



HAL
open science

Le statut du verbe en langue de spécialité: application au domaine du forage en français

Kamila Jamrz

► **To cite this version:**

Kamila Jamrz. Le statut du verbe en langue de spécialité: application au domaine du forage en français. Sciences de l'Homme et Société. 2018. dumas-01901889

HAL Id: dumas-01901889

<https://dumas.ccsd.cnrs.fr/dumas-01901889>

Submitted on 28 Jan 2019

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

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



Le statut du verbe en langue de spécialité : application au domaine du forage en français

**JAMRÓZ
Kamila Magdalena**

Sous la direction de Madame Agnès Tutin et Madame Cécile Frérot

Laboratoire : LIDILEM

UFR LLASIC
Département des Sciences du Langage et Français Langue Étrangère
Section Sciences du Langage

Mémoire de master 2 recherche - 30 crédits

Parcours : Linguistique

Année universitaire 2017-2018



Le statut du verbe en langue de spécialité : application au domaine du forage en français

JAMRÓZ
Kamila Magdalena

Sous la direction de Madame Agnès Tutin et Madame Cécile Frérot

Laboratoire : LIDILEM

UFR LLASIC
Département des Sciences du Langage et Français Langue Étrangère
Section Sciences du Langage

Mémoire de master 2 recherche - 30 crédits

Parcours : Linguistique, orientation recherche

Année universitaire 2017-2018

Remerciements

Dans un premier temps, je tiens à remercier infiniment deux personnes, Madame Agnès Tutin, ma directrice de mémoire et Madame Cecile Frérot, ma codirectrice de mémoire, pour m'avoir encouragée, guidée et conseillée tout au long de mon travail et grâce à qui j'ai pu associer la linguistique et la terminologie.

Ensuite, je souhaiterais remercier vivement Madame Małgorzata Niziołek de l'Université Pédagogique de Cracovie qui m'a encouragée à participer au programme du double diplôme de Linguistique appliquée de l'Université de Grenoble Alpes et de l'Université Pédagogique de Cracovie.

Finalement, j'adresse mes plus sincères remerciements à ma famille, en particulier à mon père qui m'a transmis l'amour des langues et qui a été mon plus grand soutien. Dziękuję Papo.



DÉCLARATION

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

NOM : **JAMRÓZ**

PRENOM : **KAMILA**

DATE : **4/09/2018**

Sommaire

Partie 1 - L'ÉTAT DE L'ART.....	8
CHAPITRE 1. TERMINOLOGIE.....	9
1. DIFFERENTES APPROCHES DE LA TERMINOLOGIE	10
CHAPITRE 2. VERBE EN TERMINOLOGIE.....	18
2. DIFFERENTES FAÇONS D'ANALYSER UN VERBE	18
CHAPITRE 3. STRUCTURES VERBALES EN LINGUISTIQUE – LES PATRONS LEXICO-SYNTAXIQUES.....	24
3. CONSTRUCTIONS DANS LA LINGUISTIQUE FRANÇAISE	24
4. CONSTRUCTIONS DANS LA LINGUISTIQUE ANGLAISE.....	28
Partie 2 - CORPUS ET MÉTHODOLOGIE	31
CHAPITRE 4. PROBLEMATIQUE	32
CHAPITRE 5. CORPUS.....	33
1. CORPUS UTILISE	33
2. OUTILS D'EXPLORATION DE CORPUS.....	38
CHAPITRE 6. METHODOLOGIE	46
1. EXTRACTION DES CANDIDATS-TERMES	46
2. SELECTION DES VERBES A ANALYSER	47
3. ANALYSE DES CONTEXTES DES VERBES CHOISIS	48
Partie 3 - ANALYSES ET RÉSULTATS	52
CHAPITRE 7. L'ONTOLOGIE.....	53
1. COLLECTION DES ARGUMENTS	53
2. CREATION DES CLASSES ET DES SOUS-CLASSES	53
CHAPITRE 8. FORMES SYNTAXIQUES SPECIFIQUES AU CORPUS DE FORAGE.....	58
1. CONSTRUCTIONS ACTIVES.....	58
2. CONSTRUCTIONS PASSIVES ET MOYENNES	61
3. CONSTRUCTION SPECIFIQUE : NOM A VERBE.....	62
4. CONSTRUCTION EN « POUR » – COMPLEMENT CIRCONSTANCIEL	63
5. PATRONS DES VERBES EXTRAITS DU CORPUS DU FORAGE.....	64
CHAPITRE 9. STATUT TERMINOLOGIQUE DES VERBES ANALYSES	66
1. MODELE DESCRIPTIF DE LA STRUCTURE ARGUMENTALE.....	67
2. RESEAU LEXICO-TERMINOLOGIQUE.....	71
Conclusion.....	75
Bibliographie.....	77
Bibliographie de corpus.....	80
Table des illustrations.....	81
Table des annexes.....	82
Table des matières.....	103

Introduction

Depuis plus de vingt ans, on observe l'importance croissante du verbe dans les travaux terminologiques. L'utilisation de la linguistique de corpus par la terminologie textuelle a permis aux terminologues de répondre à de nouveaux besoins liés à la mondialisation des échanges entre les chercheurs, spécialistes des domaines et entreprises, mais aussi d'analyser la validité de la théorie classique de la terminologie, selon laquelle le verbe ne peut pas être considéré comme un terme. En même temps, les verbes jouent un rôle fondamental dans la description des processus et c'est pourquoi ils gagnent en importance. De plus, aujourd'hui, les utilisateurs des dictionnaires, des glossaires, des bases lexicographiques ont besoin de connaître non seulement la signification d'un mot, mais surtout la façon dont ce mot peut être utilisé dans une phrase et c'est grâce aux verbes, le noyau d'une phrase, qu'on est capable de répondre à ces besoins.

Ce mémoire a pour objectif l'observation, l'analyse et la description des verbes provenant du corpus de forage français afin de trouver les spécificités de ces verbes présents dans le corpus et aussi de proposer leur codage. Afin d'atteindre cet objectif, nous nous concentrerons sur le statut terminologique des verbes et sur les divers modèles de description des verbes existants. Le but de cette analyse est de créer un tableau méthodologique qui simplifiera la recherche des équivalents de ces verbes et de leurs arguments dans la langue polonaise et d'autres langues et qui pourrait servir de base pour créer un dictionnaire terminologique multilingue du domaine du forage.

Étant donné que ce travail combine les méthodes linguistiques et terminologiques, nous commencerons par voir quelle est la place du verbe dans la terminologie et comment et pourquoi elle a changé. Nous essayerons d'établir quelles approches terminologiques nous sont les plus proches et lesquelles nous utiliserons dans notre travail. Du côté linguistique, nous nous concentrerons sur les spécificités des verbes sur le plan syntaxique et sur l'influence de la syntaxe sur la création des significations.

Dans ce mémoire, en vue de résoudre ces problématiques, nous nous servirons des méthodes de la linguistique de corpus. Nous avons décidé d'employer l'approche mixte de *corpus-based* (elle sert à confirmer ou à infirmer les hypothèses formulées au début du travail) et de *corpus-driven* (l'approche dans laquelle on tire des conclusions après avoir observé les données linguistiques). Nous avons choisi aussi de combiner les approches quantitative et qualitative. Nous voulons avoir la possibilité de voir les tendances

statistiques qui apparaissent dans notre corpus, mais en même temps, nous sommes conscients que le corpus que nous avons créé et qui compte 274 809 mots n'est pas suffisamment fiable pour être guidé par l'approche quantitative et les chiffres. Le corpus que nous utiliserons a été construit à la base des textes semi-spécialisés provenant du domaine du forage. Ceux-ci ont été choisis parce que, dans la plupart des cas, ils sont dédiés aux apprenants et par conséquent, ils sont riches en termes et leurs explications. Les verbes seront extraits à l'aide de TermoStat (Drouin, 2003), l'outil d'extraction des candidats-termes. Ensuite, nous trouverons une façon pour les modéliser et pour décrire leur statut de spécialité.

Le choix de ce sujet s'explique par l'importance de la langue française en Pologne. Ces dernières années, les employeurs, y compris les grandes compagnies pétrolières telles que Lotos ou Schlumberger, recrutent des personnes ayant une bonne connaissance (surtout celle du français technique).

Notre travail est divisé en trois grandes parties. La première partie présente différentes approches terminologiques, la façon dont elles traitent le verbe, mais aussi la description des structures verbales en linguistique et l'explication de la notion des patrons.

La deuxième partie concerne la problématique de notre analyse et la méthodologie choisie pour la résoudre. Cette partie rapprochera également les étapes de la constitution de notre corpus.

La troisième partie sera consacrée à l'analyse des verbes choisis et aux résultats de cette analyse. Nous essayerons de décrire les formes syntaxiques spécifiques au corpus de forage. Pour pouvoir le faire, nous construirons une ontologie - base structurée des arguments des verbes. Finalement, nous tenterons d'établir le statut terminologique de nos verbes.

Partie 1

-

L'ÉTAT DE L'ART

Chapitre 1. Terminologie

Selon Dubuc (1997, in TLFi - *Trésor de la Langue Française informatisé*), la terminologie est un « art de repérer, d'analyser et, au besoin, de créer le vocabulaire pour une technique donnée, dans une situation concrète de fonctionnement de façon à répondre aux besoins d'expression de l'utilisateur ». Cela veut dire que la terminologie est une discipline qui se spécialise dans la description de la langue de spécialité.

Rey dans sa *Terminologie – noms et notions* (1979, p.58) explique que :

Les besoins de description relient la terminologie au discours didactique sur chaque domaine et donc à la science, à la technologie (...); ceux de diffusion aussi, ainsi qu'à la pédagogie, à la traduction; ceux de normalisation aux théories de la connaissance, à la sociolinguistique, et encore à la traduction et à la lexicographie.

Autrement dit, la terminologie est reliée à chaque domaine de la science et peut être appliquée de façons différentes. Elle est surtout utilisée dans (Pavel&Nolet, 2001, p.XVII) :

- la lexicographie spécialisée afin de créer des dictionnaires terminologiques ;
- la traduction spécialisée qui exige la connaissance des terminologies spécialisées bilingues ou même multilingues ;
- la rédaction technique ;
- l'enseignement des langues spécialisées et des traductions.

L'activité terminologique repose donc sur la capacité de repérer les termes qui désignent les concepts propres à un domaine et « d'en attester l'emploi à l'aide de références précises, de les décrire brièvement en discernant le bon usage de l'usage erroné, et de recommander ou de déconseiller certains usages afin de faciliter une communication sans ambiguïté » (Pavel&Nolet, 2001, p.XVIII).

Dans notre mémoire, nous voulons nous concentrer sur l'approche terminologique des verbes et l'utiliser ensuite pour analyser les verbes provenant du corpus de forage. Dans ce cadre, cette partie de notre travail sera consacrée à différents modèles terminologiques, à leur approche des verbes et aux modèles linguistiques utilisés pour décrire les structures verbales.

1. Différentes approches de la terminologie

Selon Rey (1979, p. 3-4), la terminologie est une étude ancienne, mais en tant que discipline elle est née au XVIII^e siècle en Allemagne où apparaît la notion de terme, une unité lexicale dont le sens est envisagé par rapport à son domaine de spécialité (L'Homme, 2004, p.22), évoqué par Christian Gottfried Schütz. Cabré (1998, p. 21-22) ajoute que l'émergence de cette discipline est indissociable des travaux de Lavoisier et Berthollet en chimie et ceux de Linné en botanique et en zoologie et résulte de la nécessité d'établir des règles de formation des termes dans les différentes disciplines. Mais en fait, c'est au XX^e siècle où elle a connu un développement rapide et systématique. En Europe, tout a commencé dans les années 30 avec les travaux d'Eugen Wüster et ce qu'on appelle la théorie classique de terminologie. Cabré (2005) considère Wüster comme le fondateur de la théorie moderne de la terminologie.

1.1. Théorie Générale de la Terminologie (TGT)

Cabré (2012) estime que la terminologie telle que nous la connaissons aujourd'hui s'est développée, entre autres, grâce à Eugen Wüster et à la Théorie Générale de la Terminologie (TGT), théorie qui a ses racines dans le *Dictionnaire de la machine-outil* qu'il élabore en 1968 pour faciliter la communication internationale entre professionnels (Frérot, à venir). Dans son doctorat, Wüster justifie l'application de l'approche conceptuelle, en plaçant au centre de son analyse le concept, l'objet et sa désignation (figure 1). La théorie de Wüster s'inscrit dans les sciences appliquées et elle a été créée pour éliminer des ambiguïtés de la communication scientifique et technique.

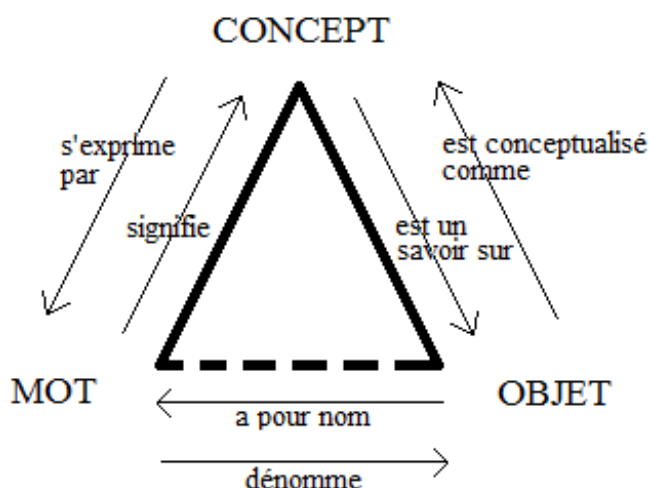


Figure 1. Triangle sémiotique d'Odgen et Richards.

Cabré (2012) montre que dans le dictionnaire sur la machine-outil de Wüster (*The Machine Tool, an Interlingual Dictionary of Basic Concepts*, 1968) on peut observer les objectifs principaux de sa théorie :

- la terminologie sert de transfert de connaissances spécialisées ;
- la terminologie normalise des concepts et des termes visant à garantir une communication sans problème ;
- la terminologie est une discipline sur les objets, les concepts et leurs désignations ;
- la terminologie s'intéresse aux termes uniquement à l'intérieur d'un domaine ;
- la terminologie est à la fois théorique et pratique.

Pour Wüster, la fonction première de la terminologie est de créer et de normaliser les dénominations des concepts. Son ouvrage a inspiré la Théorie Générale de Terminologie, définitoire des principes de base pour l'établissement et la description des données terminologiques. Bien que la Théorie Générale ait été développée par les successeurs de Wüster, elle n'a pas subi de modifications majeures. Par ailleurs, Wüster considère la terminologie et la linguistique très éloignées l'une de l'autre en raison de différences dans leurs fondements mêmes ; la terminologie est synchronique et restreinte au lexique alors que la linguistique s'intéresse à la systémativité de la langue et à son évolution et s'occupe du lexique, de la syntaxe et de la phonologie.

Le fondement de la TGT est la démarche onomasiologique. Le point de départ de cette démarche est un concept (représentation mentale d'un objet du monde réel) qui est

ensuite expliqué au moyen d'une définition (un mot). Comme l'explique et l'illustre L'Homme (2004, p. 26) « une fois fixé, ce concept fédère toutes les formes linguistiques utilisées pour le nommer ».

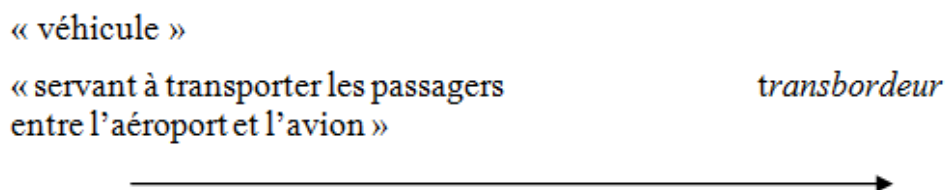


Figure 2 Le fonctionnement de la démarche onomasiologique (L'Homme 2004, p. 26)

La TGT met aussi l'accent sur l'importance du système conceptuel. L'optique conceptuelle considère que les concepts et leurs nominalisations sont le reflet de l'organisation des connaissances dans un domaine. La démarche onomasiologique a pour objectif une normalisation des connaissances et par conséquent, privilégie la bi-univocité : à un concept est attribué un seul terme (une forme) et à un terme correspond un concept.

Malgré la bonne organisation de la méthode de Wüster et ses règles précises, elle ne correspond pas aux besoins réels de la terminologie. La vision conceptuelle de la communication spécialisée est extrêmement idéalisée, elle ne peut pas répondre aux besoins de l'usage et elle ne correspond pas aux observations des textes spécialisés (L'Homme 2004, p. 29). La polysémie et la synonymie sont omniprésentes dans les textes de spécialité, même le nom de discipline, comme le montre L'Homme dans son livre (2004, p. 31), est polysémique et possède trois sens différents (figure 3). Ces phénomènes sont présents dans les textes de toutes les langues et constituent un véritable défi pour les traducteurs et les rédacteurs techniques.

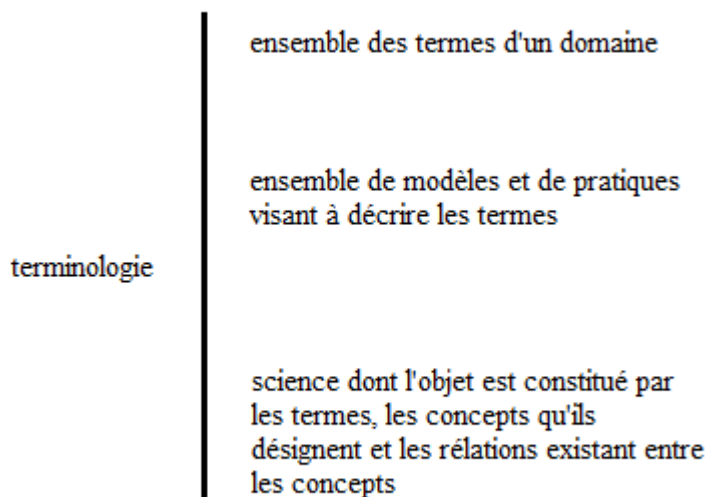


Figure 3 Polysémie de la *terminologie*.

En grande partie, le travail du terminographe¹ consiste à identifier les significations des termes qui existent déjà et non l'inverse. En ce qui concerne la normalisation, en aucun cas, elle ne reflète le fonctionnement réel des termes.

Toutefois, la situation a évolué dans les années 90. La terminologie a ouvert sa porte aux discussions théoriques et à d'autres domaines de connaissances : la sociologie, la communication et la linguistique ont ainsi donné naissance aux approches comme la socioterminologie (Gaudin 1993, 2003), le modèle sociocognitif (Temmerman 2000), la théorie communicative (Cabré 2000, 2003).

L'Homme (2004) mentionne par ailleurs que depuis quelques années, les modèles et les approches théoriques de la terminologie évoluent et s'écartent fortement de la terminologie classique pour proposer des modèles qui s'appuient sur l'analyse des textes et sur la linguistique de corpus.

1.2. Les nouvelles approches en terminologie

Les années 90 et le début de l'an 2000 constituent une période émergente des nouvelles approches en terminologie. Avec le développement d'Internet et de l'importance des échanges internationaux, la nécessité de gérer les informations pour produire, diffuser, rechercher, exploiter et traduire les documents a augmenté (Frérot, à venir). C'est la

¹ Un terminographe – personne qui décrit des termes dans les dictionnaires spécialisés ou les bases terminologiques (L'Homme 2004, p.21).

nécessité d'analyser les termes de façon plus réelle et de baser leurs descriptions sur leur emploi dans la communication qui a modifié l'approche classique de la Théorie Générale. Dans ces nouvelles approches, c'est le corpus qui se montre le plus important, donc le texte qui est un point de départ de l'analyse. De cette opposition aux principes wüstériens, plusieurs approches ont été nées : le modèle sociocognitif de Temmerman (2000), la socioterminologie de Gaudin (2003), la théorie communicative de Cabré (2003), la terminologie textuelle de Bourrigault et Slodzian (1999) et aussi l'approche lexicosémantique de L'Homme (2004).

1.2.1. Socioterminologie

Dans son article, Gaudin (1993) présente la socioterminologie comme une approche basée sur deux perspectives et deux contextes différents. L'une née en France avec les travaux de Louis Guilbert et son affirmation concernant la signification du terme qui « relève et de la rhétorique et de la grammaire et de la sociolinguistique » (1973, p. 21), d'Yves Gambier et de Louis Guespin. L'autre née au Québec, sous l'impulsion des linguistes tels que Pierre Auger, Jean-Claude Boulanger et Jean-Claude Corbeil. C'était l'analyse des mécanismes et de la réalité de l'usage des termes qui ont motivé la mise en oeuvre de l'application de l'approche sociolinguistique. Gaudin (2003, in Frérot, à venir) place le concept en discours où il est modifié en fonction de variables sociales et historiques.

1.2.2. Théorie communicative

En même temps, Maria Teresa Cabré (2003) travaille sur sa théorie communicative de la terminologie. Pour Cabré la connaissance spécialisée est présentée par des formes linguistiques appelées *unités de connaissance spécialisée*. Comme l'explique Frérot (à venir) cette unité regroupe à la fois des phrases entières, mais aussi des morphèmes dérivationnels, des unités terminologiques et des collocations.

Le terme est placé au centre de cette approche, mais on peut l'aborder de manière très diverse et ce que Cabré explique à l'aide du modèle des portes (2000, 2003, 2012) :

La maison reste toujours la même, quelle que soit la porte par laquelle on entre. On peut la parcourir entièrement et arriver à n'importe quelle chambre. Mais pour entrer dans la maison, il faut ouvrir une des portes, et pour chacune des portes, il nous faut une clé spécifique (Cabré 2012, p. 8).

Ce modèle explique comment différentes disciplines peuvent décrire le même terme et comment la signification d'un terme change si l'on change la perspective. Prenons l'exemple du terme *exploitation*. Dans le domaine des télécommunications *exploitation* signifie « ensemble des activités nécessaires pour mettre en œuvre une installation, par exemple, un réseau de télécommunication », dans le domaine de la gestion de l'entreprise signifie une « activité normale de l'entité visant à la réalisation de l'objet dominant qu'elle s'est donné » alors que dans le domaine des régimes et conditions de travail : « une utilisation de travailleurs [mal payés] tandis que l'employeur perçoit des profits de façon abusive et injuste » (définitions provenant de la base terminologique TERMIUM Plus²).

Nous pouvons donc dire « qu'un même objet (le terme) est susceptible d'être envisagé de différents points de vue (facettes d'une même unité) et que cette diversification génère d'une part une conception différente de l'objet (parce qu'on privilégie une facette ou l'autre) et d'autre part, un parcours méthodologique différent basé sur une théorie appropriée » (Cabré, 2012, p. 8).

1.2.3. Modèle sociocognitif

Dans son livre *Towards New Ways of Terminology Description : The Sociocognitive approach* (2000), Temmerman propose une attitude axée sur le potentiel cognitif de la terminologie du domaine de spécialité et sur les variations terminologiques des verbes et des contextes dans le discours et dans la communication environnementale.

Rita Temmerman critique la Théorie Générale de la Terminologie et, en utilisant des exemples issus des sciences naturelles, elle montre l'incapacité de la théorie de Wüster à décrire la langue spécialisée dans son fonctionnement, surtout les traductions spécialisées. L'auteure déclare que les principes de la Terminologie sont les suivants :

- plusieurs catégories ne peuvent pas être clairement délimitées ;
- la meilleure définition d'un terme ne peut pas se fonder sur un seul mode d'utilisation et de la forme, parce que la polysémie et la synonymie sont très présentes dans les textes spécialisés et elles doivent être intégrées dans l'analyse terminologique ;

² La description précise de la banque de données terminologiques et linguistiques TERMIUM Plus est située dans le Chapitre 7, p. 53

- les catégories, les concepts et les dénominations évoluent constamment et doivent être étudiés diachroniquement.

