre les langues de spécialité et générale se situent au niveau lexical.

1.2.2 Classifications des langues de spécialité

Étant donné que les discours spécialisés représentent les particularités conceptuelles de différents domaines de la connaissance, des distinctions claires doivent être établies en ce qui concerne la diversité des systèmes conceptuels. Laissant de côté les différentes positions théoriques, la première distinction pertinente concerne les spécialités elles-mêmes: le discours de la médecine, du droit, des sciences naturelles, etc. (Gómez, 2005).

Il faut également considérer que les langues spécialisées sont caractérisées par différents niveaux de contenu. Desmet (2006) décompose la langue de spécialité en suivant quatre critères: le degré d'abstraction, la manière naturelle / artificielle d'exprimer les éléments et la syntaxe, le support ou le type de spécialité et les participants.

Les catégories qui se retrouvent presque invariablement en consultant les études sur ce sujet peuvent être résumées comme ci-dessous:

- Discours scientifique spécialisé (*i.e.*: un manuel spécialisé ; une revue scientifique)
- Discours scientifique officiel (*i.e.*: des textes de loi)
- Discours scientifique pédagogique ou didactique (*i.e.*: un manuel pour les étudiants)
- Discours de semi-vulgarisation scientifique (*i.e.*: une revue de spécialité pour un public d'initiés)
- Discours de vulgarisation scientifique (*i.e.*: les secteurs spécialisés dans la presse générale) (Desmet, 2006)

En résumé, en ce qui concerne les langues spécialisées, leurs textes peuvent être choisis en fonction de leur appartenance à un ou plusieurs genres textuels et à un type de discours scientifique et technique correspondant à la division verticale traditionnelle de langues spécialisées (Gutiérrez Carreras, 2014).

Compte tenu de la classification précédente, la langue de spécialité qui sera analysé dans notre recherche appartient au discours scientifique spécialisé, en particulier, les mémoires liés aux domaines de la linguistique et la didactique de la langue française, écrits par des francophones dans leur langue maternelle.

1.3 Genre

Le concept de genre n'est pas nouveau, mais il s'est beaucoup développé, surtout dans les dernières décennies. En général, il se réfère aux moyens socialement reconnus d'utiliser une langue. Il est basé sur les hypothèses que les caractéristiques d'un groupe des textes dépendent du contexte social de leur création et utilisation, et que ces caractéristiques peuvent être décrites d'une manière qui établit un rapport entre un texte et d'autres textes semblables et des

choix et restrictions des producteurs de texte. La langue est considérée comme intégrée dans les réalités sociales; car c'est par l'utilisation récurrente des formes conventionnelles que les individus développent des relations, établissent des communautés, etc.

L'un des premiers hommes intéressés par l'étude du genre a été Swales dont l'approche est largement reconnue comme le prototype pour l'analyse des *moves* dans les études de genre et a été appliqué dans les études de plusieurs genres.

Swales voit le genre comme un groupe d'événements de communication dans lequel des formes similaires ou le contenu de la langue sont utilisés comme une représentation significative des objectifs de communication partagés, réalisés par les membres d'une même communauté discursive (Swales, 1990). Il dit aussi que ces objectifs entraînent certaines conventions compositionnelles, lesquelles sont changeantes et varient en fonction de la complexité des objectifs, de la structure textuelle, des moyennes, de la relation entre les rôles des individus interagissant, de la variété des langues, de l'activité et l'institution.

Selon Swales, la notion de genre est illustrée par la section d'introduction de l'article de recherche, vu en termes de « mouvements », ce qui représente l'objet social de l'écrivain, et « étapes » qui sont les éléments textuels optionnels.

Swales conceptualise le genre comme « la représentation des modèles reconnus des structures verbales qui constituent un échange social avec un but précis » reflétant « les nombreuses fonctions rhétoriques que ces textes doivent effectuer » (Swales, 1990: 48).

L'approche de Swales concernant l'analyse rhétorique se concentre sur des fins de communication et explique que la réalisation d'un but ou de plusieurs buts conduit à la réalisation d'un autre but par l'organisation rhétorique du texte.

Un autre chercheur important concernant l'étude de genre est Bhatia qui considère que la nature d'un genre « est principalement caractérisée par l'objectif(s) de communication qu'il est destiné à accomplir » (Bhatia, 1993: 13). Le (s) but (s) communicationnel (s) déterminera la structure du genre et si les objectifs changent d'une manière significative, le genre serait différent. Bhatia poursuit son élaboration avec l'observation que « l'objectif de communication est un critère assez fiable pour identifier et distinguer les sous-genres » (Bhatia, 1993: 14).

En ce qui concerne leurs critères par rapport au genre, Bhatia suit les lignes de Swales. Selon Bathia, les contraintes et les conventions ont une relation directe car elles appartiennent à un ensemble. Les contraintes sont les limites définies par le genre; si ces limites sont brisées, le texte n'appartiendra pas au genre défini par les contraintes. Les conventions sont les moyens spécifiques utilisés par l'écrivain afin de garder son texte dans les contraintes. Bhatia poursuit ses réflexions sur les contraintes de genre en commentant l'utilisation des conventions du genre par les membres des communautés professionnelles. (Bhatia, 1993).

1.3.1 Le travail de recherche en tant que genre

Comme plusieurs chercheurs l'ont souligné, le genre est une conventionnalisation des schémas textuels (structure schématique conventionnelle) et des modèles lexico-grammaticaux, qui dépendent du but ou de l'objectif de la communication d'une communauté particulière; par conséquent, il est utilisé dans une situation rhétorique donnée. Ainsi, toute typologie textuelle conventionnée et déterminée culturellement (par exemple, la communication scientifique) est considérée comme un genre. Cela indique que les genres académiques et de recherche sont des éléments d'identité de communautés discursives spécifiques (Martin, 2000, Swales, 1990, 2004, Parodi et al., 2010).

Selon Sánchez (2011: 24):

Dans le monde universitaire, la production de nouvelles connaissances et leur communication par le biais de l'utilisation de l'écriture sont obligatoires. Au niveau universitaire, l'étudiant consacre un grand effort à l'apprentissage de nouveaux genres, tels que le travail de recherche, le rapport, la monographie, etc., qui serviront d'évaluation pour les différentes matières.

Le travail de recherche est un processus qui, par l'application de la méthode scientifique, vise à obtenir des informations mesurables pour concevoir, vérifier, corriger ou élargir des connaissances (Arias, 2006).

Son objectif est de fournir les outils nécessaires pour que la recherche sur les événements, leur comportement dans le passé et leur prospective sur l'avenir produisent des résultats solides, utiles et novateurs pour une communauté discursive spécifique.

Conformément à ce qui précède, on peut affirmer que le travail de recherche, plutôt qu'un rapport écrit, est un processus de production de connaissances

intellectuel et réflexif très stimulant, qui repose sur ce qui a déjà été produit par les connaissances dans ce domaine et tente de l'étendre avec de nouvelles contributions, et peut être catalogué comme genre académique.

1.3.1.1 L'introduction en tant que sous-genre

L'élaboration de l'introduction d'un travail de recherche est considérée comme l'un des éléments vitaux, plus significatifs et principaux de tout travail, car son but est d'initier le lecteur au discours et à l'intérêt de l'étude.

Sa fonction est d'offrir au lecteur un panoramique de l'ensemble du texte. En d'autres termes, l'introduction est un point d'orientation qui indique au lecteur la direction et les sens du texte qu'il va lire.

Comme il a été indiqué par Carlino (2005), pour remplir cette fonction, l'introduction d'un document de recherche devrait être organisée comme ci-dessous: (1) décrire le sujet dans lequel le texte est encadré; (2) signaler le vide ou le manque de connaissances que le texte tente de résoudre; (3) proposer les questions ou objectifs qui guident le texte; et (4) anticiper l'organisation interne de l'ensemble du texte. Ces étapes doivent être reliées pour produire une introduction efficace.

Toutes ces caractéristiques qui décrivent l'introduction sont celles qui permettent à cette section du travail de recherche de différer des autres sections. En résumé, «l'article de recherche est un genre, mais l'introduction n'est en fait qu'une partie de ce genre et, après Ayers (1993), nous le désignerons comme un sous-genre» (Dudley Evans, 2001: 5).

1.3.2 Swales et l'analyse du genre

L'analyse des textes pour les placer dans des genres spécifiques est une discipline que existe depuis seulement quelques décennies. Différents auteurs comme Swales, Miller, Bhatia, et d'autres s'y sont intéressés. Une explication générale du concept d'analyse de genre est assez simple. Un texte donné est analysé de telle façon que l'analyste acquiert les renseignements nécessaires pour déterminer le genre auquel le texte appartient. Cette analyse ne se concentrera pas sur le

contenu informationnel du texte mais sur les caractéristiques et conventions utilisées qui font que le texte est similaire à d'autres textes. Des textes similaires appartiendraient au même genre ou, dans certains cas, les textes appartiendraient à différents sous-genres. Dans le cas particulier de notre étude, nous analyserons le sous-genre introduction.

1.3.2.1 L'analyse des mouvements rhétoriques

L'analyse de mouvements rhétoriques est récemment devenue un domaine de recherche important et constitue l'une des approches fondées sur le genre et l'analyse de genre utilisée pour identifier la structure des articles de recherche. Plusieurs chercheurs ont abordé l'étude des mouvements rhétoriques (Swales, 1990; Brett, 1994 ; Holmes, 1997 ; Nwogu, 1997 ; Posteguillo, 1999 ; Williams, 1999 ; Peacock, 2002 ; Samraj, 2002; Ruiying & Allison, 2003 ; Kanoksilapatham, 2005, 2007 ; Sabaj, 2011). Cependant, des recherches influentes sur cette forme particulière d'analyse ont été premièrement menées dans les études de Swales (1981, 1990).

Swales centre son étude sur l'article de recherche ayant la structure Introduction, Méthodes, Résultats et Discussions (IMRD). Précisément, il développe le modèle de création d'un espace de recherche (C.A.R.S. en anglais). Son modèle tente d'expliquer et de décrire le modèle organisationnel de l'écriture de l'introduction à des études de recherche savantes. De cette manière, le modèle CARS peut être utilisé pour :

- 1) commencer le processus d'écriture (le démarrage est souvent la tâche la plus difficile);
- 2) comprendre la façon dont une introduction présente la place du reste du document ;
- 3) évaluer la façon dont l'introduction s'inscrit dans la plus grande portée de l'étude.

Le modèle suppose que les écrivains suivent un modèle organisationnel général en réponse à deux types de défis concernant l'établissement d'une présence dans un domaine de recherche particulier:

- 1) la concurrence pour créer un espace rhétorique

2) la concurrence pour attirer les lecteurs dans cet espace.

Le modèle propose trois actions (Swales les appelle "mouvements") accompagnées d'étapes spécifiques, qui reflètent le développement d'une introduction efficace pour un document de recherche. Ces «mouvements» et les étapes peuvent être utilisés comme modèle pour écrire l'introduction de tous les documents de recherche en sciences sociales.

Son approche est largement reconnue comme le prototype pour l'analyse des mouvements dans les études de genre et a été appliquée dans plusieurs études de genre. Cependant, tout en distinguant «mouvement rhétorique» et «étape», en définissant «étape» comme une unité plus petite que «mouvement rhétorique», il ne définit pas clairement comment distinguer un «mouvement».

Un «mouvement» désigne un segment de discours qui effectue une fonction communicative particulière (Swales, 2004). Il représente des unités sémantiques et fonctionnelles de textes ayant des objectifs spécifiques (Biber, Connor, & Upton, 2007). L'analyse axée sur les mouvements met l'accent sur les structures schématiques hiérarchiques des textes (Nwogu, 1997). Les articles de recherche sont un genre qui a fait l'objet d'une enquête approfondie en utilisant l'approche basée sur les mouvements.

Un aspect fondamental concernant l'analyse de mouvements rhétoriques, est la fréquence avec laquelle ces segments discursifs coexistent dans le genre en question. Selon la classification de Kanoksilapatham (2005) de la fréquence d'apparition des mouvements rhétoriques dans les introductions des travaux de recherche, celles-ci peuvent être cataloguées comme obligatoires (100% des occurrences), conventionnelles (entre 60% et 99% des occurrences), et optionnelles (occurrence inférieure à 60%).

1.3.2.2 Le modèle CARS de Swales

Swales (1990) a développé les mouvements et étapes suivantes pour l'étude de l'introduction des articles de recherche :

Mouvement 1: Établir un territoire (la situation)

Ceci est généralement accompli de deux façons: en démontrant qu'un domaine général de recherche est important, critique, intéressant, problématique, pertinent ou autrement digne d'investigation ; mais aussi en introduisant et en examinant les

sources clés de recherches antérieures dans ce domaine pour montrer où se trouvent les lacunes ou lorsque des recherches antérieures ont été insuffisantes pour résoudre le problème de recherche.

Les étapes pour y parvenir seraient:

- Étape 1 – Remarquer l'importance (décrire le problème de la recherche et fournir des preuves pour soutenir l'importance qu'a l'étude du sujet)
- Étape 2 - Faire des généralisations à propos du sujet, (fournir des déclarations sur l'état actuel des connaissances, le consensus, la pratique ou la description des phénomènes)
- Étape 3 – Examiner des éléments de recherches antérieures (synthétiser des recherches antérieures qui soutiennent encore la nécessité d'étudier le problème de recherche; ce n'est pas une critique de la littérature, mais plus un reflet des études clés qui ont abordé le sujet, mais peut-être pas complètement)

Mouvement 2: Établir une niche (le problème)

Cette action fait référence à un argument clair et convaincant de la valeur et l'importance de la recherche. Cela peut se faire en indiquant un écart particulier dans les recherches antérieures, en contestant une hypothèse largement acceptée, en soulevant une question, une hypothèse ou un besoin, ou en étendant les connaissances antérieures d'une certaine manière. Les étapes pour y parvenir seraient:

- Étape 1a - Contre-réclamation, (introduire un point de vue ou une perspective opposée ou identifier un écart dans les recherches antérieures laquelle a pu raccourcir ou compromettre l'argument dominant)
- Étape 1b - Indiquer un écart, (développer le problème de recherche autour d'un espace ou d'un domaine de travail peu étudié)
- Étape 1c – Poser de questions, (similaire à l'indication d'un écart, cela implique de présenter des questions clés sur les conséquences des lacunes dans les recherches antérieures qui seront abordées par l'étude à réaliser)
- Étape 1d - Continuer une tradition (étendre les recherches antérieures pour développer ou clarifier un problème de recherche. Ceci est souvent signalé avec la terminologie de connexion logique, par exemple *donc*, *par conséquent*, *ainsi* ou un

langage qui indique un besoin. Par exemple, on pourrait affirmer: «Par conséquent, ces facteurs doivent être examinés plus en détail ...» ou «Les éléments de preuve suggèrent une corrélation intéressante, il est donc souhaitable d'enquêter sur différents répondants ...»)

Mouvement 3: Occuper la niche (la solution)

Le «mouvement» final est d'annoncer les moyens par lesquels l'étude réalisée apportera de nouvelles connaissances ou une nouvelle compréhension contrairement aux recherches antérieures sur le sujet. C'est aussi là que la structure organisationnelle restante du document est décrite.

Les étapes pour y parvenir seraient:

- Étape 1a - Définir des buts, (indiquer d'une façon claire les objectifs de l'étude)
- Étape 1b - Annoncer la recherche actuelle (décrire le but de l'étude réalisée en fonction de ce que la recherche va faire ou accomplir)
- Étape 2 – Annoncer les résultats principaux (présenter un bref résumé des principaux résultats écrits, tels que «Les résultats indiquent un besoin de ...» ou «La recherche suggère quatre approches ... »)
- Étape 3 - Indiquer la structure de l'article (indiquer comment le reste du document est organisé)

Nous allons justement utiliser le modèle CARS pour analyser les mouvements rhétoriques des travaux de recherche, même si Swales a créé ce modèle en étudiant les introductions des articles de recherche. Les deux types de texte ne diffèrent que par la structure, les objectifs et le public cible (voir les sous-sections 1.1.1.2 et 1.1.1.3).

1.4 Linguistique de corpus

L'accès actuel à de vastes ensembles de textes sous forme électronique a été une condition décisive pour le développement d'un courant linguistique récent : la linguistique à base de corpus. La linguistique de corpus fournit une base empirique pour la mise au point de matériels pédagogiques et méthodologiques de divers types, ainsi que pour la construction de grammaires, de dictionnaires, de banques des expressions, et autres, non seulement des discours généraux et spécialisés, mais en plus, des discours oraux et écrits. Elle marque les bases ou les principes

pour étudier n'importe quel domaine linguistique; elle se caractérise par le fait de nourrir la recherche sur la langue; et est également utilisé dans l'analyse de genre (Parodi, 2008).

Selon Nesselhauf (2005), la linguistique de corpus est une méthode de réalisation d'analyses linguistiques. Etant donné qu'elle très utile pour l'étude de nombreux types de questions linguistiques et qu'elle est susceptible de produire des idées nouvelles très intéressantes, fondamentales et souvent surprenantes sur le langage, elle est devenue l'une des méthodes les plus répandues des recherches linguistiques ces dernières années.

La linguistique basée sur le corpus se concentre sur la description du contenu et des relations qui relie la langue naturelle et l'étude du discours comme médiateur de la connaissance, conditionné par des implications pragmatiques fondamentales (Bybee & Hopper, 1997).

Le grand avantage de la méthode linguistique de corpus est que les chercheurs linguistiques ne doivent pas s'appuyer sur leur propre intuition ou sur d'autres locuteurs natifs, ni même sur des exemples inventés. Ils peuvent plutôt s'appuyer sur une grande quantité de données linguistiques authentiques et naturelles produites par une variété d'orateurs ou d'écrivains afin de confirmer ou de réfuter leurs propres hypothèses sur des caractéristiques linguistiques spécifiques à partir d'une base empirique.

1.4.1 Le corpus

Corpus, pl. des corpus, est emprunté au latin *corpus*, *-oris*, et veut dire masse, collection, ensemble, etc. Le mot est polysémique et signifie soit un ensemble de textes individuels aussi bien que des textes individuels ou une collection ou un ensemble de matériaux ou d'informations utilisés pour une étude. En général, le terme corpus indique toute collection de matériel linguistique, peu importe la manière dont le matériel est traité ou stocké, qui sert de source d'information non seulement pour les linguistes, mais aussi pour d'autres chercheurs d'autres domaines intéressés par la connaissance du langage et les modèles de la pensée humaine (Fillmore, 1992).

D'après Sinclair et Ball (1996 : 12), "un corpus est une collection de textes de langue sous forme électronique, choisie selon des critères externes pour représenter, dans la mesure du possible, une langue ou une variété de langue comme source de données pour la recherche linguistique". Il est important de saisir la différence du terme corpus dans son utilisation habituelle et dans ses référents concernant la linguistique de corpus. Comme on le voit dans la définition précédente, le terme corpus tel qu'il sera considéré dans ce document, ainsi que dans la linguistique de corpus, représente une grande collection de textes naturels, écrits et parlés, et ils peuvent être trouvés sous forme numérique stockés dans de vastes bases de données. Il existe des règles précises pour développer un corpus détaillant les types et les quantités de textes (Meyer, 2002) et ils sont conçus pour assurer qu'un corpus soit représentatif et équilibré (Dobrić, 2009).

Un corpus est à distinguer d'une simple collection bibliographique, il est défini usuellement par « un ensemble de documents servant de base à la description dans un domaine ou à l'étude d'un phénomène » (Mayer et Deslauriers, 2000 : 37).

1.4.2 Type de corpus

Sous le terme de corpus nous pouvons distinguer deux catégories abordées par des auteurs comme Biber, le *corpus de référence* qui est capable de fournir une vue complète d'une langue dans toutes ses variétés et ses caractéristiques linguistiques. Il peut alors servir de base pour la production de grammaires, thesaurus, dictionnaires (Biber, 1998)

Biber affirme aussi que le *corpus spécialisé* se caractérise par ses particularités lexicales et grammaticales. Il représente un langage de spécialité (vs. langage général) aussi appelé *sous langage* pour un domaine donné. Les corpus spécialisés ne contribuent donc pas à la description de la langue ordinaire.

Selon Harris (1951), un *sous langage* se définit comme un ensemble d'énoncés liés par un sujet limité, utilisé pour une fonction particulière et engendré par une grammaire et un vocabulaire spécifique.

D'autre part, par rapport à la typologie du corpus, la classification fournie par EAGLES (Sinclair et Ball, 1996), est généralement considérée comme la plus complète car six types principaux sont reproduits:

- a) corpus de référence,
- b) moniteur de corpus (corpus suivi ou veille),
- c) corpus oral,
- d) corpus de fragments textuels,
- e) corpus spécial et spécialisé,
- f) corpus bilingue ou multilingue.

