

Université de Montréal

**Partager le savoir du lexicographe: extraction et modélisation ontologique
des savoirs lexicographiques**

par
Sophie Comeau

Département de linguistique et de traduction
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)
en linguistique

décembre, 2009

© Sophie Comeau, 2009.

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:

**Partager le savoir du lexicographe: extraction et modélisation ontologique
des savoirs lexicographiques**

présenté par:

Sophie Comeau

a été évalué par un jury composé des personnes suivantes:

Marie-Claude L'Homme,	président-rapporteur
Alain Polguère,	directeur de recherche
Pascale Lefrançois,	codirecteur
Victor Boucher,	membre du jury

Mémoire accepté le:

RÉSUMÉ

Cette recherche porte sur la lexicologie, la lexicographie et l'enseignement/apprentissage du lexique. Elle s'inscrit dans le cadre du projet *Modélisation ontologique des savoirs lexicographiques en vue de leur application en linguistique appliquée*, surnommé *Lexitation*, qui est, à notre connaissance, la première tentative d'extraction des savoirs lexicographiques — i.e. connaissances déclaratives et procédurales utilisées par des lexicographes — utilisant une méthode expérimentale. Le projet repose sur le constat que les savoirs lexicographiques ont un rôle crucial à jouer en lexicologie, mais aussi en enseignement/apprentissage du lexique. Dans ce mémoire, nous décrivons les méthodes et les résultats de nos premières expérimentations, effectuées à l'aide du *Think Aloud Protocol* (Ericsson et Simon, 1993). Nous expliquerons l'organisation générale des expérimentations et comment les savoirs lexicographiques extraits sont modélisés pour former une ontologie. Finalement, nous discuterons des applications possibles de nos travaux en enseignement du lexique, plus particulièrement pour la formation des maîtres.

Mots clés: Lexicographie, Lexicologie Explicative et Combinatoire (LEC), ontologie, extraction expérimentale des connaissances, didactique des langues.

ABSTRACT

This research is about lexicology, lexicography and vocabulary teaching/learning. It is part of a project called *Ontologization of lexicographic abilities for use in the fields of applied linguistics*, nicknamed *Lexitation*, which is, to our knowledge, the first attempt at extracting lexicographic abilities using experimental techniques. The project relies on the assumption that lexicographic abilities play a role in teaching and acquisition of lexical knowledge, and not only in lexicography per se. We will describe the methods and results of our initial set of experiments, that are based on the use of so-called Think Aloud Protocol (Ericsson et Simon, 1993). We will explain how experiments have been set up and how we are currently proceeding with the extraction and modeling of various types of knowledge and strategies used by lexicographers while performing lexicographic tasks. Finally, we will present possible applications of our work in the field of language teaching, more specifically, teachers' training.

Keywords: Lexicography, Explanatory Combinatorial Lexicology (ECL), ontology, experimental knowledge extraction, language teaching.

TABLE DES MATIÈRES

RÉSUMÉ	iii
ABSTRACT	iv
TABLE DES MATIÈRES	v
LISTE DES FIGURES	viii
LISTE DES ANNEXES	ix
LISTE DES SIGLES	x
DÉDICACE	xi
REMERCIEMENTS	xii
AVANT-PROPOS	xiii
CHAPITRE 1 : INTRODUCTION	1
1.1 Lexicographie, dictionnaire, métalexigraphie : définition de l'objet d'étude	1
1.2 Métalexigraphie : état de la question	2
1.3 Lien entre la lexicographie et l'enseignement du lexique	6
1.4 Enseignement du lexique : état de la question	8
CHAPITRE 2 : PRÉSENTATION DE LA RECHERCHE : CONTEXTE ET OBJECTIFS	11
2.1 Projet Lexitation	11
2.1.1 Constat de départ et buts du projet Lexitation	11
2.1.2 Équipe de recherche	12
2.1.3 Moyens	12