L'approche sociocognitive est la première et probablement la seule approche qui tienne compte de la dimension diachronique des termes. Son modèle d'organisation des concepts est un modèle cognitif où les entités évoluent constamment dans le temps. Comme la théorie socioterminologique de Gaudin, cette approche prend en compte les dimensions sociales des termes. Une autre voie proche de ces deux théories est celle de Cabré que nous discuterons ci-après.

1.2.4. Terminologie textuelle

La terminologie textuelle (Bourigault&Slodzian, 1999) est une approche qui est née de la coopération de l'informatique, de l'exploitation de corpus et de la terminologie. Cette approche comme la socioterminologie mentionnée ci-dessus, n'accepte pas l'hypothèse wüsterienne selon laquelle le terme est une notion ayant un caractère fixe.

La terminologie textuelle analyse la place d'un terme occupé dans un corpus. Comme l'affirment Bourigault et Slodzian « la terminologie doit « venir » des textes pour mieux y « retourner » » (Bourigault&Slodzian, 1999, p. 30). Le texte est le point de départ pour un terminographe ou un linguiste qui, dans son travail, est assisté par ordinateur et par un expert de domaine analysé pour confirmer les descriptions construites par celui-ci. Dans ce contexte, un analyste retrouve des mots qui lui paraissent terminologiques et en partant de ces dénominations relevées dans un corpus, il construit des définitions diverses. Ces définitions sont basées sur les significations des mots dans leurs contextes. Après, elles sont validées par l'expert.

Cette analyse termino-linguistique peut être présentée sous plusieurs formes (comme un glossaire, une liste, un tableau, un réseau, etc.) et elle décrit non seulement les noms, mais aussi d'autres catégories du discours (les adjectifs, les verbes, les adverbes) et les syntagmes nominaux, verbaux, adjectivaux, etc. (Bourigault&Slodzian, 1999).

1.2.5. Approche lexico-sémantique

L'Homme dans *La terminologie : principes et technique* (2004) a présenté l'approche lexico-sémantique qui « aborde le terme comme une unité lexicale dont la particularité est d'avoir un sens qu'on peut associer à un domaine de la connaissance » (L'Homme, 2004, p. 37). C'est une approche qui tient compte de la structure argumentale du verbe et de son appartenance à un réseau lexical (L'Homme 2012).

Cette approche nous servira de base pour établir le taux de spécialisation de nos verbes, c'est pourquoi sa description détaillée est présentée dans le Chapitre 2.

Afin d'établir les fondements pour construire un dictionnaire terminologique multilingue basé sur les verbes, nous avons décidé d'utiliser l'approche de la terminologie textuelle et, à partir de corpus, d'analyser les structures argumentales et grâce à celles-ci de décrire les significations des verbes choisis dans le corpus de forage et de les comparer à une ressource générale.

Comme nous pouvons le voir, l'impact de la linguistique de corpus et de la sémantique lexicale sur la terminologie classique a mené à l'augmentation de l'importance du verbe dans des analyses terminologiques.

Nous avons présenté plusieurs approches (comme la théorie wüsterienne, la terminologie textuelle, l'approche basée sur les travaux de sémantique lexicale). Ces approches montrent clairement que la terminologie ne peut pas se limiter aux principes de l'approche conceptuelle afin de répondre à tous les besoins liés à l'exploitation des textes spécialisés. Selon notre opinion, les mêmes principes s'appliquent au caractère nominal des termes. La clef de la compréhension de chaque phrase est un verbe qui est son noyau et forme son sens. C'est pourquoi nous considérons que les verbes sont aussi essentiels que les noms qui, dans la phrase, exercent bien la fonction de ses arguments, c'est-à-dire en sont dépendants.

Dans le Chapitre 2, nous présenterons comment les approches déjà présentées perçoivent le verbe et nous discuterons également les modèles linguistiques d'analyse des verbes afin de trouver un modèle convenable à notre analyse.

Chapitre 2. Verbe en terminologie

La terminologie est une discipline scientifique dont l'objet d'étude est le terme. Pendant longtemps, cette discipline s'est intéressée surtout aux termes de nature nominale et de leurs caractéristiques. Nous pouvons l'observer pendant la consultation d'un dictionnaire spécialisé ou d'une banque de terminologie où environ 90% des entrées sont des noms ou des syntagmes nominaux. Cela s'explique par une influence de l'ouvrage de Wüster (Wüster, 1968 in L'Homme, 2004) et de sa terminologie classique ainsi que par le fait que l'extraction d'informations de corpus se fasse principalement à partir de substantifs. Toutefois, depuis plus de 25 ans, on peut observer de nouvelles approches qui se basent sur la linguistique de corpus et qui s'inspirent de la sémantique lexicale en mettant le verbe (mais aussi l'adjectif et même l'adverbe) au centre de leurs travaux. Elles montrent que l'exclusion des verbes mène à la perte d'une partie d'informations (Tateisi *et al.*, 2004), au caractère incomplet des dictionnaires terminologiques et, par conséquent, rend le travail des traducteurs et terminologues plus difficile.

Cependant, tous les verbes présents dans les textes spécialisés ne sont pas, par définition, eux-mêmes spécialisés. Pour pouvoir établir le taux de spécialisation d'un verbe, il faut analyser sa signification dans le contexte (dans le corpus). Nous présentons, ci-dessous, comment les différentes approches procèdent pour identifier et décrire les verbes spécialisés.

2. *Différentes façons d'analyser un verbe*

2.1. *L'approche conceptuelle*

Nous avons déjà souligné que la plupart des entrées dans l'approche conceptuelle sont de nature nominale. Néanmoins, le verbe n'est pas absolument exclu de cette perspective. Comme Rey l'a écrit dans son livre :

La terminologie ne s'intéresse aux signes (mots et unités plus grandes que le mot) qu'en tant qu'ils fonctionnent comme des noms dénotant des objets et comme des « indicateurs de notions » (de concepts). Dans cette optique, les verbes sont des noms de processus, d'actions (Rey 1979, p. 25).

Dans ce cas, le verbe devient terminologique quand il désigne le même concept que le nom, comme c'est le cas de *développer* et *développement*. Cela explique partiellement une faible participation des verbes dans les dictionnaires spécialisés (L'Homme, 2012).

2.2. *L'approche lexico-sémantique*

Comme nous l'avons mentionné ci-dessus, la démarche onomasiologique est appropriée dans le cas des néologismes, comme le terminographe cherche à nommer un nouveau concept. Toutefois, la plus grande partie du travail du terminographe consiste à repérer des termes existants dans des textes et à en appréhender leur sens. Dans ce cas, le terminographe introduit une démarche inverse de celle recommandée par l'optique conceptuelle, dans une démarche sémasiologique empruntée à la sémantique lexicale où le terme est considéré comme une unité lexicale.

Le point de départ pour l'analyse et la description des termes est le texte. Dans cette approche, la signification d'un terme est définie par rapport à sa signification dans le texte, un terme est le résultat de l'analyse d'un corpus. Le terme, lui-même est défini comme un type d'unité lexicale³ et donc, comme des unités lexicales, les termes résultent de l'association d'une forme linguistique et d'un sens lexical.

L'approche lexico-sémantique utilise la démarche sémasiologique où le point de départ de l'analyse est la forme. L'Homme (2004, p. 33-34) explique que selon cette démarche les sens sont dégagés et distingués à partir des connaissances du terminographe qui travaille pour décrire un texte. Mais les distinctions peuvent aussi être faites à travers l'observation des interactions entre les termes et leur environnement dans les contextes.

L'Homme (2004) affirme que les relations entre les termes interviennent sur deux plans principaux : paradigmatique et syntagmatique. Le plan paradigmatique concerne les liens (aussi appelés verticaux) que partagent les unités lexicales substituables dans le lexique, comme *forer*, *forage* et *foration*. Le plan syntagmatique comprend les liens (appelés horizontaux) que partagent les unités lexicales qui se combinent dans une phrase.

En grande partie, notre travail sera fondé sur l'approche lexico-sémantique de Marie-Claude L'Homme (2004). Nous nous servirons de l'observation de corpus et de l'analyse de la structure argumentale des verbes choisis et sur leurs réseaux lexicaux afin de pouvoir établir le taux de spécialisation de nos verbes.

³ L'unité lexicale « appelée tantôt mot, lexème ou lexie par les lexicologues » (L'Homme 2014, p. 33) est une minimale de signification appartenant au lexique.

2.2.1. La structure argumentale

La structure argumentale d'un verbe est l'ensemble de ses arguments (ou actants). L'analyse de la structure argumentale est l'une des méthodes pour montrer la nature terminologique du verbe, comme L'Homme l'explique: « le verbe est spécialisé ou non dans la mesure où l'ensemble de sa structure argumentale est prise en compte » (2012).

Plusieurs travaux montrent que dans les cas d'incertitude sur le caractère du verbe analysé, la simple observation de la nature des arguments peut résoudre le problème.

L'Homme (1998) propose un simple test basé sur la nature des arguments d'un verbe qui est intuitivement perçu comme étant spécialisé. Elle donne l'exemple d'un verbe *installer* (L'Homme, 2012) dans la phrase :

L'utilisateur installe la nouvelle version du traitement du texte sur son PC.

Comme nous pouvons le voir, les arguments du verbe *installer* sont les termes qui appartiennent au domaine de l'informatique :

<i>utilisateur</i>	Installe	<i>version</i>	sur	<i>PC</i>
		<i>tableur</i>		<i>portable</i>
		<i>système</i>		<i>ordinateurs</i>
		<i>d'exploitation</i>		
		<i>etc.</i>		<i>etc.</i>

Pimentel (2012), dans son travail sur l'identification des verbes spécialisés dans un corpus bilingue de textes juridiques, se base sur la structure argumentale des verbes pour confirmer leur statut spécialisé après l'extraction automatique à l'aide de TermoStat⁴ (Drouin, 2003). Pour organiser les arguments des verbes, l'auteur propose de s'appuyer sur la Sémantique des cadres (l'idée générale de la Sémantique des cadres de Fillmore sera examinée dans la suite de ce mémoire, dans le chapitre « FrameNet »). C'est aussi le cas de Tellier (2008) qui, dans son travail sur les verbes spécialisés dans un corpus médical, se base sur les travaux de Fellbaum (1990), de Condamines (1993) et de L'Homme (1998) pour établir le statut d'un verbe en organisant ses arguments dans les classes sémantiques.

⁴ L'outil TermoStat est décrit en détail dans le Chapitre 5, p. 38

Dans son article sur le verbe terminologique, L'Homme (2012) réunit trois modèles de structure argumentale :

Verbe	Structure argumentale	Auteur
contaminer	Classes sémantiques : X (AGENT INFECTIEUX) ~ Y (ANIMAL; PRÉPARATION; SUBSTANCE BIOLOGIQUE; SUBSTANCE COMPLEXE, MATÉRIAU)	Tellier (2007)
acquitt ₁	Étiquettes d'éléments d'un cadre : Judge ~ Defendant of Charges	Pimentel (2010)
télécharger	Rôles sémantiques : Agent ~ Patient de Source à Destination Termes typiques : Utilisateur ~ fichier; logiciel de réseau, ordinateur à ordinateur	L'Homme - DiCoInfo (2010)

Tableau 1 Représentations diverses de la structure argumentale des verbes spécialisés (L'Homme 2012 : 9)

La nature des arguments est une solution alternative à l'optique conceptuelle et son idée que seuls les verbes liés strictement à un nom peuvent avoir un statut terminologique. Comme nous pouvons le voir dans les exemples mentionnés plus haut, la nature des actants sémantiques peut servir comme indice pour confirmer le sens spécialisé d'un verbe. Si ses actants sont déjà admis comme étant des termes, le verbe risque d'être spécialisé lui-même (L'Homme, 2004, p. 64)

2.2.2. Réseau lexico-terminologique

Dans son livre *La terminologie : principes et techniques* (2004), L'Homme présente les critères pour identifier les termes et l'un des critères mentionnés est le fait que les dérivés des unités lexicales qui étaient déjà admis comme des termes soient inévitablement des termes. Elle donne l'exemple du mot *compilateur* qui appartient au domaine de l'informatique, défini comme un « programme dont la fonction est de convertir un code source en code machine » (L'Homme, 2004, p. 65). Selon ce principe, les verbes *compiler* et *recompiler* doivent être aussi des termes.

On peut distinguer deux cas : le premier, quand un nom et un verbe ont la même signification - tous les deux décrivent une action (L'Homme, 2012, p. 99) :

développement *développer*

téléchargement *télécharger*

réchauffement *réchauffer*

Et le deuxième, quand les sens sont reliés, mais distincts, comme dans *programme* et *programmer* ou *programme* est le résultat de l'activité de *programmer* (L'Homme 2012, p. 100) :

<i>Programme</i>	<i>programmer</i> (programme est le résultat de l'activité de <i>programmer</i>)
<i>Programmeur</i>	<i>programmer</i> (<i>programmeur</i> est un l'agent typique de <i>programmer</i>)

Dans la plupart des cas, c'est le nom qui sert de base pour retrouver les verbes qui lui sont liés, mais il est possible de commencer l'analyse avec un verbe. Ainsi, on peut découvrir non seulement certains noms indissociables sémantiquement et morphologiquement des verbes, mais aussi d'autres unités telles que les adjectifs ou les adverbes, ainsi que les unités lexicales qui ne sont pas visibles morphologiquement, comme dans l'exemple (L'Homme, 2012, p.100) :

<i>NOM</i>		<i>VERBE</i>	
<i>compilateur</i>	<i>compiler</i>	<i>compiler</i>	<i>compilateur</i>
	<i>compilation</i>		<i>compilation</i>
			<i>compilable</i>
			<i>recompiler</i>
			<i>recompilable</i>
			<i>décompiler</i>
			<i>etc.</i>

Dans cette approche, très souvent, ce sont les noms qui nous servent de point de départ pour découvrir des verbes en fonction des liens entre eux. Mais il est possible, ce qui est aussi très intéressant, de renverser cette analyse et de commencer par un verbe. Cela peut nous amener à découvrir des réseaux beaucoup plus riches.

Cette perspective donne aussi la possibilité de découvrir des liens entre les verbes et d'autres unités lexicales non liés morphologiquement. Sur la figure 4, nous pouvons voir que le verbe *programmer* mentionné ci-dessus est lié à ces dérivations comme *programmation* ou *programme*, mais en même temps à *informaticien*, *langage*, *logiciel*, *écrire* et *développer*.

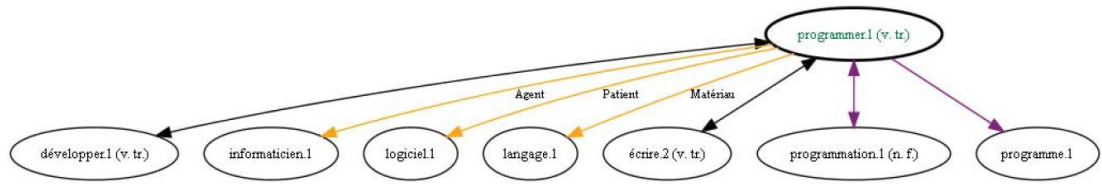


Figure 4 Une partie du réseau lexical de programmer1 (DicoInfo visuel, Robichaud 2011, in L'Homme 2012, p. 100)

Comme le montre la figure 4, cette approche a été utilisée lors de la création du dictionnaire DicoInfo. C'est une base de données lexicales qui contient les termes du domaine de l'informatique et de l'internet. Pour chaque terme, on obtient différents types d'informations : les liens lexicaux, la réalisation des actants, les contextes, entre autres (Wandji, 2014).

Chapitre 3. Structures verbales en linguistique – les patrons lexico-syntaxiques

Dans notre étude, pour pouvoir identifier le sens des verbes et décrire leur taux de spécialisation, nous utiliserons les patrons des verbes. Ces éléments préconstruits se situent à la limite de la phraséologie⁵ et allient les champs duexique et de la syntaxe (Yan, 2017).

Dans ce chapitre, nous présenterons les constructions des phrases dans la langue française en nous inspirant du travail de Yan (2017). Nous nous concentrons en particulier sur les patrons verbaux pour pouvoir les appliquer à notre corpus de forage.

3. Constructions dans la linguistique française

Les chercheurs de la linguistique française relient les formes du langage à leur sens, ce qui est visible dans les travaux de Gross (1975), de Willems (1981) et de Dubois et Dubois-Charlier (1997).

3.1. Le lexique-grammaire

La théorie du lexique-grammaire (Gross, 1975) repose sur l'hypothèse que « l'unité significative du lexique est la phrase simple » (Gross, 1986, p. 299). Cette hypothèse est confirmée par la nature du verbe parce que le verbe est indissociable de son sujet et de ses compléments (s'ils sont présents).

Gross met en contraste toutes les constructions verbales identiques associées à des verbes différents. Il observe que ces constructions se différencient selon une série de propriétés syntaxiques qui possèdent « une fonction classifiante qui semble être liée aux propriétés sémantiques du langage » (Gross, 1986, p. 171, in Yan, 2017), comme : temps, voix de la phrase (active ou passive), présence d'une complétive *que*, types de prépositions utilisés, nombre et nature des compléments ou réalisations du sujet (humain, non humain, nom au pluriel). À la suite de ces observations, Gross classe les constructions syntaxiques de 6000 verbes dans 81 tableaux.

Mais comme le souligne Yan (2017), la polysémie des verbes, collectés selon leurs propriétés syntaxiques, n'est pas précisée. Elle donne l'exemple du verbe *dire* qui

⁵ La phraséologie : « domaine qui traite les séquences lexicales perçues comme préconstruites » (Legallois et Tutin, 2013, p. 3)

appartient à trois tableaux différents et que chacune désigne un sens. Le problème est que le sens n'est pas explicite et l'utilisateur est obligé de le déduire à partir d'un exemple.

Cela ne change pas le fait que cette méthode est la première étude systématique sur les constructions verbales en français et est devenue une base pour la création d'autres ressources des constructions verbales.

3.2. Les constructions verbales de Willems

Au même titre que Gross, Willems (1981) place le verbe au centre du classement des constructions verbales. Comme le décrit Yan (2017, p.10), dans ce modèle, le verbe est un noyau qui permet d'établir le lien entre la syntaxe et la sémantique. Les classes des verbes créées par Willems sont plus petites que ceux de Gross, mais en même temps, elles restent plus homogènes. Willems classe 3000 verbes selon trois principes :

- (1) Les relations sémantiques générales, qui sélectionnent
- (2) un nombre restreint de constructions syntaxiques
- (3) le choix d'une construction implique le choix d'un certain lexique. (Willems, 1981, p. 239, in Yan, 2017, p. 12)

Son travail montre que le sens d'une phrase est lié à une construction concrète. Selon elle, tous les verbes peuvent devenir des « verbes déclaratifs » (*ibid*, p. 12), au moment où ils adoptent les propriétés syntaxiques caractéristiques de cette construction, donc de cette classe verbale. Willems a identifié cinq types essentiels de complémentation (Yan, 2017, p. 12-13) :

- absence de compléments « complémentation intransitive » (ex. *il pleut, il marche*)
- compléments nominaux « complémentation transitive » (ex. *il ressemble à Paul, le jardin pullule d'abeilles, il donne un livre à Marie*)
- compléments adjectivaux ou substantivaux « complémentation attributive » (ex. *il devient grand, il rend sa mère malade*)
- compléments infinitifs « complémentation infinitive » (ex. *il vient jouer, il cherche à réussir, il envisage de partir demain*)
- compléments phrastiques « complémentation complétive » (ex. *je sais que tu pars*).

En ce qui concerne les arguments nominaux, Willems distingue cinq sous-catégories : nom de personne, nom de chose, nom de lieu, nom désignant une partie du corps et nom d'action.

3.3. *Les Verbes Français (LVF)*

Inscrits dans la continuité des deux travaux précédents, *Les Verbes Français* de J. Dubois et F. Dubois Charlier est un ouvrage qui propose la classification des verbes autour de leurs propriétés sémantico-syntaxiques (classes sémantiques définies par la syntaxe). Il comprend « 25 610 entrées verbales simples représentant 12 310 verbes différents, dont 4 188 à plusieurs entrées » (Dubois&Dubois-Charlier, 1997, p. 1). Il existe en version électronique et il est accessible en ligne gratuitement (<http://rali.iro.umontreal.ca/rali/?q=fr/lvf>).

La base du classement sémantique constitue les arguments syntaxiques qui associent à différents emplois d'un verbe leur sens et chaque cas est représenté par une phrase simple pour illustrer son sens et sa construction (François, 2007).

Les constructions sémantiques sont décrites à l'aide d'un « opérateur » qui « interprète sémantiquement les schèmes syntaxiques » (Dubois et Dubois-Charlier, 1997, p.7). Chaque entrée possède son opérateur complexe qui est composé d'opérateurs simples.

Les constructions syntaxiques sont décrites à l'aide d'une suite de chiffres et de lettres dénommées « schème ». À chaque entrée verbale correspondent un ou deux schèmes de construction (le deuxième est souvent la description de l'emploi pronominal).

La compréhension de ces étiquettes ne pose aucun problème, car dans la version électronique, nous pouvons vérifier directement la signification de chaque schème et en outre, la liste exhaustive est disponible dans *Les Verbes Français* de Dubois et Dubois-Charlier (1997).

Prenons l'exemple du verbe *produire* (tableau 2) : 10 emplois différents, chacun a son schème attribué, noté *T1306*, *P3006*, *T3306*, etc., qui indique :

- type de construction syntaxique (intransitive *A*, transitive directe *T*, transitive indirecte *N*, pronominale *P*)
- type de constituant sujet/objet (humain *1*, animal *2*, chose *3*, complétive *4/5*)

- type de complément prépositionnel et circonstant (prépositionnel *a-q*, locatif *l-4*, à modalité *6*, causatif *7*, instrumental *8*, etc.).

Donc, pour l'entrée numéro 1 :

- domaine : *IND* « industrie » ;
- opérateur : *fab* « réaliser un objet » produit ;
- sous-type syntaxique : *R3a.1* « faire quelque chose, un objet » ;
- schème : *T1306* « transitif direct, humain, chose, modalité » et *P3006* « pronominal, chose, modalité ».

No	DOM	SCS	Opérateur	Sens	Phrases	Schémes
1	IND	R3a.1	fab produit	fournir, fabriquer	On p~ du vin dans cette région. Cette usine p~ des pneus.	T1306, P3006
2	BOT	R3a.1	qc fab qc naturel	donner naturellement	L'arbre p~ des fruits.	T3306
3	SOC	R2a.1	fab personnage	donner naissance à	On p~ de grands savants dans ce pays.	T1106
4	LIT	R4a.1	fab texte	rédiger, créer	Cet auteur p~ plusieurs romans par an.	T1306, P3000
5	TEC	R3a.1	qc fab qc	réaliser	Ce frottement p~ l'usure. La buée se p~ par condensation.	T3306
6	PSY	E4f	abstrait f.ire qc	provoquer, susciter	Cette nouvelle p~ une grande émotion.	T3300
7	DRO	C3a.1	ind qc/qn A qn	exhiber, présenter	On p~ un témoin devant le tribunal. On p~ un certificat.	T19a0
8	CIN	R4a.1	fab produit spectacle	réaliser	Le producteur p~ une émission de télé, un film.	T1306, P3000
9 (se)	TPS	L4a.1	qc lc.qp temps actuel	se rencontrer	Un grand silence se p~. Un événement s'est p~ hier.	P3005
10 (se)	LOC	L1a.1	lc.qp spectacle	se montrer	L'acteur se p~ dans plusieurs films.	P1001

Tableau 2 Les emplois du verbe *produire* répertoriés dans LVF

Le tableau 2 montre bien l'exhaustivité de l'ouvrage de Dubois et Dubois-Charlier. Les relations entre la syntaxe et la sémantique sont visibles grâce aux schèmes de construction, aux opérateurs et à leurs sens.