Les critères qui correspondent le mieux à nos intérêts sont ceux de Bowker et Pearson qui se concentrent essentiellement sur les textes spécialisés. Celles auteures ont établi une série de dichotomies étant donné qu'elles considèrent qu'il existe autant de types de corpus qu'il existe d'études (Bowker et Pearson, 2002). Parmi les types les plus généraux, elles citent : *les corpus de références*, corpus très larges, représentatifs d'un langage dans son ensemble, et *les corpus d'études*, corpus créés afin d'observer un aspect particulier du langage; *les corpus écrits* qui sont les corpus composés de textes écrits et *les corpus oraux*, corpus composé de transcription de matériel oral; *les corpus synchrones* : contenant des documents datant d'une période restreinte, permettant d'étudier le langage sur une période précise et *les corpus diachroniques*, rassemblant des écrits de périodes différentes, afin d'observer l'évolution du langage (Kyto, 1994); *les corpus ouverts* (« monitor corpus»): corpus auxquels des textes sont constamment ajoutés et *les corpus fermés*; et *les corpus spécialisés* : plus petits qu'un corpus de référence générale, sont basés sur un but spécifique et ne sont pas utilisés pour décrire le langage standard.

1.4.3 Critères pour la construction de corpus

À l'inverse de grands corpus (corpus de référence), qui visent à être représentatifs de la langue en général, les corpus spécialisés, se caractérisent par leur plus petite taille et la précision de leur but. Il faut déterminer, par conséquent, quels sont les critères qui permettront de définir l'objet de l'étude, ceux-ci pouvant être très divers. Cependant, malgré différents critères (Meyer et Mackintosh, 1996 ;

Browker, 1996 ; Pearson 1998), les discussions autour des critères pour le développement d'un corpus ont essentiellement tourné autour de trois piliers: la taille, l'équilibre et la représentativité (Auría, 2004).

Il n'y a malheureusement pas de consensus sur la taille souhaitable pour un corpus, ni en langue générale, ni en langue spécialisée. Le seul point certain semble être qu'un corpus général doit être beaucoup plus grand qu'un corpus spécialisé. Pearson (1998) indique que, la plupart du temps, la taille d'un corpus spécialisé est déterminée selon l'intuition des chercheurs, le caractère des textes inclus et leur domaine.

La représentativité d'un corpus est sa capacité à fonctionner comme une base fiable pour des généralisations sur une langue particulière (générale ou spécialisée). Ceci implique à la fois une taille et une variété de textes adéquates.

Comme le souligne Sinclair (1991), l'équilibre et la variété du contenu d'un corpus sont importants pour neutraliser les particularités des auteurs, sauf si le chercheur désire les analyser. Selon les critères d'Atkins, Clear & Ostler (1991: 14), l'équilibre du corpus « offre un modèle maniable à petite échelle du matériau linguistique que les constructeurs de corpus souhaitent étudier ».

1.5 Les banques des expressions linguistiques

La banque des expressions linguistiques (BEL) s'appuie largement sur une approche d'analyse de textes académiques initiés par Swales dans les années 1980. Cet auteur a non seulement identifié les mouvements couramment utilisés dans les introductions d'articles, mais il souhaitait montrer le type de langage utilisé pour atteindre l'objectif de communication de chaque mouvement et comment cette unité d'analyse rhétorique est utilisée comme l'une des principales catégories des BEL. Une grande partie de cette langue était de nature phraséologique. La ressource s'appuie également sur des connaissances psycholinguistiques, sur la manière dont le langage est appris et produit. Il est maintenant admis qu'une grande partie du langage utilisé est de nature phraséologique; acquises, stockées et récupérées sous forme de constructions pré-formulées (Bolinger, 1976, Pawley et Syder, 1983). Ces connaissances ont commencé à être attribuées de manière empirique à la technologie informatique dans l'identification de modèles

phraséologiques récurrents dans de très grands corpus à l'aide de logiciels spécialisés (par exemple, Sinclair, 1991).

La BEL reconnaît qu'il existe une dimension phraséologique importante dans le langage académique et tente d'en rendre des exemples explicites. C'est une source générale pour les auteurs universitaires. Son objectif est de fournir des exemples de certains aspects phraséologiques de base de l'écriture organisée, en tenant compte des sections principales des travaux de recherche ou des mémoires. Pour la plupart, les composants de la BEL présentent un contenu de nature neutre et générique; de sorte que leur utilisation ne plagie pas le travail des autres auteurs (Morley, 2014).

1.5.1 Types des banques des expressions linguistiques

Même s'il y a un manque considérable de bibliographie sur ce sujet en raison de sa nouveauté, nous avons décidé d'adopter les classifications des banques des expressions définies par Amed Leiva Mederos dans sa thèse de doctorat. En ce sens, Leiva-Mederos et al. (2011) a suggéré trois types de banques des expressions linguistiques qui les englobent en termes généraux. Ces trois types sont:

1. Banques des expressions basées sur la rhétorique d'un texte à finalité scientifique.
2. Banques des expressions pour la construction de textes artistiques ou littéraires.
3. Banques des expressions regroupant des expressions formelles ou dialectiques.

Cette recherche est centrée sur la création d'une banque des expressions basées sur la rhétorique d'un texte à finalité scientifique, c'est-à-dire, notre banque appartiendra au premier classement déjà établi.

1.5.2 Importance des banques des expressions linguistiques

John Morley est le plus grand spécialiste et même créateur de l'une des banques des expressions linguistiques les plus reconnues. Selon lui, la banque des expressions linguistiques est un outil très efficace non seulement pour les

étudiants mais aussi pour les enseignants. Comme elle se construit à travers l'analyse des mouvements rhétoriques qui composent les différentes parties des textes académiques et rassemble les expressions qui composent ces mouvements, la banque des expressions linguistiques est très pratique car les sections et les détails qu'elle fournit peuvent aider ses utilisateurs en train d'analyser le contenu et l'organisation de leur texte, tandis que les exemples de phrases peuvent être facilement intégrés au produit, le cas échéant, pendant le processus de rédaction (Morley, 2014). De plus Morley affirme que la banque des expressions linguistiques encourage et régleme la rédaction de textes en fonction des techniques qui y figurent. Dans un autre sens, de nombreuses recherches ont montré qu'il existait une condition appelée blocage mental de l'écrivain et que la banque des expressions linguistiques pouvait fonctionner comme le stimulus nécessaire à ce blocage (Morley, 2014). Il exprime aussi que cette banque travaille également comme un élément pour provoquer l'esprit des écrivains les plus expérimentés, car ils connaissent souvent les phrases mais ne s'en souviennent pas.

D'autre part, Morley soutient que la banque des expressions linguistiques aide à comprendre la critique littéraire d'un point de vue différent. De même, selon lui, cet instrument est précieux pour aider ses consommateurs à organiser et à structurer leurs textes écrits, essentiellement les sections les plus complexes des textes académiques tels que l'Introduction.

1.5.3 Méthodes pour la construction des banques des expressions linguistiques

L'analyse syntaxique, qui permet d'obtenir la structure d'une phrase selon une description formelle du langage sous forme de grammaire, est une étape d'une grande pertinence tant pour la compilation de programmes que pour le traitement du langage naturel. À partir des travaux de Chomsky qui ont formalisé la notion de grammaire indépendante du contexte dans les années 50, différents algorithmes ont été développés pour mener à bien cette tâche. Beaucoup de ces algorithmes diffèrent grandement par la manière dont les utilisateurs procèdent à l'analyse, en utilisant différentes approches pour obtenir des résultats identiques ou similaires, et chacun d'entre eux est plus adapté à un type de situation spécifique.

Les schémas d'analyse syntaxique fournissent une méthode uniforme pour décrire, analyser et comparer différents algorithmes d'analyse. Pour ce faire, ils s'appuient sur le fait de considérer cette analyse comme un processus de génération de résultats intermédiaires appelé items. La phrase qui est analysée génère un ensemble initial d'items. L'analyse consiste en l'application d'une série de règles permettant de générer de nouveaux items contenant des informations sur la structure de la phrase, jusqu'à atteindre celle contenant explicitement l'arborescence syntaxique, ou garantit son existence et permet de l'obtenir facilement (Meneses et al. 2014). Ce sont les phrases qui formeront la banque des expressions linguistiques. Afin de mener l'analyse sémantique de ces phrases et de les compiler, il y a différentes méthodes divisées en deux classes fondamentales correspondant au déroulement du processus de compilation.

1.5.3.1 Manuel (compilation manuelle)

Comme son nom l'indique, dans le cas d'une compilation manuelle, c'est l'homme lui-même qui compile les données ou, dans notre cas particulier, les expressions linguistiques. C'est-à-dire qu'il n'y a pas de médiation d'aucun type d'instrument électronique.

1.5.3.2 Numériques

Les algorithmes de regroupement numérique sont des techniques qui ont contribué à la mise au point de systèmes d'extraction de textes, dans la mesure où ils permettent la classification et le regroupement de documents. Ce sont des opérations mathématiques qui reflètent les possibilités exactes d'effectuer une extraction et une classification avec fiabilité.

La variété des méthodes de regroupement de texte a été telle qu'il est très difficile pour l'auteur de trouver une classification qui puisse décrire toutes les procédures méthodologiques existantes pour le développement de groupes.

Ils sont classés en algorithmes hiérarchiques, ceux qui forment des partitions et se concentrent sur de petits modèles ou des prototypes simulés. Ces modèles-ci commencent par une division ou partition d'une entrée qui peut être aléatoire ou non, et aboutissent à un raffinement. Les algorithmes hiérarchiques construisent une structure hiérarchique d'objets.

D'autre part, les algorithmes de division (top-down) partent de l'hypothèse de la concordance d'un groupe général, c'est pourquoi tous les objets y sont inclus. Cet algorithme segmente les groupes jusqu'à ce qu'ils contiennent un seul objet.

Un type de *cluster* est le *clustering* basé sur la densité, qui regroupe les objets d'un voisinage donné en un ensemble de données en utilisant les propriétés de la densité. Ce type d'algorithme comprend DBSCAN et MajorClusf, Réurrence de Lance & Williams et Expectation-Maximization (EM). D'autres classifications déclarent des algorithmes itératifs similaires à ceux qui décrivent des itinéraires ou des distances. Celles-ci sont divisées en: Base exemple et Commutation de base. Ils appartiennent à ce type d'organisation: K-means (McQueen, 1967), Kmedoid, Kerninghan-Lin, DK-Means, KNN, etc.

Du point de vue de la possibilité de la recherche métacontrôlée, les algorithmes sont divisés en méthodes descendantes et méthodes compétitives. Un exemple de ces méthodes est *Simulated Annealing* et les algorithmes génétiques. Pour des domaines spécifiques, les algorithmes de factorisation des concepts et de classification en arborescence de suffixes apparaissent.

Ces procédures mathématiques ont conduit au développement d'algorithmes flous, parmi lesquels: FUZZY SkWIC, SKWIC et MLRul permettant d'obtenir et de regrouper des données dont la structure n'est pas définie. Il est important de souligner que, dans ce travail, seuls les algorithmes dont la pertinence influe le développement de notre étude sont décrits (Leiva-Mederos, 2011).

1.6 Ontologie

Le terme "ontologie" (utilisé en philosophie pour parler d'une théorie de l'existence) a été adopté par la communauté scientifique de l'intelligence artificielle pour définir une catégorisation et les relations entre ses termes (Berners-Lee et al. 2001).

La catégorisation définit les classes et les relations entre elles. Les éléments concrets sont des "instances" de ces classes, également appelées termes de l'ontologie. Comme dans toute classification, la relation fondamentale entre les termes est celle de l'héritage, où une classe A (sous-classe) est un type de classe B (superclasse), de sorte qu'elle a toutes ses caractéristiques.

Les relations sémantiques définies entre les classes d'une ontologie peuvent être considérées comme des prédicats qui, associés à un ensemble de règles d'inférence, permettent de construire des logiciels (agents) effectuant un raisonnement automatique. Lorsque qu'une relation de ce type est définie, il peut être spécifié, parmi de nombreuses autres propriétés, l'ensemble domaine de la relation et l'ensemble image, c'est-à-dire les classes d'où la relation part, et les classes avec lesquelles elle est associée en fonction du sens de celle-ci.

Pour Senso (2011), plusieurs ontologies peuvent être regroupées selon leurs types dans quatre catégories:

Niveau élevé: conçu pour décrire tous les concepts généraux tels que l'espace, le temps, la matière, l'objet, le fait, l'action, etc.

De domaines: décrivent le vocabulaire lié à un domaine générique.

De tâches: décrivent les activités pouvant être utiles aux organisations.

D'applications: décrivent les concepts selon un domaine ou des tâches spécifiques.

Les ontologies et toutes les structures ou outils qui y sont utilisés offrent de nouvelles façons de représenter et de partager les connaissances grâce à l'utilisation d'un langage commun conçu antérieurement pour son application (Leiva-Mederos, 2011). L'un de ces outils est le RAOP (Rhetorical Annotation Ontology Project) qui vise à apporter un outil de numérisation des formes de discours en utilisant les technologies associées aux données liées au *Linked data* et au Web sémantique. Le RAOP a pour but de soutenir l'analyse littéraire numérique rhétorique et syntaxique ainsi qu'une approche théorique savante de ce domaine. Il cherche à créer le terrain pour la représentation de structures de texte complexes en tant que figure rhétorique impliquant de multiples structures d'annotation coexistantes, hiérarchiques et non hiérarchiques (Marchesini et Mancinelli, 2017).

1.6.3 Knowledge graphs

L'ontologie est la base fondamentale des *knowledge graphs*. Selon Paulheim (2017 :6), ces *knowledge graphs* «décrivent principalement des entités du monde réel et leurs interrelations, organisées en graphe, définissent les classes et

relations possibles d'entités dans un schéma, permettent potentiellement l'interrelation d'entités arbitraires entre elles et couvrent divers domaines d'actualité».

De plus, Färber et al. (2018) définissent un *knowledge graph* «comme un graphe RDF. Un graphe RDF consiste en un ensemble de triplets RDF où chaque triple RDF (s, p, o) est un ensemble ordonné des termes RDF suivants: un sujet $s \in U \cup B$, un prédicat $p \in U$ et un objet $U \cup B \cup L$. Un terme RDF est soit URI $u \in U$, un nœud vide $b \in B$, ou un littéral $l \in L$.»

Pour nous, la définition qui convient le mieux à cette recherche est celle proposée par Lisa Ehrlinger et Wolfram Wöß (2016: 3) dans l'article *Towards a definition of knowledge graphs* qui écrit comme suit: « Un *knowledge graph* acquiert et intègre des informations dans une ontologie et s'applique à un raisonnement connaissance», parce que ce concept décrit précisément les processus qu'on va entreprendre dans cette recherche.

L'intégration de graphes est l'une des procédures les plus efficaces pour résoudre le besoin de construire des systèmes linguistiques supportés automatiquement par RDF en tant que banques des expressions (Leiva-Mederos, 2018).

1.7 Conclusions partielles

L'objectif de ce premier chapitre a consisté à démontrer que tous les concepts présentés ont contribué à soutenir les éléments théoriques de cette recherche. Les concepts d'écriture académique, genre, langues de spécialité, linguistique de corpus, corpus, banques des expressions linguistiques, ontologie et *Knowledge graphs* ont été décrits et liés à cette recherche. On a fait allusion aux différents types de textes académiques, aux classifications des langues de spécialité, au travail de recherche et à l'introduction en tant que genre et subgenre respectivement, aux types de corpus, et, aux méthodes pour la construction des banques des expressions linguistiques. En outre, on a expliqué l'analyse des mouvements rhétoriques et le modèle CARS de Swales pour faire cette analyse.



Deuxième

Chapitre

CHAPITRE II: RÉFÉRENCES MÉTHODOLOGIQUES DE LA RECHERCHE

Dans ce chapitre, les référents méthodologiques de la recherche sont explorés. On aborde le contexte, étapes, type et portée de notre recherche. De même, on fait référence aux méthodes théoriques, empiriques et mathématiques qu'ont été utilisées. On présente aussi l'opérationnalisation des variables, la population, et l'échantillon, les outils, la révision bibliographique et les limites de notre étude.

2.1 Contexte de la recherche

La recherche actuelle a été réalisée à la Faculté de Sciences Humaines, à l'Université Centrale «Marta Abreu» de Las Villas. La banque des expressions linguistiques, qui en est le résultat, est destinée aux étudiants de la Licence ès Langue Anglaise, option Langue Française de cette université qui font leur travail de recherche en français. Ils ne disposent pas d'une bibliographie suffisante et de documents de référence qui leur permettent de clarifier leurs doutes lors de la rédaction de l'introduction de leur travail de recherche. En raison des caractéristiques de la banque des expressions, la recherche est développée dans les domaines de la linguistique et de la didactique, car les étudiantes choisissent de faire leur travail de recherche dans un de ces deux domaines.

2.2 Étapes de la recherche

La recherche a été menée selon les étapes suivantes:

- **Élaboration du cadre théorique de la recherche:** consiste à regrouper toutes les informations bibliographiques sur le sujet de la recherche, en reconnaissant et en distinguant ses antécédents et les référents théoriques et méthodologiques associés au sujet, de la linguistique informatique et de la linguistique de corpus.
- **Application des instruments:** à ce stade, les instruments sont appliqués pour obtenir des informations sur le domaine de recherche: étudiants et enseignants.

- **Élaboration de la banque des expressions linguistiques:** processus par lequel sont créées les étapes permettant de générer la banque d'expressions.

2.3 Type de recherche

La recherche est non expérimentale, descriptive. Elle présente une approche mixte avec une prédominance quantitative, parce que l'étude est centrée sur une procédure semi-automatique de construction d'une banque des expressions dans le cadre du Web sémantique. Cette étude a été réalisée à l'aide de méthodes et de techniques permettant de collecter des informations conjointement ou indépendamment sur les concepts ou les variables auxquels il est fait référence. L'approche mixte de la recherche est illustrée par l'utilisation de méthodes mathématiques et statistiques démontrant la fiabilité de l'enquête appliquée, des entretiens et des techniques semi-automatiques de construction de banques des expressions, de l'accord entre les travailleurs, et donc, des articles et des dimensions les plus importants. Il est séquentiel et probant puisque chaque étape précède l'autre et aucune étape n'est ignorée.

2.4 Portée de la recherche: descriptive

Comme l'a expliqué Sampieri et al. (2010), le chercheur doit souvent décrire des phénomènes, des situations, des contextes et des événements, détailler leur nature et leur manifestation. Dans le cas des études descriptives, celles-ci visent à spécifier les propriétés, caractéristiques et profils de personnes, groupes, communautés, processus, objets ou tout autre phénomène analysé. En d'autres termes, l'objectif est uniquement de mesurer ou de collecter des informations indépendamment ou conjointement sur les concepts ou les variables auxquels elles se rapportent. Comme l'explique Sampieri, les études descriptives sont utiles pour montrer avec un degré de précision les angles ou les dimensions d'un phénomène, d'un événement, d'une communauté, d'un contexte ou d'une situation. Dans ce type d'étude, le chercheur doit avoir la capacité de définir, ou, au moins, de visualiser, ce qui sera mesuré (concepts, variables, composants, etc.) et qui collectera les données (personnes, groupes, communautés, etc.). Le degré de profondeur de la description peut varier dans une moindre ou majeure mesure, cependant, il doit être basé sur la mesure d'un ou de plusieurs attributs du phénomène considéré (Hernández Sampieri, et al., 2010).

Dans notre cas, la recherche est descriptive avec des limitations parce que la variable de recherche est décrite, puisqu'il y a suffisamment d'antécédents du sujet. Mais on n'aura pas la possibilité de constater si notre hypothèse est accomplie ou non. Des concepts, caractéristiques ainsi que le développement associé au web sémantique, sont inclus.

2.5 Opérationnalisation des variables

L'étape suivante de la recherche est l'opérationnalisation des variables, car comme Gil-Leiva (2008) l'a exprimé, sans définition des variables il n'y a pas de recherche. Celles-ci doivent être définies de manière conceptuelle et opérationnelle (tableau 1).