2.2	Objectifs de la présente recherche	14
2.3	Cadre théorique : la lexicologie explicative et combinatoire	16
2.4	Où en sommes-nous ?	18
CHAPITRE 3 : APPROCHE EXPÉRIMENTALE DE L'EXTRACTION DES SAVOIRS LEXICOGRAPHIQUES		20
3.1	Nature du travail lexicographique	20
3.2	Utilisation du <i>Think Aloud Protocol</i>	21
3.3	Organisation des expérimentations	25
3.3.1	Activités lexicographiques	25
3.3.2	Participants impliqués dans les expérimentations	26
3.3.3	Aménagement expérimental au Laboratoire-Observatoire de Recherche en Ingénierie du Téléapprentissage (LORIT)	28
CHAPITRE 4 : ANALYSE ET ENCODAGE DES DONNÉES		30
4.1	Analyse des données	30
4.1.1	Approche inductive	30
4.1.2	Notions de base	32
4.1.3	Division de l'activité en tâches	36
4.1.4	Division des tâches en opérations	37
4.1.5	Présentation des classes d'opérations identifiées pour l'Expérimentation 3	38
4.1.6	Transcription d'une tâche et son analyse	47
4.1.7	Cas difficiles	49
4.2	Encodage des données dans l'ontologie	52
4.3	Analyse de l'Expérimentation 5	60
4.3.1	Nouvelles classes de tâches et d'opérations	60
4.3.2	Classes de tâches et d'opérations réutilisées	68
CHAPITRE 5 : PERSPECTIVES SUR LES APPLICATIONS POSSIBLES EN DIDACTIQUE DES LANGUES		72

5.1	Entraînement à la lexicographie pour les maîtres	72
5.1.1	Orientation qualitative plutôt que quantitative	73
5.1.2	Le rôle de notre ontologie dans un entraînement à la lexicographie	75
5.1.3	Les deux applications d'un entraînement à la lexicographie . . .	78
5.1.4	Activités décrochées et intégrées sur le vocabulaire	81
5.2	Théorisation : où les apprentissages lexicographiques se situent-ils dans les programmes scolaires ?	84
5.2.1	Contenus enseignés	84
5.2.2	Lien avec le programme de formation scolaire	87
5.2.3	Compétences professionnelles mobilisées	96
5.3	Conclusion	103
	BIBLIOGRAPHIE	105

LISTE DES FIGURES

3.1	Participants impliqués dans les expérimentations	27
4.1	Hiérarchie des opérations lexicographiques	39
4.2	Instance d'expérimentation	53
4.3	Instance de participant	55
4.4	Instance de tâche	56
4.5	Instance d'opération	58
4.6	Opérations concernant la rédaction de définitions	64
4.7	Opérations concernant le statut d'une lexie	66
4.8	Opérations concernant la structure polysémique d'un vocable	66
4.9	Opérations concernant la consultation d'un dictionnaire	67
4.10	Hiérarchie des tâches lexicographiques	69
4.11	Hiérarchie des opérations lexicographiques	70
4.12	Hiérarchie des opérations lexicographiques (suite)	71
I.1	Modèle de fiche lexicographique utilisé lors de l'Expérimentation 3	xv

LISTE DES ANNEXES

Annexe I : Modèle de fiche lexicographique xv

LISTE DES SIGLES

CREFD	<i>Collins-Robert English-French Dictionary</i>
DEC(FC)	<i>Dictionnaire explicatif et combinatoire (du français contemporain)</i>
DiCo	<i>Dictionnaire de combinatoire</i>
GTN	<i>Gros Tas de Notions</i> (Ontologie des notions lexicologiques et lexicographiques)
LEC	Lexicologie explicative et combinatoire
LICEF	Laboratoire en informatique cognitive et environnements de formation
LORIT	Laboratoire-Observatoire de Recherche en Ingénierie du Téléapprentissage
OLST	Observatoire de linguistique Sens-Texte
TLFi	<i>Trésor de la langue française informatisé</i>

Je dédie ce mémoire à trois êtres chers qui sont
disparus pendant sa rédaction, Raymond, Ursule et Marcel.

REMERCIEMENTS

Ce mémoire a été dirigé par Alain Polguère et co-dirigé par Pascale Lefrançois. Ce fut un privilège et un plaisir de travailler avec chacun d'eux.

Alain est le maître d'œuvre du projet Lexitation, dont mon projet de recherche fait partie. C'est grâce à son énergie si je me suis lancée dans ce projet, et grâce à son talent pour résoudre des problèmes s'il y a aujourd'hui un mémoire à déposer. Merci pour ta rigueur, ta générosité et ta bonne humeur.

Pascale Lefrançois a gentiment accepté de codiriger ce travail, dont une portion traite de didactique des langues. Ses commentaires, toujours judicieux, m'ont permis d'améliorer de façon considérable plusieurs aspects de ce travail. Merci pour ta disponibilité et tes encouragements.

Un savant « fou mais tout à fait inoffensif » (selon ses dires) est à l'origine de mon intérêt pour l'étude scientifique de la langue : Igor Mel'čuk, qui m'a appris à travailler et qui m'a donné ma première chance. Je tiens à lui exprimer ici toute ma reconnaissance.