La grande couverture du nombre de verbes français de la langue générale et de leurs propriétés fait que cette ressource est une référence parfaite pour notre travail et sera adoptée pour montrer les différences des utilisations des verbes et pour faciliter la division des verbes dans des classes dans le Chapitre 9.

4. Constructions dans la linguistique anglaise

4.1. FrameNet

FrameNet est un projet basé sur la théorie des cadres sémantiques, élaboré par Charles J. Fillmore et ses collègues (Fillmore 1976, 1977, 1982, 1985, Fillmore&Baker 2001). L'idée de base est que les significations de la plupart des mots peuvent être accessibles grâce aux « connaissances d'arrière-plan » (L'Homme, 2016, p.3) faisant référence à ces mots (nous avons accès à ces informations directement dans les phrases, mais les dictionnaires classiques nous en privent). Dans les cadres sémantiques ces informations ont été consignées dans les frames qui décrivent un type d'événement, de relation ou d'entité et ses participants (*frame elements*, FEs) et qui grâce à ces éléments, donnent accès à l'utilisation d'un mot recherché et aussi à son interprétation.

Prenons l'exemple mentionné sur le site officiel de FrameNet : cuisiner [Cook]. Le concept de cuisiner implique généralement une personne qui va cuisiner, la nourriture qui va être cuisinée [Food], quelque chose pour maintenir la nourriture pendant la cuisson [Container] et une source de chaleur [Heating_instrument]. Dans FrameNet, cette situation est représentée par un frame nommé « Apply_heat » et les éléments *Cook*, *Food*, *Heating_instrument* et *Container* sont nommés frames éléments (FEs). Les mots qui évoquent ce frame comme *fry* (*frire*), *bake* (*cuire*), *boil* (*bouillir*), et *broil* (*griller*) sont nommés **lexical units (LUs)** de « Apply_heat » frame.

Puisque les frames sont fondamentalement sémantiques, ils peuvent être comparés dans diverses langues. Bien sûr, cette tâche n'est pas toujours facile et peut poser des problèmes étant donné que la Sémantique des cadres est une approche très complexe. Néanmoins, plusieurs projets utilisent l'approche des cadres sémantiques pour identifier les équivalents dans les langues française et italienne, dans le domaine de la médecine (Riccio, 2014) et portugaise, dans le domaine judiciaire (Pimentel, 2012).

Il est bien visible que la théorie des cadres sémantiques est une approche prometteuse et qui a fonctionné dans divers domaines et dans plusieurs langues. Néanmoins, elle n'est pas très intuitive et peut poser des problèmes, ce qui constitue un défi majeur dans notre travail. Notre objectif est de construire une méthode de description des

verbes spécialisés qui sera claire et facile à développer. C'est pourquoi nous avons décidé de ne pas utiliser la Sémantique des cadres, mais nous la mentionnons ici afin de montrer les possibilités qu'elle offre. Dans notre travail, nous nous appuyerons sur le modèle *Corpus Pattern Analysis* que nous présenterons dans la section suivante.

4.2. *Corpus Pattern Analysis (CPA)*

Proche des travaux sur le lexique de Sinclair et Halliday, le modèle *Corpus Pattern Analysis* est un modèle lexicographique qui se focalise sur les patrons syntagmatiques basés sur les corpus. Selon Hanks (2004), les patrons des noms et les patrons des verbes varient considérablement et ces derniers contiennent beaucoup plus d'informations parce qu'ils prennent en considération aussi d'autres éléments de la phrase analysée (sujet, objet, etc.).

Cette approche, basée sur *Theory of Norms and Exploitations* (TNE - la théorie des normes et des utilisations), est une approche *corpus-driven*, donc la construction des patrons se fonde sur l'analyse des concordances de corpus. Dans le CPA le sens d'un verbe est associé à son contexte. Cela élimine les ambiguïtés liées à la polysémie des verbes.

Selon Hanks (2004), le verbe est le fondement d'une phrase et la structuration des verbes mène à la structuration semi-automatique des noms. La structure des patrons des verbes est basée sur la grammaire systématique de Halliday (1961) : sujet, objet, complément de sujet, complément d'objet et complément circonstanciel.

Le modèle CPA a été utilisé dans la construction du dictionnaire anglais *Patterns Dictionary of English Verbs* (PDEV). Les patrons ont été conçus à la base du British National Corpus. C'est un dictionnaire accessible gratuitement en ligne (<http://pdev.org.uk>).

Comme nous l'avons déjà dit, les patrons CPA sont constitués à partir de leurs caractéristiques sémantiques. Une ontologie de 200 mots regroupe les catégories sémantiques (types sémantiques) comme [Human] ou [Animal] qui appartiennent à un groupe plus général [Animate].

Examinons donc l'exemple du verbe *polish*₁ et sa signification :

Pattern¹: [Human] or [Material] polishes [Inanimate]

Implicature: [Human] causes the surface of [Inanimate = Surface] to be smooth and shiny by rubbing it vigorously with [Material = Abrasive] (un être humain ou un matériau rend la surface d'un objet inanimé lisse et brillante en le frottant vigoureusement avec un matériau)

Example: Luigi continued polishing his new coffee machine. (Traduction de l'auteur : Luigi continue à cirer sa nouvelle machine à café.)

Dans ce cas, les types sémantiques du patron du verbe *polish* sont : ses sujets - *un humain* ou *un matériau* et son objet - *un objet inanimé*. De plus, les types sémantiques peuvent aussi posséder les rôles sémantiques, par exemple [Inanimate = Surface], identifiés grâce au contexte.

À part l'application de CPA dans le dictionnaire PDEV, donc dans l'anglais langue générale, cette approche a été aussi utilisée dans la création du DicSci - *E-Advanced Learner's Dictionary of Verbs in Science* (Alonso, Millon&Williams, 2011), un dictionnaire spécialisé des verbes en science, mais aussi dans d'autres langues, notamment en espagnol et dans le domaine de l'environnement (Alonso&Renau, 2013) et en français - l'analyse des verbes du lexique scientifique transdisciplinaire (Yan, 2017).

Partie 2

-

CORPUS ET MÉTHODOLOGIE

Chapitre 4. Problématique

Comme nous avons évoqué dans le Chapitre 1, la terminologie textuelle (Bourigault&Slodzian, 1999) a mis l'accent sur l'analyse des unités lexicales dans leurs contextes. L'observation des contextes des termes a permis de constater que non seulement les noms peuvent transférer le sens spécialisé, comme il est indiqué dans l'approche classique de la terminologie (Wüster, 1968), mais également d'autres catégories grammaticales telles que les verbes, les adverbes ou les adjectifs (L'Homme, 1998).

En examinant les mots dans les contextes, on constate que les noms ne jouent pas le rôle le plus important dans la phrase. Dans la perspective linguistique, la catégorie primordiale dans le discours est le verbe, parce que, comme le dit Hanks (2010, p. 3) : « meanings are constructed around the verb, the pivot of the cause ». Le verbe exerce la fonction de liaison, il relie les sujets avec les objets en créant des cadres du texte et cela contribue à l'organisation du discours (Alonso&Renau, 2013). En prenant en considération les arguments exposés ci-dessus, nous avons décidé de nous concentrer sur les verbes comme les potentiels termes.

Dans cette partie, nous présenterons la constitution de notre corpus : les raisons du choix des textes, sa source et sa taille. Ensuite, nous montrerons le travail de l'extraction des potentiels verbes-termes, les critères de la sélection des verbes à analyser et aussi les étapes de l'analyse de la liste des verbes choisis, y compris : l'étude et la création de l'ontologie (sous-catégorisation des sujets et des objets des verbes), les analyses sémantique et syntaxique des verbes (création des patrons selon le CPA). Finalement, cette analyse nous servira comme base de données pour établir le taux de spécialisation de nos verbes de forage.

Chapitre 5. Corpus

1. Corpus utilisé

Dans ce chapitre, nous présenterons notre corpus et tous ses détails : sa taille, les raisons du choix des textes, sa source et sa constitution. Nous préciserons également notre approche de cette étude.

1.1. Choix et constitution du corpus

La constitution du corpus a été une tâche très complexe. Le problème du choix des textes utilisés pour l'analyse résulte du fait qu'il a fallu joindre deux domaines de notre analyse. Le présent travail réunit les propriétés linguistiques et celles terminologiques, c'est pourquoi, initialement, nous nous sommes concentrés sur la partie terminologique. Nous avons décidé de travailler sur des textes spécialisés, des articles scientifiques destinés aux spécialistes, mais également sur des brevets. Nous avons adopté cette démarche, parce que les articles de ce genre contiennent en particulier le lexique spécialisé qui nous intéressait le plus en tant que terminologue. Le problème de ce choix découlait surtout de la difficulté à trouver un bon nombre d'articles pour constituer un corpus représentatif. Les textes spécialisés portant sur le forage sont d'un accès difficile: dans la majorité des cas l'accès aux articles qui nous intéressaient était payant. Nous avons aussi essayé de contacter des entreprises francophones se spécialisant dans la géophysique, la géo-ingénierie ou le forage, mais sans aucun effet. Les facteurs en question ont fait que nous n'avons pas réussi à rassembler une quantité d'articles qui soit satisfaisante. De surcroît, le forage en tant que domaine scientifique se divise en sous-domaines, qui ne sont toujours pas liés entre eux. La première version du corpus contenait peu de textes où chacun d'entre eux appartenait à un autre sous-domaine. C'est pour cette raison que nous avons décidé de changer notre approche et de nous diriger vers la linguistique et d'utiliser les ressources accessibles à tout le monde et destinées aux élèves d'écoles techniques et aux étudiants.

Tout au début, notre travail était censé être une analyse contrastive des deux langues : française et polonaise. Nous avons planifié de construire deux corpus comparables afin de construire le début d'un dictionnaire terminologique du forage. Mais en raison des problèmes concernant la constitution d'un corpus qui a pris beaucoup plus de temps que prévu, ainsi que d'une multitude de travail lié à l'analyse sémantique et syntaxique de la partie française, nous avons décidé d'abandonner la partie polonaise de notre travail et de nous concentrer sur l'analyse de la partie française et sur la constitution

d'une méthode visant à faciliter l'analyse et la classification des verbes nous intéressant et à créer un dictionnaire multilingue dans le futur.

Comme nous avons déjà mentionné ci-dessus, dans la deuxième version, celle convenable de notre corpus, nous nous sommes concentrée sur les textes destinés aux élèves et aux étudiants. Nous avons choisi cette solution pour deux raisons. Premièrement, par égard pour un accès plus facile aux textes qui nous intéressaient, ce qui nous a permis de constituer un corpus suffisamment grand. Deuxièmement, le présent travail a pour objectif une analyse terminologique, ainsi que linguistique des verbes dans les textes spécialisés, dans le cas présent liés au domaine du forage, d'où notre décision que les textes semi-spécialisés sont plus appropriés et permettent d'effectuer une analyse plus approfondie.

Tous les articles et les livres que nous utiliserons tout au long de notre travail proviennent de deux sources : Techniques de l'Ingénieur et ScholarVox.

1.1.1. Techniques de l'Ingénieur

Les Éditions Techniques de l'Ingénieur (Éditions T.I.) appartiennent au groupe Weka et ont été fondées en 1946. L'une des parties de cet éditeur constitue une base documentaire technique et scientifique en langue française du même nom qui est accessible en ligne : <https://www.techniques-ingenieur.fr/>.

La base est constituée de plus de 10 000 articles de références et fiches pratiques qui ont été vérifiés et validés par les spécialistes - chercheurs, professeurs de l'enseignement supérieur et ingénieurs de l'industrie. Les principaux domaines traités dans cette ressource documentaire sont :

- Automatique-Robotique,
- Biomédical-Pharma,
- Construction et travaux publics,
- Électronique-Photonique,
- Énergies,
- Environnement-Sécurité,
- Génie industriel,

- Ingénierie des transports,
- Innovation,
- Matériaux,
- Mécanique,
- Mesures-Analyses,
- Procédés chimie-bio-agro,
- Sciences fondamentales,
- Technologies de l'information.

1.1.2. ScholarVox

ScholarVox est une plateforme qui fait partie de Cyberlibris : une bibliothèque de livres électroniques. ScholarVox est dédiée aux étudiants de niveaux Licence et Master. Elle est constituée de plus de 35000 ouvrages et manuels principalement en français qui appartiennent aux domaines de :

- Sciences,
- Économie, gestion, droit,
- Santé, science de la vie, médecine;
- Sciences humaines et sociales.

Pour profiter librement de ces ressources, il faut avoir un compte. Les étudiants de l'UGA ont un accès gratuit aux textes qui s'y trouvent. Vu que ce sont des articles destinés aux étudiants, donc aux apprenants, le langage utilisé est plus accessible pour faciliter la compréhension, et ils possèdent une grande quantité de mots spécialisés de base avec leurs traductions. Les textes que nous avons choisis sont par cela semi-spécialisés.

En ce qui concerne les domaines choisis, nous avons décidé de nous limiter à deux sous-domaines qui sont : énergies et constructions et travaux publics. Ces deux sous-domaines sont proches, mais, en même temps, ils se complètent et tous les deux, ils constituent un nombre important de données. Le sous-domaine des énergies est composé de 10 articles et celui des constructions et travaux publics contient 9 articles destinés aux étudiants, aux élèves ingénieurs, mais aussi aux géologues et géophysiciens travaillant dans le secteur du forage. Ces articles sont issus de la base Techniques de l'Ingénieur.

Notre corpus est également construit sur le livre de Biteau et Baudin : *Géologie du pétrole - Historique, genèse, exploration, ressources* (2017), accessible à la bibliothèque électronique *ScholarVox*. C'est un livre qui décrit en détail les notions liées au système pétrolier et au traitement du pétrole, il relie les deux sous-domaines que nous avons choisis.

Le corpus se compose de 274 809 mots et de 9516 phrases, dont le livre *Géologie du pétrole* est composé de 91 534 mots. Tous les articles ont été publiés entre les années 1992 et 2015, mais ils ont été vérifiés et revalidés par les experts travaillant pour les Techniques de l'Ingénieur après 2014. De cette façon, les verbes et leurs constructions que nous examinerons sont suffisamment récents.

Tous les textes analysés ont été collectés sous le format .pdf. La première partie de notre travail a consisté donc à modifier le format des fichiers. Les outils d'extraction et d'exploitation que nous aborderons dans un instant et utilisés dans notre travail se servent des fichiers en format .txt. Le changement du format a été effectué à l'aide d'AntFile Converter, un outil gratuit qui permet de convertir les fichiers .pdf et Word en texte brut. Puis, nous avons fait le nettoyage complet des fichiers .txt et les avons préparés à l'analyse ultérieure.

1.2. Étude de corpus

Notre approche des verbes fait partie du cadre de la linguistique de corpus. Dans la linguistique de corpus, le corpus, comme l'a décrit Sinclair, est « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon au langage » (Habert, Nazarenko&Salem, 1997, p. 11). On distingue deux méthodes dans le cadre de cette approche : *corpus-based* et *corpus-driven*. La méthodologie de *corpus-based* part du principe qu'on confirme ou infirme les hypothèses constituées au départ du travail. En ce qui concerne la méthodologie de *corpus-driven*, elle permet de parvenir à des conclusions grâce à l'observation du corpus.

Nous avons décidé que pour effectuer la meilleure analyse de nos données nous avons besoin de nous servir de ces deux approches (*corpus-based* et *corpus-driven*) en même temps. Cela est dû au fait que, d'un côté, nous avons construit l'hypothèse que les verbes présents dans les textes de spécialité peuvent être spécialisés, donc être des termes eux-mêmes. Comme nous avons évoqué dans le Chapitre 2 (p. 9), selon Rey (1979, p. 24), les verbes peuvent être terminologiques uniquement lorsqu'ils désignent les mêmes

concepts que les noms (des verbes nominalisables). L'Homme dans son *Statut du verbe en langue de spécialité* (1998) affirme qu'un simple test des arguments d'un verbe est capable de confirmer le statut spécialisé de ce verbe. Alors que Lerat (2002, in L'Homme, 2012), en se basant sur les classes d'objets de Gross (1994, in L'Homme, 2012), classe les verbes du domaine juridique dans trois groupes :

1. les verbes très spécialisés ;
2. les verbes supports ;
3. les verbes polysémiques dont au moins un sens est spécialisé.

En s'appuyant sur ces trois travaux, nous avons décidé d'effectuer une analyse sémantique des verbes sélectionnés, de leurs arguments et sur cette base construire et répartir les deux en classe. Nous estimons que cette analyse nous permet de construire un modèle d'analyse termino-linguistique facilitant la construction des dictionnaires spécialisés multilingues. D'un autre côté, nous essaierons de trouver et de décrire les régularités en syntaxe présentes dans notre corpus, en observant toutes les occurrences des verbes choisis pour notre analyse.

Dans notre travail, nous avons besoin de ces deux approches parce que, comme le constate Storjohann (2005), *corpus-based* est surtout utilisé dans le cas de l'analyse sémantique alors que *corpus-driven* est employé pendant la description grammaticale. Ce qui est la raison pour laquelle nous utiliserons le mélange de ces deux approches.

Nous utiliserons aussi, à la fois, l'analyse quantitative et l'analyse qualitative. Quantitativement, notre corpus nous donne accès à un nombre assez grand de textes qui nous permettra de trouver des régularités au niveau de l'analyse sémantique et syntaxique des verbes choisis et de construire les spécificités des éléments étudiés. Qualitativement, le corpus constitue un contexte où se produisent des phrases et des combinaisons lexicales qui nous intéressent. Il est évident que les données statistiques que nous mettrons en relief risquent de ne pas être objectives, mais nous croyons qu'elles peuvent présenter un bon point de départ et donner des indices pour savoir quels résultats valent une analyse plus approfondie. Elles nous permettront aussi de dégager quelques conclusions finales concernant notre corpus et notre travail. Bien sûr, l'élément le plus important de notre travail est l'analyse sémantique et syntaxique de nos données et non les résultats statistiques.

2. Outils d'exploration de corpus

Après la préparation du corpus, nous commençons le processus de son analyse. Au cours de notre travail, nous avons utilisé principalement deux outils : TermoStat et AnaText.

2.1. TermoStat

TermoStat est un outil d'extraction des candidats termes qui utilise une méthode de comparaison de deux corpus : spécialisés et non-spécialisés en vue de l'identification des termes. La version de TermoStat disponible en ligne (<http://termostat.ling.umontreal.ca/>) prend en charge le français, l'anglais, l'espagnol, l'italien et le portugais. TermoStat reçoit un texte en entrée (TermoStat analyse les textes en format texte brut) et retourne comme résultat principal une liste des candidats termes (simple - un mot - cela pourrait être un nom, un verbe, un adverbe ou un adjectif; ou complexe - une suite de mots) tirés du texte.

Ce qui est très important, ce que les textes soumis à TermoStat sont étiquetés par un logiciel d'étiquetage externe (TreeTagger). Le rôle de l'étiqueteur est de désambigüiser les mots qui peuvent recevoir plus d'une catégorie syntaxique. Grâce à cela, tous les mots du texte possèdent une étiquette syntaxique.

Comme l'explique Drouin (2003), des candidats-termes sont sélectionnés en fonction d'un score reçu à la suite d'un test des mesures statistiques. Les candidats termes ayant reçu les scores les plus hauts sont considérés comme les plus pertinents dans le texte. Ce qui permet d'exclure des mots non considérés comme terminologiques dans le texte, c'est un seuil d'acceptabilité. TermoStat utilise cinq tests différents.

Pour comparer les fréquences des candidats termes (CT) dans le corpus de référence (CR) et le corpus d'analyse (CA), TermoStat utilise la table située au-dessous qui présente les divers scénarios possibles :

REPRÉSENTATION DES FRÉQUENCES DES UNITÉS			
Corpus	CR	CA	Total
Fréquence CT	a	b	a+b
Fréquences autres mots	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

Tableau 3 : Tableau de contingence pour représentation des fréquences des unités (Drouin, 2003).

2.1.1. Fréquence

La première mesure est directement observable dans les corpus et c'est la fréquence brute - nombre d'occurrences d'un terme dans le corpus analysé. Daille(1994) a observé que la fréquence représente un bon critère et un point de départ pour l'identification des termes dans les corpus spécialisés.

2.1.2. Spécificité

Le calcul de spécificité a été proposé par Lafon (1980) afin de cerner le vocabulaire spécifique à un sous-corpus par rapport à un corpus de référence.

$$\log P(X=b) = \log (a+b)! + \log (N-(a+b))! + \log (b+d)! + \log (N-(b+d))! - \log N! - \log a! - \log b! - \log (N-(a+b+d))!$$

Cette approche permet de comparer le comportement des unités lexicales en fonction de critères variables. Termostat a adapté cette démarche à travers la fusion du corpus de référence et du corpus d'analyse. Grâce à cela, nous sommes capables de vérifier si le lexique de ce dernier se comporte comme le lexique du premier. Le calcul des spécificités facilite le classement des candidats termes les uns par rapport aux autres.

2.1.3. X²

Le test du X² est utilisé par TermoStat pour comparer les fréquences d'occurrence de candidats termes. Les unités qui se démarquent le plus se verront attribuer une valeur plus élevée.

$$X^2 = N(ad-bc)^2 / ((a+b)(c+d)(a+c)(b+d))$$

2.1.4. Log-likelihood

Proposé par Dunning (1993), le log-likelihood a été utilisé, entre autres, par Rayson et Garside (2000) pour la comparaison de corpus.

$$E1 = ((a+c)(a+b)) / ((a+c)(b+d))$$

$$E2 = ((b+d)(a+b)) / ((a+c)(b+d))$$

$$LL = 2 * ((a * \ln(a/E1)) + (b * \ln(b/E2)))$$

Grâce au log-likelihood on obtient un poids de terme. Plus l'écart entre la fréquence relative observée dans le corpus analysé et celle que l'on pourrait prédire à partir du corpus de référence est important, plus le score est grand.

2.1.5. Log-odds ratio

Le log-odds ratio se calcule ainsi :

$$\text{log-odds} = \log\left(\frac{(b+0.5)(c+0.5)}{(d+0.5)(a+0.5)}\right);$$

Plus le log-odds ratio est élevé, plus le candidat terme est potentiellement intéressant.

2.1.6. Résultats d'analyse dans TermoStat

Les résultats d'analyse sont affichés sous diverses formes :

- liste des termes - option standard, les termes extraits du texte analysé sont listés et triés en fonction du score obtenu par chaque terme, de leur fréquence ou de leur matrice;

Liste des termes Nuage Statistiques Structuration Bigrammes				
Candidat de regroupement	Fréquence	Score (Spécificité)	Variante orthographiques	Matrice
forer	34	45.76	forer fore forent	Verbe
permettre	529	37.77	permettre permette permets permettraient	Verbe
nécessiter	100	34.11	nécessiter nécessitera nécessitent nécessitait nécessitaient	Verbe

Figure 5 : Liste de candidats termes pour les verbes provenant du corpus de forage.