Tableau 1. Opérationnalisation des variables

Variables	Définition conceptuelle	Définition opérationnelle	Dimension	Indicateurs	Articles
Banque des expressions	La banque des expressions donne des exemples explicites de la dimension phraséologique importante qui existe dans le langage académique (Morley, 2014).	Enquête pour vérifier les problèmes dans la rédaction en français de la section Introduction du travail de recherche des étudiants en raison de l'absence d'outils lexicaux pour écrire ce type de texte	Expressions	Types des expressions composant la banque selon l'emploi	Marqueurs discursifs : -Structurateurs de l'information : .commentateurs .ordonnateurs .marqueurs de digression -Connecteurs : .d'addition .consécutifs .contre-

		académique. Interview de professeurs de la faculté pour évaluer les difficultés qui apparaissent dans le processus éducatif en raison de l'absence d'outils spécialisés pour aider aux étudiants dans la rédaction de leur travail de recherche en français			argumentatifs -Reformulateurs: .explicatifs .rectificatifs .de distanciation .récapitulatifs -Opérateurs argumentatifs : .de renforcement argumentatif .de concrétion -Marqueurs conversationnels: .de modalité épistémique .de modalité déontique .concentrateurs de l'altérité .métadiscursifs conversationnels
		Interview de professeurs de la faculté pour évaluer leur emploi des	Méthodes de Construction	Emploi des méthodes de construction manuels, automatiques	Méthodes de construction : Manuels Automatiques

		méthodes de construction manuels, automatiques ou les deux pour l'élaboration de la banque des expressions linguistiques.		, ou les deux, pour l'élaboration de la banque des expressions	*Algorithme de clustering * Techniques de <i>stemming</i> * Distances géométriques
<i>Knowledge graphs</i>	Un graphe de connaissances acquiert et intègre des informations dans une ontologie et s'applique à un raisonnement connaissance (Ehrlinger & Wöß, 2016).	Interview de professeurs de la faculté pour évaluer leur emploi des recours linguistiques pour l'organisation de la banque des expressions linguistiques.	Recours linguistiques	Emploi des recours linguistiques pour l'organisation de la banque des expressions	1. Ontologies -RAOP -Lemon -SKOS 2. Lexiques - Wordnet -Eurowordnet
		Interview de professeurs de la faculté pour évaluer leur emploi des outils d'extraction des phrases et les outils de	Outils	Emploi des outils d'extraction des phrases Outils de visualisation	Logiciel pour la construction des banques des expressions : *Protégé *GraphDB *AntConc

		visualisation.			*Graphviz *OpenRefine
		Interview de professeurs de la faculté pour évaluer leur emploi des méthodes d'intégration de <i>graphs</i> et des mécanismes d'incrustation de <i>graphs</i> .	<i>Knowledge graphs</i>	Emploi des méthodes d'intégration de <i>graphs</i> Emploi des mécanismes d'incrustation de <i>graphs</i>	- Big data - Incrustation de <i>graphs</i> - Sémantique navigationnelle - Similitude : Cosinus, Manhattan, Distance Euclidienne
Rédaction des textes académiques	Celle qui se produit dans les domaines universitaire et scientifique (Ramírez, 2007).		Structures textuelles	Emploi des différents types des textes académiques Emploi des mouvements rhétoriques	Types des textes académiques <ul style="list-style-type: none"> • Article de recherche • Mémoires • Travail de recherche
	Un «mouvement» désigne un segment de discours qui effectue une fonction communicative				Mouvements rhétoriques : Mouvement 1: Établir un territoire Étape 1: Remarquer

	<p>particulière. Il représente des unités sémantiques et fonctionnelles de textes ayant des objectifs spécifiques (Swales, 1990).</p>			<p>l'importance</p> <p>Étape 2: Faire des généralisations à propos du sujet</p> <p>Étape 3: Examiner des éléments de recherches antérieures</p> <p>Mouvement 2 :</p> <p>Établir une niche</p> <p>Étape 1a: Contre-réclamation, ou</p> <p>Étape 1b: Indiquer un écart, ou</p> <p>Étape 1c: Poser des questions, ou</p> <p>Étape 1d: Continuer une tradition</p> <p>Mouvement 3: Occuper la niche</p> <p>Étape 1a: Définir des buts, et/ou</p> <p>Étape 1b: Annoncer la</p>
--	---	--	--	---

					recherche actuelle Étape 2:Annoncer les résultats principaux Étape 3: Indiquer la structure de l'article
--	--	--	--	--	--

L'objet de cette recherche est les banques des expressions linguistiques à travers du *knowledge graph*. **Le domaine** est les banques des expressions linguistiques.

2.6 Méthodes de la recherche

Pour la confection de ce travail, différentes méthodes de recherche ont été employées.

- Méthodes théoriques:

Historique-logique: méthode employée pour étudier le développement logique et historique de principaux critères sur le thème. Dans le sujet qui nous concerne, on fait référence au développement des critères tels que l'écriture académique, le genre et les mouvements rhétoriques, la linguistique de corpus, les banques des expressions linguistiques, l'ontologie, et les *knowledge graphs*.

Analyse-synthèse: méthode employée pour l'analyse de la littérature afin de contraster, analyser, évaluer et généraliser les principales tendances par rapport à la notion de genre et mouvements rhétoriques, la linguistique de corpus, les banques des expressions linguistiques, l'ontologie, et les *knowledge graphs*.

Induction-déduction: méthode employée pour analyser les aspects dès la perspective la plus générale jusqu'à un élément particulier qui permet de tirer des conclusions ou de déterminer parmi de nombreux éléments les plus importants. Dans la présente étude, à travers du processus d'induction analytique, des

déductions, analyses et abstractions, l'analyse du corpus et l'élaboration de la banque des expressions ont été faites.

- Méthodes empiriques:

Analyse des documents traditionnels: Afin de réaliser l'analyse documentaire, le modèle utilisé dans les travaux de Ruiz Pérez et Yunta (1994) a été utilisé, dans lequel le problème principal de l'analyse documentaire a été défini, en précisant ses phases. L'auteure a examiné la documentation pertinente sur le sujet et a procédé à la réalisation des actions suivantes avec les informations pertinentes:

- ❖ **Déterminer les objectifs de l'étude documentaire:** lors de cette phase, les sources d'information de l'univers documentaire sont déterminées et les actions réalisées permettent de discriminer les sources dans l'univers entourant le thème.
- ❖ **Établir un échantillon des documents qui seront utilisés:** Le deuxième aspect de l'analyse documentaire classique est la recherche dans différentes sources en utilisant différentes stratégies de recherche permettant de déterminer les aspects liés aux langues contrôlées, leurs méthodologies de construction automatique et les techniques d'origine informatique et linguistique qui affectent ces processus.
- ❖ **Déterminer les unités d'analyse dans lesquelles le contenu est divisé pour étudier les documents:** Cette phase a consisté à déterminer les éléments d'étude à partir du champ et de l'objet de l'enquête qui sont ceux qui génèrent ultérieurement les catégories d'analyse dans la recherche. Les unités d'analyse retenues étaient les suivantes: vocabulaire, vocabulaires contrôlés, modèles informatiques de construction de banques des expressions, modèles basés sur RDF, méthodologies de génération automatique et semi-automatique de banques des expressions.
- ❖ **Réaliser l'étude documentaire en enregistrant les informations:** Les informations ont été enregistrées en respectant les exigences des processus de structuration des connaissances, en tenant compte des questions multidisciplinaires de la recherche.

- ❖ **Évaluation des informations obtenues:** Des évaluations critiques ont été réalisées sur la base des aspects théoriques et méthodologiques sélectionnés, en présentant des arguments, des critiques, des suggestions et en tirant des conclusions sur les services d'information et les ontologies.

Tous les documents ont été enregistrés dans le gestionnaire bibliographique EndNote et décrits par le style bibliographique d'APA (6^e édition).

Questionnaire: Dans tout type de recherche, l'utilisation de cette méthode permet d'obtenir des informations au sein de la communauté des utilisateurs qui font partie de l'étude à travers une série de questions orales et écrites. Dans notre étude, cette méthode a permis de poser une série de questions aux étudiants de la Licence ès Langue Anglaise, option Langue Française, pour obtenir des informations sur la nécessité d'utiliser des banques des expressions linguistiques dans la rédaction de l'introduction de leur travail de recherche en français.

Les techniques utilisées pour le questionnaire ont été :

- ❖ **Interview:** elle est déroulée avec cinq professeurs liés à l'enseignement du français, à la méthodologie de la recherche et à la linguistique, pour vérifier leurs expériences en travaillant avec ce type de banque des expressions
- ❖ **Enquête semi-directive:** cette méthode a permis de poser une série de questions aux étudiants de la Licence ès Langue Anglaise, option Langue Française, pour obtenir des informations sur la difficulté posée par l'écriture en français de la section Introduction du travail de recherche.
- Méthodes mathématiques

Selon Arco et al. (2005) le regroupement est une tâche fondamentale pour le bon fonctionnement de plusieurs systèmes d'extraction de textes et de récupération d'informations, car il peut être utilisé pour chercher les plus proches voisins d'un document, pour atteindre la qualité des systèmes de récupération d'informations, de l'organisation et de la personnalisation des informations dans les moteurs de recherche, pour contrôler l'homogénéité d'un corpus textuel dans le résumé de la collection de documents et dans la catégorisation des termes.

Le regroupement de documents a pour but de former une collection de *clusters* (sous-ensembles, groupes, classes) répondant aux propriétés suivantes:

- ❖ L'homogénéité au sein des *clusters*, c'est-à-dire les documents appartenant au même groupe doivent être aussi similaires que possible.
- ❖ Hétérogénéité entre les *clusters*, c'est-à-dire les documents appartenant à différents groupes doivent être aussi différents que possible. Un *cluster* est une collection d'objets similaires les uns aux autres, basés sur les valeurs de leurs attributs, et pouvant être traités collectivement en tant que groupe. Le regroupement est utile pour obtenir une distribution interne de l'ensemble de données (Fernández-Pampillón Cestreros, 2010).

Mesures de comparaison d'objets

Lors du regroupement des objets d'un ensemble de données, certaines mesures sont nécessaires pour quantifier le degré d'association entre eux. À cette fin, nous pouvons utiliser une mesure de distance, ou une mesure de similitude ou de dissimilitude. Certains algorithmes de regroupement ont une exigence théorique pour l'utilisation d'une mesure spécifique, mais le plus commun est que le chercheur sélectionne la mesure à utiliser avec une certaine méthode (Fernández-Pampillón Cestreros, 2010).

Techniques de sélection des fonctionnalités utilisées dans la construction de banques des expressions

Des techniques de sélection de caractéristiques ont été utilisées dans la construction de lexiques par la linguistique informatique. La sélection des caractéristiques a permis la sélection des termes pour la construction de lexiques qui, dans le calcul, permettent d'obtenir le meilleur sous-ensemble de mots dans tous les ensembles ou sous-ensembles possibles.

- ❖ **Stop word elimination:** Cette technique consiste à éliminer les mots sans signification indépendante aussi fréquemment appelés mots d'arrêt (Salton et McGillm, 1983). En général, ces mots ont des niveaux élevés d'apparition dans les textes, car ils sont couramment utilisés comme liens et compléments prépositionnels. Les mots vides ont une fréquence d'occurrence élevée et sont difficiles à distinguer (Rijsbergen, 1979 ;

Sahami et al., 1993) par conséquent, il est recommandé leur élimination dans la construction des banques des expressions.

- ❖ **Seuil de fréquence des termes et Loi de Zipf:** c'est une technique très simple qui est responsable de la sélection des mots les plus répétés dans un domaine linguistique ou dans un corpus documentaire. On sait que ses principes sont associés à des activités de calcul réalisées par Peter Luhn (1958) en calculant la fréquence des occurrences de mots pour déterminer la représentativité de ces derniers dans une communauté ; de cette façon, les mots qui ont moins la fréquence d'occurrence dans un corpus peuvent être éliminés pour être peu représentatifs. Sahami (1988) a étudié la Loi de Zipf (1949) notant qu'il est possible de sélectionner avec 50% d'efficacité si les mots qui ont une fréquence très élevée ou très faible sont supprimés, cela est en fonction du seuil délimité, c'est-à-dire, le nombre de mots que l'on souhaite obtenir.
- ❖ **Méthode N-grams:** Cette méthode a été utilisée pour construire des expressions composées à partir de la taille des chaînes de caractères de longueur fixe. Avec N-grams un filtrage de mots est réalisé, très similaires à ceux réalisés avec la loi de Zipf et TDF-IDF, de sorte que la fréquence des chaînes de caractères dans une longueur prédéterminée est calculée avant.
- ❖ **Seuil de fréquence des documents:** Si l'on sait que $n(t)$ est le nombre de documents dans un domaine où il apparaît au moins un mot t , est très avantageux l'emploi d'une heuristique où ces mots dont la fréquence d'occurrence est inférieure à un certain seuil soient éliminés à l'avance. Les mots à faible fréquence d'apparition n'auront pas une charge sémantique très élevée, et ne seront pas représentatifs du domaine des connaissances à structurer. Selon Arco, Hatmann et al. (2005), utiliser l'occurrence rare de termes n'est pas statistiquement fiable. En éliminant ces termes, le pouvoir discriminant est maintenu et l'efficacité du regroupement et de la classification textuelle est améliorée.
- ❖ **TF-IDF (fréquence du terme - fréquence inverse du document) :** Cette mesure utilisée dans les travaux de Gil-Leiva (2010) et de Urdiciain et Sánchez (2014) diffère du cas précédent dans TDF-IDF. La valeur

discriminante d'un terme est obtenue non pas de ceux qui ont une fréquence élevée, mais de ceux qui ont une fréquence inverse, c'est-à-dire, des termes les moins fréquents.

Algorithmes de *stemming*

Une autre des techniques utilisées pour obtenir des groupes efficaces, et très présents dans les systèmes d'analyse, sont les algorithmes de lematization. Selon Rolke et al. (2006), le processus de *stemming* est responsable de la normalisation et de la représentation correcte des descripteurs de caractéristiques sémantiques des documents, de manière généralisée. Les processus de *stemming* ont été appliqués à la récupération d'informations, et leur utilisation est générée dans les années 60, selon Baeza-Yates (Frakes et Baeza-Yates, 1992) afin de réduire la taille des index de texte, comme un moyen de normaliser les termes. La technique peut être réalisée de deux manières:

- En associant les mots à son lemme à travers un lexique ou une base lexicale, dans ce cas Word-net.
- En utilisant les algorithmes de thématization avec des règles de dérivation morphologiques.

Une fois analysé l'univers des techniques et procédures utilisées pour travailler avec des banques d'expressions, nous avons sélectionné les différentes techniques d'origine mathématique.

- Stop word elimination: permettra d'éliminer les mots qui ont très peu ou pas de contexte sémantique significatif dans une phrase.
- Loi de Zipf: sera utilisé pour obtenir la fréquence des mots dans un corpus d'une spécialité.
- N-Grams: facilitera l'obtention de phrases substantielles et de phrases composées pour les banque des expressions.
- TDF-IDF: permettra de connaître la qualité des mots dans un corpus.

2.7 Outils

AntConc: outillage pour analyser le corpus conçu spécifiquement pour être utilisé en classe. Il s'agit d'une application gratuite, idéale pour les écoles ou les collèges disposant d'un budget limité. Il fonctionne à la fois sur des systèmes Windows et

Linux / Unix. Bien qu'il soit gratuit, il comprend une interface d'utilisateur graphique intuitive et facile à utiliser et offre une puissante concordance, des générateurs de fréquence de mots et de mots clés, des outils d'analyse de groupes et de lexiques, ainsi qu'un diagramme de distribution des mots (Diniz, 2005).

AntConc effectue toutes les opérations directement sur les textes bruts du corpus. Cela est utile dans la mesure où l'utilisateur change ou modifie souvent le corpus cible pour un besoin particulier, car le programme ne nécessite aucun traitement préalable des données, par exemple pour créer un index (Anthony, 2004). L'utilisateur peut également travailler sur plusieurs textes ou extraits et mots-clés en comparant plusieurs textes ou bien en comparant un corpus d'analyse avec un corpus de référence qu'il aura constitué préalablement.

Protégé: éditeur d'ontologies et de bases de connaissances qui a été produit par l'Université de Stanford. C'est un outil qui permet la construction d'ontologies de domaine, et de formulaires de saisie de données personnalisés. De plus, il permet la définition de classes, de hiérarchies de classes, de variables, de restrictions de valeurs variables, ainsi que les relations entre les classes et les propriétés de ces relations. Il s'agit de la technologie la plus largement utilisée, indépendante du domaine, librement disponible et indépendante de la plateforme, pour le développement et la gestion de terminologies, d'ontologies et de bases de connaissances dans un large éventail de domaines d'application. Protégé est livré avec des packages de visualisation tels que OntoViz; tout cela aide l'utilisateur à visualiser les ontologies à l'aide de diagrammes.

Le principal point fort de Protégé est qu'il aide à la fois les constructeurs d'outils, les ingénieurs du savoir et les spécialistes principaux (Malviya, et al., 2011). Protégé peut lire et sauvegarder des ontologies dans la plupart des formats d'ontologies : RDF, RDFS, OWL, etc. Une ontologie est une base de connaissance qui décrit un domaine. Elle décrit les concepts généraux d'un domaine et les relations qui peuvent lier ces concepts. Ces concepts et liens sont décrits à l'aide d'un certain nombre de primitives inspirés de la logique de description.

RAOP (Rhetorical Annotation Ontology Project) : vise à apporter un outil de numérisation des formes de discours en utilisant les technologies associées aux données liées au *Linked data* et au Web sémantique. Le RAOP a pour but de

soutenir l'analyse littéraire numérique rhétorique et syntaxique ainsi qu'une approche théorique savante de ce domaine. Il cherche à créer le terrain pour la représentation de structures de texte complexes en tant que figure rhétorique impliquant de multiples structures d'annotation coexistantes, hiérarchiques et non hiérarchiques.

Ce projet vise à mettre en place un système permettant à l'utilisateur de détecter les expressions parlées afin de démêler les mécanismes, règles grammaticales et stylistiques, qui construisent un discours ou les aspects de la rhétorique qui rendent la communication plus persuasive. Il soutient que l'annotation de figures rhétoriques est importante non seulement pour des raisons linguistiques, mais également pour la découverte de différents styles d'écriture, de finalités et de styles / typologies de documents écrits et de genres (Marchesini et Mancinelli, 2017).

GraphDB: Modèle de données qui intègre une modélisation explicite des structures de graphes sans à-coups dans un environnement «standard» de modélisation et de requête orienté vers l'objet. L'objectif principal de GraphDB est qu'il soit capable de formuler des opérations sur les graphes et de les intégrer de manière claire dans les requêtes. Cela est possible car la base de données a une structure de graphique explicite bien définie.

Le modèle est couplé à un concept de mise en œuvre qui offre des structures de données spéciales pour la représentation des graphes et des algorithmes de graphes efficaces pour les opérations sur les graphes. En plus d'être convenable pour les applications standards, GraphDB est particulièrement adapté à la modélisation et à la manipulation sophistiquée de réseaux intégrés dans l'espace (Hartmut Güting, 1994).

2.8 Population et échantillon

Population :

Il s'agit de 63 étudiants de la 2^e jusqu'à la 5^e année, à la Faculté des Sciences Humaines de l'UCLV dans l'année scolaire 2018-2019.

Échantillon :

9 étudiants qui font leur travail de recherche en français, ce qui justifie un échantillonnage non probabiliste intentionnel car la chercheuse définit à l'avance les individus qu'elle inclura dans son étude à sa convenance.

2.9 Révision bibliographique

À travers les différentes techniques de révision bibliographique, un résumé des auteurs principaux, des années et des types de documents faisant partie de la bibliographie consultée dans l'élaboration de cette recherche, a été obtenu. Les résultats sont montrés ci-dessous (tableaux 2, 3, 4):

Tableau 2. Auteurs principaux

Auteurs	Citations bibliographiques
Bressia, Rocío	7
Sinclair, John	5
Bowker, Lynne	4
Swales, John	4
Biber, Douglas	3
Gil Leiva, Isidoro	3
Parodi, Giovanni	3
Allison, Desmond	2
Hernández Sampieri, Roberto	2
Leiva-Mederos, Amed	2
Vargas Sierra, Chelo	2

Tableau 3. Anées principales

Anées	Citations bibliographiques
2005	7
1996	6
2008	6
2010	6
2014	6
2004	5
2007	5
1994	4
2000	4
2002	4
2011	4
1998	3
2006	3
2009	3
2017	2
2018	2

Tableau 4. Types de documents

Types de documents	Citations bibliographiques
Article de revue	60
Livre	23
Document de conférence	5

Générique	3
Page Web	2
Article de Magazine	1
Thèse	1
Travail non publié	1

2.10 Limites de notre étude

L'étude est limitée car le travail concerne seulement la section d'Introduction des travaux de recherche. Une autre limitation est le peu d'utilisation de banques des expressions linguistiques et de *knowledge graphs*, et la presque absence d'études précédentes liées à ces sujets au niveau national.

2.11 Conclusions partielles

Dans ce deuxième chapitre, on a exploré les référents méthodologiques de la recherche. On a décrit le contexte, les étapes, le type et la portée de notre recherche. De même, on a fait référence aux méthodes théoriques, empiriques et mathématiques qu'ont été utilisées. On a présenté aussi l'opérationnalisation des variables, la population, et l'échantillon, les outils, la révision bibliographique et les limites de notre étude.



Troisième

Chapitre

CHAPITRE III: BANQUE DES EXPRESSIONS LINGUISTIQUES: ANALYSE DES DONNÉES ET RÉSULTATS

Dans ce chapitre, la méthodologie employée dans notre recherche pour élaborer la banque des expressions linguistiques est explorée. On présente également les étapes principales suivies pendant ce processus jusqu'à la présentation finale de la banque.