- nuage - liste alphabétique des 100 termes dont le score est le plus élevé, l'impression de nuage est donnée par la différence de taille de caractère des candidats ;

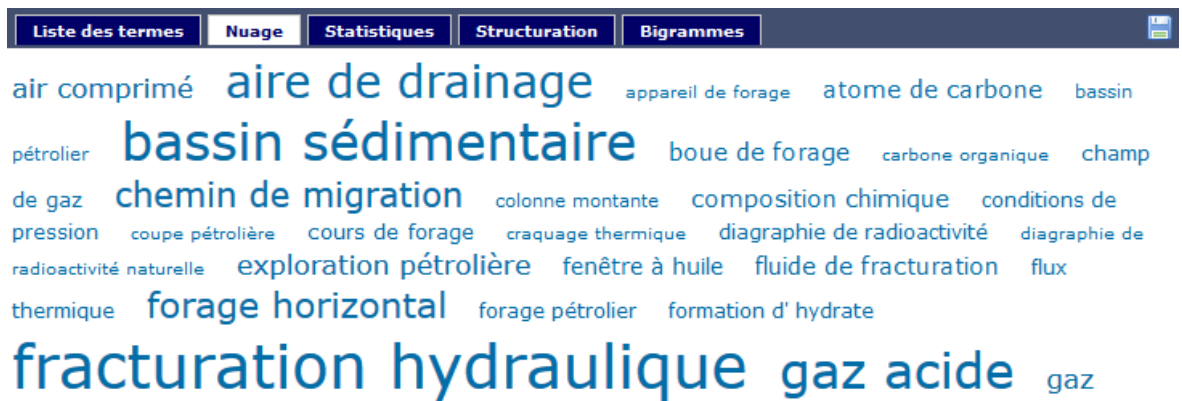


Figure 6 : Liste de candidats termes des termes complexes provenant du corpus de forage sous forme de nuage.

- statistiques - tout en haut on trouve le nombre de candidats sélectionnés pour le texte, après TermoStat affiche le nombre de candidats pour chaque matrice (les 10 premiers termes correspondant à chaque matrice sont visibles en cliquant sur le nombre) ;

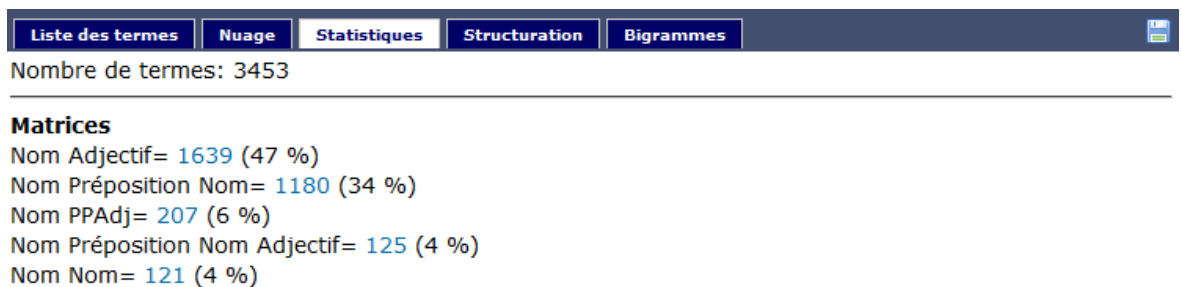


Figure 7 : Liste des matrices des termes complexes provenant du corpus de forage.

- structuration – liste des candidats termes placés dans un tableau, pour chaque candidat terme on trouve sa fréquence (seulement les candidats termes avec une fréquence égale ou supérieure à 4 sont affichés) et la liste des candidats qui l’incluent. Certains candidats termes possèdent une information sur leur structuration. Ils sont marqués d’une icône cliquable qui ouvre un nouvel onglet sur la page de décomposition ;

Liste des termes			Nuage	Statistiques	Structuration	Bigrammes
Candidat de regroupement	Fréquence	Terme inclus				
gaz naturel	260	production mondiale de gaz naturel - consommation de gaz naturel - gaz naturel liquéfié - production de gaz naturel				
matière organique	129	type de matière organique - matière organique produite - matière organique contenue				
système pétrolier	126	notion de système pétrolier				
fracturation hydraulique	96	fracturation hydraulique naturelle				

Figure 8 : Liste des candidats termes et des termes qui les incluent provenant du corpus de forage.

- bigrammes - les bigrammes de TermoStat sont composés d'un verbe et d'un nom (sujet ou objet du verbe), accompagnés de leur fréquence et d'un score (log-likelihood) qui donne la force de l'association entre les mots qui les composent.

Liste des termes				Nuage	Statistiques	Structuration	Bigrammes
Verbe	Nom	Fréquence	Score d'association				
fondre	trou-O	12	109.35				
présenter	avantage-O	17	85.87				
présenter	forme-O	11	54.08				
coupler	couverture-O	5	53.74				
citer	procédé-O	7	51.53				

Figure 9 : Liste des bigrammes des termes simples provenant du corpus de forage.

Les termes sont classés selon un score basé sur la fréquence du terme dans le corpus analysé, et sa fréquence dans un autre corpus prétraité, dans ce cas un corpus de référence de la langue française (il est composé d'environ 28 500 000 occurrences et c'est un corpus non technique composé d'articles de journaux portant sur des sujets variés tirés du quotidien français Le Monde 2002).

Tout d'abord, TermoStat en tant qu'outil d'exploration automatique est vulnérable à de nombreuses erreurs liées aux étiquetages syntaxiques, mais aussi à cause de recherches basées sur la statistique, les simples concordances sont très souvent prises comme les termes. Ces erreurs influencent les résultats obtenus, mais TermoStat donne la possibilité d'examiner le contexte dans le texte et de vérifier des incertitudes possibles.

Pour notre recherche, nous utiliserons seulement quelques fonctionnalités de TermoStat, entre autres : une liste des termes, une statistique et des bigrammes. Le test que

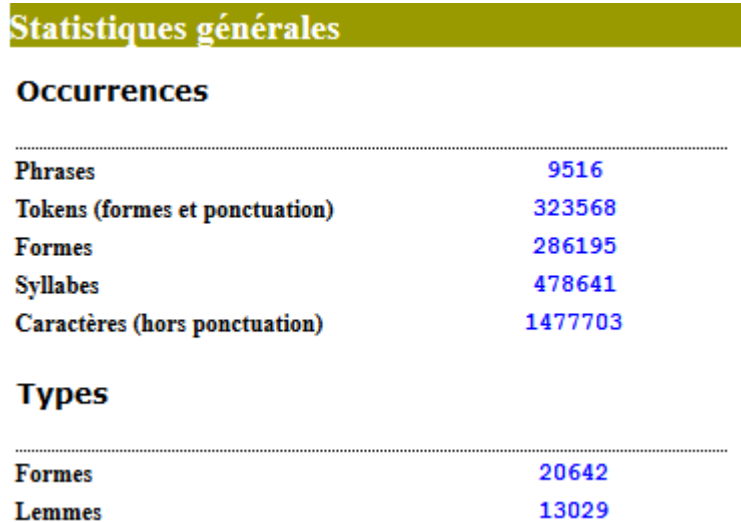
nous avons choisi pour notre analyse est le test de spécificité et sur cette base nous choisirons des verbes à analyser.

2.2. AnaText

AnaText est un outil de textométrie, disponible gratuitement en ligne (<http://phraseotext.u-grenoble3.fr/anaText/index.php>) qui était créé par Olivier Kraif. Le texte soumis est analysé et étiqueté par TreeTagger, ce qui permet d'effectuer des recherches. AnaText donne la possibilité d'interroger des textes français, anglais, allemands, italiens, espagnols, latins, grecs anciens, portugais et russes.

AnaText comporte plusieurs fonctionnalités statistiques et permet de :

- compter des phrases, des mots et des syllabes ;



The image shows a screenshot of the AnaText interface. At the top, there is a green header with the text "Statistiques générales". Below this, the section "Occurrences" is displayed, followed by a table with two columns: the category name and the corresponding count. The counts are: Phrases (9516), Tokens (formes et ponctuation) (323568), Formes (286195), Syllabes (478641), and Caractères (hors ponctuation) (1477703). Below the "Occurrences" table, the section "Types" is displayed, followed by another table with two columns: the category name and the corresponding count. The counts are: Formes (20642) and Lemmes (13029).

Statistiques générales	
Occurrences	
Phrases	9516
Tokens (formes et ponctuation)	323568
Formes	286195
Syllabes	478641
Caractères (hors ponctuation)	1477703
Types	
Formes	20642
Lemmes	13029

Figure 10 : Statistiques générales créées par AnaText pour le corpus de forage.

- étudier le vocabulaire (trié par ordre de fréquence) ;
- extraire le vocabulaire spécifique ;

Formes spécifiques								
Show	10	entries	Search:	<input type="text"/>	Copy	CSV	Excel	Print
Rang	Lemme	Fréquence	CorpusRef (par million)	LogLike (spécificité)				
1	hydrocarbure	895	0.23	16819.264				
2	forage	856	0.645	14244.613				
3	gaz	1541	34.835	13530.293				
4	production	794	15.34	7424.028				
5	hydrate	304	0.12	5451.394				
6	gisement	347	0.955	4874.831				
7	pression	592	31.03	3853.107				
8	bassin	443	14.74	3382.569				
9	puits	460	19.07	3252.915				
10	sismique	230	1.11	2972.714				

Showing 1 to 10 of 101 entries

◀ First ◀ Previous Next ▶ Last ▶

Figure 11 : Liste des formes spécifiques provenant de corpus de forage extraits par AnaText

- extraire le vocabulaire catégorisé par nom, verbe, adjectif, adverbe ;

Verbes lemmatisés classés par fréquence décroissante								
Show	10	entries	Search:	<input type="text"/>	Copy	CSV	Excel	Print
Rang	Lemme	Fréquence						
1	être	7246						
2	avoir	1381						
3	pouvoir	1377						
4	permettre	677						
5	utiliser	543						
6	devoir	462						
7	faire	419						
8	figurer	381						
9	mettre	293						
10	produire	268						

Showing 1 to 10 of 1,697 entries

◀ First ◀ Previous Next ▶ Last ▶

Figure 12 : Liste des verbes lemmatisés provenant de corpus de forage extraits par AnaText

- rechercher des concordances ;
- rechercher des patterns (du type : DET ADJ NOM).

Recherche de patterns

Entrez un pattern composé de formes ou d'étiquettes

Mot Entier Sensibilité à la casse Recherche de lemme

Etiquettes reconnues : NOM, VER, ADJ, ADV, DET, PRO, PRE, CON, NUM (nombre), PHR (fin de phrase), PON (ponctuation)

Show entries Search:

Segment	Fréquence
gaz naturel	246
matière organique	143
système pétrolier	88
fracturation hydraulique	80
bassins sédimentaires	41
masse volumique	41
hydrocarbures liquides	37
systèmes pétroliers	35
gaz acide	34
offshore profond	33

Showing 1 to 10 of 8,680 entries

Figure 13 : Liste des patterns du type : NOM ADJ trouvés par AnaText dans le corpus de forage.

Les fréquences de référence ont été tirées de Lexique.org pour le français. Son corpus est construit de deux sous-corpus : le corpus de textes de Frantext et le corpus de sous-titres de films. Le corpus de textes est composé de 218 textes littéraires (romans) publiés entre 1950 et 2000. Il est donc composé de 14,7 millions d'items. Le corpus de sous-titres est composé des sous-titres de 9474 films ou saisons de séries qui représentent en totalité 50 millions de mots.

Nous avons déjà mentionné que TermoStat donne la possibilité d'examiner des mots analysés dans leurs contextes, mais nous trouvons que ce n'est pas suffisant dans notre cas. C'est pour cette raison que nous avons décidé d'utiliser AnaText. Pour nous assurer que TermoStat n'omettra aucune des phrases contenant les verbes qui nous intéressent et que nous étudierons tous.

Chapitre 6. Méthodologie

Dans ce chapitre, nous présenterons la méthodologie de notre travail. Nous expliquerons comment nous avons choisi les verbes et nous donnerons tous les détails de notre analyse.

1. *Extraction des candidats-termes*

Nous avons soumis notre corpus à TermoStat et nous avons obtenu trois listes de candidats : verbes (195 propositions de TermoStat), noms simples (88 propositions) et noms complexes (188 propositions). Comme nous avons déjà mentionné, TermoStat est un outil automatique, donc il peut engendrer des erreurs. Pour cette raison, en étant que spécialiste du domaine analysé, nous avons examiné toutes les trois listes obtenues pour retenir les termes valables.

Pendant l'évaluation des listes, nous avons rencontré deux problèmes majeurs :

1. très souvent, les textes portant sur le forage possèdent une quantité significative des mots d'origine anglaise qui ne sont pas évidemment présents dans le corpus de référence de TermoStat. À cause de cela, les verbes comme ceux énumérés ci-dessous ont été choisis par TermoStat comme les candidats-termes et placés très haut dans son classement - ils se trouvaient dans les quinze premiers :
 - a. *figurer* (le cas de nom *figure* erronément réduit aux formes verbales),
 - b. *per* (provenant de forme latine comme *per ascensum* ou *per laterum*),
 - c. *naphta* (ici dans le sens le *naphte*),
 - d. *riser* (le nom anglais pour *tube prolongateur*) ;
2. l'étiquetage erroné, dans ce cas le comptage des participes passés à valeur d'adjectif comme les verbes. C'est surtout le cas du verbe *dissoudre*. Nous regarderons ce problème plus précisément dans la partie *Analyse des contextes des verbes choisis*.

Enfin, après notre intervention, nous avons obtenu la liste qui comprend 153 verbes. En ce qui concerne deux autres listes : de termes simples et de termes complexes, nous les avons extraites à des fins de comparaison avec les verbes choisis, donc leur évaluation exacte n'est pas requise.

2. *Sélection des verbes à analyser*

Nous avons décidé de limiter notre analyse à 19 verbes. Nous avons choisi les verbes selon deux critères principaux. Le facteur déterminant est le taux de spécificité calculé par TermoStat. Le calcul de spécificité permet d'observer et de comparer le comportement du lexique de corpus analysé avec le comportement du lexique de corpus de référence (TermoStat, Drouin 2003). Pour notre corpus, le taux de spécificité le plus élevé est égal à 45,76. C'est le résultat de verbe « forer » qui appartient à la même famille de mots que le nom de domaine analysé « forage ». Pour la spécificité de la valeur la plus faible, nous avons décidé de ne pas dépasser 10. Au final, les verbes que nous avons choisis appartiennent aux quarante premiers dans la liste de TermoStat, avec le taux de spécificité inférieur à 13.

Deuxièmement, nous avons choisi les verbes qui sont fortement liés au domaine, comme *forer*, *exploiter* ou *excaver* et donc peuvent être spécialisés, ainsi que les verbes qui, selon notre opinion, peuvent être intéressants et se spécialiser grâce à leur contexte et aux mots présents dans leur voisinage, comme *produire*, *générer* ou *migrer*.

De cette manière, nous avons obtenu la liste finale que nous présentons dans le tableau suivant.

	verbe	spécificité	nombre d'occurrences (selon TermoStat / après vérification)	équivalent nominal	nombre d'occurrences
1.	forer	45,76	34 / 34	foration, forage	10, 856
2.	varier	22,5	65 / 63	-	-
3.	produire	21,79	132 / 126	production	794
4.	optimiser	19,72	22 / 22	-	-
5.	diminuer	18,75	62 / 60	-	-
6.	dissoudre	18,09	34 / 5	-	-
7.	adsorber	18	6 / 6	adsorption	24
8.	générer	17,93	44 / 44	-	-
9.	désacidifier	16,93	6 / 6	désacidification	33
10.	augmenter	16,66	107 / 105	-	-
11.	sédimenter	15,73	5 / 5	sédimentation	24
12.	visualiser	15,06	12 / 12	-	-
13.	exploiter	13,5	42 / 42	exploitation	206

14.	carotter	13,44	6 / 6	-	-
15.	excaver	13,40	3 / 3	excavation	13
16.	injecter	13,34	17 / 17	injection, réinjection	167, 12
17.	migrer	13,31	9 / 9	migration	204
18.	déshydrater	13,19	4 / 4	déshydratation	31
19.	extraire	13,07	20 / 19	extraction	61

Tableau 4 : Liste des verbes sélectionnés pour l'analyse.

3. Analyse des contextes des verbes choisis

L'étape suivante a consisté à examiner les contextes de toutes les occurrences des verbes choisis, la subdivision des éléments de ces occurrences en catégories : sujet, verbe, objet, complément et l'élaboration d'un modèle pour classer les verbes afin de faciliter la recherche de leurs équivalents dans d'autres langues. Dans cette optique, nous avons utilisé deux modèles d'analyse des verbes : la ressource lexicographique *Les verbes du français* (LVF) et le modèle *Courpus Pattern Analysis* (CPA).

3.1. LVF

Comme nous l'avons déjà dit, Dubois et Dubois-Charlier, dans leur dictionnaire *Les verbes français*, proposent la classification des verbes français à la base de leurs propriétés sémantiques et syntaxiques. Dans la mesure où nous ne savions pas encore quelle était la nature des constructions et des sens des verbes qui font l'objet de notre analyse, nous avons décidé que ce serait cette liaison de deux propriétés que nous utiliserions pour notre analyse. LVF montre la relation entre les schèmes de construction de verbe et leur sémantique. Toutes les deux contribuent à la construction du sens (chapitre 3, tableau 2, p. 27).

Nous avons commencé par vérifier si tous les verbes choisis étaient présents dans la ressource LVF. Après avoir confirmé cet état, nous avons listé toutes les entrées des verbes et décodé leurs abréviations pour être sûrs d'avoir bien compris le système.

Par la suite, nous avons préparé un document pour chaque verbe avec toutes ses occurrences. Nous avons vérifié ces occurrences pour nous assurer que chaque exemple est l'exemple d'un verbe et pas sa dérivation. Malgré la présence de la liste des variantes orthographiques, nous pourrions nous trouver face à l'étiquetage erroné (l'adjectif classé dans un groupe de verbes) et, ce qui est plus important dans notre cas, la polysémie des verbes. Il est donc très important d'identifier les sens de chaque verbe de notre liste. Dans le cas des doutes liés à la fixation de la signification des verbes et quand certaines subtilités

étaient difficiles à appréhender pour nous en tant qu'une personne pour qui le français n'est pas la langue maternelle, nous avons recouru au dictionnaire TLFi *Trésor de la Langue Française informatisé*.

Chaque phrase a été examinée selon les critères de LVF et à chaque phrase a été attribué :

- Domaine - domaine principal d'emploi du verbe et niveau de langue ;
- Classe : classe sémantique de l'entrée verbale ;
- Opérateur : interprétation au moyen d'un schéma syntactico-sémantique ;
- Sens : sens donné par un synonyme, un parasynonyme, une définition ou une explication ;
- Phrase : phrase simple illustrant le sens et la construction syntaxique ;
- Construction : schèmes syntaxiques du verbe ;

Pour chaque verbe, nous avons obtenu un tableau (annexe 1) contenant les entrées du dictionnaire LVF et les entrées résultant de notre propre analyse.

Ce système nous donne la possibilité d'observer les différences entre les constructions et les significations existant dans le LVF et observées dans notre corpus. Nous avons cherché toutes les différences possibles à repérer, en particulier les domaines, les sens et les schèmes. Ces différences nous ont aidés à trouver tous les exemples spécifiques au corpus du forage analysé.

Voyons l'exemple du verbe *générer* dans le tableau 5. LVF propose seulement une signification de ce verbe, appartenant au domaine de la sociologie qui est *produire* dans le sens de réaliser une action abstraite. Les schèmes proposés sont : *transitif direct humain chose* et *pronominal chose*. Par contre, dans notre corpus, pour le verbe *générer* nous n'avons pas trouvé l'exemple indiqué dans le LVF. Dans notre cas, *générer* fait préférentiellement partie des domaines de la géologie, du pétrole et de l'industrie. De plus, nous ne trouvons pas d'exemples de phrases pronominales et dans la plupart des cas le sujet de ce verbe n'est pas humain, ce sont les procédés et les substances qui prennent le rôle d'un agent.

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	sociologie : littéraire	réaliser une	réaliser une action	produire	On g~ l'inflation en haussant les	transitif direct

	(SOct)	action (R4b.1)	(fac) abstraite		prix.	humain chose (T1300), pronominal chose (P3000)
Corpus de forage	géologie	réaliser une action	fabriquer un produit	produire	Ils g~ des microfissures dans la roche en injectant de l' eau sous pression dans le puits : c'est la fracturation.	transitif direct humain chose
					La matière organique g~ des hydrocarbures.	transitif direct chose chose

Tableau 5 : L'analyse du verbe « générer » selon le modèle lexicographique LVF.

LVF est une ressource linguistique qui décrit les emplois des verbes dans la langue générale. Il est idéal comme le point de départ. Il nous a permis de remarquer certaines spécificités de chaque verbe et de confirmer ou de lever nos soupçons sur leur statut spécialisé. Cependant, ce système est centré sur les verbes et ignore leurs arguments. Dans le cas de *générer*, nous avons noté que sur base d'observations menées nous sommes capables de décrire ce verbe avec le schème suivant :

suj[CHOSE] (POUVOIR) **générer**_verbe obj[CHOSE]

Comme il a été dit précédemment, le sujet de verbe *générer* le plus présent dans le corpus du forage est un sujet non humain représenté par les noms comme *roche*, *roche-mère*, *matière organique*.

De plus, les schèmes utilisés dans la ressource LVF ne nous donnent pas la possibilité de décrire toutes les dépendances observées dans notre corpus, telles que la présence des modélisateurs comme *pouvoir* ou *permettre de*.

Pour ces raisons, nous avons choisi de changer notre approche. En nous inspirant du travail de Yan (2017), nous utiliserons LVF comme une base lexicographique de la langue générale pour pouvoir plus tard tirer des conclusions sur le statut terminologique de nos verbes. Et pour la description des verbes, nous appliquerons un autre modèle, le modèle CPA que nous présenterons dans la section suivante.

3.2. CPA

Rappelons que le modèle CPA est un modèle utilisé pour analyser les propriétés sémantiques et syntaxiques des verbes dans leurs contextes (un modèle *corpus-driven*). À l'aide des patrons, il offre la possibilité de décrire les emplois exacts des verbes utilisés dans les discours analysés. Pour identifier les patrons dans un corpus, nous sommes obligés de prendre en compte les constructions syntaxiques récurrentes et aussi les cooccurrences entourant les verbes. Celle-ci est effectuée par les catégories sémantiques qui caractérisent des arguments nominaux des verbes analysés dans un corpus. Ces deux caractéristiques font que le modèle CPA nous semble approprié à nos besoins : à la détermination du statut terminologique des verbes que nous avons choisis pour notre analyse et à la préparation du modèle d'un tableau méthodologique applicable à d'autres langues et donc facilitant la traduction des expressions typiques du domaine du forage.

Comme explique Yan (2017, p.109), un rôle important du modèle CPA est d'identifier les préférences sémantiques des verbes par rapport à leurs arguments, ce qui permet de comprendre le sens du verbe utilisé dans la phrase. Dans ce but, il faut trouver les unités lexicales se trouvant en position de sujet et d'objet d'une phrase et leur attribuer un trait sémantique. Les unités lexicales qui se répètent, qui ont les caractères communs, et qui sont des cohyponymes, on peut les mettre dans un groupe et leur donner des étiquettes sémantiques. Lors de son analyse, Hanks a utilisé le programme SketchEngine qui cherche automatiquement les cooccurrents d'un verbe dans une position donnée, et ensuite il leur a choisi et donné des étiquettes convenables.

Comme dans le CPA, nous construirons aussi des types sémantiques, une ontologie pour pouvoir après construire des patrons. Nous effectuerons notre travail manuellement. Au départ, nous identifierons tous les sujets et les objets des mêmes phrases que nous avons utilisés pour l'analyse selon les objectifs de la ressource LVF. Sur cette base, nous construirons notre propre ontologie, spécifique à notre corpus (Chapitre 7). Ensuite, nous identifierons les structures syntaxiques de chaque phrase et les décrirons (Chapitre 8). Nous calculerons le pourcentage des structures identifiées pour les verbes donnés et sur cette base nous créerons des patrons pour chaque verbe, lesquels seront présentés à la fin du Chapitre 8. Finalement, nous analyserons le taux de spécification de chaque verbe à la base de la ressource LVF et des patrons établis (Chapitre 9).

Partie 3

-

ANALYSES ET RÉSULTATS

Chapitre 7. L'Ontologie

L'ontologie ou la classification sémantique des arguments des verbes est indispensable et elle nous intéresse sur deux plans : sur le plan linguistique, elle nous servira pour pouvoir construire et décrire une base syntaxique des verbes analysés à la manière du modèle CPA ; sur le plan terminologique, elle nous permettra d'analyser le taux de spécialisation de ces verbes.

1. *Collection des arguments*

Nous avons commencé notre travail par regrouper tous les sujets et les objets de nos phrases dans un document et par former une liste de mots. Chaque entrée a été lemmatisée et ramenée à sa version la plus simple, comme dans le cas du mot *densité*. Ce mot est compris dans plusieurs constructions privilégiées comme *densité API*, *densité de boue* ou *densité de fluide*. Ces expressions sont plus précises, mais elles appartiennent au même concept général. Après ce nettoyage, nous avons obtenu une liste de 375 arguments, dont ceux avec le plus grand nombre d'occurrences sont : *gaz* (59 occurrences), *hydrocarbure* (41 occurrences) et *production* (16 occurrences).

2. *Création des classes et des sous-classes*

Nous avons décidé de suivre le travail de Yan (2017) et de créer deux niveaux de notre ontologie : classes et sous-classes. Le premier pas a été la construction des sous-classes. À cet égard, nous avons établi des définitions de chaque nom figuré sur notre liste pour pouvoir choisir un mot désignant notre argument.

2.1. *Création des sous-classes*

Dans un premier temps, nous avons vérifié des définitions de nos mots dans le dictionnaire TLFi *Trésor de la Langue française informatisé*. Dans la plupart des cas, ces définitions ont été trop spécifiques et au lieu de nous donner une idée générale d'un mot recherché, elles ont créé plus de doute. C'est pour cela que nous avons aussi utilisé la banque de données terminologiques et linguistiques TERMIUM Plus.

TERMIUM Plus, originellement développée à l'Université de Montréal, est une base de données du gouvernement du Canada qui depuis 2009 est accessible gratuitement en ligne (<http://www.btb.termiumplus.gc.ca/>). Cette banque possède près de quatre

millions de termes et donne la possibilité d'interroger des fiches terminologiques dans les deux langues officielles du Canada : l'anglais et le français.

Pour nous assurer que nos sous-classes ont été établies proprement et qu'elles cohérentes sémantiquement à la fin de cette étape nous avons consulté les listes de l'ontologie de *Pattern Dictionary of English Verb* (<http://pdev.org.uk/#onto>) et l'ontologie créée par Yan (2017) pour observer les ressemblances entre ces deux ressources et notre travail. Aussi pour cette raison, nous trouvons que les mots appartenant à la même classe (Hatier *et al.*, 2016 in Yan, 2017) doivent être en relation de cohyponymie et correspondre à la définition donnée pour une classe.

Pour visualiser le travail que nous avons effectué, nous présentons l'analyse du mot *densité* et *masse*:

1. La définition de *densité* selon le TLFi :

PHYSIQUE – MÉCANIQUE

– *Densité (relative)*. Rapport qui existe entre la masse du volume d'un corps et la masse du même volume d'un corps de référence (eau pour les corps solides et liquides, air pour les corps gazeux), dans les mêmes conditions de température et de pression. *La densité du métal déposé [avec le procédé Schoop] est toujours inférieure à celle du fil employé* (Gasnier, Dépôts métall., 1927, p. 151).

– *Densité (absolue)*. Rapport de la masse d'un corps à son volume. Synon. *masse spécifique*

2. La définition de *masse* selon TLFi :

[Caractéristique physique des choses]

1. *La masse d'une chose, d'un objet*. L'ensemble physique que constitue dans l'espace cette chose, cet objet, caractérisé par son volume, sa rigidité, sa densité, etc. *Le choc, très doux par ce calme, était tout à fait amorti; il avait été si faible même, que vraiment il semblait que cet autre navire n'eût pas de masse et qu'il fût une chose molle, presque sans poids* (Loti, *Pêch. Isl.*, 1886, p. 188). *Un grand viaduc enjambe, pour le chemin de fer, la vallée (...). Sa réalité, sa masse, se dissout dans la brume* (J.-R. Bloch, *Dest. du S.*, 1931, p. 47)

3. La définition de *densité* selon TERMIUM Plus :

FICHE TERMINOLOGIQUE 4

Domaine(s) : Mesures et analyse (Sciences) ; Poids et charges (Transports)

Synonyme : masse volumique / masse spécifique / densité absolue / poids spécifique

Définition : **Grandeur** qui correspond à la masse [...] d'un corps divisée par son volume.

Observation : L'unité de mesure de la masse volumique dans le Système international est le kilogramme par mètre cube (1 kg/m^3) ; densité : terme que certains auteurs emploient comme synonyme de «masse volumique» même si le terme «densité» désigne d'abord la notion de «densité relative» (rapport d'une masse volumique d'un corps à la masse volumique d'un corps de référence).

4. La définition de *masse* selon TERMIUM Plus :

Domaine(s) : Physique ; **Mesures** et analyse (Sciences)

Définition : **Quantité de matière** d'un corps; rapport constant qui existe entre les forces qui sont appliquées à un corps et les accélérations correspondantes.

Observation : Aux États-Unis plus qu'ailleurs, on confond souvent poids et masse; [...] le poids est une grandeur de la nature d'une force, c'est la force due principalement au champ gravitationnel de la Terre et à la force centrifuge due à la rotation de la Terre, variables d'un lieu à l'autre [...] La masse d'un corps au contraire reste constante [...] en tout point de l'Univers [...] ; Le symbole SI de cette grandeur est un «m» minuscule italique ; l'unité SI de masse est le kilogramme (kg)

Comme nous pouvons le voir, dans le cas de *densité*, la définition du TLFi est juste, mais elle inclut seulement les informations détaillées, à part de nom du domaine général: mécanique. Par contre, la fiche de TERMIUM Plus nous fournit : le nom du domaine, des synonymes, la définition et les observations, et elle contient deux informations sur la globalité de *densité*, c'est le nom du domaine : *mesures* et le premier mot de la définition : *grandeur*. Dans ce contexte, nous avons qualifié la densité dans *mesures (quantités)* et plus précisément, dans *propriétés physiques*. En ce qui concerne *masse*, la tâche a été plus simple, parce que les deux définitions comportent les indications sur le groupe auquel le mot appartient.

2.2. *Création des classes*

Les classes ont été construites de même façon que les sous-classes, à la base des définitions de TLFi et de TERMIUM Plus. Il a fallu uniquement réunir les sous-classes dans les groupes et trouver pour eux des noms communs.

En dernier ressort, le classement construit comporte 12 classes et 57 sous-classes. Le nombre des exemples figurant dans différentes sous-classes n'est pas le même, il y a des sous-classes qui contiennent moins que 2 exemples. Pour cela, et aussi parce que notre corpus n'est pas grand : environ 275 000 mots, nous avons choisi d'inclure dans l'ontologie seulement les sous-classes qui possèdent le seuil de fréquence à plus de dix

occurrences. Cette liste pourrait être enrichie avec l'élargissement du corpus et complétée par des équivalents d'autres langues. L'ontologie complète se trouve dans l'annexe 3.

CLASSE	SOUS-CLASSE	EXEMPLE
espace	lieu, partie	<i>milieu, trou, zone, chemin de migration</i>
humain	collectivité, individu	<i>entreprise, industrie, France, usine, explorateur, géologue, opérateur, taillant</i>
objet scientifique	cas, donné, méthode_1, relation,	<i>cas, voie, phénomène, projet, méthode, modèle, effet</i>
outil	appareil, construction, création, élément d'appareil, équipement, machine,	<i>outil, FLNG, gas lift, plate-forme, unité de séparation, tamis, tarière, tube, navire de forage</i>
procédé	action, augmentation, changement, création, enregistrement, forage, méthode_2, réaction chimique_1, procédé technologique	<i>échauffement, craquage, moussage, fracturation hydraulique, production, auscultation, underbalanced drilling, forage, réaction, combustion, extraction, récupération</i>
processus	changement, processus chimique, processus physique, processus évolutif, processus géologique, réaction chimique_2	<i>gazéification, accroissement, augmentation, développement, diminution, multiplication, catagenèse, érosion, migration, absorption, dégagement de chaleur</i>
qualité	caractéristique, possibilité, surnombre	<i>absence, efficacité, froid, déséquilibre, présence, trajectoire, nuisance, option, risque, redondance des capteurs</i>
quantité	collection, entité, grandeur, mesure, quantité abstraite	<i>quantité, rapport, énergie, rendement, teneur, pression, composition, coût, niveau, nombre, PCI, valeur, surface de contact, débit, densité, flux, température, vitesse, proportion, part</i>
sol	formation géologique, roche, trou	<i>terrain, champ, gisement, découverte, couche, réservoir, ressource, poussière, roche, roche-mère, microfissure</i>
substance	combustible, composé	<i>carburant, CO₂, eau,</i>

	chimique, élément chimique, fluide, liquide, solide	<i>fraction, hydrocarbure, méthane, kérogène, gaz, filtrat, matière organique, catalyseur</i>
--	--	---

Tableau 6 : Les types sémantiques des noms utilisés dans l'analyse des verbes.

Chapitre 8. Formes syntaxiques spécifiques au corpus de forage

Dans la plupart des cas, la structure syntaxique d'un patron CPA est une forme standard du verbe sous forme de la voix active. Mais ce modèle permet de décrire toutes les structures syntaxiques rencontrées dans un texte, parce que, comme l'a dit El Maarouf (2011, in Yan 2017, p.115), un patron du verbe peut « correspondre à de nombreuses réalisations (variations) en contexte, mais conserve le même sens ». Dans ce chapitre, nous présentons toutes les constructions syntaxiques que nous avons trouvées lors de l'analyse de nos 19 verbes.

1. *Constructions actives*

La voix active est le trait grammatical le plus fréquent dans notre corpus. Les constructions actives s'opposent à la voix passive et possèdent la forme basique :

sujet (agent) + verbe + objet (patient)

Tous nos verbes possèdent des exemples d'utilisation de cette construction. On distingue trois types de constructions actives : transitive, intransitive et attributives.

1.1. *Sujet – Verbe – Objet (SVO)*

Les constructions transitives sont construites avec un complément d'objet. Ce type de construction est divisé en constructions transitives directes, indirectes et ditransitives.

On parle de construction transitive directe si le complément est directement relié au verbe – complément d'objet direct (*Le taillant pilote et son aléteur forent un trou.*) et de construction transitive indirecte si le complément est introduit par une préposition – complément d'objet indirect (*Il pense aux vacances ; Il va à Paris*). La construction ditransitive est tout simplement construite avec deux compléments dont l'un est généralement direct et l'autre indirect (*Grammaire méthodique du français*, 1997 p. 394).

25% des phrases que nous avons analysées sont des phrases de type transitives directes. Chaque verbe, à part *sédimenter* et *varier*, possède au moins une phrase de construction SVO.

1.1.1. Constructions causatives

Parmi nos occurrences nous avons remarqué des exemples de constructions causatives. Une phrase où le référent du sujet représente la cause ou l'agent du procès décrit par la phrase originale (*Grammaire méthodique du français*, 1997). Cette construction soumet une phrase de départ à une double modification. Le sujet nominal de la phrase est postposé au verbe qui lui-même se met à l'infinitif : *Le géologue américain Dad Joiner fait forer la découverte d' East Texas*. Les verbes de constructions infinitives sont : *faire, voir, entendre, sentir, laisser*.

Dans notre corpus, trois verbes possèdent des constructions causatives : *forer, varier* et *diminuer*. Le verbe *diminuer* apparaît avec le verbe causatif *voir*. Dans le cas de *forer* et de *diminuer*, les phrases sont construites à l'aide du verbe *faire*. Regardons l'exemple suivant :

[1] L'huile devenue plus légère voit sa viscosité diminuer. (*Geologie du pétrole*, Biteau&Baudin, 2017)

Cette phrase peut être décrite avec le cadre :

suj[NON-hum] VOIR obj[NON-hum] diminuer_verbe

Selon *Grammaire méthodique du français* (1997, p. 411-414) dans les constructions causatives [*faire* + VInf] *faire* et le verbe à l'infinitif ont des sujets distincts, mais ils ne peuvent être séparés ni par le sujet ni par le complément du second.

1.1.2. Construction périphrastique causative *permettre de*

Une autre construction causative découverte pendant notre analyse des verbes est la construction *permettre de*. Cette construction comporte un sujet inanimé qui renvoie souvent aux activités par la préposition *de*, comme le décrit Yan (2017). Les constructions périphrastiques correspondent à « une relation de cause à effet » (Bak& Novakova, 2013, p. 101), comme nous pouvons le voir dans les exemples suivants :

[2] Un **forage horizontal** permet de devoir **forer** un nombre plus restreint de **puits**. (*Gaz non conventionnels*, Favreau&Rouaud, 2012)

[3] Les **caractéristiques physico-chimiques** des hydrocarbures « lourds » ne permettent pas de **les produire** avec des méthodes conventionnelles. (*Hydrocarbures de roche-mère*, Vially, 2014)

- [4] La **connaissance de la composition** d'un pétrole brut est importante pour le raffineur, car elle lui permet d'optimiser le choix du pétrole brut. (*Pétrole*, Travers&Tocqué, 2016)

Le sujet inanimé, souvent un objet (*caractéristique, connaissance*) ou un procédé (*forage*) détermine les actions à réaliser, comme *forer, produire* ou *optimiser*. La fonction de cette construction est donc de fournir des explications pour justifier les actions entreprises et celle-ci peut être décrite comme suit :

suj[NON-hum] PERMETTRE DE forer|produire|optimiser_verbe obj[NON-hum]

1.2. Sujet – Verbe

Les verbes des constructions intransitives n'ont pas de complément d'objet. Ils ne s'emploient qu'avec des compléments circonstanciels. Dans notre corpus, cinq verbes ont la forme intransitive, ce sont :

- *sédimenter* : 100% de ses phrases ont la forme intransitive et il possède seulement six occurrences dans notre corpus ;
- *varie* : 95% (60/63 occurrences) de ses phrases ont la forme intransitive ;
- *augmenter* : 41% (43/105 occurrences) ;
- *diminuer* : 40% (24/60 occurrences) ;
- *migrer* : 33,33% (3/9) ;

1.3. Sujet – Verbe – Attribut

Les constructions attributives sont des constructions formées d'un verbe introducteur qui est le verbe *être* ou un verbe d'état et d'un attribut du sujet. Un attribut du sujet est un prédicat qui exprime une propriété, un état ou une catégorisation du sujet comme *Les gaz sont dissous* ou *Le prix est suffisant*. Certains verbes transitifs peuvent aussi établir le même type de rapport (*on trouve les prix suffisants.*), mais ils ne sont pas présents dans notre corpus, donc nous ne les discutons pas.

Nous mentionnons cette construction verbale à cause du verbe *dissoudre*. TermoStat a choisi ce verbe comme un candidat terme en lui accordant vingtième place (seizième après le nettoyage) sur la liste des candidats termes avec 34 occurrences. Après avoir effectué le nettoyage, nous avons constaté que parmi les phrases choisies par TermoStat seulement cinq contiennent le verbe *dissoudre*, les autres incluent la forme

adjectivale de ce verbe, sont participe passé *dissous*. Une chose nous a frappé, cependant : cinq phrases attributives *être dissous*. Dans ces phrases le verbe exerce une fonction d'adjectif et décrit l'état de sujet :

[5] Les bruts **sont** généralement **dissous** dans des solvants organiques puis analysés soit par chromatographie en phase gazeuse (GC, gas chromatography en anglais) , soit par chromatographie en phase gazeuse couplée à la spectrométrie de masse (GC-MS, gas chromatography - mass spectrometry en anglais). (*Geologie du pétrole*, Biteau&Baudin, 2017)

[6] Le gaz peut **être dissous** dans le pétrole. (*Gaz naturel*, Kimmerlin, 2010)

Pour ce cas, nous avons construit le patron suivant :

[**substance=gaz**] est dissous DANS_complément_circonstanciel

Ce qui est aussi intéressant, le verbe *dissoudre* se lie toujours avec un complément circonstanciel de lieu *dans*.

2. *Constructions passives et moyennes*

Les constructions passives ne sont pas très présentes dans nos verbes choisis. En général, la voix passive est utilisée pour mettre en évidence la personne ou l'objet qui subit une action, l'objet devient le sujet de la phrase. Nous pensons que cette absence des constructions passives parmi nos verbes est le résultat de présence de la voix moyenne.

Il n'est possible de trouver la notion de la voix moyenne ni dans les manuels traditionnels de la langue française ni dans la *Grammaire méthodique du français* (1997) parce que selon les traditionalistes la voix moyenne n'existe pas en français et ces exemples sont qualifiés comme des types de passifs non-agentifs. Néanmoins, dans notre corpus nous avons trouvé des phrases dites pronominales, mais dans lesquelles nous avons constaté l'absence de sujet réel pour le verbe, en d'autres mots l'action d'un sujet n'est pas le résultat de son intention (*la contamination se produit à l'occasion de, les roches bitumineuses s'exploitent en carrières*).

Frias-Conde (2006) constate que quand on parle de la voix moyenne dans les langues romanes, on parle des phrases qui sont construites avec un morphème libre *se* d'une apparente nature pronominale, mais qui dans ce cas fonctionne comme une particule verbale. Ce sont donc les phrases conjuguées à la troisième personne du singulière ou du plurielle, accompagnées d'une particule *se* et suivies d'un objet.

L'utilisation de la voix moyenne provoque une réorganisation de la relation entre les fonctions syntaxiques et sémantiques d'une phrase. Elle cumule les rôles agent et patient sur le sujet, en soulignant par là même l'importance du sujet.

Dans notre corpus seulement deux verbes possèdent les constructions de la voix moyenne : *produire* et *exploiter*. Dans le cas du verbe *produire*, la voix moyenne constitue 15% (20/126) des phrases analysées et est liée à des sujets décrivant les actions, comme : *réaction, changement, création, migration, ségrégation*. Cette opération met en position de sujet ces actions, produisant un effet de renforcement de cette action en insistant sur sa signification sémantique :

[7] Une ségrégation chimique **se produit** entre la roche-mère et les niveaux réservoirs qui lui sont contigus. (*Geologie du pétrole*, Biteau&Baudin, 2017)

[8] Cette contamination **se produit** à l'occasion du percement d'un forage sous l'effet de la corrosion. (*Geologie du pétrole*, Biteau&Baudin, 2017)

Ces phrases peuvent être décrites à l'aide du cadre suivant :

subj[NON-hum] SE produire

La même chose s'applique au verbe *exploiter*. Cependant, le nombre d'exemples n'est pas si représentatif, on note ici seulement deux cas d'utilisation de la voix moyenne :

[9] Il doit ainsi intégrer les métiers de mineur et de carrier par cette évolution récente de la recherche et de l'exploitation des roches bitumineuses (qui **s'exploitent** en carrières). (*Geologie du pétrole*, Biteau&Baudin, 2017)

[10](...) et les charbons qui **s'exploitent** sous forme de mine et sont ensuite utilisés comme combustible. (*Gaz non conventionnels*, Favreau&Rouaud, 2012)

3. Construction spécifique : nom à Verbe

Un autre type de construction relevé pendant l'analyse de notre corpus est une construction *nom à verbe*. Cette construction inclut un objet (patient) inanimé qui à chaque fois renvoie aux objets appartenant aux classes *sols* et *espace*. Cette construction apparaît dans les phrases des verbes : *excaver, forer, carotter, exploiter, extraire*. Pour cette construction, les verbes qui en possèdent dans leurs occurrences sont des cohyponymes, mais aussi des synonymes, sauf *exploiter*, et ils peuvent être définis comme : opération consistant à creuser dans le sol, comme le témoignent les exemples ci-dessous :

[11] Une fois le titre minier obtenu, la décision principale sera de faire un inventaire exhaustif et de choisir les prospects à **forer**. (*Geologie du pétrole*, Biteau&Baudin, 2017)

[12] La machine comporte une scie qui se déplace sur un châssis adapté à la forme de la section à **excaver**. (*Tunneliers*, Gesta, 1992)

[13] La grosseur des pierres est surtout choisie pour tenir compte de la dureté des terrains à **carotter**. (*Forage et carottage dans les roches*, Robert, 2010)

Le sens du verbe *exploiter* dans ce contexte peut être expliqué par un mot *utiliser*.

[14] Dans ce dernier cas, elles permettent d'obtenir des données très importantes concernant le réservoir géothermal à **exploiter**. (*Énergie géothermique*, Laplaige&Lemale, 2001)

Tous ces exemples s'inscrivent dans le cadre comme suit :

obj[**NON-hum**] À **forer|excaver|carotter|exploiter**_verbe

4. Construction en « pour » – complément circonstanciel

Les constructions en *pour* qui figurent dans notre corpus, construisent des compléments circonstanciels de but, mais nous les listons dans ce chapitre de notre mémoire à cause de leur fréquence. Ces constructions liées aux verbes choisis constituent 20% des phrases analysées.

Le complément circonstanciel de but nous renseigne sur la finalité et répond aux questions : *pourquoi ?*, *dans quel but ?* La fonction principale de cette construction est d'exprimer l'objectif que l'on souhaite atteindre, mais aussi de faire référence au thème évoqué dans le discours, parce qu'il ne faut pas oublier que les textes analysés sont les textes dans la grande partie dédiés aux apprenants de ce domaine. Il est introduit par les prépositions telles que : *pour*, *afin de*, *en vue de*, *destiné à*.

Comme nous l'avons mentionné, nous avons trouvé 127 cas des verbes que nous avons choisis pour analyser sous forme de complément circonstanciel. Voici quelques exemples :

[15] Ces outils sont recommandés **pour forer** dans les terrains tendres. (*Forage et carottage dans les roches*, Robert, 2010)

[16] Le pétrole peut subir dans un réservoir une altération thermique appelée craquage secondaire, qui correspond à la rupture des liaisons chimiques des hydrocarbures de poids moléculaire élevé **pour produire** des hydrocarbures de plus en plus légers. (*Geologie du pétrole*, Biteau&Baudin, 2017)

[17] Plusieurs grandes catégories d'additifs chimiques sont utilisées : - additifs permettant de contrôler le pH, la teneur en oxygène du mélange afin d'optimiser les qualités du fluide de fracturation. (*Gaz non conventionnels*, Favreau&Rouaud, 2012)

[18] Pour diminuer cette érosion, un deuxième tube (intérieur) est positionné à l'intérieur du carottier (tube extérieur). (*Forages et sondages*, Robert, 2011)

Étant donné que cette construction n'occupe qu'une place secondaire dans la phrase et peut être déplacée ou supprimée, elle n'a pas de sujet et s'inscrit dans le cadre :

POUR forer|produire|optimiser_ verbe obj[NON-hum]

Il est intéressant de remarquer que ces constructions ne sont presque pas présentes parmi les cooccurrences des verbes : *varier, adsorber, générer, sédimenter, excaver, injecter*, donc les verbes qui ont le taux élevé des constructions intransitives.

5. *Patrons des verbes extraits du corpus du forage*

Notre analyse nous a permis de distinguer 27 patrons des verbes analysés. Chaque verbe, à l'exception du verbe *visualiser*, possède au moins un patron. Nous n'avons pas pu créer de patron pour le verbe *visualiser*, parce qu'il présente une diversité trop grande au niveau des constructions syntaxiques et aussi des types sémantiques de ses arguments.