3.1 Résultats des techniques de collecte d'informations

Une enquête (Annexe I) a été menée auprès des 9 étudiants de la Licence ès Langue Anglaise option Langue Française qui ont décidé de faire leurs travaux de recherche en français. Le but de cette enquête était de savoir s'ils connaissaient la structure et les informations que l'introduction de plans de recherche ou de travaux de recherche dans cette langue devrait présenter, et quelle importance ils accordent à cette section. En outre, il était prévu de connaître le degré de difficulté qu'ils rencontrent quand ils écrivent l'introduction en raison de l'absence d'instruments lexicaux servant de guide pour mener à bien ce processus. Les résultats de l'enquête sont résumés ci-dessous :

Tableau 5. Résultats de l'enquête

Importance accordée à la section d'introduction	Peu important	Important	Très important		
	-	44.44%	55.56%		
Degré de complexité de l'écriture des introductions (1 le moins complexe et 5 le plus complexe)	1	2	3	4	5
	11%	22%	22%	44%	-
Connaissance de la structure des introductions	Oui	Non			
	89%	11%			

travaux de recherche					
Connaissance des mouvements rhétoriques	Oui	Non			
	33%	67%			
La banque des expressions linguistiques sera-t-elle utile?	Oui	Non			
	100%	-			

De même, une interview a été déroulée avec cinq professeurs liés à l'enseignement du français, à la méthodologie de la recherche et à la linguistique, pour vérifier leurs expériences avec le travail de ce type de banque des expressions.

Tous les professeurs ont convenu que les marqueurs discursifs les plus fréquemment utilisés dans l'introduction des travaux de recherche sont ceux qui ont pour but de mettre en contraste ce qui existe réellement avec ce qui devrait exister, dans l'objectif de mettre en évidence la présence du problème à analyser et à résoudre dans le reste du travail. Face à la question sur les méthodes ou opérations de base utilisées par les étudiants de la licence au cours du processus d'écriture de l'introduction de leurs travaux de recherche, les professeurs affirment qu'aucune méthode n'est utilisée, car dans l'introduction, nous devons réviser toute bibliographie existante sur le sujet à étudier ; et vérifier dans la pratique, en fonction de notre expérience personnelle ou de ce que l'enseignant indique, quel aspect en particulier doit être examiné. Dans le cas des opérations de base, elles coïncident avec ceux qui sont réalisés dans tous les processus de recherche: analyse, synthèse, induction, déduction, et qui correspondent à des processus de pensée logiques, ont-ils ajouté.

Tous les enseignants sont en conformité avec l'idée de que l'utilisation de mouvements rhétoriques est très importante car ils structurent le texte et guident le lecteur dans l'organisation interne qui présente le travail, et que l'expéditeur et le destinataire doivent être conscients de ce que sont les mouvements rhétoriques,

pour que l'écriture soit correcte et qu'il y ait une meilleure compréhension. L'utilisation de ces mouvements a une grande valeur car, après en avoir pris connaissance, l'étudiant émetteur est mieux placé pour écrire un texte cohérent et compréhensible et le lecteur en tant que destinataire discerne mieux le but de l'auteur.

Il existe un manque de connaissances quasi généralisé en ce qui concerne les méthodes utilisées dans la construction des banques des expressions linguistiques puisqu'il s'agit d'un sujet relativement nouveau et sa mise en œuvre au niveau national n'est pas connue. Malgré la méconnaissance presque totale de l'utilisation des ontologies pour la construction de marques discursives et les relations entre termes ou grands lexiques, ou pour la création des outils linguistiques, certains professeurs reconnaissent leur utilité dans ces processus.

Les enseignants interrogés conviennent que l'outil le plus utilisé dans notre licence, est AntConc, car il est facile à utiliser et gratuit. En ce qui concerne l'influence des *knowledge graphs* dans la création des banques des expressions, les professeurs manquent de connaissances en raison de la nouveauté du sujet et de l'absence de bibliographie, même sur Internet, à cet égard.

Enfin, selon les professeurs, au niveau universitaire, il existe deux types principaux de textes académiques: les textes écrits et les textes oraux. Le projet de recherche, les travaux de cours, les travaux de diplôme, le mémoire de maîtrise et le doctorat appartiennent au premier type. Dans le second type, nous trouvons les discours, les rapports oraux et la soutenance des thèses de maîtrise et de doctorat.

3.2 Étapes pour l'élaboration de la banque des expressions linguistiques

Cette section décrit le processus d'élaboration d'une banque des expressions linguistiques basée sur des mouvements rhétoriques présents dans les introductions des travaux de recherche. Pour l'accomplissement de notre tâche, on a suivi le cadre théorique fourni par la théorie de Vargas (2005) qui a proposé sept étapes de base qui, selon elle, devraient guider, en général, l'élaboration de la banque des expressions. Ces étapes sont elles-mêmes basées sur la méthodologie proposée par Schwarz (1990) et sur l'adaptation aux nouvelles

technologies présentée par Gómez et Vargas (2004). L'auteure de cette recherche a décidé d'adapter ces stages-là aux caractéristiques du travail décrit ici. Ces étapes sont: (1) sélection du corpus; (2) analyse des mouvements rhétoriques; (3) extraction des expressions à travers les outils; (4) traitement et stockage des données; (5) normalisation, et (6) édition de la banque des expressions linguistiques.

3.2.1 Sélection du corpus

En tenant compte des critères exposés au chapitre I concernant la sélection des corpus, un corpus spécialisé a été compilé pour cette thèse. Ce corpus a été créé intentionnellement pour servir à des fins de recherche spécifiques, afin qu'il puisse être considéré comme raisonnablement représentatif de l'utilisation écrite de la langue dans les domaines de la linguistique et de la didactique de la langue française. De plus, le corpus sélectionné est cohérent avec la notion d'équilibre (Sinclair, 1991; Atkins, Clear & Ostler, 1991) abordé dans le chapitre I, car il « offre un modèle maniable à petite échelle du matériel linguistique que les constructeurs de corpus souhaitent étudier » (Atkins, Clear & Ostler, 1992 :5). Quant à la taille, le corpus spécialisé sélectionné a 47,878 mots. Ce genre de corpus est généralement plus petit que ceux du langage général. Par conséquent, le corpus sélectionné était conforme aux idées de Pearson (1998) selon lesquelles la taille du corpus dépend de la disponibilité de l'information électronique et des critères établis pour la sélection des corpus.

D'autres éléments ont également été pris en compte lors de la compilation du corpus. Tous les textes ont été tirés de sources authentiques, c'est-à-dire cinquante mémoires publiés et spécialisés dans la linguistique et la didactique françaises. De plus, tous les mémoires ont été rédigés en français. Un autre élément était la date de publication qui allait de 2000 à 2018. Ce point a été jugé pertinent pour l'étude parce qu'il assurait que la langue des mémoires était mise à jour.

Les caractéristiques du corpus sélectionné, suivant les critères de compilation de Bowker et Pearson (2002), ont également été prises en compte. Elles sont résumées comme suit:

Tableau 6. Caractéristiques du corpus sélectionné

Taille	47,878
Mode de transmission	Écrit transformé en format .txt
Niveau de spécialisation	Textes spécialisés du domaine spécifique de la linguistique et de la didactique
Type de textes	Mémoires
Auteurs	Tous les textes ont été créés par des spécialistes de la linguistique et de la didactique
Langue du corpus	Textes écrits à l'origine en français
Date de publication	Textes publiés de 2000 à 2018

Une fois le corpus a été finalement compilé, tous les documents *pdf* ont été convertis en *format Word* afin de commencer l'étape d'étiquetage.

3.2.2 Analyse des mouvements rhétoriques

Comme l'auteure l'avait déjà expliqué dans le cadre théorique de la dissertation, l'analyse des mouvements est considérée comme un outil pour analyser des textes. Il fournit au chercheur des buts communicatifs caractérisant un genre (Swales 1990, Bathia 1993). Le présent travail a pris en compte les modèles d'analyse basés sur les mouvements proposés par Swales (1990) pour étiqueter les fragments textuels du corpus dans la recherche de mouvements.

Après avoir sélectionné le modèle rhétorique pour mener l'analyse de genre du corpus, il a été nécessaire de coder les échantillons de texte composant la section de l'introduction. Par conséquent, chaque texte a reçu un code unique (par exemple I-1...I-50). Par la suite, l'étiquetage du corpus a été réalisé afin d'identifier la structure rhétorique caractérisant cette partie du mémoire du domaine de la linguistique et de la didactique.

Le processus d'étiquetage était l'une des étapes les plus complexes de la recherche. Certains fragments textuels contenaient plus d'un mouvement

rhétorique et il était difficile de distinguer un mouvement de l'autre. Dans ce cas, comme il a été indiqué dans d'autres études (Ruiying et Allison, 2003), le mouvement le plus saillant du fragment textuel a été étiqueté. En outre, pas tous les mouvements ne sont présents dans les textes et dans certains cas, ils n'étaient pas représentés de manière ordonnée.

Comme indiqué précédemment, l'étiquetage des fragments textuels a été effectué en prenant en compte le modèle rhétorique de Swales (1990). Une fois qu'on a analysé des mouvements, on a conclu que la section de l'introduction des travaux de recherche appartenant aux domaines de la linguistique et de la didactique comprenait trois mouvements (tableau 7).

Tableau 7. Mouvements et étapes identifiés dans la section Introduction des mémoires

Nombre d'identification	Mouvement	Étapes
1	Établir un territoire (la situation)	1. Remarquer l'importance 2. Faire des généralisations à propos du sujet 3. Examiner des éléments de recherches antérieures
2	Établir une niche (le problème)	1a. Contre-réclamation 1b. Indiquer un écart 1c. Poser de questions 1d. Continuer une tradition
3	Occuper la niche (la solution)	1a. Définir des buts 1b. Annoncer la recherche actuelle 2. Annoncer les résultats principaux 3. Indiquer la structure du mémoire

Le modèle proposé dans cette thèse ne diffère pas du tout du modèle C.A.R.S. utilisé comme cadre pour analyser la section de l'Introduction. Comme mentionné dans la revue de la littérature, le modèle de Swales (1990) comportait trois mouvements dans la section de l'Introduction et de différentes étapes dedans chaque mouvement.

Après l'analyse des mouvements rhétoriques, on a conclu que certains étaient plus fréquents que d'autres. Ainsi, conformément à la classification de la fréquence des mouvements de Kanoksilapatham (2005), les mouvements ont été classés comme obligatoires (100%), conventionnels (60% à 99%) et optionnels (moins de 60%) (Tableau 8). On a trouvé que seulement l'étape 3 du mouvement 3, « Indiquer la structure du mémoire », était obligatoire puisqu'elle était présente dans l'ensemble du sous-corpus. De plus, toutes les étapes du mouvement 2, et l'étape 2 appartenant au mouvement 3, étaient classées comme optionnelles. Les autres étapes ont été classées comme conventionnelles.

Tableau 8. Fréquence des mouvements dans la section de l'introduction

Mouvement	Étapes	Classification des mouvements en accord avec la fréquence d'occurrence (%)		
		Obligatoire	Conventionnel	Optionnel
1	1		76%	
	2		60%	
	3		76%	
2	1a			42%
	1b			54%
	1c			44%
	1d			50%
3	1a		70%	
	1b		60%	
	2			38%

	3	100%		
--	----------	-------------	--	--

En conclusion, dans la section d'introduction, seulement une étape a été trouvée obligatoire, cinq ont été conventionnelles et cinq ont été considérées optionnelles.

Mouvement 1: *Établir un territoire (la situation)*. Comme on a déjà démontré dans le travail, ce mouvement est composé par trois étapes différentes :

Étape 1: Remarquer l'importance:

La linguistique fournit une part importante des fondements théoriques de l'enseignement des langues. (I-1)

Ce mémoire porte sur l'évolution de la réalisation syntaxique de la négation en français entre les XVIe et XIXe siècles. (I-23)

Étape 2: Faire des généralisations à propos du sujet:

Parallèlement, les théoriciens se sont également joints à cet engouement en proposant des théories et des pistes d'action pour l'intégration « scientifique » du jeu dans l'éducation. (I-7)

Pour remédier aux difficultés des élèves, diverses mesures ont été prises, divers programmes adoptés. (I-35)

Étape 3: Examiner des éléments de recherches antérieures:

En ce qui concerne l'enseignement de l'oral au primaire, les auteurs, didacticiens et linguistes, expliquent que ce dernier repose sur des compétences langagières que l'on retrouve dans la vie sociale. (I-24)

Ce qui est rapporté ici en exergue au sujet des prépositions, par exemple, trouve écho dans plusieurs recherches en linguistique cognitive. (I-42)

Mouvement 2: *Établir une niche (le problème)*. Les étapes qui conforment ce mouvement sont:

Étape 1a: Contre-réclamation:

En effet, à notre connaissance, aucune étude n'a rendu compte de façon quantitative, avec une composante qualitative, des effets d'un enseignement grammatical coordonné et simultané. (I-9)

Par contre, apprendre une langue étrangère c'est apprendre une culture nouvelle, des modes de vivre, des attitudes, des façons de penser nouvelles, et différentes. (I-19)

Étape 1b: Indiquer un écart :

Ainsi, jusqu'à présent, les phénomènes de contact des langues ont rarement été envisagés de façon exhaustive, c'est-à-dire en tenant compte des spécificités et de la complexité des systèmes linguistiques considérés, sur de multiples dimensions. (I-3)

En travaillant auprès des étudiants universitaires éprouvant des difficultés à l'écrit, nous avons souvent pu constater que des étudiants, même s'ils appliquaient correctement la norme grammaticale, n'arrivaient pas à produire des énoncés qui soient toujours cohérents dans un texte. (I-43)

Étape 1c: Poser de questions:

En d'autres termes, est-ce que le manuel scolaire de la 3ème AS, tel qu'il est conçu actuellement, contribue à l'acquisition et au développement d'une compétence culturelle chez les apprenants du FLE ? (I-38)

Notre question de recherche est la suivante : doit-on distinguer les COD des CM? (I-49)

Étape 1d: Continuer une tradition:

À la lumière d'une synthèse de certaines idées directrices, nous proposons un modèle dérivé d'une théorie originale, formalisé en Grammaires de Propriétés (ci-après GP, cf. par exemple [Blache, 2001] ou plus récemment [Blache, 2005]) (I-11)

L'optique dans laquelle j'aimerais entreprendre cette étude correspond à une approche interdisciplinaire tenant compte de l'évolution de différentes conceptions. (I-27)

Mouvement 3 : *Occuper la niche (la solution)*. Il est composé par des étapes suivantes :

Étape 1a: Définir des buts :

*L'objectif de ce mémoire est historique : trouver les étapes qui ont mené au classement actuel des pronoms dans les grammaires. (I-15)
En fait, nous montrerons que l'existence du « de pro » est difficilement vérifiable, échappant à la pronominalisation par en ou par dont, et problématique. (I-30)*

Étape 1b: Annoncer la recherche actuelle :

Notre recherche de type quantitatif vise à comparer le comportement orthographique d'élèves de différents niveaux scolaires d'écoles pluriethniques d'une part, et d'autre part, d'écoles à majorité francophone de la région de Montréal. (I-4)

Autrement dit, dans le présent travail, je souhaite étudier si les variantes affriquées susciteront (ou pas) des remarques à partir de leur simple présence. (I-44)

Étape 2 : Annoncer les résultats principaux :

Dans le cadre de ce mémoire, une description des tâches réalisées avec les TIC, conçues par des enseignants de français langue seconde (FLS) au secondaire, est proposée. (I-16)

C'est du moins le projet que nous nous sommes fixé, à savoir, faciliter l'apprentissage de stratégies de réécriture en contexte de fiction dans le but d'opérer un traitement sur le texte dans une optique d'amélioration. (I-46)

Étape 3: Indiquer la structure de l'article:

Les différentes parties du mémoire qui décrivent notre étude se présenteront de la manière suivante: nous exposerons dans le chapitre I, la problématique afin de cerner notre domaine d'étude et de formuler notre question de recherche. (I-12)

Après la définition des concepts à l'étude, nous abordons le troisième chapitre où nous présentons en premier lieu l'étude faisant l'objet de reprise. (I-50)

3.2.3 Extraction des expressions à travers les outils

Bien qu'une banque des expressions puisse être compilée et produite manuellement, cette procédure n'est pas recommandée. Si nous voulons augmenter l'efficacité et éviter des erreurs graves, nous devons utiliser un logiciel spécialisé dans la gestion des banques des expressions. Cependant, il est nécessaire de mentionner que le processus d'extraction des expressions pour former une banque se produit de manière semi-automatique (rélation homme-ordinateur).

Il a été confirmé que les programmes de concordance constituent une aide efficace pour l'apprentissage d'une langue seconde ou d'une langue étrangère de manière générale, car ils facilitent l'acquisition du vocabulaire, de la grammaire et des styles d'écriture. Par exemple, plusieurs études ont montré que le vocabulaire ne peut être appris qu'en trouvant des mots ou des expressions dans leur contexte naturel et dans différentes situations. Sur la seule base de l'intuition, il est presque impossible de trouver un nombre suffisant d'expressions spécifiques pouvant satisfaire à ces conditions.

Après avoir analysé les exemples de mouvements et d'étapes, les expressions linguistiques les plus récurrentes ont été identifiées à travers un logiciel. AntConc 3.4.3 5 w (Windows) s'agit d'un programme de concordance conçu pour être facile à utiliser et efficace pour exécuter des tâches telles que des listes de mots, des analyses de concordance, des cooccurrences et des listes de mots clés, entre autres.

L'outil *Concordance* était particulièrement utile pour identifier les expressions linguistiques les plus récurrentes des mouvements identifiés dans la section codée. Cet outil a présenté des groupes de mots entourant un terme de recherche et les a classé par ordre alphabétique ou par fréquence. De plus, cet outil a permis d'observer la variation des temps dans les différents mouvements.

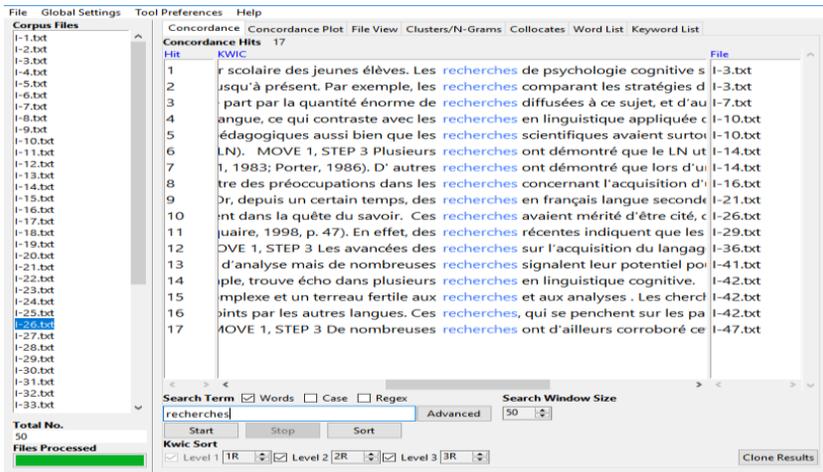


Fig. 1. Outil de Concordance appliqué à un élément linguistique récurrent

L'outil *Word List* du logiciel AntConc 3.4.3 5, était utile pour déterminer les mots qui se sont produits le plus. Il est important de noter que l'auteure de cette recherche est concentrée sur des mots qui étaient pertinents au contenu et à la signification, par exemple, les sujets, les adjectifs et les verbes.

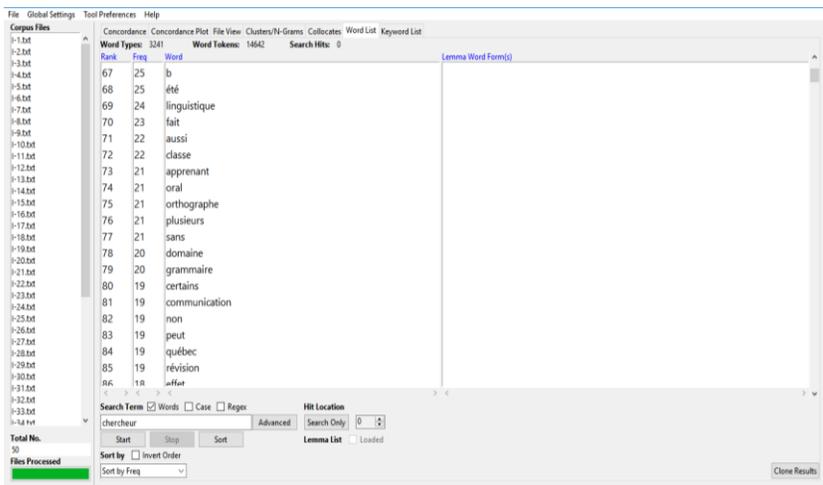


Fig. 2. Outil *Word List* appliqué à la section Introduction

En plus, les *Word Clusters/NGrams* facilitent également l'identification des mots qui apparaissent ensemble très fréquemment dans la section de l'Introduction.

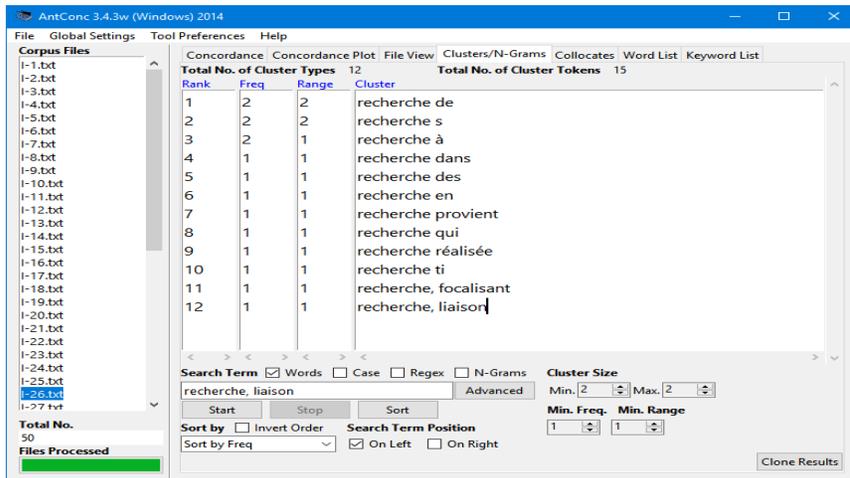


Fig. 3. Outil *Word Clusters / NGrams* appliqué à la section Introduction

3.2.4 Traitement et stockage des données

Après l'extraction des expressions du corpus à travers les outils, l'étape suivante consiste à traiter et stocker ces expressions pour les inclure dans la banque des expressions linguistiques. Tous les mots ont été analysés dans leur contexte, dans le mouvement rhétorique et l'étape auxquels ils appartenaient.

Le stockage des données a été réalisé à partir de fichiers indépendants au format *.txt. Une DTD (Document Text Definition) a été utilisée pour placer les phrases grammaticales de l'introduction au format XML afin de faciliter le traitement de l'information. Une fois que les phrases qui caractérisent chaque mouvement sont obtenues, elles sont placées dans le logiciel Protégé afin de relier sémantiquement toutes les relations établies entre les mouvements et les exemplifications de celles-ci au niveau du texte en vue de l'intégration d'un *knowledge graph*, démontrant que les mouvements appartiennent vraiment au domaine qui est analysé en question.

3.2.5 Normalisation

La phase de normalisation comprend deux éléments élémentaires. Le premier consiste en la structuration du texte à traiter par les applications informatiques ; et le second s'agit de délimiter toute la symbolisation et les étiquettes nécessaires pour que la présentation et la visualisation des mouvements rhétoriques suivent les normes internationales. Pour ce cas, 3 hiérarchies nommées Mouvements 1, 2

et 3 ont été placées avec leurs exemples et les marqueurs de discours qui identifient les axes rhétoriques des ensembles propositionnels du texte.

La terminologie utilisée est basée sur le modèle lexical utilisé par Lemon où les expressions lexicales sont considérées comme une entrée d'un lexique. Ce modèle est suggéré par l'Association des Linguistes d'Angleterre et le Consortium de la Web Sémantique pour la normalisation des processus linguistiques dans les environnements numériques.

3.2.6 Édition de la banque des expressions linguistiques

La dernière étape pour l'élaboration de la banque des expressions linguistiques a été son édition. Pour le faire, on a vu la nécessité de concevoir sa macrostructure et microstructure. De même, on a décidé d'ajouter un prologue pour exprimer l'objectif de la banque des expressions, la description de son organisation, et ses possibles utilisateurs.

3.2.6.1 Macrostructure

La macrostructure est orientée de manière à décrire la structure globale de la banque des expressions linguistiques. Premièrement, la sémantique, la syntaxe et la position pragmatique des expressions sont identifiées. Ces résultats sont entrés dans la base de données traitée par l'ontologie RAOP (*Rhetorical Annotation Ontology Project*).

La détection des axes sémantiques est développée à travers l'idée « L'organisation du discours n'est ni plate ni linéaire; elle est hiérarchique, c'est-à-dire que les clauses forment des structures d'ordre supérieur, des paragraphes, qui à leur tour se combinent pour former des épisodes ou des sections majeures du discours » (Tomlin et al., 2003 : 83).

La notion de macrostructure formulée par Van Dijk (1980) est en relation avec le sujet général. La macrostructure est la structure sémantique globale d'un discours et peut être exprimée par son titre, son gros titre ou par des phrases de synthèse. Les propositions macrostructurales sont dérivées de règles macro, c'est-à-dire en éliminant les propositions qui ne sont pas pertinentes pour l'interprétation d'autres propositions (élision), en convertissant une série de propositions

spécifiques en une proposition plus générale (généralisation), par le biais de la construction d'une proposition à partir d'un certain nombre de propositions du texte (construction) (Tomlin et al., 2003). Dans le cas de cette recherche, la macrostructure est composée des phrases rhétoriques les plus fréquentes tirées des travaux de recherche constituant le corpus compilé et analysé.

3.2.6.2 Microstructure

La microstructure est développée pour définir les structures syntaxiques des éléments grammaticaux trouvés dans les phrases du texte. Ici, toutes les structures grammaticales correspondant à la première entrée de termes dans le système ontologique sont identifiées. L'essentiel dans ce processus est de déterminer les unités syntaxiques (verbes, noms, pronoms, etc.). Ce sont les premières entrées de mots dans la banque des expressions, celles-ci sont formalisées dans l'ontologie.

En revanche, dans le cas de notre banque des expressions linguistiques, la phase de marquage a consisté à placer dans un fichier les éléments qui caractérisent les mouvements rhétoriques, c'est-à-dire le mouvement rhétorique ordonné plus les étapes qui le composent. Ensuite, les relations qui sont données au niveau de la logique au sein du mouvement rhétorique ont été présentées, en analysant en particulier les éléments de la microstructure, par exemple, les verbes, les adjectifs, les noms, les adverbes, les pronoms et les articles, et leur relation avec la phrase rhétorique où ils sont utilisés.

Enfin, on a conclu que chacun de ces éléments sont des instances et que celles-ci sont liées aux éléments macro qui peuvent être n'importe laquelle des étapes constituant chaque mouvement.



Conclusions

CONCLUSIONS

Étant donné le manque de méthodes pour écrire la section d'introduction du travail de recherche en français qui existe actuellement, notre recherche représente une contribution importante à ce processus. Par conséquent, les conclusions suivantes peuvent être tirées:

- ❖ Les fondements théoriques et méthodologiques du développement de cette recherche ont été analysés à travers différentes études et points de vue concernant l'écriture académique, le genre et les mouvements rhétoriques, la linguistique de corpus et les banques des expressions linguistiques. De plus, cette dissertation a aussi approfondi dans l'analyse rhétorique d'un genre en termes de mouvements et d'étapes. En outre, on a décidé d'étudier quelques sujets qui sont plutôt nouveaux et intéressants : les ontologies et les *knowledge graphs*.
- ❖ À travers l'utilisation de la méthode **Questionnaire**, on a pu obtenir des informations au sein de la communauté des utilisateurs qui font partie de l'étude à travers une série de questions orales et écrites. Grâce à cette méthode, on peut conclure qu'il y a une nécessité d'utiliser des banques des expressions linguistiques dans la rédaction de l'introduction de textes académiques en français.
- ❖ Les mouvements rhétoriques ont été analysés pour la création d'une banque des expressions linguistiques décrivant la structure rhétorique et les éléments linguistiques les plus récurrents qui caractérisent la section Introduction des travaux de recherche sur la didactique et la linguistique françaises. La procédure méthodologique a consisté en six étapes majeures: (1) sélection du corpus; (2) analyse des mouvements rhétoriques; (3) extraction des expressions à travers les outils; (4) traitement et stockage des données; (5) normalisation, et (6) édition de la banque des expressions linguistiques. De plus, des différents outils ont été employés: l'outil AntConc 3.4.3 5 w (Windows) a été utile pour l'étiquetage du corpus; le RAOP a été l'ontologie employé; l'outil Protégé a été la base de la banque et a été utilisé pour traiter l'ontologie et établir les relations entre ses éléments; le GraphDB a été le mécanisme de validation de la banque des expressions.



Recommendations

RECOMMANDATIONS

Le présent mémoire s'encadre dans le domaine de banques des expressions linguistiques et représente un pas importante pour contribuer à améliorer l'écriture de l'introduction du travail de recherche en français, dans les domaines de la didactique et de la linguistique, des étudiants de la Licence ès Langue Anglaise, option Langue Française. Pour cette raison, nous proposons les recommandations suivantes :

- ❖ Élargir le corpus de la présente recherche afin d'améliorer la banque des expressions linguistiques élaborée.
- ❖ Réaliser des études similaires sur les autres sections des mémoires sur la didactique et la linguistique françaises.
- ❖ Valider l'efficacité de la banque des expressions après chaque année scolaire.
- ❖ Introduire des thèmes de linguistique computationnelle dans le sujet Méthodologie de la recherche.



Bibliographie

BIBLIOGRAPHIE

- Adelstein, A., & Kuguel, I. (2005). *Los textos académicos en el nivel universitario*: Universidad Nacional de General Sarmiento.
- Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *Proceedings of IWLeL*, 7-13.
- Arco, L., Bello, R., Valdés, L., Mederos, J. M., & Pérez, Y. (2007). CorpusMiner 1.0: Herramienta para el agrupamiento de documentos. *Revista Cubana de Ciencias Informáticas*, 1(2).
- Arias, F. (2006). El proyecto de investigación: Introducción a la metodología. *Caracas, Episteme*.
- Atkins, S., Clear, J., & Ostler, N. (2010). Corpus Design Criteria, 1991. *British National Corpus Archive*.
- Auger, P. (2008, Hiver 2008). Pour une typologie des langues de spécialité. *Circuit*, 36.
- Auría, C. P.-L. (2004). The epistemic nature of disciplinary discourses: echoing postmodern literary practices. *Estudios Ingleses de la Universidad Complutense*, 12, 25-41.
- Ayers, M. (1993). Locke. Volume I: Epistemology. Volume II: Ontology.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2009). The Semantic Web Scientific American Magazine: March.
- Bhatia, V. K. (2014). *Analysing genre: Language use in professional settings*: Routledge.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes*, 26(3), 263-286.
- Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure* (Vol. 28): John Benjamins Publishing.
- Bolinger, D. (1976). Gradience in entailment. *Language Sciences*, 41, 1-13.
- Bowker, L. (1996). Towards a corpus-based approach to terminography. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 3(1), 27-52.
- Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical*

- guide to using corpora*: Routledge.
- Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for specific purposes*, 13(1), 47-59.
- Bybee, J. L., & Hopper, P. J. (2001). *Frequency and the emergence of linguistic structure* (Vol. 45): John Benjamins Publishing.
- Cabré, M. T. (1993). La terminología: teoría, metodología, aplicaciones. Barcelona: Editorial Antártida. *Empúries*.
- Carlino, P. (2005). *Escribir, leer, y aprender en la universidad. Una introducción a la alfabetización académica*: Fondo de cultura económica.
- Desmet, I. (2006). Variabilité et variation en terminologie et langues spécialisées: discours, textes et contextes. *Mots, termes et contextes. Actualité scientifique*, 235-247.
- Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., . . . Tomlin, J. A. (2003). *SemTag and Seeker: Bootstrapping the semantic web via automated semantic annotation*. Paper presented at the Proceedings of the 12th international conference on World Wide Web.
- Dillon, M. (1983). Introduction to modern information retrieval: G. Salton and M. McGill. McGraw-Hill, New York (1983). xv+ 448 pp., \$32.95 ISBN 0-07-054484-0: Pergamon.
- Diniz, L. (2005). Review of TextStat 2.5, AntConc 3.0, and Compleat Lexical Tutor 4.0. *Language Learning & Technology*, 9(3), 22-27.
- Dobric, N. (2009). Corpus Linguistics-The Basic Form of Linguistic Analysis. *Philologia* (7), 31-41.
- Dudley-Evans, T. (2001). Team-teaching in EAP: Changes and adaptations in the Birmingham approach. *Research perspectives on English for academic purposes*, 225-238.
- Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*, 48.
- Färber, M., Bartscherer, F., Menne, C., & Rettinger, A. (2018). Linked data quality of dbpedia, freebase, opencyc, wikidata, and yago. *Semantic web*, 9(1), 77-129.
- Fernández Fastuca, L., & Bressia, R. (2009). Definiciones y características de los

principales tipos de texto. *Buenos Aires, Argentina. Obtenido de http://www.uca.edu.ar/uca/common/grupo18/files/Definicion_generos_discursivos_abril_2009.pdf.*

- Fernández-Pampillón Cesteros, A. (2010). *La construcción de tesauros académicos: un modelo general y un método inductivo con aplicación al " e-learning"*. Universidad Complutense de Madrid, Servicio de Publicaciones.
- Fillmore, C. J. (1992). *Corpus Linguistics' or 'Computer-aided armchair linguistics'*. Paper presented at the Directions in corpus linguistics. Proceedings of Nobel Symposium.
- Frakes, W. B., & Baeza-Yates, R. (1992). *Information retrieval: Data structures & algorithms* (Vol. 331): prentice Hall Englewood Cliffs, NJ.
- Galisson, R., & Coste, D. (1976). *Dictionnaire de didactique des langues: la conception de l'ensemble de l'ouvrage. Paris: Hachette.*
- Gil Leiva, I. (2010). *La automatización de la indización: propuesta teórica-metodológica. Aplicación en el área de biblioteconomía y documentación. Proyecto de investigación.*
- Gómez González-Jover, A. (2005). *Terminografía, lenguajes profesionales y mediación interlingüística: aplicación metodológica al léxico especializado de la industria del calzado y las industrias afines*: Universidad de Alicante.
- Gómez González-Jover, A., & Vargas Sierra, C. (2004). *Aspectos metodológicos para la elaboración de diccionarios especializados bilingües destinados al traductor.*
- Gutiérrez Carreras, Ó. (2014). *Glosario terminológico sobre Internet y contenidos multimedia.*
- Gütting, R. H. (1994). *GraphDB: Modeling and querying graphs in databases.* Paper presented at the VLDB.
- Haché, L., Ferron, B., & Khelil, Z. (2014). *Guide de présentation d'un travail universitaire selon la méthodologie APA (6e édition 2010)*: Moncton: Université de Moncton, Campus de Shippagan.
- Harris, Z. S. (1951). *Methods in structural linguistics.*
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la investigación* (Vol. 3): México: McGraw-Hill.

- Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for specific purposes*, 16(4), 321-337.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for specific purposes*, 27(1), 4-21.
- ISO, E. (2000). 9001: 2000. *Quality management systems-Requirements (ISO 9001: 2000)*.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for specific purposes*, 24(3), 269-292.
- Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. *Discourse on the move: Using corpus analysis to describe discourse structure*, 73-119.
- Kay, H., & Dudley-Evans, T. (1998). Genre: What teachers think.
- Kytö, M. (1994). Be vs. have with intransitives in Early Modern English. Amsterdam Studies in the Theory and History of Linguistic Science Series 4, 179-179.
- Lavagnino, E. (2012). Les Langues de Spécialité et la variation discursive: une étude contrastive de la réduction des termes complexes. *Synergies Espagne*(5).
- Leiva, I. G. (2008). *Manual de indización: teoría y práctica*: Trea.
- Leiva-Mederos, A. (2018). *Desarrollo de grafos de conocimiento en sistemas de información (inédito)*.
- Leiva-Mederos, A., Domínguez-Fernández, S., & Senso, J. A. (2012). PuertoTex: un software de minería textual para la creación de resúmenes automáticos en el dominio de ingeniería de puertos y costas basado en ontologías. *Transinformação*, 24(2).
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Paper presented at the Proceedings of the fifth Berkeley symposium on mathematical statistics and probability.
- Malo, M. (1996). *Guide de la communication écrite au cégep, à l'université et en entreprise*: Éditions Québec/Amérique.

- Malviya, N., Mishra, N., & Sahu, S. (2011). Developing university ontology using protégé owl tool: Process and reasoning. *International Journal of Scientific & Engineering Research*, 2(9), 1-8.
- Marchesini, A. e. M., Tiziana. (2017). RAOP: Rethorical Annotation Ontology Project. Retrieved 01/20, 2019
- Martin, J. R. (2000). Design and practice: Enacting functional linguistics. *Annual Review of Applied Linguistics*, 20, 116-126.
- Mayer, R., & Deslauriers, J.-P. (2000). Quelques éléments d'analyse qualitative: l'analyse de contenu, l'analyse ancrée, l'induction analytique et le récit de vie. *Méthodes de recherche en intervention sociale*, 159-189.
- Meneses, J., Fàbregues, S., Jacovkis, J., & Rodríguez-Gómez, D. (2014). La introducción de las TIC en el sistema educativo español (2000-2010): un análisis comparado de las políticas autonómicas desde una perspectiva multinivel= The introduction of ICT in Spanish education (2000-2010): a comparative analysis of regional policies from a multi-level approach. *Estudios sobre educación*, 27, 63-90.
- Meyer, C. F. (2002). *English corpus linguistics: An introduction*: Cambridge University Press.
- Meyer, I., & Mackintosh, K. (1996). The corpus from a terminographer's viewpoint. *International Journal of Corpus Linguistics*, 1(2), 257-285.
- Morley, J. (2014). Academic phrasebank. *Manchester: University of Manchester*. [online] Available at: <http://www.phrasebank.manchester.ac.uk/> [Accessed 19.01. 2019].
- Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14): John Benjamins Publishing.
- Nwogu, K. N. (1997). The medical research paper: Structure and functions. *English for specific purposes*, 16(2), 119-138.
- Paltridge, B. (1996). Genre, text type, and the language learning classroom.
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de lingüística teórica y aplicada*, 46(1), 93-119.
- Parodi, G. (2010). *Lingüística de corpus: de la teoría a la empiria*: Iberoamericana.
- Parodi, G., Ibáñez, R., Venegas, R., & González, C. (2010). Identificación de

- géneros académicos y géneros profesionales: principios teóricos y propuesta metodológica. *Alfabetización académica y profesional en el siglo XXI: Leer y escribir desde las disciplinas*, 249-289.
- Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3), 489-508.
- Pawley, A., & Syder, F. H. (1983). Natural selection in syntax: Notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics*, 7(5), 551-579.
- Peacock, M. (2002). Communicative moves in the discussion section of research articles. *System*, 30(4), 479-497.
- Pearson, J. (1998). *Terms in context* (Vol. 1): John Benjamins Publishing.
- Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for specific purposes*, 18(2), 139-160.
- Puiatti de Gómez, H. (2005). El artículo de investigación científica. *Los textos de la ciencia*, 23-85.
- Ramírez Gelves, S. (2007). Géneros discursivos y tipos de textos en el discurso académico. *CAICYT Curso virtual para editores de revistas científicas, módulo, 3*.
- Rijsbergen, C. (1979). v.(1979). *Information retrieval*, 2.
- Rölleke, T., Tsirikia, T., & Kazai, G. (2006). A general matrix framework for modelling information retrieval. *Information processing & management*, 42(1), 4-30.
- Ruiying, Y., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for specific purposes*, 22(4), 365-385.
- Ruiz Pérez, R., & Yunta, L. R. (1994). El análisis documental. Bases terminológicas. Conceptualización y estructura operativa. *Revista Española de Documentación Científica*, 17(2), 234.
- Sabaj Meruane, O. (2011). Construcción de un modelo de movidas retóricas para el análisis de artículos de investigación en español= Construction of a model of rhetorical moves for the analysis of research articles in Spanish.
- Sager, J. C., Dungworth, D., & McDonald, P. F. (1980). *English special languages: principles and practice in science and technology*: John Benjamins Pub Co.

- Sahami, M. (1993). *Learning non-linearly separable boolean functions with linear threshold unit trees and madaline-style networks*. Paper presented at the AAAI.
- Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for specific purposes*, 21(1), 1-17.
- Sánchez Jiménez, D. (2011). Las funciones retóricas de la citación en la escritura académica universitaria. Estudio comparado del género de memorias de máster en nativos españoles y estudiantes filipinos de ELE. *Trabajo de investigación DEA Madrid: Universidad Antonio de Nebrija*.
- Schwarz, E. (1990). Traducción de textos sectoriales: importancia de la terminología. *Actas de los II Encuentros Complutenses en torno a la traducción, celebrados en el IULMT*, 12, 12.
- Senso, J. A. (2011). Gestión del conocimiento y tecnologías semánticas en inteligencia y defensa. *Inteligencia y Seguridad: Revista de Análisis y Prospectiva*(10).
- Séris, J.-P. (2000). *La technique*. Paris: PUF.
- Sinclair, J. (1991). *Corpus, concordance, collocation*: Oxford University Press.
- Sinclair, J., & Ball, J. (1996). EAGLES preliminary recommendations on text typology. *Birmingham, Corpus Linguistics Group, School of English, Birmingham University*.
- Sinclair, J., & Ball, J. (1996). Preliminary recommendations on text typology. *EAGLES Document EAG-TCWGTTYP/P*.
- Swales, J. (1981). DEFINITIONS IN SCIENCE AND LAW: EVIDENCE FOR SUBJECT SPECIFIC COURSE COMPONENTS?
- Swales, J. (1990). *Genre analysis: English in academic and research settings*: Cambridge University Press.
- Swales, J. (2004). *Research genres: Explorations and applications*: Ernst Klett Sprachen.
- Swales, J. M., & Feak, C. B. (2004). *Academic writing for graduate students: Essential tasks and skills* (Vol. 1): University of Michigan Press Ann Arbor, MI.
- Urdiciain, B. G., & Sánchez Jiménez, R. (2014). Técnicas de recuperación de

información aplicadas a la construcción de tesauros. *Transinformação*, 26(1).