Nous présentons dans le tableau 7 les patrons de 18 verbes avec leurs types sémantiques et leurs fréquences dans le corpus. Dans les cas où c'était possible, à côté des types sémantiques, nous avons indiqué leurs exemples les plus fréquents.

patron	fréquence
[substance] se adsorbe	2/6 (33,3%)
[quantité=température] augmente	43/105 (41%)
[sol=sol, terrain] à carotter	3/6 (50%)
pour désacidifier [substance=gaz]	3/6 (50%)
désacidifier [substance=gaz]	3/6 (50%)
[procédé outil] permet de déshydrater [substance=gaz]	2/4 (50%)
[quantité] diminue	24/60 (40%)
[substance=gaz] est dissous	5/8 (62,5%)
[sol espace] à excaver	2/3 (66,7%)

[sol=ressource, réservoir] à exploiter	6/42 (6,5%)
pour exploiter [sol=gisement, découverte substance=gaz]	12/42 (28,6%)
construction impersonnelle extraire [substance=hydrocarbure, gaz]	7/19 (36,8%)
[substance sol] à extraire	6/19 (31,6%)
pour extraire [substance]	3/19 (15,8%)
[humain outil procédé] fore [espace quantité sol=puits]	8/34 (23,5%)
[espace quantité sol=puits] à forer	11/34 (32,4%)
pour forer [espace=terrain quantité sol]	7/34 (20,6%)
[procédé=réaction, forage substance=matière organique, kérogène] génère [substance=hydrocarbure, gaz]	30/44 (88,8%)
construction impersonnelle injecter [substance]	7/17 (41,2%)
[substance=hydrocarbure, gaz] migre	8/9 (88,9%)
pour optimiser [procédé quantité]	10/22 (45,5%)
[procédé=auscultation quantité] permet d'optimiser [procédé=production]	9/22 (86,4%)
[humain procédé processus sol substance] produit [substance=hydrocarbure, gaz]	47/125 (37,6%)
[procédé=réaction processus=migration] se produit	17/125 (13,6%)
pour produire [substance=hydrocarbure, gaz quantité=énergie, chaleur]	29/125 (23,2%)
[substance=matière organique] sédimente	6/6 (100%)
[quantité qualité] varie	58/63 (92,1%)

Tableau 7 Les patrons relevés du corpus du forage

Chapitre 9. Statut terminologique des verbes analysés

De nombreux auteurs ont proposé de classer les verbes selon leurs « niveaux de spécialisation ». Ces classifications s'appuient sur les classes des objets des arguments des verbes créés par Gross (1994). Dans notre mémoire, nous nous servons du travail de L'Homme (2012) sur les verbes originaires des corpus d'informatique et d'environnement.

L'Homme (2012) se base sur la proposition de Condamines (1993, in L'Homme 2012) d'utiliser la caractérisation syntaxico-sémantique des verbes et des noms. Condamines applique cette méthode au domaine bancaire, ce qui a permis de découvrir deux groupes de verbes :

- des verbes spécifiques au domaine, par exemple *débiter* ;
- des verbes connus, autrement dit généraux, qui possèdent un sens lié au domaine, par exemple *retirer*.

Lerat (2002, in L'Homme 2012) classe les verbes juridiques dans trois groupes : des verbes très spécialisés, par exemple *adjuger*, *abroger* ; des verbes supports, par exemple *mettre* dans *mettre aux voix* ; des verbes polysémiques dont au moins un sens est spécialisé, par exemple 'HUMAIN contracter (v. intr.)' et 'HUMAIN contracter (v. tr.) N <obligation>' relèvent du domaine juridique. Dans son travail, Lorente (2002, in L'Homme 2012) propose des critères syntaxiques, lexico-sémantiques et discursifs pour l'élaboration d'une typologie des verbes qui contiennent quatre groupes distincts :

1. les verbes discursifs (permettant d'articuler le discours) ;
2. les verbes connecteurs (exprimant des attributs, des équivalences, des relations de similarité ou de dépendance) ;
3. les verbes phraséologiques (se combinant avec un terme du domaine et exprimant avec ce terme, des connaissances spécialisées) ;
4. les verbes terminologiques (ayant un sens spécifique à l'intérieur d'un domaine spécialisé).

En ce qui concerne les critères, ce sont : « la fonction discursive du verbe, sa valeur spécialisée, le contenu sémantique du groupe verbal, la formation morphologique des verbes, leur sous-catégorisation, la relation sémantique qu'ils partagent avec le sujet et la sélection lexicale dont ils sont le pivot » (L'Homme, 2012, p. 101). Mais deux paramètres

de cette liste sont suffisants pour établir un statut de verbe, ce sont : le lien entre un verbe et des termes de nature nominale et la classe sémantique des verbes.

Après avoir comparé nos verbes aux entrées verbales répertoriées dans le LVF, nous avons constaté que les verbes forer, varier, optimiser, dissoudre, désacidifier, carotter et excaver ne se différencient pas en aucune manière des acceptions du LVF. Dans tous les autres, nous avons trouvé des différences au moins sur l'un des éléments suivants : domaine, sens ou schèmes. Cependant, ce constat n'est pas déterminant pour le statut de notre verbe. Il nous semble que les verbes comme forer, désacidifier ou excaver possèdent seulement une signification qui est la signification spécialisée et ils ne peuvent pas être utilisés différemment que dans un domaine de spécialité. Pour pouvoir définitivement confirmer ou infirmer le statut de nos verbes, dans la section suivante nous ferons appel à la structure argumentale de ces verbes et de leur appartenance à un réseau lexico-terminologique.

1. Modèle descriptif de la structure argumentale

En vue de créer un dictionnaire terminologique utile aux étudiants et aux traducteurs, nous avons dû nous concentrer sur les relations existant entre nos verbes et leurs arguments. Dans cette optique, nous considérons que l'analyse des associations lexicales est indispensable.

Sur le niveau des relations sémantiques, on distingue deux types de relations lexicales : relation syntagmatique et relation paradigmaticque. Comme explique Victorri (2010), si les mots apparaissent souvent ensemble, on parle de relation syntagmatique (*livre – lire* ou *livre – page*), et s'ils pouvaient être remplacés l'un par l'autre on parle de relation paradigmaticque (*livre – journal* ou *livre – bouquin*). Dans notre travail, nous sommes surtout intéressés par les relations syntagmatiques. Nous voulons savoir à quels arguments nos verbes ont tendance à s'associer.

Dans chaque langue, certains mots ont tendance à apparaître ensemble (*rendre visite, prêter attention*). L'interprétation sémantique de ces expressions se fait de façon compositionnelle, c'est-à-dire leur sens est prédictible à partir du sens des parties. Très souvent, ces expressions sont erronément appelées collocations. En cette matière, nous sommes fondamentalement d'accord avec Tutin et Grossmann (2002) et c'est pourquoi nous avons choisi de parler de cooccurrence lexicale privilégiée.

Comme nous avons déjà indiqué, TermoStat permet une recherche des bigrammes – une composition d’un verbe et d’un nom accompagnée de leur fréquence. Malheureusement, dans notre cas, TermoStat a effectué une analyse des bigrammes seulement pour des termes simples et il a trouvé des relations entre huit verbes et leurs objets.

Verbe	Nom	Fréquence	Score d'association
adsorber	-	-	-
augmenter	densité (objet)	4	18,61
carotter	-	-	-
désacidifier	gaz (objet)	5	30,51
déshydrater	-	-	-
diminuer	-	-	-
dissoudre	-	-	-
excaver	-	-	-
exploiter	gisement (objet)	5	25,99
extraire	-	-	-
forer	puits (objet)	4	17,47
générer	hydrocarbure (objet)	6	17,93
injecter	-	-	-
migrer	-	-	-
optimiser	production (objet)	4	17,4
produire	huile (objet)	4	12,42
	gaz (objet)	12	23,24
	hydrocarbure (objet)	12	32,93
sédimenter	-	-	-
varier	nature (objet)	4	16,04
visualiser	-	-	-

Tableau 8 Les bigrammes calculés par TermoStat pour le corpus du forage.

Nous n’avons pas été capables de calculer le log-likelihood nous-mêmes parce que nous n’avions pas accès aux informations et aux chiffres de corpus de référence. Néanmoins, nous souhaitons savoir quels arguments sont les plus souvent sélectionnés par nos verbes. De ce fait, nous avons utilisé les patrons établis au cours de notre analyse.

Pour pouvoir constater que le verbe est spécialisé ou non à la base de la structure argumentale, il faut définir quels mots peuvent être considérés comme des termes spécifiques à un domaine choisi. En conséquence, nous avons choisi plusieurs noms provenant de notre corpus que, selon notre opinion, sont étroitement liés au domaine du forage et qui sont eux-mêmes des termes (leurs fiches terminologiques extraits de TERMIUM Plus sont disponibles dans l’annexe 2). Les mots que nous avons choisis sont : *hydrocarbure, gaz, pétrole, matière organique, roche-mère, ressource, gisement,*

découverte, puits, forage et fracturation hydraulique. Si un verbe se lie préférentiellement à ces termes (on les trouve dans la position de ses sujets, ses objets ou ses compléments), ils peuvent être un indice sur le taux de spécialisation de ce verbe. Nous pouvons regarder de près ces verbes:

1. forer – comme le montre les bigrammes de TermoStat ainsi que nos patrons, *forer* choisit préférentiellement comme son objet le terme *puits* ;
2. varier – ce verbe se lie préférentiellement à des grandeurs physiques, comme *densité* ou *diamètre*, il ne présente aucun trait de spécialités ;
3. produire – dans la ressource LVF, nous n’avons pas trouvé de signification applicable à notre corpus. *Produire* montre le taux élevé de spécification. Comme le montrent les bigrammes et nos patrons, ses objets préférés sont *gaz*, *huile* et *hydrocarbure*. Aussi parmi ses sujets, on trouve des traces de domaines, comme des outils spécialisés, comme *FLNG* ou *plate-forme* ;
4. optimiser – ce verbe se lie préférentiellement à des procédés, surtout de *production* : *production de champs*, *production perdure*, mais il ne présente aucun trait de spécialité ;
5. diminuer – ce verbe se lie préférentiellement à des grandeurs physiques, comme *pression*, *température* ou *vitesse*, il ne présente aucun trait de spécialité ;
6. dissoudre – le verbe fortement lié au domaine de la chimie, ses sujets préférés ce sont *gaz (naturel)* et *brut* ;
7. adsorber – comme le verbe *dissoudre* il appartient au domaine de la chimie, mais dans ce cas-ci on ne trouve pas nos termes parmi ses arguments ;
8. générer – ce verbe est entouré des termes *forage* (sujet) et *hydrocarbure*, *gaz* (objet) ;
9. désacidifier – 100% des objets de ce verbe constituent le terme *gaz*. Selon le LVF, l’objet de *désacidifier* est toujours un mot *gaz*, mais dans notre cas le terme *gaz* est lié au domaine du forage et signifie (TERMIUM Plus) : « corps gazeux utilisé comme combustible ou comme carburant » ;
10. augmenter – comme *varier*, *diminuer* et *optimiser*, son sujet privilégié est *température* et il ne présente aucun trait de spécialité;

11. sédimenter – dans chaque phrase analysée son sujet est *matière organique* ;
12. visualiser – nous n'avons été en mesure de trouver aucune régularité dans le cas de ce verbe, ce qui est visible dans le registre de nos patrons ;
13. exploiter – ses sujets préférés sont *ressource*, *gisement* et *découverte* ;
14. carotter – son objet préféré est le mot *terrain*, mais il est lié au domaine des mines et ses arguments ne montrent pas de connexions au domaine du forage ;
15. excaver – comme carotter, il est très fortement lié au domaine des mines. Mais au cours de notre analyse, nous avons rencontré un autre problème : une très faible occurrence de ce verbe. Dans le corpus entier, il apparaît seulement 3 fois dans un article. Nous avons essayé de trouver d'autres articles qui nous permettraient de l'étudier plus profondément, sans succès ;
16. injecter – ce verbe comme ses sujets choisit en général les noms appartenant à la classe *procédé*, parmi lesquels on trouve *fracturation hydraulique* et comme ses objets les noms provenant de la classe *substance*. Ses arguments ne montrent pas clairement leurs liaisons avec le domaine analysé, c'est pourquoi nous avons examiné ses compléments circonstanciels. Dans la plupart des cas, ils contiennent les mots *puits* ou *forage* ;
17. migrer – 100% des sujets de ce verbe sont les termes *hydrocarbure* et *gaz* ;
18. déshydrater – 100% des objets de ce verbe constituent le terme *gaz (naturel)*;
19. extraire – ce verbe comme ses objets choisit préférentiellement les termes *hydrocarbure* et *gaz*, ce qui est aussi intéressant, ce sont ses compléments circonstanciels parmi lesquels le terme *puits* se répète ;

Tenu de ce qui précède, nous sommes capables de constater quel verbe est spécialisé et lequel ne se spécialise pas, mais pour nous assurer dans nos choix, nous avons décidé de confirmer nos prévisions en consultant encore le réseau lexico-terminologique de nos verbes.

2. Réseau lexico-terminologique

Examinons de près le tableau 4 (p. 47) où, à côté des verbes listés qui sont soumis à l'analyse, nous avons aussi mis leurs équivalents nominaux repérés par TermoStat comme les termes. Nous y trouvons des couples tels que :

verbe	équivalent nominal	domaine d'un équivalent (TERMIUM Plus)
forer	foration, forage	sondage et forage
produire	production	forage, énergie géothermique
adsorber	adsorption	raffinage du pétrole (chimie)
désacidifier	désacidification	chimie
sédimer	sédimentation	géologie, hydrologie
exploiter	exploitation	extraction du pétrole et du gaz naturel
excaver	excavation	constructions
injecter	injection, réinjection	extraction du pétrole et du gaz naturel, production pétrolière
migrer	migration	géologie, gisement pétrolifère
déshydrater	déshydratation	géochimie, pétrographie
extraire	extraction	exploitation minière

Tableau 9 Les verbes possédant des équivalents nominaux relevés par TermoStat

Nous avons effectué la recherche des fiches terminologiques de ces noms dans le TERMIUM Plus (annexe 2) en nous concentrant sur leurs domaines. Nous avons marqué en jaune les équivalents nominaux qui appartiennent au domaine du forage ou aux sous-domaines qui en font partie.

Pour pouvoir définir facilement le taux de spécialisation de nos verbes, nous avons décidé de créer un tableau (tableau 11) dans lequel nous avons regroupé toutes les analyses effectuées.

	TermoStat (taux de spécificité)	acceptions LVF (OUI/NON)	présence des termes liés au domaine dans les patrons (OUI/NON)	équivalents nominaux au domaine du forage (OUI/NON)
forer	45,76	N	O	O
varier	22,50	N	N	N
produire	21,79	O	O	O
optimiser	19,72	N	N	N
diminuer	18,75	O	N	N
dissoudre	18,09	N	O	N
adsorber	18,00	N	N	N
générer	17,93	O	O	N

désacidifier	16,93	N	O	N
augmenter	16,66	O	N	N
sédimenter	15,73	O	O	N
visualiser	15,06	O	N	N
exploiter	13,50	O	O	O
carotter	13,44	N	N	N
excaver	13,40	N	N	N
injecter	13,34	O	N	O
migrer	13,31	O	O	O
déshydrater	13,19	O	O	N
extraire	13,07	O	O	N

Tableau 10 Repérage des traits de spécialisations des 19 verbes choisis pour l'analyse.

En tenant compte des résultats, nous avons décidé de répartir nos verbes en trois groupes :

- verbes spécialisés, verbes qui possèdent une signification et pour lesquels c'est la même signification que celle évoquée dans la ressource LVF et qui possèdent dans leur structure grammaticale et leur réseau lexico-terminologiques les termes nominaux faisant partie du domaine du forage. Cette solution nous a semblé juste, parce que, si le sens d'un verbe coïncide au sens du LVF et ses arguments sont des termes du domaine du forage, cela veut dire qu'il possède juste un sens et il est spécialisé ;
- verbes généraux, verbes qui ne possèdent aucun point commun avec notre corpus – les réponses négatives pour les questions liées aux patrons et aux équivalents nominaux ;
- verbes généraux qui se spécialisent, ce sont les verbes qui appartiennent à la langue générale, mais qui possèdent une signification spéciale dans notre corpus et qui tendent à se relier avec les termes provenant du domaine du forage.

Les résultats de notre analyse sont présentés dans le tableau 12 ci-dessous :

verbes spécialisés dans le domaine du forage	verbes généraux qui se spécialisent	verbes généraux
forer	produire	varier
carotter	sédimenter	optimiser
excaver	exploiter	diminuer
	migrer	augmenter
	générer	visualiser
	extraire	
	piéger	
	tester	
	injecter	
	déshydrater	
	dissoudre	
	désacidifier	

Tableau 11 Classification des verbes selon leur taux de spécialisation

Comme nous pouvons le voir, il y a un verbe n'a pas été inclus dans le tableau 12, c'est le verbe *adsorber*. Cela est dû au fait que ce verbe ne possède pas de points communs avec notre corpus. Le verbe *adsorber* est un verbe qui appartient au domaine de la chimie. Dans notre corpus, nous avons trouvé seulement 6 occurrences de ce verbe. Dans tous les cas, il se lie aux arguments qui, dans notre ontologie, figurent dans la classe substance, sous-classe substance chimique, comme dans les exemples suivants :

[19] Un catalyseur de type alumine adsorbe l' H_2S et le SO_2 . (*Traitement du gaz naturel*, Magné-Drisch&Gonnard, 2015)

[20] Les tamis adsorbent peu le COS et le CS_2 . (*Traitement du gaz naturel*, Magné-Drisch&Gonnard, 2015)

En ce qui concerne les verbes *carotter* et *excaver*, malgré le fait que les arguments de ces deux verbes ne soient pas les termes de notre domaine et leur signification soit la

même que celle indiquée dans la ressource LVF, nous avons décidé de les mettre dans le groupe des verbes spécialisés pour deux raisons. Premièrement, les domaines auxquels appartiennent ces deux verbes sont : le domaine du pétrole pour *carotter* et le domaine des travaux publics pour *excaver*. Deuxièmement, nous avons examiné des contextes plus larges de ces verbes, parce que, selon notre avis, ces verbes sont très spécialisés et il nous a semblé impossible qu'ils puissent appartenir à la langue générale. Dans le cas du verbe *carotter* qui possède seulement 6 occurrences dans notre corpus, nous avons découvert qu'il est utilisé comme synonyme et substitut du verbe *forer*, comme nous pouvons le voir dans l'exemple 21.

[21] **Carotter**, ou **forer**, la plus grande longueur avec la même couronne, ou le même outil, et ce, dans le temps le plus court possible, est le but premier d'une entreprise qui réalise ce genre d'activité. (*Forage et carottage dans les roches*, Robert, 2010)

Le verbe *excaver* ne possède pas de termes liés au forage dans son voisinage immédiat, mais si on regarde des contextes plus larges, on trouve des mots comme :*tête d'excavation circulaire, boue (de forage) ou haveuse* (exemple 22).

[22] Les **haveuses** sont des machines de prédécoupage mécanique, essentiellement françaises (Perforex) qui ont été utilisées avec succès depuis 1974. (...) La machine comporte une scie qui se déplace sur un châssis adapté à la forme de la section **à excaver** (figure 4). (*Tunneliers*, Gesta, 1992)

Conclusion

L'analyse que nous avons effectuée nous a permis de constater que la combinaison des deux disciplines scientifiques peut être avantageuse et mener à des conclusions très intéressantes. Dans notre cas, la combinaison de la linguistique et de la terminologie a causé la fusion des approches sémantique et syntaxique afin d'établir le sens des verbes analysés. Elle nous a montré la possibilité d'utiliser dans la linguistique des ressources terminologiques accessibles gratuitement en ligne, telles que les bases terminologiques. Cette démarche s'est montrée en particulier utile lors de la création des types sémantiques. Elle peut être aussi utilisée en cas de continuation de ce travail pour agrandir l'ontologie et le corpus d'autres langues suivantes, parce que beaucoup de bases terminologiques sont multilingues. Quant à la terminologie, nous avons appris que le modèle CPA (*Corpus Pattern Analysis*) peut être utilisé avec succès en vue d'établir le taux de spécialisation des verbes dans un corpus de spécialité et d'écrire les exemples donnés de leur emploi.

Notre analyse nous a permis de distinguer l'emploi des verbes de façon spécialisée et non spécialisée et de déterminer si leur sens est dépendant d'un contexte ou non. Sur cette base, nous pouvons en arriver à la conclusion que les changements de l'approche des verbes en tant que termes potentiels que nous observons depuis de plus de vingt ans sont justes. Nos recherches ont confirmé que les verbes jouent un rôle crucial dans une phrase, ils donnent un sens à la phrase et ils méritent plus d'attention, si nous voulons bien saisir le message du discours spécialisé.

CPA est un outil bien utile. Il permet définitivement d'établir le taux de spécialisation d'un verbe donné et de décider s'il devrait faire partie d'un dictionnaire en construction, parce qu'il montre lesquels de ces emplois sont essentiels. L'importance d'une structure donnée est confirmée à l'aide de la linguistique de corpus et de l'approche de *corpus-driven*, ce qui permet d'évaluer le pourcentage de représentativité d'une structure donnée dans le corpus. Les résultats de nos recherches ont confirmé la légitimité de l'emploi de la linguistique de corpus et son importance dans les travaux linguistiques et terminologiques. En ne nous basant que sur le corpus, nous sommes capables de créer un dictionnaire fonctionnel, utile pour les étudiants, les chercheurs et les traducteurs.

Les résultats de notre travail ont démontré que le modèle CPA est une base idéale pour mettre en place un dictionnaire terminologique. Les principes de description des

verbes et de leurs structures sont clairs, et la base des types sémantiques construite peut être facilement agrandie et modifiée.

Nous avons aussi réussi à déterminer le statut terminologique de nos verbes. Nous l'avons atteint en créant trois taux de spécialisation des verbes : verbes spécialisés, verbes généraux et verbes généraux qui se spécialisent. Quant aux verbes eux-mêmes, comme nous nous doutions avant d'entamer nos recherches, le verbe dont le taux de spécialisation est le plus grand est le verbe *forer*, qui est le dérivé direct du nom du domaine analysé *forage*. Malgré les tentatives de compléter notre corpus de textes ayant des verbes d'un petit nombre d'occurrences, il nous a été difficile de le faire. Ce fait témoigne aussi de la grande spécialisation de ces verbes : ils ne sont même pas courants dans les textes spécialisés. Nous avons créé 27 patrons verbaux qui ont montré les spécificités liées au comportement lexico-syntaxique des verbes dans le domaine du forage. L'analyse menée a montré que chaque verbe ne se comporte pas de la même manière. Chaque verbe a ses patrons, propres à lui-même, parfois se trouvant également dans le langage général. La construction syntaxique digne d'être signalée est celle *nom à verbe*. Dans notre corpus, elle apparaît dans le cas des verbes signifiant *creuser le sol*.

Bien que notre travail soit assez complexe, nous ne nous sommes concentrés que sur 19 verbes parmi plus de 150 proposés par TermoStat. Cela montre que ce n'est qu'un début du travail qui devrait être poursuivi pour décrire tous les patrons appartenant au domaine du forage. Néanmoins, notre mémoire a démontré comment ce travail pourrait être effectué, ainsi que la simplicité de cette méthode. Cette étude pourrait être développée à travers l'augmentation de la taille du corpus de textes originaires d'autres langues pour pouvoir créer ainsi un dictionnaire spécialisé multilingue.