Van Dijk, T. A. (1980). Text and context explorations in the semantics and pragmatics of discourse.

Van Dijk, T. A. (1998). *Texto y contexto: semántica y pragmática del discurso*: Cátedra.

Vargas Sierra, C. (2008). La sistematización terminográfica: una propuesta metodológica para la elaboración de diccionarios traductológicos.

Williams, M., & Burden, R. L. (1999). *Psicología para profesores de idiomas: Enfoque del constructivismo social*. Cambridge University Press.



Annexes

ANNEXES

Annexe I

Encuesta

Estimado alumno(a):

Estamos desarrollando una investigación relacionada con la estructura y las informaciones que deben aparecer en la introducción de su diseño de investigación (2º año) o trabajo de curso (3º, 4º años) en francés. La presente encuesta tiene el propósito de conocer si ustedes, los estudiantes, conocen dicha estructura e informaciones y qué importancia le atribuyen. Para ello pedimos su colaboración.

Muchas gracias

Preguntas

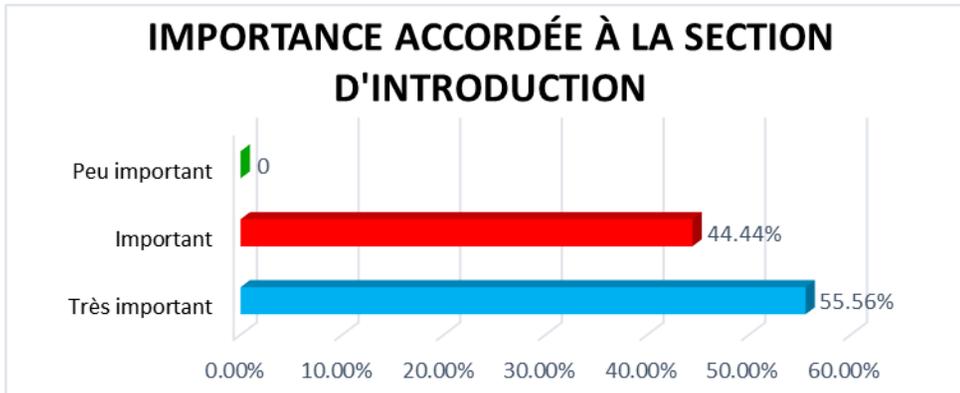
- 1- ¿Qué importancia le atribuye a la introducción de los diseños de investigación o trabajos de curso en francés?
__Muy importante __Importante __Poco importante
- 2- Marque con una cruz el nivel de complejidad que usted le atribuye a la escritura de la introducción de los diseños de investigación o trabajos de curso en francés (siendo 1 el nivel mínimo y 5 el máximo).

Grado de complejidad				
1	2	3	4	5

- 3- ¿Conoce la estructura que debe seguir la introducción de los diseños de trabajos de investigación o trabajos de curso en francés? __ Sí __No
 - 4- ¿Sabe lo que es una movida retórica? __Sí __No
 - 5- ¿Considera que un banco de expresiones que contenga frases para facilitar la escritura de la introducción de los diseños de investigación o trabajos de curso en francés sería útil para los estudiantes? __ Sí __No
- Fundamente su respuesta.

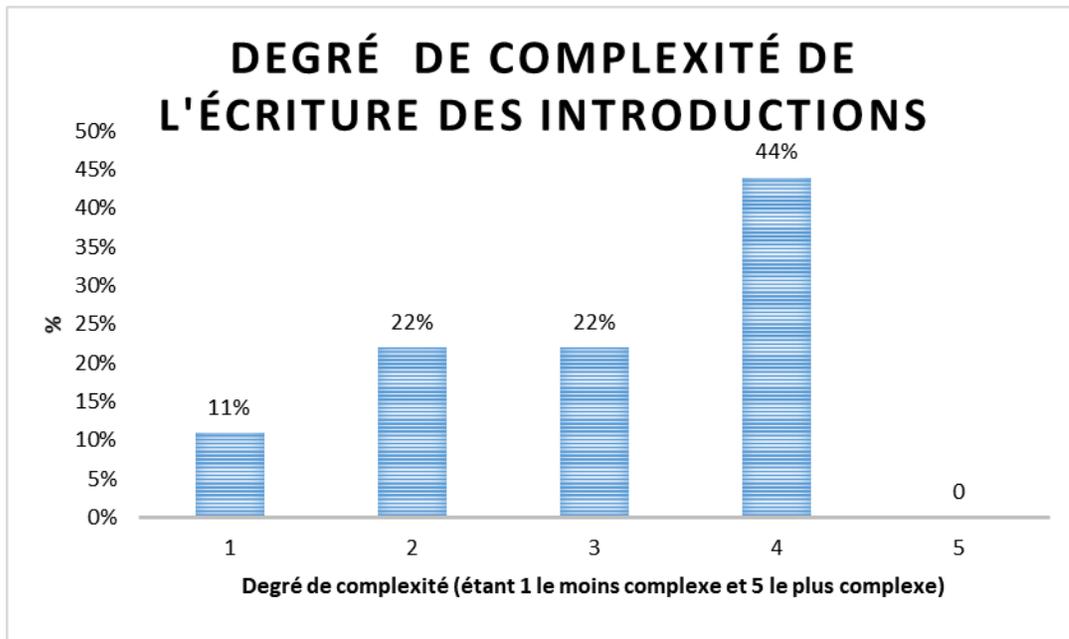
Annexe II: Résultats de l'enquête

1. ¿Qué importancia le atribuye a la introducción de los diseños de investigación o trabajos de curso en francés?



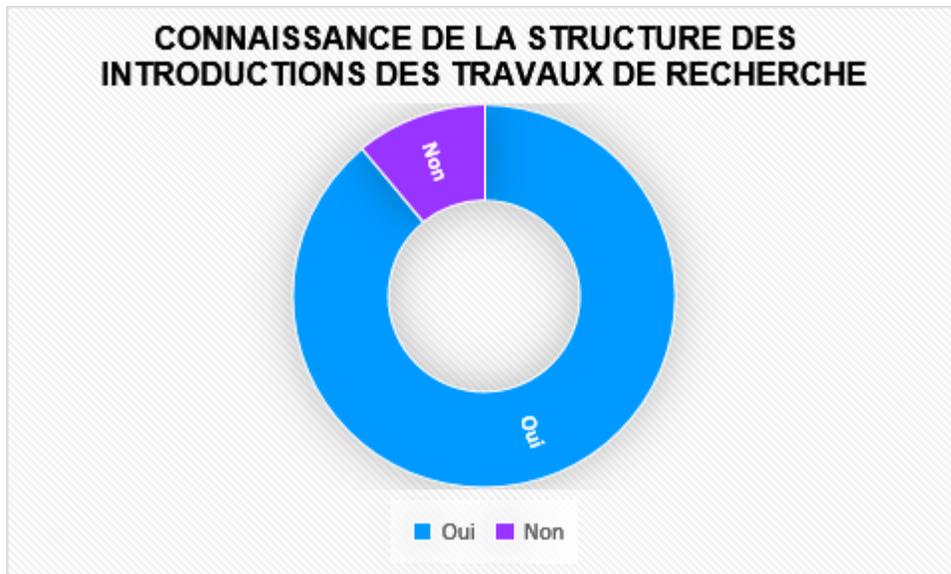
Graphique 1. Importance accordée à la section d'introduction

2. Nivel de complejidad que se le atribuye a la escritura de la introducción de los diseños de investigación o trabajos de curso en francés (siendo 1 el nivel mínimo y 5 el máximo).



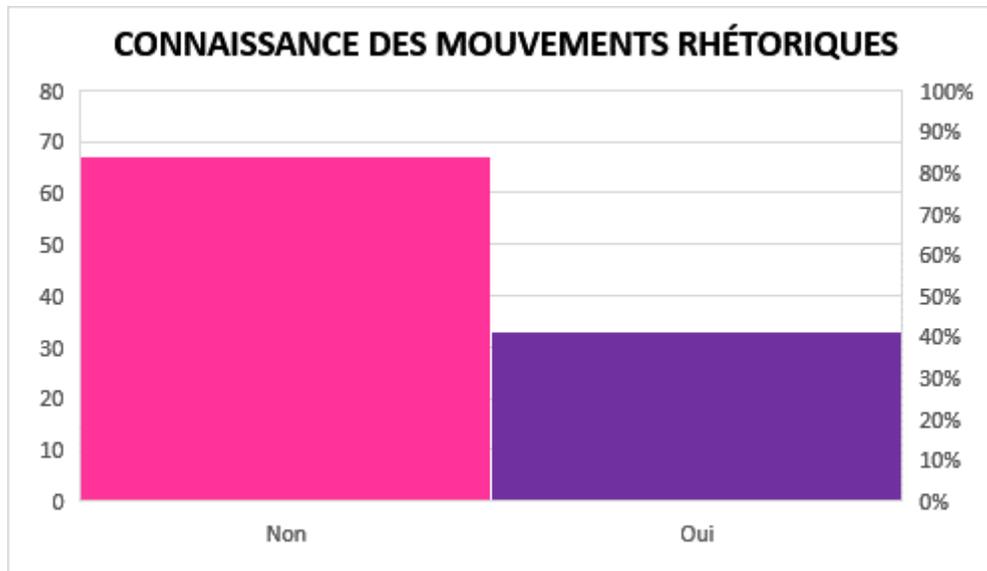
Graphique 2. Degré de complexité de l'écriture des introductions

3. ¿Conoce la estructura que debe seguir la introducción de los diseños de trabajos de investigación o trabajos de curso en francés?



Graphique 3 Connaissance de la structure des introductions des travaux de recherche

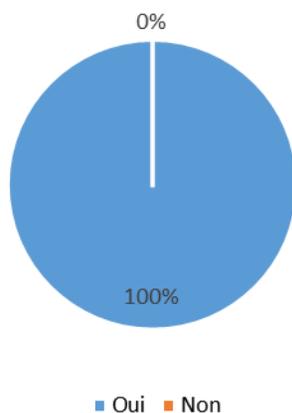
4. ¿Sabe lo que es una movida retórica?



Graphique 4 Connaissance des mouvements rhétoriques

5. ¿Considera que un banco de expresiones que contenga frases para facilitar la escritura de la introducción de los diseños de investigación o trabajos de curso en francés sería útil para los estudiantes?

**LA BANQUE DES EXPRESSIONS LINGUISTIQUES
SERA-T-ELLE UTILE?**



Graphique 5 La banque des expressions linguistiques sera-t-elle utile?

Annexe III

Entrevista

Estimado profesor (a):

Esta entrevista se aplica como parte del desarrollo de la tesis de diploma de una estudiante de la carrera Lengua Inglesa de la Universidad Central «Marta Abreu» de las Villas con el objetivo de conocer su opinión acerca el empleo de herramientas léxicas como los bancos de expresiones u otros recursos, así como la utilización de las movidas retóricas en los textos académicos. Para ello pedimos su colaboración.

Muchas gracias.

Identificación

Nombre: _____

Lugar de trabajo: _____

Cargo que ocupa: _____

1. ¿Cuáles son los marcadores discursivos que se usan con más frecuencia en la introducción de los trabajos de curso de la carrera?
2. ¿A qué se debe el uso frecuente de estos marcadores en el discurso?
3. ¿Cuáles son los métodos o las operaciones básicas que se realizan con frecuencia por los estudiantes de la carrera de Lengua Inglesa cuando van a redactar las introducciones de sus trabajos de curso?
4. ¿Qué importancia usted le da al uso de las movidas retóricas en este tipo de trabajo?
5. ¿Qué valor científico o educativo tiene para la Lengua Inglesa el uso de estas movidas retóricas?
6. ¿Qué métodos se han empleado a lo largo del tiempo en la construcción de bancos de expresiones?

7. ¿Qué opina sobre el uso de ontologías que permiten construir marcas discursivas y relaciones entre términos, como por ejemplo: Lemon, SKOS, u otras que son grandes léxicos como: Wordnet, Eurowordnet?
8. ¿Qué valor le otorga a dichas ontologías para la creación de herramientas lingüísticas y para la compilación?
9. ¿Qué herramientas ha visto usted en la literatura que se utilicen para el desarrollo y cuáles opina que son mejores y por qué? (ejemplos: Protégé, Graphdb, AntConc, Graphviz, OpenRefine)
10. Sabemos que el Big data es un fenómeno que está afectando a toda la red, por ejemplo hay grandes cantidades de texto, lo que ha llevado a la aparición de los grafos de conocimiento que están ayudando a los sistemas a construirse sobre la semántica navegacional, a incrustar sentido en los grafos para que trabajen a nivel de texto infiriendo cosas, entonces, ¿cómo han influido estas técnicas en la creación de bancos de expresiones?
11. ¿Cuáles son los tipos de textos académicos más utilizados a nivel universitario y por qué?

Annexe IV

Guide de l'utilisateur de la banque des expressions linguistiques

Aplication BANQUE DES EXPRESSIONS LINGUISTIQUES pour ordinateur

Disponible pour Windows 7, 8 et 10

Guide de l'utilisateur

Auteure : Aniabet Alfonso Valdés

Université Centrale "Marta Abreu" de Las Villas

À propos de l'application :

La Banque des expressions linguistiques est une application destinée aux étudiants de la 2^e jusqu'à la 5^e année de la Licence ès Langue Anglaise, option Langue Française pour mener à bien leurs projets ou travaux de recherche en langue française. Les utilisateurs peuvent y trouver certaines des expressions les plus fréquemment utilisées dans les différents mouvements qui structurent l'introduction de travaux de recherche dans les domaines de la linguistique et de la didactique. Son objectif est de faciliter la rédaction de l'introduction des travaux de recherche en français.

Note explicative :

Lors de la création de l'application Banque des expressions linguistiques, le programme GraphDB a été utilisé pour valider l'exactitude des relations établies dans le Protégé avec la création de la banque des expressions. Le GraphDB a pu créer des classes de contexte, c'est-à-dire trouver des textes sur Internet qui ont suivi les mouvements rhétoriques établis lors de la programmation de la banque, une fois que le projet de l'application a été introduit dans Protégé.

Ce manuel fournit une description détaillée de l'utilisation de l'application Banque des expressions linguistiques.

- La section 1 traite du début dans l'application (étape 1-7).
- La section 2 traite de la recherche dans l'application (étape 8-17).

Section 1 :

1. Allumez votre ordinateur et copiez une version de l'application « Protégé » (Fig.1) et la application « Banque des expressions linguistiques » (Fig. 2) via une clé USB provenant d'une source donnée (un autre ordinateur ou le réseau).

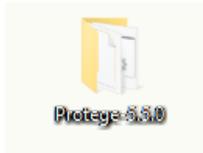


Fig.1 Fichier contenant l'application Protégé (version utilisée : 5.5.0)



Fig. 2 Application Banque des expressions linguistiques

2. Une fois que vous avez copié l'application, ouvrez-la



Fig. 3 Application Protégé

3. Après ouvrir l'application Protégé, cherchez dans la barre supérieure l'icône « File » et cliquez sur « Open... » (Fig. 4)

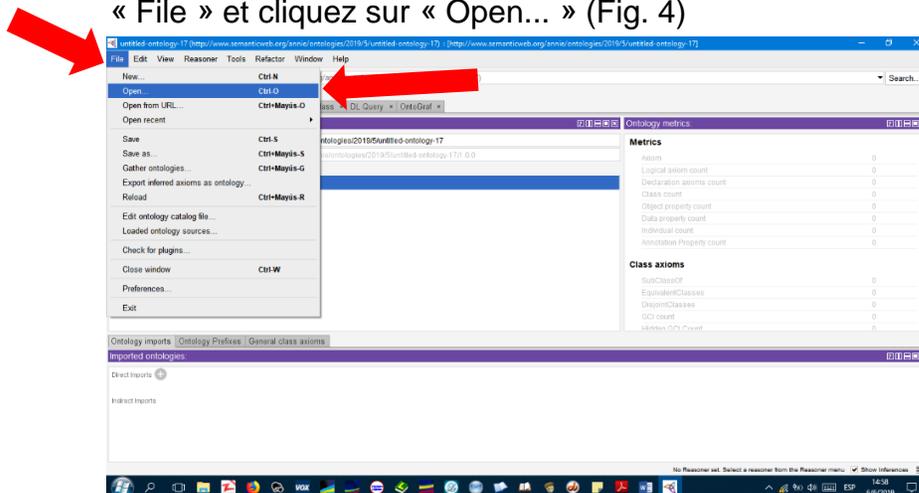


Fig. 4

4. Dans la fenêtre qui apparaît, cherchez l'endroit où vous avez précédemment copié l'application « Banque des expressions linguistiques » qui apparaîtra comme un fichier *.owl (Fig. 5)

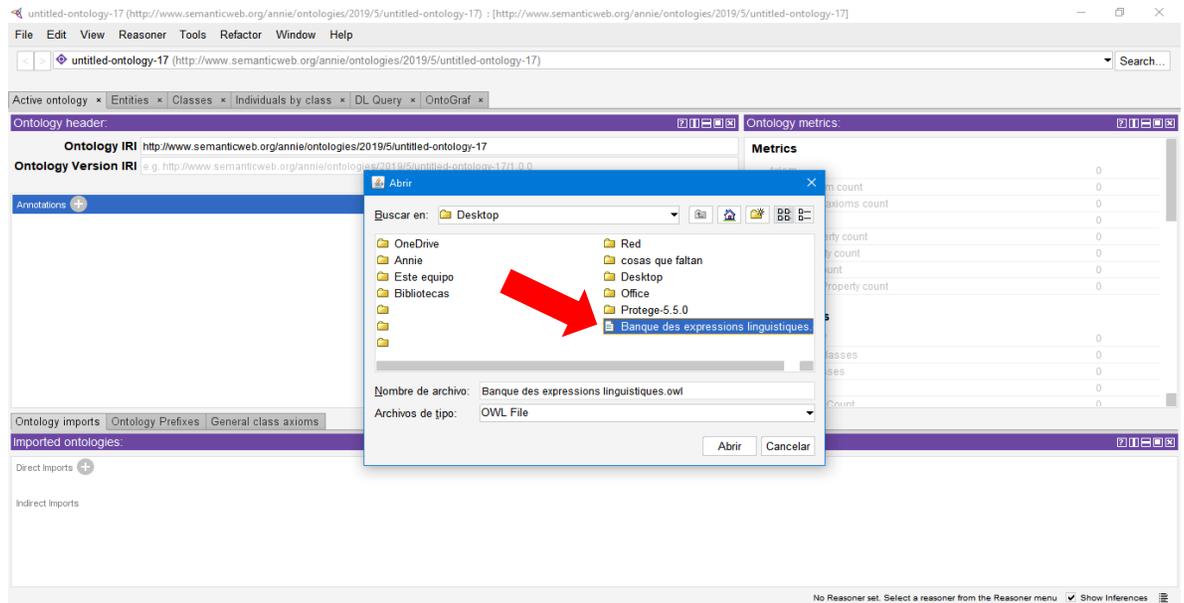


Fig. 5

5. Une fois que vous ouvrez l'application, vous verrez une fenêtre presque identique à celle qui apparaît lorsque vous ouvrez « Protégé » pour la première fois. Pour vous assurer que l'application « Banque des expressions linguistiques » a ouvert, les valeurs « Ontology metrics » se trouvant sur le côté droit de l'écran ne doivent pas être « 0 ». (Fig. 6)

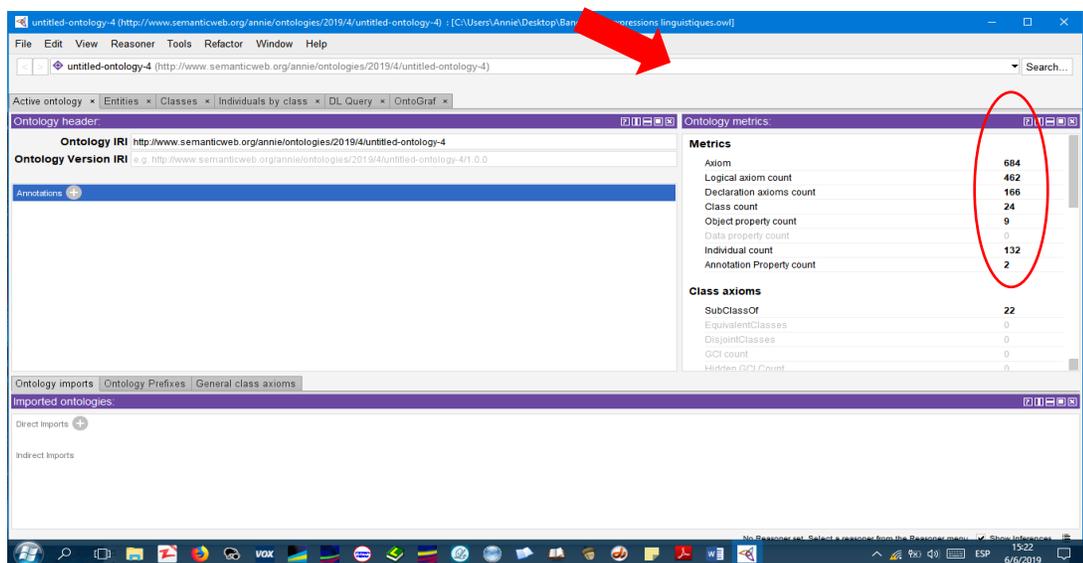


Fig. 6

6. À gauche, vous trouverez une série d'onglets, appuyez sur le dernier, appelé « OntoGraph ». (Fig. 7)

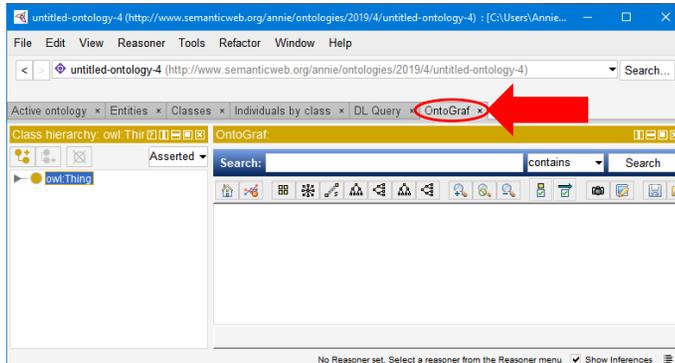


Fig. 7

7. Une fois que vous avez atteint cette image, vous vous trouvez déjà face à la banque des expressions linguistiques. Tout d'abord, vous devez savoir que dans notre banque des expressions, nous avons représenté les éléments qui caractérisent les mouvements rhétoriques, à savoir les mouvements rhétoriques organisés et les étapes qui les composent. En outre, les relations entre les instances grammaticales (nom, adjectif, verbe, adverbe, etc.) et la rhétorique où elles sont utilisées ont été représentées.

Section 2 :

8. Vous pouvez effectuer la recherche de deux manières différentes:
Option 1: Écrivez dans l'espace situé à droite de « Search » ce que vous souhaitez chercher. Par exemple, « Mouvement_1 » (si vous voulez rechercher plusieurs mots, vous devez entrer un trait d'union-sous « _ » entre chaque terme recherché). (Fig. 8)

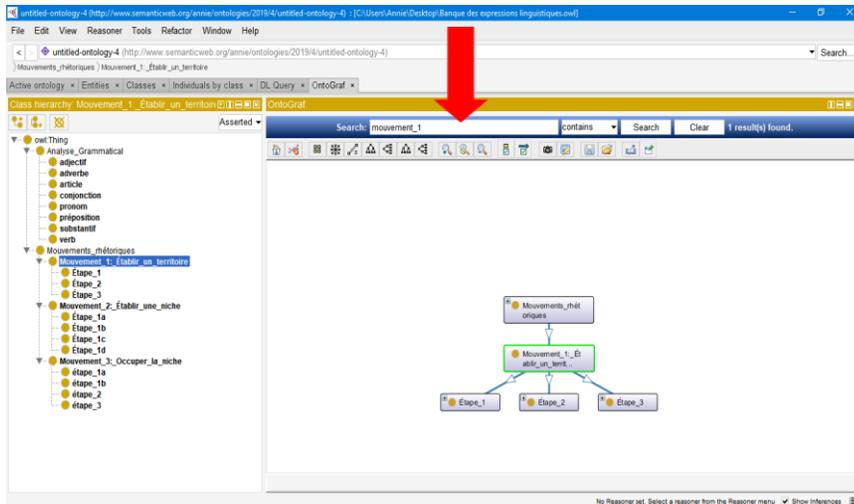


Fig. 8

9. Une fois que vous avez écrit le ou les mots que vous souhaitez rechercher, un diagramme apparaît à l'écran. Il comprend l'élément à chercher contenu dans un rectangle avec une bordure verte, la classe à laquelle cet élément appartient (dans ce cas, « Mouvements_rhétoriques ») et les classes qui le composent (dans ce cas « Étape_1 », « Étape_2 », « Étape_3 »).

10. Dans la barre qui apparaît sous l'espace pour la recherche, vous pouvez modifier la disposition du diagramme, augmenter ou réduire sa taille, parmi d'autres options. (Fig.9)

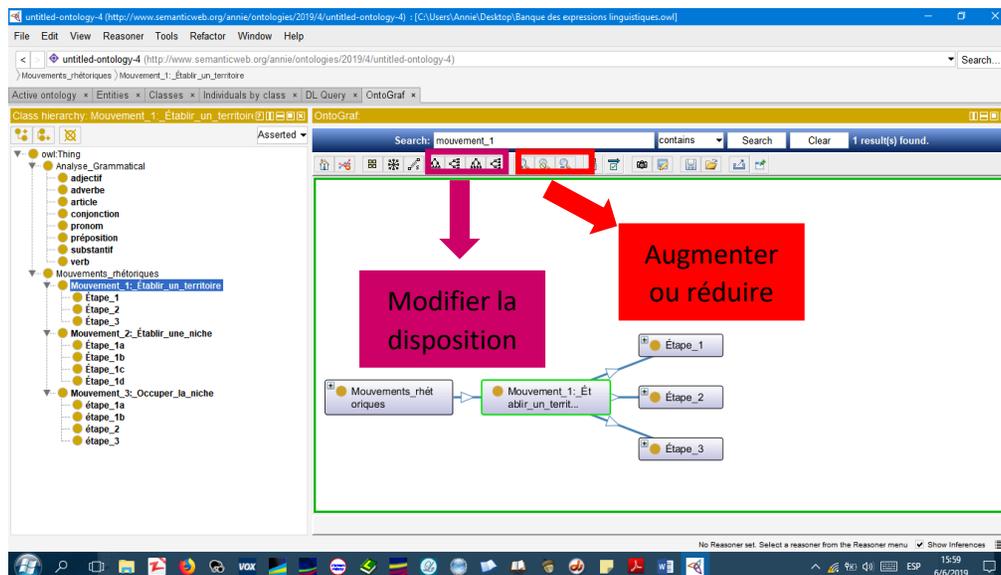


Fig. 9

11. Lorsque vous placez le curseur sur les étapes, un signe jaune apparaît, indiquant des caractéristiques générales de l'étape, des exemples concrets des expressions linguistiques et des fragments contenant les exemples. (Fig.10)

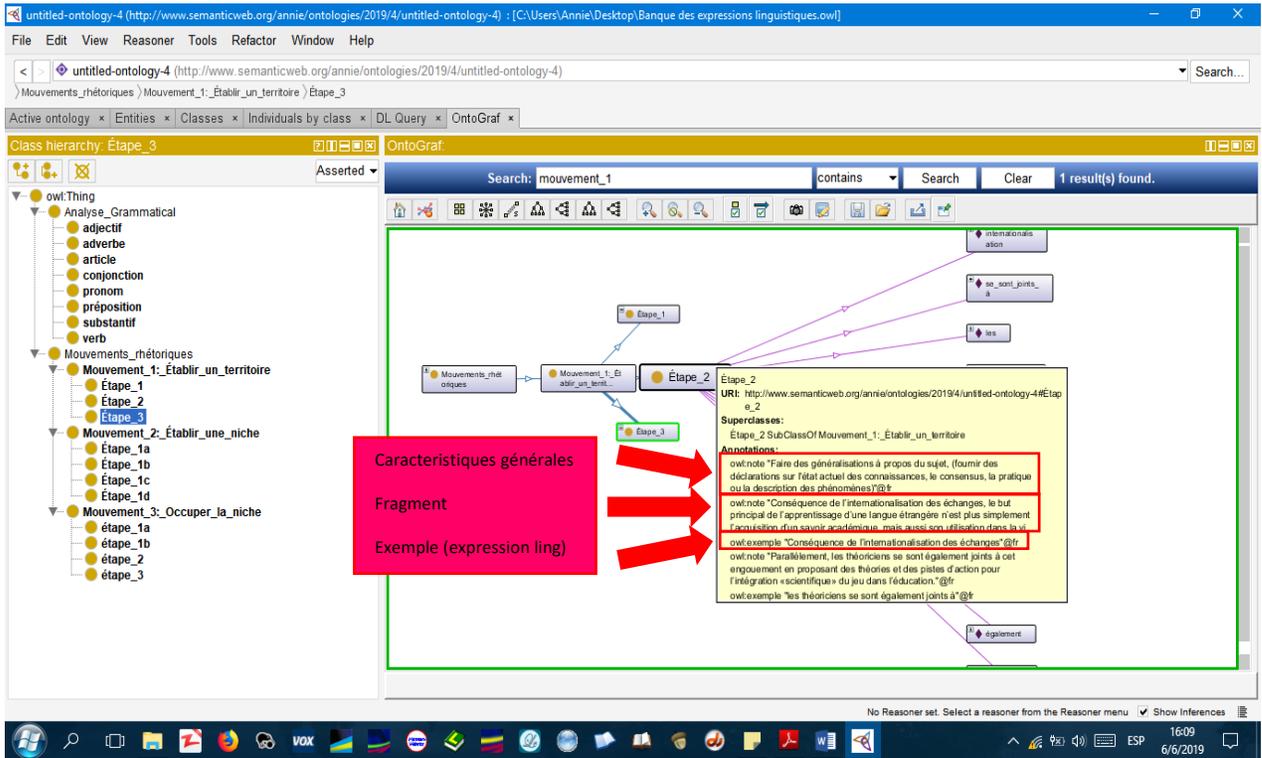


Fig. 10

12. Lorsque vous cliquez sur une étape, les exemples spécifiques et chacune des instances grammaticales qui la composent apparaissent. Lorsque vous placez le curseur sur une instance, sa classification et d'autres mots appartenant à la même classification apparaissent. (Fig. 11)

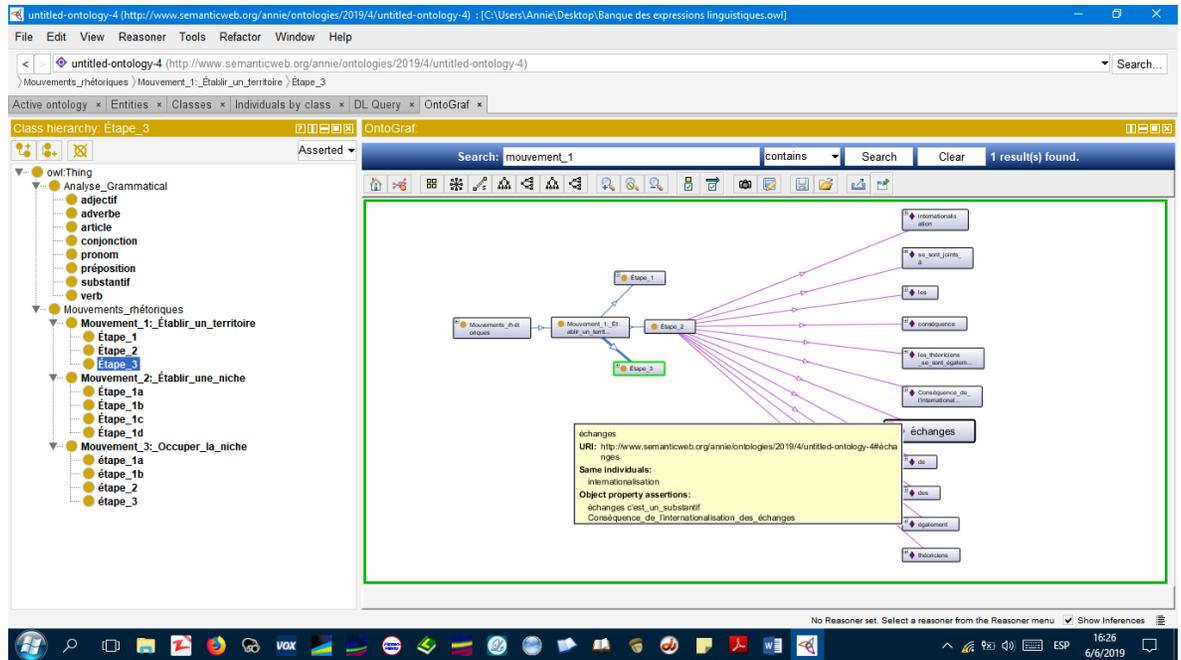


Fig. 11

13. Option 2: Faites défiler les menus qui apparaissent sous « owl: Thing » en cliquant dessus. (Fig. 12)

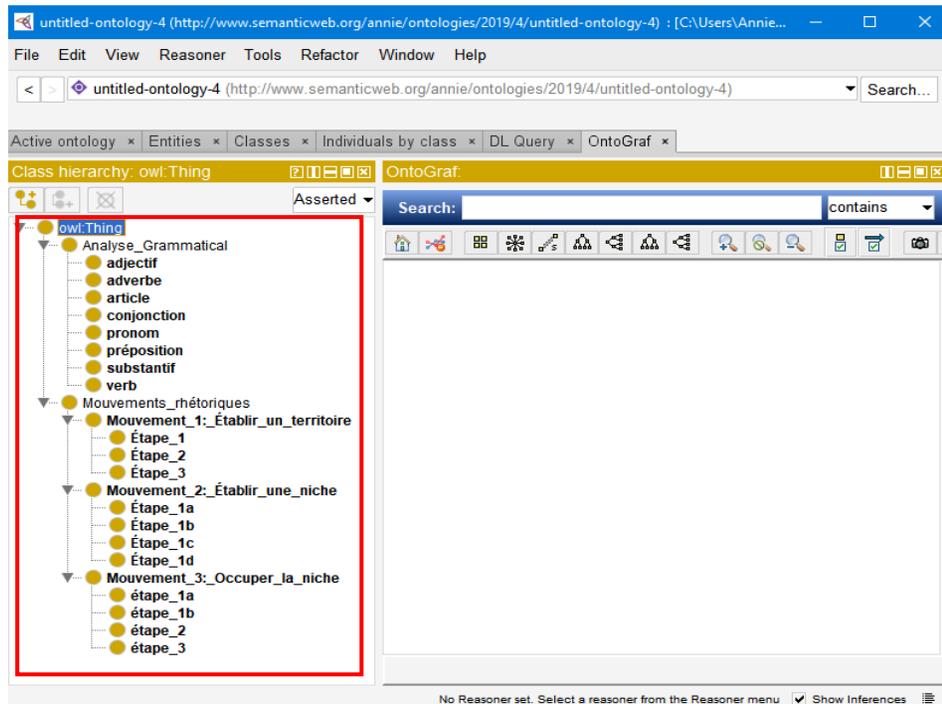


Fig. 12

14. Pour que ces menus apparaissent sous forme de diagramme, vous devez double-cliquer sur le bouton jaune  qui apparaît à gauche de chacun des éléments.
15. La flèche discontinue  indique la phrase à laquelle appartient l'instance. La flèche violette  indique que l'instance à laquelle pointe la flèche appartient à l'élément d'où provient cette flèche. La flèche bleue  établit les relations entre les éléments les plus généraux.
16. Une fois que les éléments apparaissent sous forme de diagramme, les étapes 10, 11 et 12 s'appliquent quelle que soit l'option choisie pour effectuer la recherche.
17. Les relations qui peuvent être établies avec l'application sont illustrées à la Fig. 13.

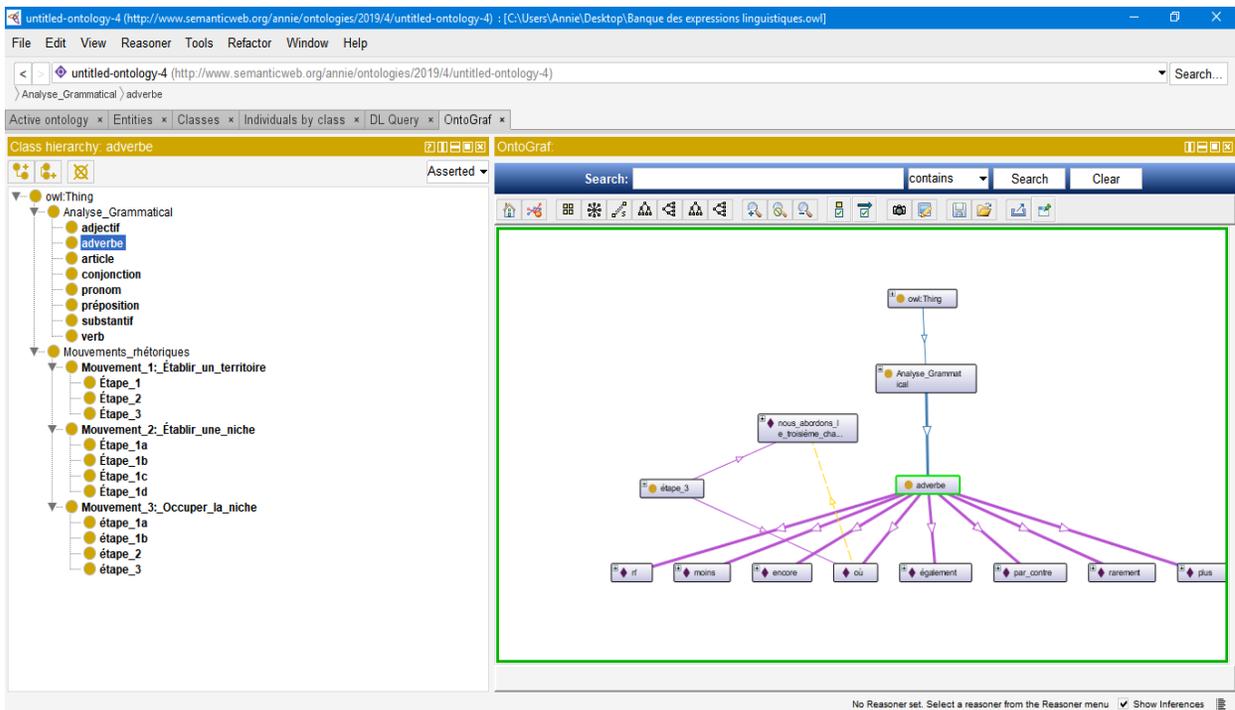


Fig. 13

Annexe V

Guía del usuario del banco de expresiones lingüísticas

Aplicación BANCO DE EXPRESIONES LINGÜÍSTICAS para computadora

Disponible para Windows 7, 8, y 10

Guía del usuario

Autora: Aniabet Alfonso Valdés

Universidad Central "Marta Abreu" de Las Villas

Sobre la aplicación

El **Banco de expresiones lingüísticas** es una aplicación diseñada para los estudiantes de segundo a quinto año de la Licenciatura en Lengua Inglesa con segunda lengua: Francés, que realicen su proyecto de investigación o trabajo de curso en la lengua francesa. Los usuarios pueden encontrar en ella, algunas de las expresiones más frecuentes utilizadas en las diferentes movidas que estructuran la introducción de los trabajos de curso del campo de la lingüística y la didáctica. El objetivo de esta aplicación es facilitar la redacción de la introducción de los trabajos de investigación en francés.

Nota : Cuando se construyó la aplicación **Banco de expresiones lingüísticas**, se utilizó el programa GraphDB para validar que las relaciones establecidas en el Protégé al crear el banco fueran correctas. El GraphDB fue capaz de construir clases de contexto, o sea, encontrar textos en internet que seguían las movidas retóricas establecidas durante la programación del banco, una vez introducido el proyecto de la aplicación en Protégé.

Este manual provee una descripción detallada para la utilización de la aplicación **Banco de expresiones lingüísticas**.

La sección 1 trata sobre el comienzo en la aplicación (Paso 1-7).

La sección 2 trata sobre la búsqueda en la aplicación (Paso 8-17).

Sección 1:

1. Encienda su computadora y copie cualquier versión de la aplicación Protégé (Fig.1) y la aplicación «Banco de expresiones lingüísticas» (Fig. 2) a través de un dispositivo USB desde una fuente dada (otra computadora o la red).