Bibliographie

- Alonso, A., & Rene, I. (2013). Corpus Pattern Analysis in determining specialised uses of verbal lexical units. *Terminalia* 7, 26-33.
- Alonso, A., Millon, C., & Williams, G. (2011). Collocational networks and their application to an E-Advanced Learner's Dictionary of Verbs in Science (DicSci). *Electronic Lexicography in the 21st Century: New Applications for New Users* (pp. 12-22). Proceedings of eLex.
- Bak Sienkiewicz, M., & Novakova, I. (2013). Le raisonnement dans les textes scientifiques : le cas des verbes causatifs. Dans A. Tutin, & F. Grossmann, *L'écrit scientifique : du lexique au discours. Autour de Scientext*. (pp. 101-122). Rennes : Presses Universitaires de Rennes.
- Bourigault, D., & Slodzian, M. (1999). Pour une terminologie textuelle. *Terminologies Nouvelles*, 19, pp. 29-32.
- Cabré, & M.T. (2000). Terminologie et linguistique : la théorie des portes. *Les cahiers du Rifal*, 21, pp. 10-15.
- Cabré, M. T. (1998). A propos de la notion de qualité en terminologie. *La Banque des mots: revue semestrielle de terminologie française* 8, 7-34.
- Cabré, M. T. (2003). Theories of terminology: Their description, prescription and explanation. *Terminology*, 163-199.
- Cabré, M. T. (2005). La terminologie, une discipline en évolution: le passé, le présent et quelques éléments prospectifs. Dans M.-C. L'Homme, & S. Vandaele, *Lexicographie et terminologie: compatibilité des modèles et des méthodes* (pp. 79-109). Ottawa: Les Presses de l'Université d'Ottawa.
- Cabré, M. T. (2012). Disciplinarisation de la terminologie: contribution de la linguistique. *Dossiers d'HEL-SHESL*, 1-10.
- Condamines, A. (1993). Un exemple d'utilisation de connaissances de sémantique lexicale : acquisition semi-automatique d'un vocabulaire de spécialité. *Cahiers de lexicologie* 62, 25-65.
- Daille, B. (1994). Study and implementation of combined techniques for automatic extraction of terminology. *The balancing act: Combining symbolic and statistical approaches to language*.
- Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology*, vol. 9, no 1, 99-117.
- Dubois, J., & Dubois-Charlier, F. (1997). *Les Verbes français*. Larousse.
- Dubuc, R. (1997). *Terminology: A Practical Approach*. Québec: Linguatech.
- Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, 19(1), 61-74.
- EL Maarouf, I. (2011). Formalisation de connaissances à partir de corpus : modélisation linguistique du contexte pour l'extraction automatique de relations sémantiques. *Thèse de doctorat. Université de Bretagne Sud*.
- Fellbaum, C. (1990). English verbs as a semantic net. *International Journal of Lexicography*, 3(4), 278-301.
- Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1), 20-32.
- Fillmore, C. J. (1977). Scenes-and-frames semantics. *Linguistic structures processing*, 59, 55-88.
- Fillmore, C. J. (1982). Frame Semantics. *Cognitive linguistics: Basic readings*, 373-400.
- Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di semantica*, 6(2), 222-254.

- Fillmore, C. J., & Baker, C. F. (2001). Frame semantics for text understanding. *Proceedings of WordNet and Other Lexical Resources Workshop*. NAACL.
- François, J. L. (2007). Présentation de la classification des Verbes français de Jean Dubois et Françoise Dubois-Charlier. *Langue française, (1)*, 3-19.
- Frérot, C. (à venir). Enseignement de la terminologie appliquée à une formation universitaire professionnalisante : illustration d'une collaboration avec l'Organisation Mondiale de la Propriété Intellectuelle.
- Frias-Conde, X. (2006). La voix moyenne dans les langues romanes. *Revista de Filología Románica 23*, 215-228.
- Gaudin, F. (1993). Socioterminologie : du signe au sens, construction. *META, 38(2)*, 293-301.
- Gaudin, F. (2003). *Socioterminologie. Une approche sociolinguistique de la terminologie*. Bruxelles: De Boeck/Duculot.
- Gross, M. (1975). *Méthodes en syntaxe : régime des constructions complétives*. Paris: Hermann.
- Gross, M. (1986). *Grammaire transformationnelle du français : Syntaxe du verbe*. Paris: Cantilène.
- Gross, M. (1994). Lexique-grammaire et adverbes : deux exemples. *Revue québécoise de linguistique, 15(2)*, 299-310.
- Guilbert, L. (1973). La spécificité du terme scientifique et technique. *Langue française (17)*, 5-17.
- Habert, B., Nazarenko, A., & Salem, A. (1997). *Les linguistiques de corpus*. Paris: Armand Colin & Masson.
- Halliday, M. A. (1961). Categories of the theory of grammar. *Word, 17(2)*, 241-292.
- Hanks, P. (2004). Corpus pattern analysis. *Euralex Proceedings Vol. 1*, 87-98.
- Hanks, P. (2010). Terminology, phraseology, and lexicography. *In Proceedings of the XIV Euralex International Congress*, (pp. 6-10). Leeuwarden.
- Hatier, S., Augustyn, M., Tran, H., Yan, R., Tutin, A., & Jacques, M.-P. (2016). French cross-disciplinary scientific lexicon extraction and linguistic analysis. *Proceedings Of The XVII Euralex International Congress Lexicography & Linguistic Diversity* (pp. 355-365). Tbilisi: Ivane Javakhishvili Tbilisi State University.
- Lafon, P. (1980). Sur la variabilité de la fréquence des formes dans un corpus. *Mots. Les langages du politique, 1(1)*, 127-165.
- Lerat, P. (2002). Qu'est-ce qu'un verbe spécialisé? Le cas du droit. *Cahiers de lexicologie, (80)*, 201-211.
- L'Homme, M.-C. (1998). Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de lexicologie, 73 (2)*, 61-84.
- L'Homme, M.-C. (2004). *La terminologie: principes et techniques*. Montréal: Presse de l'Université de Montréal.
- L'Homme, M.-C. (2012). Le verbe terminologique : un portrait de travaux récents. *SHS Web of Conferences (Vol. 1, pp. 93-107)*. EDP Sciences.
- L'Homme, M.-C. (2016). Terminologie de l'environnement et Sémantique des cadres. *SHS Web of Conferences (Vol. 27)*. EDP Sciences.
- Lorente, M. (2002). Verbos y discurso especializado. *Estudios de lingüística española (ELiEs) 16*.
- Pavel, S., & Nolet, D. (2001). *Précis de terminologie*. Québec: Ministre des Travaux publics et Services gouvernementaux Canada.
- Pimentel, J. (2012). Identifying Equivalents of Specialized Verbs in a Bilingual Comparable Corpus of Judgments: A Frame-based Methodology. *In LREC, (1791-1798)*.

- Rayson, P., & Garside, R. (2000, Octobre). Comparing corpora using frequency profiling. *Proceedings of the workshop on Comparing corpora*(Volume 9), 1-6.
- Rey, A. (1979). *La terminologie : noms et notions*. Paris: Presse Universitaire de France.
- Riccio, A. (2014). Analyzing Specialized Verbs in a French-Italian-English Medical Corpus: A Frame-based Methodology. *Proceedings of the XVI EURALEX International Congress: The User in Focus* (pp. 1049-1058). Bolzano/Bozen: EURAC research.
- Riegel, M., Pellat, J.-C., & Rioul, R. (1997). *Grammaire méthodique du français*. Paris: Presses Univesitaire de France.
- Robichaud, B. (2011). A Graph Visualization Tool for Terminology Discovery and Assessment. In Boguslavsky, I. et L.Wanner (eds.). *Actes de la 5e Conférence internationale sur la Théorie Sens-Texte*. Barcelone, Espagne.
- Slodzian, & M. (2000). L'émergence d'une terminologie textuelle et le retour du sens. Dans H. Béjoint, & P. Thoiron, *Le sens en terminologie* (pp. 61-85). Presse Universitaire de Lyon.
- Storjohann, P. (2005). Elexiko: A corpus-based monolingual German dictionary. *Hermes*, 34, 55-73.
- Tateisi, Y., Yakushiji, A., Miyao, Y., & Tsujii, J. (2004). Finding anchor verbs for biomedical ie using predicate-argument structures. *roceedings of the ACL 2004 on Interactive poster and demonstration sessions*. Barcelona: Association for Computational Linguistics.
- Tellier, C. (2008). *Verbes spécialisés en corpus médical: une méthode de description pour la rédaction d'articles terminologiques*. Technical report, Université de Montréal.
- Temmerman, R. (2000). *Towards new ways of terminology description: The sociocognitive-approach (Vol. 3)*. John Benjamins Publishing.
- Tutin, A., & Grossmann, F. (2002, 1). Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée*, VII, pp. 7-25.
- Victorri, B. (2010, mai 8). Quand les mots s'organisent en réseaux. *L'Archicube*, pp. 53-59.
- Wandji, O. T. (2014). Les modèles de description du verbe dans les travaux de Linguistique, Terminologie et TAL. *Terminologie et TAL*, 21ème TAL, 37-48. Marseilles.
- Willems, D. (1981). *Syntaxe, lexicque et sémantique des constructions verbales*. Gent: Rijksuniversiteit.
- Wüster, E. (1968). *The Machine Tool, an Interlingual Dictionary of Basic Concepts*. Londres: Technical Press.
- Yan, R. (2017). Étude des constructions verbales scientifiques dans une perspective didactique: utilisation des corpus dans le diagnostic des besoins langagiers en FLE à l'aide des techniques de TAL. *Thèse de doctorat. Université Grenoble Alpes*.

Bibliographie de corpus

Energies

- Biteau, J.-J., & Baudin, F. (2017). *Géologie du pétrole - Historique, genèse, exploration, ressources*. Dunod.
- Favreau, D., & Rouaud, T. (2012). Gaz non conventionnels. *Techniques de l'Ingénieur*.
- Guibet, J.-C., & Montagne, X. (2010). Carburants liquides - Caractéristiques et principes généraux. *Techniques de l'Ingénieur*.
- Kimmerlin, G. (2010). Gaz naturel. *Techniques de l'Ingénieur*.
- Laplaige, P., & Lemale, J. (2001). Energie géothermique. *Techniques de l'Ingénieur*.
- Magné-Drisch, J., & Gonnard, S. (2015). Traitement du gaz naturel. *Techniques de l'Ingénieur*.
- Pilenko, T. (2011). La production des hydrocarbures en offshore profond. *Annales des Mines-Responsabilité et environnement*, 17-23.
- Saint-Macroux, J.-F. (2017). Production de pétrole en mer. *Techniques de l'Ingénieur*.
- Saint-Macroux, J.-F. (2017). Production de gaz naturel en mer. *Techniques de l'Ingénieur*.
- Tonnac, A. D., & Perves, J.-P. (2013). Le gaz de schiste, une aubaine pour la France ? *Groupe de Reflexion sur l'energie Et l'Environnement au 21eme siecle*. Paris: Societe Francaise d'Energie Nucleaire - SFEN.
- Travers, C., & Tocqué, E. (20016). Pétrole. *Techniques de l'Ingénieur*.
- Vially, R. (2014). Hydrocarbures de roche-mère. *Techniques de l'Ingénieur*.
- Vially, R. (2016). Hydrates de méthane - Ressource et enjeux environnementaux. *Techniques de l'Ingénieur*.

Constructions et travaux publics

- Amar, S., & Jézéquel, J.-F. (1998 (validé 2015)). Propriétés mécaniques des sols déterminées en place. *Techniques de l'Ingénieur*.
- Briançon, L., Cazeadumec, B., & Pincent, B. (2016). Auscultation géotechnique. *Techniques de l'Ingénieur*.
- Gesta, P. (1992 (validé 2015)). Tunneliers. *Techniques de l'Ingénieur*.
- Héraud, H., & Durville, J.-L. (1995 (validé 2015)). Description des roches et des massifs rocheux. *Techniques de l'Ingénieur*.
- Lagabrielle, R. (2007). Diagraphies et géophysique de forage. *Technique de l'Ingénieur*.
- Magnan, J.-P., & Pilot, G. (1988 (validé 2015)). Amélioration des sols. *Techniques de l'Ingénieur*.
- Maquennehan, F. (2001). Mini et microtunneliers. *Techniques de l'Ingénieur*.
- Robert, J.-P. (2010). Forage et carottage dans les roches. *Techniques de l'Ingénieur*.
- Robert, J.-P. (2011). Forages et sondages. *Techniques de l'Ingénieur*.

Table des illustrations

Figure 1. Triangle sémiotique d'Odgen et Richards.....	11
Figure 2 Le fonctionnement de la démarche onomasiologique (L'Homme 2004, p. 26)	12
Figure 3 Polysémie de la <i>terminologie</i>	13
Figure 4 Une partie du réseau lexical de programmer1 (DicoInfo visuel, Robichaud 2011, in L'Homme 2012).....	23
Figure 5 : Liste de candidats termes pour les verbes provenant du corpus de forage.	40
Figure 6 : Liste de candidats termes des termes complexes provenant du corpus de forage sous forme de nuage.	41
Figure 7 : Liste des matrices des termes complexes provenant du corpus de forage.	41
Figure 8 : Liste des candidats termes et des termes qui les incluent provenant du corpus de forage.	42
Figure 9 : Liste des bigrammes des termes simples provenant du corpus de forage.....	42
Figure 10 : Statistiques générales créées par AnaText pour le corpus de forage.	43
Figure 11 : Liste des formes spécifiques provenant de corpus de forage extraits par AnaText.....	44
Figure 12 : Liste des verbes lemmatisés provenant de corpus de forage extraits par AnaText.....	44
Figure 13 : Liste des patterns du type : NOM ADJ trouvés par AnaText dans le corpus de forage.	45

Table des annexes

Annexe 1 L'analyse des 19 verbes selon les objectifs du LVF	83
Annexe 2 Les fiches terminologiques extraits de TERMIUM Plus	90
Annexe 3 L'ontologie - types sémantiques des noms utilisés dans l'analyse des verbes	99

Annexe 1
L'analyse des 19 verbes selon les objectifs du LVF

FORER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	pétrole	faire un trou dans qc (R3f.4)	être ou mettre en morceaux (rag) qc en creux	creuser	Les ouvriers f~ un puits, un trou.	transitif direct chose moyen (T1308), pronominal chose moyen (P3008)
	technique		être ou mettre en morceaux (rag) qc par trou	percer	Les voleurs f~ une porte blindée avec une perceuse.	
Corpus de forage	pétrole	faire un trou dans qc	être ou mettre en morceaux qc en creux	creuser	Les opérateurs f~ un grand nombre de puits pour atteindre un niveau de production important	transitif direct humain chose

VARIER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	quantitatif	se transformer, devenir autre (T4d.1)	qc indique une transformation (mut+qt)	évoluer, changer	Les prix v~ du simple au double.	intransitif chose modalité (A36)
Corpus de forage	quantitatif	se transformer, devenir autre	qc indique une transformation	évoluer, changer	D'une densité moyenne voisine de 5, sa température v~ de 1 000 à 3 000 °C.	intransitif chose modalité

PRODUIRE

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	industrie	faire qc, un objet (R3a.1)	fabriquer (fab) produit	fournir, fabriquer	On p~ du vin dans cette région.	transitif direct humain chose modalité (T1306), pronominal chose modalité (P3006)

Corpus de forage	géologie	faire qc, un objet	qc fabriquer qc naturel	créer	Une roche-mère p~ des hydrocarbures.	pronominal chose
					Le processus de compaction du sédiment se p~ par une forte diminution de la porosité.	transitif direct chose chose
	pétrole	faire sortir qc de lieu	faire sortir qc de terre	sortir de	Le puits p~ alors plus d'hydrocarbures au départ de sa vie.	transitif direct humain chose, transitif direct chose chose

OPTIMISER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	technique	rendre qc tel (T4b)	rendre / devenir tel (r/d) optimum abstrait	donner optimum à appareil	On cherche à o~ le rendement, cette machine.	transitif direct chose moyen (T1308), pronominal chose moyen (P3008)
Corpus de forage	technique	rendre qc tel	rendre / devenir tel optimum abstrait	donner optimum à appareil	Le raffineur o~ le choix du pétrole brut.	transitif direct humain chose

DIMINUER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	quantitatif	augmenter ou diminuer de tant)	qc faire un mouvement moins	baisser	L'intérêt d~ de beaucoup.	intransitif chose modalité (A36)
Corpus de forage	quantitatif	augmenter ou diminuer de tant	qc faire un mouvement moins	baisser	La pression du gisement d~ progressivement avec l'exploitation.	intransitif chose
					Un tube lisse d~ les pertes de charge.	transitif direct chose chose

DISSOUDRE

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	chimie	détruire qc (F3c.1)	donner un coup (ict) qc totalement	fondre	On d~ du Sucre dans l'eau.	transitif direct chose moyen (T1308), pronominal chose moyen (P3008)

Corpus de forage	chimie	détruire qc	donner un coup qc totalement	fondre	Des gaz se d~ dans les aquifères.	pronominal chose instrumental moyen
-------------------------	--------	-------------	------------------------------	--------	--	-------------------------------------

ADSORBER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	chimie	Absorber (<i>S4g</i>)	qc tenir ou prendre (<i>grp</i>) gaz	absorber	Le pot catalytique a~ les gaz.	transitif direct chose chose modalité (<i>T3306</i>),
Corpus de forage	chimie	absorber	qc tenir ou prendre gaz	absorber	Les tamis a~ peu le COS et le CS2.	transitif direct chose chose modalité
					Les molécules lourdes s'a~ sur les particules minérales.	pronominal chose

GÉNÉRER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	sociologie : littéraire (<i>SOct</i>)	réaliser une action (<i>R4b.1</i>)	réaliser une action (<i>fac</i>) abstrait	produire	On g~ l'inflation en haussant les prix.	transitif direct humain chose chose (<i>T1300</i>), pronominal chose (<i>P3000</i>)
Corpus de forage	géologie	réaliser une action	fabriquer un produit	produire	Ils g~ des microfissures dans la roche en injectant de l'eau sous pression dans le puits : c'est la fracturation.	transitif direct humain chose
					La matière organique g~ des hydrocarbures.	transitif direct chose chose

DÉSACIDIFIER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	chimie	faire perdre le caractère donné (<i>T3e.2</i>)	transformer en sens inverse (<i>dmut</i>) qc DE acide	ôter acides, acidité	Le chimiste d~ les huiles.	transitif direct humain chose moyen (<i>T1300</i>)
					Le vin se d~ .	pronominal chose (<i>P30b0</i>)
Corpus de forage	chimie	faire perdre le caractère donné	transformer en sens inverse qc DE acide	ôter acides, acidité	Le procédé permet de d~ le gaz.	transitif direct humain chose moyen

AUGMENTER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	quantitatif	augmenter ou diminuer qc (M4a.4)	abstrait faire mouvement (f.mvt) plus	accroître	Les prix, les pouvoirs a~, s'a~.	intransitif chose modalité (A36)
					On a~ l'intérêt, les pouvoirs.	transitif direct humain chose moyen (T1300)
Corpus de forage	quantitatif	se transformer, devenir autre	qc indique une transformation	évoluer, changer	Le nombre d'isomères a~ rapidement.	intransitif chose modalité (A36)
					Le désasphaltage a~ la densité API.	transitif direct chose chose moyen (T1300)

SÉDIMENTER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	géologie	se changer en qc , prendre telle forme, tel caractère (T3b)	qc indique une transformation (mut) EN sédiment	former des sédiments	Les dépôts végétaux s~. L'évolution a s~ le dépôt.	transitif direct chose moyen (T1308), pronominal chose moyen (P3008)
Corpus de forage	géologie	se changer en qc , prendre telle forme, tel caractère	qc indique une transformation EN sédiment	former des sédiments	Le flux de matière organique s~.	pronominal chose instrumental moyen

VISUALISER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	médecine	montrer qc qp (C3a.2)	montrer (ind) qc par radio	montrer, rendre visible	On v~ sur la radio la place de la tumeur.	transitif direct humain chose locatif (T1301)
Corpus de forage	mesure	montrer qc	montrer qc	montrer, rendre visible	On v~ ici les trois états des pétroles.	transitif direct humain chose locatif,

					Cette courbe v~les fréquences des températures.	transitif direct chose chose
--	--	--	--	--	---	---------------------------------

EXPLOITER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	économie	utiliser qc (S4h.3)	tenir ou prendre (grp) pour usage	faire valoir	On e~ une ferme, un brevet, un journal. La mine est e~.	transitif direct humain chose pour (T130q), pronominal chose (P3000)
Corpus de forage	économie	utiliser qc	tenir ou prendre pour usage	faire valoir	On e~ la matière organique dans les mines de charbon.	transitif direct humain chose

CAROTTER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	pétrole	faire un trou dans qc (R3f.4)	être ou mettre en morceaux (rag) qc avec carotte	creuser avec carotte	Les ouvriers c~ le sol pour trouver du pétrole.	transitif direct transitif direct chose moyen (T1308), pronominal chose moyen (P3008)
Corpus de forage	pétrole	faire un trou dans qc	être ou mettre en morceaux qc avec carotte	creuser avec carotte	Les outils au carbure de tungstène sont surtout utilisés pour c~ des terrains tendres.	transitif direct humain chose

EXCAVER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	travaux publicse	faire un trou dans qc (R3f.4)	être ou mettre en morceaux (rag) qc avec carotte	creuser avec carotte	Les ouvriers e~ le terrain avec un excavateur.	transitif direct transitif direct chose moyen (T1308), pronominal chose moyen (P3008)
Corpus de forage	travaux publics	faire un trou dans qc	être ou mettre en morceaux qc avec carotte	creuser avec carotte	La composition de la boue dépend très directement de la nature	-

					du sol à e~.	
--	--	--	--	--	--------------	--

INJECTER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	médecine	faire aller dans (E3f.4)	faire aller quelque part (f.ire) qc DANS som	infiltrer	On i~ un médicament dans les artères. Le sérum s'i~, est i~.	transitif direct humain chose en, dans (T13j0), pronominal chose en, dans (P30j0)
	économie	mettre qc dans (L4b.2)	mettre à telle place(lc) argent DANS affaire	investir argent dans	Les banques i~ des capitaux dans l'affaire.	
	pathologie	être dans tel état physiologique (H1b.1)	qc donne tel état au corps (f.som)sang yeux	être coloré par sang	Ses yeux s'i~, sont i~ de sang. Le sang i~ le visage	pronominal chose (P3000), transitif direct chose chose (T3300)
Corpus de forage	pétrole	mettre qc dans	être, mettre à telle place chose DANS roche	introduire qc sous pression dans roche	On i~ un fluide homogène d'eau et de sable	transitif direct humain chose en, dans
	outillage		être, mettre à telle place chose	introduire qc sous pression	On i~ en permanence une graisse épaisse spéciale entre une série de brosses métalliques	

MIGRER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	locatif, lieu	faire un trou dans qc (E1d.2)	aller quelque part (ire) qp de lieu vers pays	émigrer	On m~ dans les pays riches	intransitif humain double locatif (A14)
Corpus de forage	pétrole	aller vers un lieu	faire un mouvement vers puits/ réservoir	effectuer migration vers	Les hydrocarbures m~ vers les points hauts	intransitif chose locatif

DÉSHYDRATER

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	culturem cultivateur	démunir qc (objet, partie de l'objet, substance, liquide) de qc (N3b.1)	dmu qc DE hydrate	ôter eau de	Les techniciens d~ le fourrage avec une déshydrateuse.	transitif direct humain chose de moyen (T13b8), pronominal chose de (P30b0)
	pathologie	démunir, priver qn, animal de qc (partie du corpus, mal) (N1d.1)	dmu qn DE liquide	perdre eau	On d~ l'enfant en le laissant dans la voiture. P se d~, est d~ :	transitif direct humain chose de moyen (T19b8), pronominal humain de (P10b0)
Corpus de forage	pétrole	démunir qc de qc	démunir gaz DE hydrate	ôter eau de	Le reflux en tête de la colonne Sprex® permet de d~ le gaz brut entrant dans la colonne	transitif direct chose chose

EXTRAIRE

	domaine	classe	opérateur	sens	phrases	schèmes
LVF	mines	faire sortir qc de lieu (E3c)	faire sortir (f.ex) qc de lieu terre	sortir de	Le mineur e~ du charbon de la mine. Le fer s'e~, est e~ d'ici.	transitif direct humain chose de (T13b0)
Corpus de forage	mines	faire sortir qc de lieu	faire sortir qc de lieu terre	sortir de	On e~ le gaz d'une zone charbonneuse.	transitif direct humain chose de

Annexe 2

Les fiches terminologiques extraits de TERMIUM Plus

Adsorption
Domaine(s) <ul style="list-style-type: none">• Raffinage du pétrole• adsorption CORRECT, FÉM
CONTEXTE <p>L'adsorption est utilisée dans le raffinage du pétrole, pour améliorer la couleur des huiles et en extraire les composants asphaltiques ou résineux indésirables.</p>

Découverte
Domaine(s) <ul style="list-style-type: none">• Minerais divers (Mines métalliques)• Prospection géologique• Recherches et prospections géologiques• découverte CORRECT, FÉM
OBSERVATION <p>d'un gisement.</p>

Désacidification
Domaine(s) <ul style="list-style-type: none">• Chimie• Entretien des livres• Restauration d'œuvres d'art (Muséologie et Patrimoine)• désacidification CORRECT, FÉM
DEFINITION <p>[Processus] complexe qui permet, grâce à un agent alcalin (gazeux ou liquide), de neutraliser les acides présents dans le papier et de fournir une réserve alcaline dans le but de prolonger la durée de vie des documents.</p>
CONTEXTE <p>[Toute] restauration devrait permettre d'obtenir une stabilisation mécanique et chimique ainsi qu'une meilleure lisibilité du document, et permettre d'en faciliter la consultation et la mise en exposition. Quelques-uns des traitements, les plus courants, sont le dépoussiérage, le nettoyage à sec, le marouflage (action qui consiste à coller du papier ou de la toile sur un support pour le consolider), le traitement aqueux, la désacidification et la pose de renforts.</p>

Déshydratation

Domaine(s)

- Géochimie
- Pétrographie
- **déshydratation**
CORRECT, FÉM

CONTEXTE

On distingue deux modes de consolidation [des roches]: par recristallisation ou déshydratation.