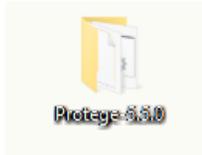


Fig.1 Carpeta que contiene la aplicación Protégé (versión utilizada : 5.5.0)



Banque des
expressions linguistiques

Fig. 2 Aplicación Banque de expresiones linguistiques

2. Una vez que la aplicación ha sido copiada, ábrala.



Fig. 1. Aplicación Protégé

3. Después de abrir la aplicación Protégé, busque en la barra superior el ícono « File » y allí presione « Open... » (Fig. 4)

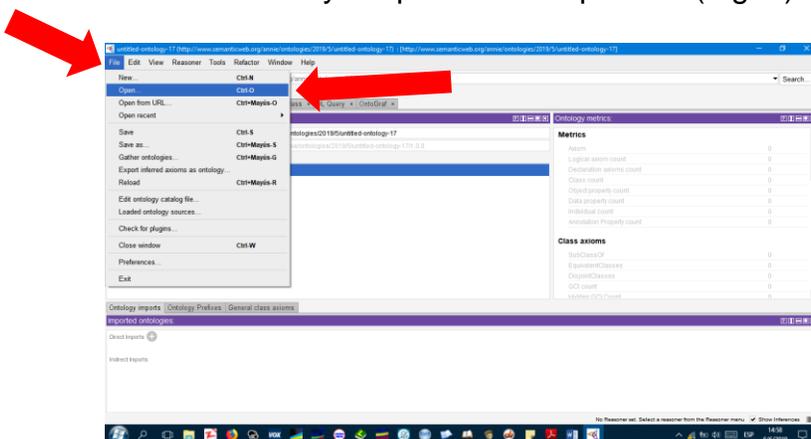


Fig. 4.

4. En la ventana que aparece busque el lugar donde copió previamente la aplicación « Banco de expresiones lingüísticas » la cual es un fichero *.owl (Fig.5)

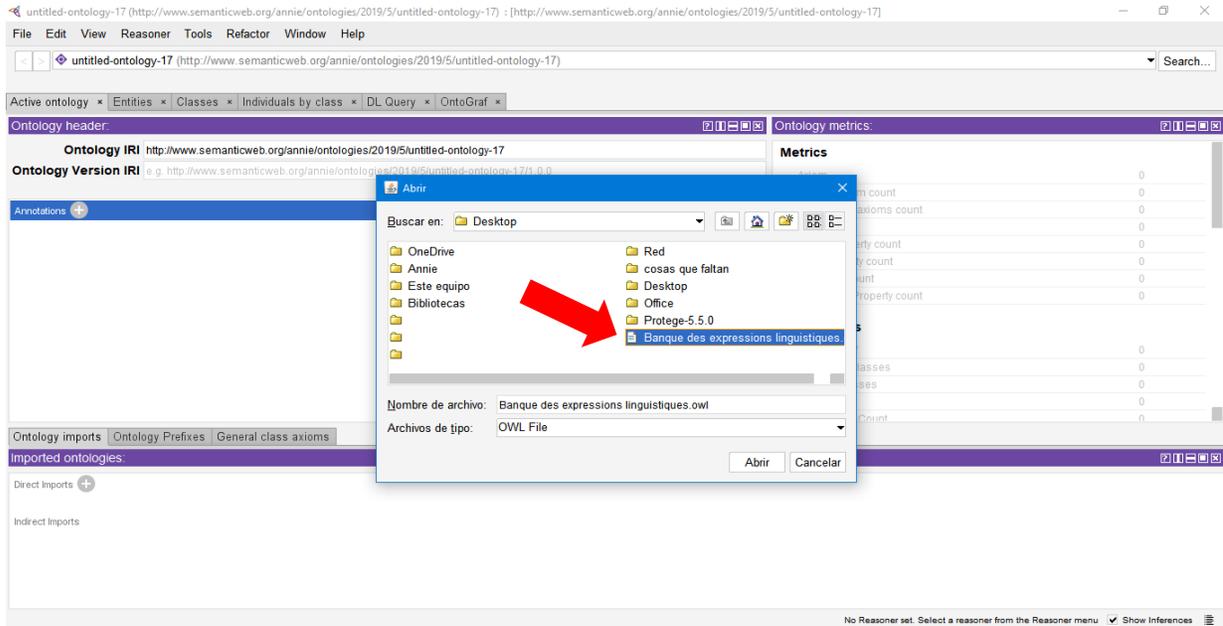


Fig. 5

5. Una vez que abra la aplicación le aparecerá una ventana casi idéntica a la que le sale al abrir « Protégé » por primera vez. Para tener la certeza de que la aplicación del banco de expresiones se abrió, los valores « *Ontology metrics* » que aparecen en la parte derecha de la pantalla no deben ser «0». (Fig. 6)

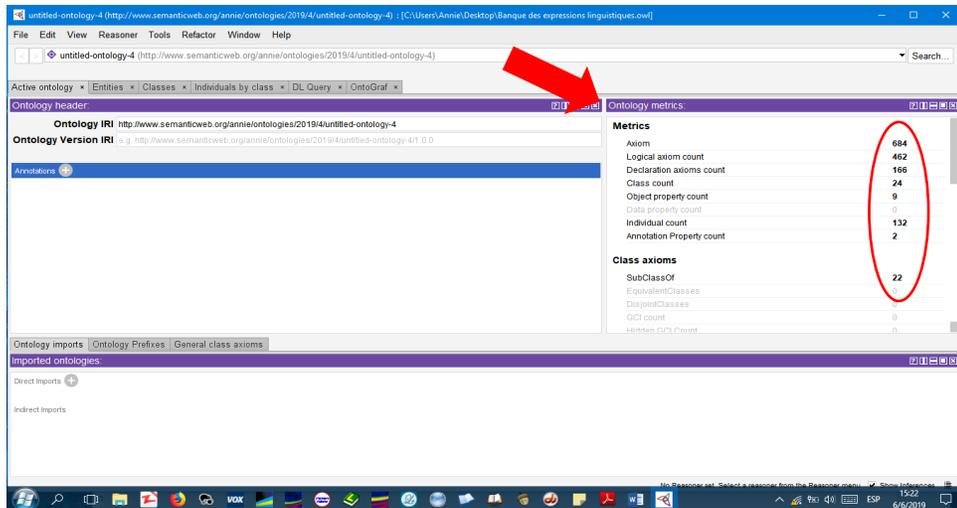


Fig. 6

6. A la izquierda se encuentra una serie de pestañas. Presione la última, esta se llama « OntoGraph ». (Fig. 7)

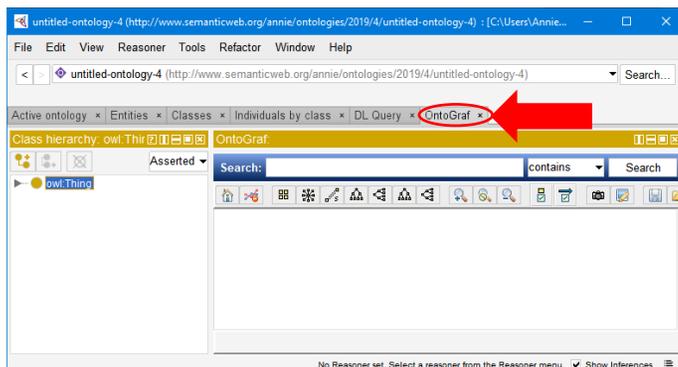


Fig. 7

7. Una vez que haya llegado a esta imagen, ya se encuentra frente al banco de expresiones lingüísticas. Primero que todo debe saber que en nuestro banco de expresiones hemos representado los elementos que caracterizan las movidas retóricas, o sea, las movidas retóricas organizadas y las etapas que las componen. Además, se representaron las relaciones entre las instancias gramaticales (sustantivo, adjetivo, verbo, adverbio, etc.) y la frase retórica donde se utilizan.

Sección 2:

8. Para realizar la búsqueda puede hacerlo de dos formas diferentes:

Opción 1: Escribiendo en el espacio a la derecha de « Search » lo que desee buscar. Por ejemplo « Mouvement_1 » (cuando busque más de una palabra debe escribir guión bajo « _ » entre cada término de la búsqueda). (Fig. 8)

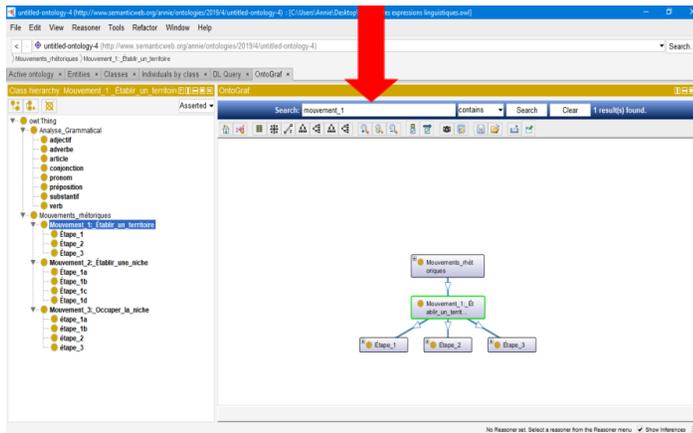


Fig. 8

9. Tras haber escrito la palabra o palabras que desea buscar se le creará en la pantalla un diagrama que incluye al elemento dentro de un rectángulo con un borde verde, la clase a la que pertenece (en este caso « Mouvements_rétoriques ») y las clases que lo conforman (en este caso « Étape_1 », « Étape_2 », « Étape_3 »).
10. En la barra que aparece debajo del espacio para la búsqueda se puede cambiar la distribución del diagrama, aumentar o disminuir su tamaño, entre otras opciones. (Fig. 9)

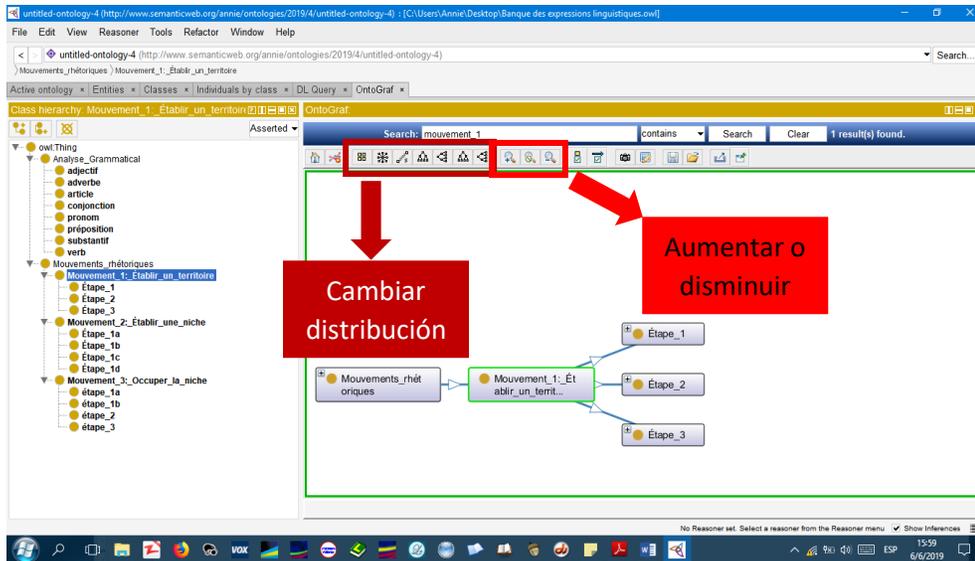


Fig. 9

11. Al ubicar el cursor sobre las etapas aparecerá un cartel amarillo indicando las características generales de la etapa, ejemplos concretos de expresiones lingüísticas y fragmentos donde aparecen los ejemplos. (Fig.10)

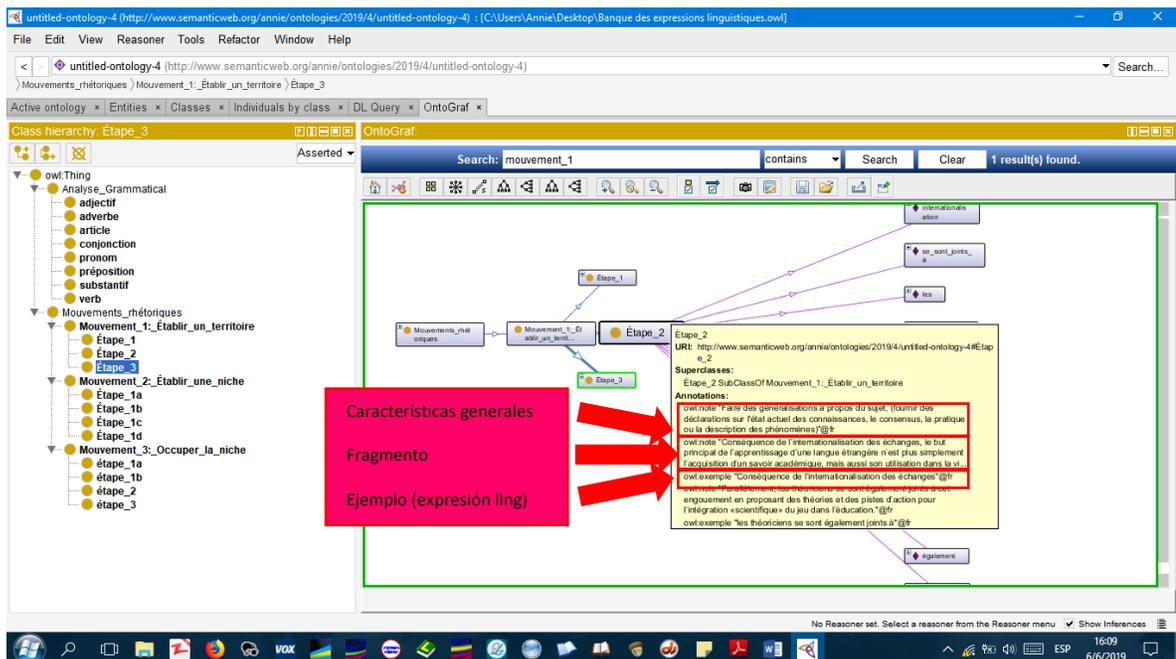


Fig. 10

12. Al dar clic sobre cualquier etapa aparecerán los ejemplos específicos y cada una de las instancias gramaticales que lo componen. Cuando ubica el

cursor sobre una instancia aparecen su clasificación y otras palabras que pertenecen a la misma clasificación. (Fig. 11)

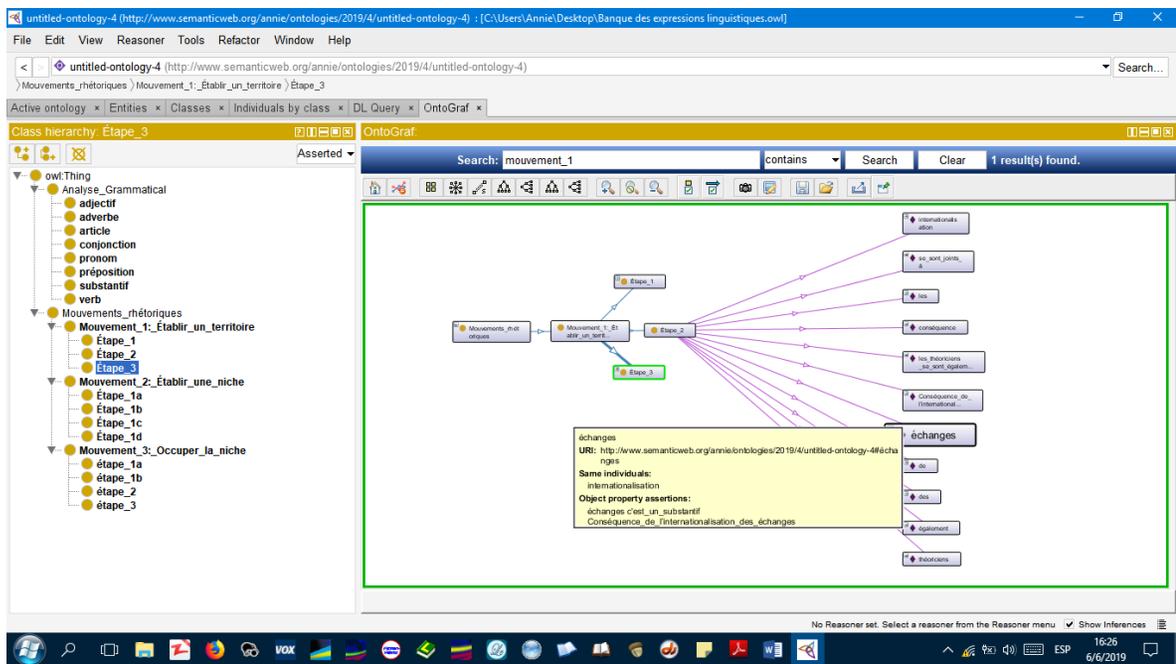


Fig. 11

13. **Opción 2:** Desplazando los menús que aparecen debajo de « owl: Thing » al clicar sobre ellos. (Fig.12)

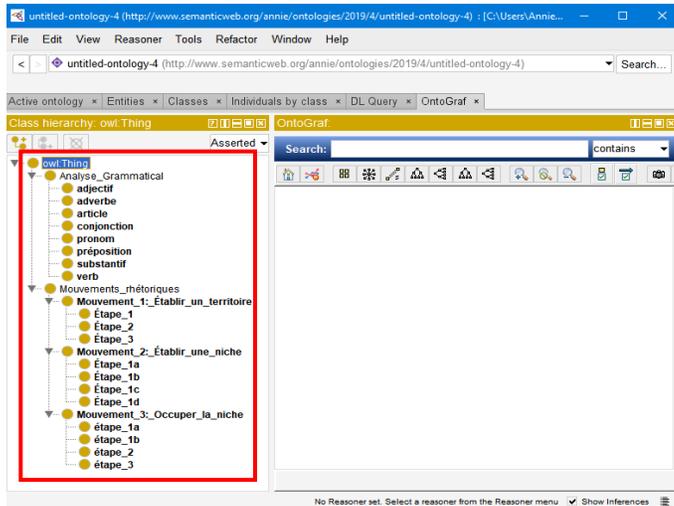


Fig. 12

14. Para que estos menús aparezcan representados en forma de diagrama, debe dar doble clic sobre el botón amarillo  que aparece a la izquierda de cada uno de los elementos.

15. La flecha discontinua  indica la frase a la que pertenece la instancia. La flecha violeta  indica que la instancia a la que apunta la saeta pertenece al elemento de donde sale dicha flecha. La flecha azul  establece las relaciones entre los elementos más generales.
16. Una vez que los elementos aparezcan en forma de diagrama, los pasos 10, 11 y 12 son aplicables sin importar la opción elegida para hacer la búsqueda.
17. En la Fig. 13 se muestran las relaciones que se pueden establecer con la aplicación.

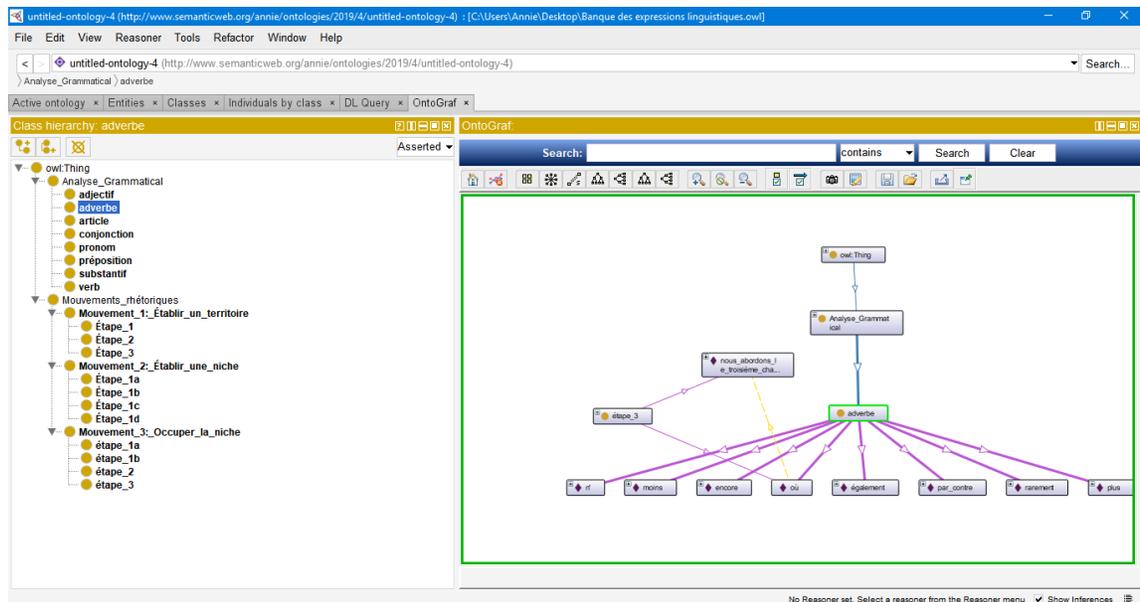


Fig. 13