OBSERVATIONS

Il peut sembler assez bizarre qu'une roche se déshydrate en milieu marin ou lacustre, mais il ne faut pas oublier qu'il s'agit là de transformations d'ordre chimique. Les argiles peuvent alors se transformer en schistes, et ceux-ci en phyllades.

PHRASÉOLOGISME

Déshydratation métamorphique

Excavation

Domaine(s)

- Fouilles (Construction)
- fouille
CORRECT, VOIR OBS, FÉM, NORMALISÉ

DÉFINITION

Toute excavation faite dans la terre (pour les constructions, travaux publics, etc.).

DEFINITION

Excavation destinée à recevoir les parties en infrastructure d'une construction. [Définition normalisée par l'ISO.]

OBSERVATION

Au-dessus de cette largeur [2 m], la fouille est dite «excavation»; c'est une fouille d'importantes dimensions [...] dans laquelle les tombereaux, aujourd'hui les camions automobiles, peuvent entrer directement pour l'enlèvement des terres.

OBSERVATION

Noter que c'est abusivement que les lexicographes utilisent le terme «excavation» pour définir le terme «fouille», car techniquement, en français universel, le premier est un spécifique par rapport au second; l'usage canadien, cependant, tend à faire coïncider les deux termes.

OBSERVATION

fouille : terme normalisé par l'ISO.

Terme(s)-clé(s)

excavation

Exploitation
<p>Domaine(s)</p> <ul style="list-style-type: none"> Extraction du pétrole et du gaz naturel exploitation CORRECT, FÉM <p>DEFINITION</p> <p>Consécutivement à une découverte confirmée [...] d'un gisement [...] ensemble des opérations qui permettent : d'extraire la plus grande partie possible des réserves [...], de collecter les produits et les rendre aptes à la livraison aux clients, les stocker s'il y a lieu, les expédier.</p> <p>PHRASÉOLOGISME</p> <p>Exploitation d'un champ pétrolier.</p>
Extraction
<p>Domaine(s)</p> <ul style="list-style-type: none"> Exploitation minière Tiroir - Commission canadienne de sûreté nucléaire extraction CORRECT, FÉM extraction minière FÉM
Forage
<p>Domaine(s)</p> <ul style="list-style-type: none"> Sondage et forage (Construction) Égouts et drainage Adduction d'eau (Alimentation en eau) forage CORRECT, MASC <p>DEFINITION</p> <p>Ensemble des techniques permettant de creuser un puits de faible diamètre à de très grandes profondeurs.</p>
Foration
<p>Domaine(s)</p> <ul style="list-style-type: none"> Fonçage des puits de mines Travail aux explosifs (Exploitation minière) Puits autres qu'artésiens (Sources d'eau) Forage des puits de pétrole foration CORRECT, FÉM, SPÉCIFIQUE

DÉFINITION

Opération qui consiste à forer des trous de mine, en principe avec une foreuse rotative (la perforation étant réalisée avec une foreuse percutante).

OBSERVATIONS

foration : terme d'utilisation plus restreinte que le terme «forage» couramment utilisé dans plusieurs secteurs d'activité du génie civil.

Fracturation hydraulique

Domaine(s)

- Forage des puits de pétrole
- **fracturation hydraulique**
CORRECT, FÉM
- hydrofracturation
CORRECT, FÉM

DEFINITION

Procédé qui consiste à créer artificiellement, par pression hydraulique, des fractures afin d'augmenter la perméabilité au voisinage d'un puits.

Terme(s)-clé(s)

hydro-fracturation

Gaz

Domaine(s)

- Gaz naturel et dérivés
- **gaz naturel**
CORRECT, MASC, NORMALISÉ

DEFINITION

Mélange complexe d'hydrocarbures, composé principalement de méthane, mais comprenant généralement aussi, de l'éthane, du propane, des hydrocarbures supérieurs, et quelques gaz non combustibles tels que l'azote et le dioxyde de carbone

DEFINITION

Hydrocarbures gazeux (principalement du méthane) provenant de gisements souterrains et dont la production peut être associée à celle du pétrole brut.

OBSERVATION

gaz naturel : terme et définitions normalisés par l'ISO.

OBSERVATION

gaz naturel : terme recommandé par le Comité de terminologie et normalisé par le Comité de validation du Lexique de la mobilité durable.

OBSERVATION

gaz naturel : terme normalisé par Environnement Canada et le Bureau de la traduction dans le cadre de l'Inventaire des émissions de polluants atmosphériques.

Gisement
<p>Domaine(s)</p> <ul style="list-style-type: none"> • Gisements pétrolifères • gisement CORRECT, MASC <p>DEFINITION</p> <p>Accumulation commerciale d'huile ou de gaz occupant un réservoir indépendant et se trouvant sous un système de pression.</p> <p>PHRASÉOLOGISME</p> <p>Gisement continu, disséminé, horizontal, interrompu, oblique, pétrolifère.</p>

Hydrocarbure
<p>Domaine(s)</p> <ul style="list-style-type: none"> • Éléments et composés chimiques • Agents de pollution • hydrocarbure CORRECT, MASC, NORMALISÉ, UNIFORMISÉ • HC CORRECT, MASC, NORMALISÉ • carbure d'hydrogène MASC <p>DEFINITION</p> <p>Composé organique ne contenant que du carbone et de l'hydrogène.</p> <p>OBSERVATION</p> <p>Les hydrocarbures contribuent largement au réchauffement climatique. Le méthane, après le dioxyde de carbone, est le principal gaz à effet de serre. Mentionnons également le propane, l'éthane et l'acétylène.</p>

Injection (réinjection)
<p>Domaine(s)</p> <ul style="list-style-type: none"> • Extraction du pétrole et du gaz naturel • Gisements pétrolifères • Production pétrolière • injection CORRECT, FÉM • réintégration PROPOSITION, FÉM • réinjection À ÉVITER, FÉM <p>OBSERVATIONS</p>

Réinjection. Ce terme est souvent employé à tort, en lieu et place du terme: injection. Réinjecter signifie: injecter à nouveau (une deuxième fois, par exemple). Or l'eau et le gaz naturel associés [...] à la production du pétrole quand ils sont renvoyés par injection dans le gisement le sont une seule et unique fois, en toute rigueur il n'y a donc pas «réinjection» et cela a fortiori quand on injecte de l'eau prise à la surface [...]. Le terme réintégration (dans le gisement) serait semble-t-il plus approprié.

Matière organique

Domaine(s)

- Chimie
- Science du sol
- Énergie de la biomasse
- Géochimie
- **matière organique**
CORRECT, VOIR OBS, FÉM
- matériau organique
MASC

DEFINITION

Dans les sols, ensemble de composés celluloseux, de lignine et (en plus faible quantité) de protéines d'origine essentiellement végétale fournis par les litières de feuilles et rameaux et par les racines (ou litières racinaires).

CONTEXTE

La matière organique, qui se présente habituellement sous la forme d'hydrocarbures goudronneux («pyrobitume»), est un constituant commun dans certains districts à minéralisations de type Mississippi-Valley [...] Dans quelques cas, on a relevé la présence de méthane ou de pétrole liquide dans des inclusions fluides [...]

OBSERVATION

Sous l'influence de l'activité biologique (faune et micro-organismes), cette matière organique fraîche se décompose partiellement, et plus ou moins vite, en dioxyde de carbone, ammoniac et vapeur d'eau.

OBSERVATION

matière organique : terme en usage à Ressources naturelles Canada - Secteur des sciences de la Terre.

Migration

Domaine(s)

- Géologie
- Gisements pétrolifères
- **migration**
CORRECT, FÉM

DEFINITION

Déplacement des hydrocarbures de la roche-mère vers la roche-réservoir.

Pétrole

Domaine(s)

- Pétroles bruts et dérivés
- **pétrole**
CORRECT, MASC, NORMALISÉ
- huile
CORRECT, VOIR OBS, FÉM
- naphte
VOIR OBS, MASC, VIEILLI

DÉFINITION

Huile minérale naturelle, de couleur très foncée, dotée d'une odeur caractéristique plus ou moins prononcée, d'une densité variant de 0,8 à 0,95, composée en presque totalité d'hydrocarbures paraffiniques, naphthéniques et aromatiques, et souvent d'un peu de soufre à l'état de combinaisons organiques et de traces de composés oxygénés et azotés.

CONTEXTE

Le degré de raffinage que devra subir un pétrole brut pour répondre aux caractéristiques du marché ne peut se définir [sans] une analyse des fractions obtenues, à la manière des alchimistes d'autrefois qui chauffaient l'huile de roche dans un alambic pour en vaporiser progressivement les parties légères.

OBSERVATION

[...] dans les opérations d'exploration et de production, [on dit aussi] huile.

OBSERVATION

naphte : Autrefois, pétrole brut ou raffiné.

OBSERVATION

pétrole; pét : terme et abréviation d'usage obligatoire au ministère de la Défense nationale du Canada et les Forces canadiennes.

OBSERVATION

pétrole : terme normalisé par l'ISO

Production

Domaine(s)

- Énergie géothermique
- Forage des puits de pétrole
- Extraction et transport du minerai
- **production**
CORRECT, FÉM
- rendement
CORRECT, MASC

OBSERVATIONS

En dehors de son sens général s'appliquant à la quantité d'hydrocarbures extraite, le mot production est souvent utilisé pour désigner l'ensemble des techniques d'extraction des hydrocarbures.

Puits

Domaine(s)

- Exploitation minière souterraine
- **puits**
CORRECT, MASC

OBSERVATION

[...] les mines sont exploitées à partir de puits qui débouchent à intervalles réguliers [...] sur des niveaux.

Ressource

Domaine(s)

- Géologie économique
- Exploitation de minerais organiques divers
- Exploitation commerciale (Pétrole et gaz naturel)
- ressources
CORRECT, FÉM, PLUR

OBSERVATION

Pour éviter la confusion qui règne entre les ressources (à définition très large) et les réserves (part des ressources connue avec une certaine précision et jugée économiquement exploitable), différentes classifications ont été proposées. La plus générale, émanant des Nations Unies et inspirée de celle du Bureau des mines des États-Unis [...] toutes les quantités indiquées désignent des ressources en place, dont la définition est géologique.

OBSERVATION

ressources : terme rarement utilisé au singulier (ressource).

Roche-mère

Domaine(s)

- Péetrographie
- **roche mère**
- CORRECT, FÉM
- roche source
CORRECT, FÉM

DEFINITION

Roche qui, par érosion mécanique, a fourni les éléments détritiques d'un sédiment.

CONTEXTE

Par soulèvement, exposition à l'air libre et altération de ce protolite, les concentrations [...] de substances utiles augmentent jusqu'à présenter une teneur rentable ou améliorée dans les gîtes

résiduels sus-jacents ou adjacents aux roches mères.

OBSERVATION

roche mère : terme en usage à Ressources naturelles Canada - Secteur des sciences de la Terre.

Sédimentation

Domaine(s)

- Géologie
- Hydrologie et hydrographie
- **sédimentation**
CORRECT, FÉM

DEFINITION

Ensemble des processus conduisant à la formation de sédiments.

Annexe 3
L'ontologie - types sémantiques des noms utilisés dans l'analyse des verbes

CLASSE	SOUS-CLASSE	EXEMPLE
espace	lieu	<i>milieu, trou, rayon d'investigation</i>
	partie	<i>atelier, block, chemin de migration, terrain, zone</i>
événement particulier	besoin	<i>demande</i>
humain	colléctivité	<i>entreprise, industrie, Afrique, France, Moyen-Orient, usine, pays</i>
	individu	<i>aléreur, explorateur, géologue, homme, opérateur, pionnier, taillant</i>
objet scientifique	cas	<i>cas, voie,</i>
	donné	<i>phénomène, problème, projet, signal, source d'énergie,</i>
	formule	<i>chaîne carbonée</i>
	méthode_1	<i>méthode, modèle, système</i>
	relation	<i>effet, résultat</i>
	support	<i>support, document, index de production</i>
	usage abstrait	
organisme vivant	faune	<i>bactérie</i>
	flore	<i>plante</i>
outil	appareil	<i>centrale à cycle combiné, condenseur, dispositif de dépollution, outil, FLNG, garnissage, installation de fond, IPB, reflux, upgrader</i>
	construction	<i>plate-forme, unité de séparation,</i>
	création	<i>portail web</i>
	élément d'appareil	<i>couronne, emprise au sol des installation, engin, tamis, tarière, tête d'excavation, tube</i>
	équipement	<i>instrumentation</i>
	installation	<i>vapocraqueurs</i>
	machine	<i>machine à pression de boue, navire de forage</i>
procédé	action	<i>allée-venu de camion, choix, collaboration, consommation, construction, disposition, enregistrement, envoie, exportation, intervention (exploitant intervenir), investissement, maintien, placement de condenseur, réalisation, réduction des coûts, report, travail</i>

	ajout	<i>adjonction, ajout</i>
	création	<i>architecture des drains, création de porosité, emplacement d'autres sondage, formulation du fluide, production</i>
	méthode_2	<i>fonctionnement du Claus, gestion du fluide, simulation dynamique, technique technologie, utilisation (de couplage / de cluster)</i>
	procédé technologique	<i>additivé la boue, auscultation, cisaillement, craquage, diagraphie, diversification des sources, échauffement, extraction, forage, fracturation (hydraulique), gas lift, hydrocraquage final, injection, lubrification, moussage, séparation, sismique 4D, technologie rotary, télédétection, terrassement, underbalanced drilling</i>
	réaction chimique_1	<i>réaction, combustion, désasphaltage, pyrolyse, stripage azéotropique, synthèse Fischer-Tropsch</i>
processus	changement,	<i>allongement de longueur de la vis, changement de phase, déstabilisation, évolution de la maturité, mise en charge d'un réservoir, perte de charge, repousser les frontières</i>
	dégagement	<i>émission de polluants</i>
	processus chimique	<i>gazéification, ségrégation chimique</i>
	processus évolutif	<i>accroissement, altération thermique, augmentation, croissance optimale, développement, diminution, gain, multiplication, perte, progrès en méthode</i>
	processus géologique	<i>catagenèse, compaction, délavage, diagenèse, éboulement, effondrement, érosion, lessivage, métagenèse, migration, mouvement latéral, processus de compaction du sédiment, sédimentation</i>
	processus physique	<i>usure excessive, absorption, rupture de liaisons</i>
	réaction chimique_2	<i>absorption, rupture de liaisons, dégagement de chaleur, précipitation des asphaltes</i>
qualité	anomalie	<i>erreur</i>
	caractéristique	<i>absence, caractère abrasif, caractéristique, condition opératoire,</i>

		<i>contamination, courbe, déséquilibre de compaction, différence de maturité, disponibilité en naphta, efficacité, état, fiabilité, fraction (liquide / gazeuse), froid, instabilité, nuisance, phase gaz, porosité, présence, propriété des fluides, structure, tortuosité, trajectoire, variation</i>
	état	<i>problématique, mode</i>
	état physique	<i>vapeur, forma gazeuse</i>
	importance	<i>importance</i>
	possibilité	<i>option, possibilité, risque</i>
	surnombre	<i>redondance des capteurs</i>
quantité	collection	<i>base de données</i>
	dimension	<i>rapport</i>
	entité,	<i>électricité, énergie</i>
	grandeur	<i>acidité, amplitude de signaux, baril de pétrole, chaleur, charge, chiffre d'affaire, coefficient de compressibilité, composition, concentration, conductivité thermique, corrosivité, courant, coût, critère de rétention, débit, densité, diamètre, distance, durée de vie, effort de poussée, épaisseur, flux, fréquence, GOR, hauteur, kWh produit, linéaire, longueur, longévité de l'auscultation, masse, mesure, montant, niveau, nombre, ordre de grandeur, PCI, performance, perméabilité, pH, poids, points de rosée, potentiel, poussée, pression, prix, productivité du puits, profondeur, puissance, ratio, rendement, sévérité, solubilité, surface de contact, taille, taux, température, teneur, toxicité, tranche de terrain, valeur, viscosité, vitesse, volatilité, volume</i>
	mesure	<i>molécule, particule</i>
	quantité abstraite	<i>capacité, connaissance, étape, paramètre, part, partie, période, proportion, répartition, représentation des hétérogénéités, réserve, ressource</i>
	sol	formation géologique
roche		<i>granulat, grès fluviatiles, poussière, roche, roche mère, sable, sol</i>
trou		<i>drain, microfissure, puits</i>

substance	combustible	<i>biocarburant, carburant, coal mine, méthane, condensat, gaz, GNL, kérogène, pétrole, charbon, tar mats</i>
	composé chimique,	<i>acide, benzène, butane, carbonate, CO2, COS, coupe, CS2, dioxyde de carbone, eau, essence, éthane, éthylène, éthyl-tertio-butyl-éther, gazoline, gel de silice, glucide, glycol, goudron, H2, H2S, hydrate, hydrocarbure, mercaptan, méthane, Nox, paraffine, propane, propylène, SO2, SOx</i>
	élément chimique,	<i>hydrogène, soufre</i>
	fluide,	<i>boue, saumure</i>
	liquide,	<i>breaker, filtrat, fluide, huile, solvant</i>
	matière	<i>ciment, matière organique, minéral de la roche-mère</i>
	solide	<i>adsorbant, graisse, épaisse, proppant</i>

Table des matières

Remerciements	3
Sommaire	5
Introduction	6
PARTIE 1 - L'ÉTAT DE L'ART	8
CHAPITRE 1. TERMINOLOGIE	9
1. Différentes approches de la terminologie	10
1.1. Théorie Générale de la Terminologie (TGT)	10
1.2. Les nouvelles approches en terminologie.....	13
1.2.1. Socioterminologie	14
1.2.2. Théorie communicative.....	14
1.2.3. Modèle sociocognitif	15
1.2.4. Terminologie textuelle.....	16
1.2.5. Approche lexico-sémantique.....	17
CHAPITRE 2. VERBE EN TERMINOLOGIE.....	18
2. Différentes façons d'analyser un verbe.....	18
2.1. L'approche conceptuelle.....	18
2.2. L'approche lexico-sémantique.....	19
2.2.1. La structure argumentale.....	20
2.2.2. Réseau lexico-terminologique.....	21
CHAPITRE 3. STRUCTURES VERBALES EN LINGUISTIQUE – LES PATRONS LEXICO-SYNTAXIQUES	24
3. Constructions dans la linguistique française	24
3.1. Le lexique-grammaire	24
3.2. Les constructions verbales de Willems	25
3.3. Les Verbes Français (LVF).....	26
4. Constructions dans la linguistique anglaise	28
4.1. FrameNet	28
4.2. Corpus Pattern Analysis (CPA).....	29
PARTIE 2 - CORPUS ET MÉTHODOLOGIE	31
CHAPITRE 4. PROBLEMATIQUE	32
CHAPITRE 5. CORPUS.....	33
1. Corpus utilisé.....	33
1.1. Choix et constitution du corpus.....	33
1.1.1. Techniques de l'Ingénieur	34
1.1.2. ScholarVox	35
1.2. Étude de corpus.....	36
2. Outils d'exploration de corpus	38
2.1. TermoStat.....	38
2.1.1. Fréquence.....	39
2.1.2. Spécificité	39
2.1.3. X^2	39
2.1.4. Log-likelihood	39
2.1.5. Log-odds ratio.....	40
2.1.6. Résultats d'analyse dans TermoStat	40
2.2. AnaText	43
CHAPITRE 6. METHODOLOGIE	46
1. Extraction des candidats-termes	46
2. Sélection des verbes à analyser.....	47
3. Analyse des contextes des verbes choisis	48
3.1. LVF.....	48
3.2. CPA.....	51
PARTIE 3 - ANALYSES ET RÉSULTATS	52

CHAPITRE 7. L'ONTOLOGIE.....	53
1. Collection des arguments.....	53
2. Création des classes et des sous-classes.....	53
2.1. Création des sous-classes	53
2.2. Création des classes	55
CHAPITRE 8. FORMES SYNTAXIQUES SPECIFIQUES AU CORPUS DE FORAGE.....	58
1. Constructions actives	58
1.1. Sujet – Verbe – Objet (SVO)	58
1.1.1. Constructions causatives	59
1.1.2. Construction périphrastique causative <i>permettre de</i>	59
1.2. Sujet – Verbe.....	60
1.3. Sujet – Verbe – Attribut	60
2. Constructions passives et moyennes	61
3. Construction spécifique : nom à Verbe.....	62
4. Construction en « pour » – complément circonstanciel	63
5. Patrons des verbes extraits du corpus du forage	64
CHAPITRE 9. STATUT TERMINOLOGIQUE DES VERBES ANALYSES	66
1. Modèle descriptif de la structure argumentale	67
2. Réseau lexico-terminologique	71
Conclusion.....	75
Bibliographie.....	77
Bibliographie de corpus.....	80
Table des illustrations.....	81
Table des annexes.....	82
Table des matières.....	103

MOTS-CLÉS : verbe, statut terminologique, CPA, forage

RÉSUMÉ

Le présent mémoire a pour but d'analyser les constructions verbales provenant de corpus semi-spécialisé de forage afin de trouver une méthode qui permette la désambiguïsation de leur sens et leur description et d'établir leur statut terminologique. Les critères sémantiques et syntaxiques ont été pris en considération, ce qui a permis, à la base du modèle *Corpus Pattern Analysis* (CPA), de dégager des constructions typiques du discours de ce domaine. Notre objectif final est de fournir les informations pour pouvoir créer dans l'avenir un dictionnaire spécialisé de forage accessible en ligne.

KEYWORDS : verb, terminological status, CPA, drilling

ABSTRACT

The aim of this master's thesis is to analyse the verb constructions originating de semi-specialized drilling corpus in order to find a way to disambiguate the meanings of the verbs and their description and to establish their terminological status. The semantic and syntactic criteria have been taken into account resulting in, on the basis of *Corpus Pattern Analysis* (CPA), a generation of verbal constructions typical for the rhetoric of this field. Our ultimate goal is to provide information in order to create a specialised drilling dictionary available on-line.