Le travail de recherche décrit dans ce mémoire n'aurait pas pu être réalisé sans la contribution de Jacqueline Bourdeau, que je remercie spécialement. Au nom de toute l'équipe du projet Lexitation, je remercie également tous ceux qui ont participé aux expérimentations : Jean Fontaine, Pierre Martel, Pascale Bernard, Mélissa Samson-Fauteux, Anne-Laure Jousse, Sara-Anne Leblanc, Ophélie Tremblay, Lucie Barque et les membres de l'équipe technique du LORIT.

Je tiens enfin à remercier le Conseil de recherches en sciences humaines du Canada (CRSH), qui m'a accordé une bourse de recherche en 2008-2009, ainsi que le Département de Linguistique et de Traduction de l'Université de Montréal, qui m'a octroyé une bourse d'études pendant la rédaction de ce travail.

AVANT-PROPOS

The craft of lexicography demands not only the ability to collect data, but also the ability to make sense of it.

SUE ATKINS, *Then and Now : Competence and Performance in 35 Years of Lexicography*

La lexicographie a vu sa pratique transformée dans les dernières décennies par l'apparition de l'informatique, qui a permis d'améliorer le processus de collecte de données linguistiques grâce à l'utilisation de vastes corpus informatisés (Atkins, 2008a). Par contre, comme la citation en exergue l'indique, la pratique de la lexicographie demande plus que l'habileté d'amasser des données : il faut aussi être capable de les rendre significatives. En effet, à partir d'une grande quantité d'exemples d'utilisation d'une lexie (une *lexie*, aussi appelée *unité lexicale*, est soit un *lexème* soit une *locution*)¹ et d'indications sur son comportement syntaxique, le lexicographe doit repérer les éléments dont les utilisateurs d'un dictionnaire auront besoin pour comprendre le sens de cette lexie ou pour l'utiliser correctement. Ce mémoire cherche à mettre en lumière les savoirs mobilisés par les lexicographes pour effectuer cette tâche particulière (appelée « analyse ») ainsi que les autres tâches reliées au travail lexicographique. La recherche qui fait l'objet de ce mémoire s'inscrit dans cadre du projet *Lexitation*, dont le but est d'extraire et de modéliser sous la forme d'une ontologie les savoirs des lexicographes professionnels, pour en faire bénéficier un plus grand nombre de professionnels de la langue. En effet, nous croyons que les connaissances lexicographiques pourraient être transmises à d'autres lexicographes, bien sûr, mais aussi que ces connaissances ont un rôle à jouer entre autres en enseignement des langues, particulièrement dans le domaine de l'enseignement/apprentissage du vocabulaire. Des méthodes structurées d'enseignement des notions lexicales ont été suggérées par quelques auteurs (Picoche, 1993), (Tremblay, 2003) ; pour notre part, nous croyons que le potentiel applicatif de notre ontologie à la

¹ Voir (Polguère, 2008) pour les notions de *lexème* et de *locution*.

didactique des langues se situe en grande partie dans les interventions spontanées à teneur lexicale que les enseignants font naturellement en classe, quotidiennement, pour guider la lecture ou pour répondre à des questions. Ces interventions ponctuelles ne sont peut-être pas suffisantes, mais elles sont nécessaires et, à l'heure actuelle, les enseignants sont rarement outillés pour mener rapidement et efficacement ces interventions, comme l'a remarqué Alain Polguère (2004). Celui-ci suggère aussi que, pour être efficaces, les enseignants doivent être capables d'analyser et de modéliser des phénomènes lexicaux avec une certaine aisance, et que l'acquisition des savoirs et des techniques propres au travail lexicographique permet de maîtriser les tâches de nature métalinguistique impliquées dans l'enseignement du lexique. L'objectif du projet Lexitation est donc d'accroître l'autonomie de l'ensemble des professionnels du langage, dont les enseignants, qui bénéficieront en particulier du volet pédagogique du projet. En effet, des outils pédagogiques pourront être dérivés de l'ontologie des savoirs lexicographiques. La recherche faisant l'objet de ce mémoire ne prétend pas élaborer complètement de tels outils ; elle se concentre plutôt sur les premières étapes de ce projet, c'est-à-dire l'extraction des savoirs lexicographiques par une méthode expérimentale (Chapitre 3), l'analyse des données recueillies lors des premières expérimentations et l'encodage de ces données dans une ontologie (Chapitre 4). Une mise en contexte par la présentation de la problématique (Chapitre 1) et du projet Lexitation (Chapitre 2) sera toutefois faite au préalable, et une réflexion sur les possibles applications de notre ontologie en didactique des langues terminera le mémoire (Chapitre 5).

CHAPITRE 1

INTRODUCTION

1.1 Lexicographie, dictionnairique, métalexigraphie : définition de l'objet d'étude

Dans cette section, nous définissons notre objet d'étude et effectuons quelques clarifications terminologiques. D'emblée, remarquons que le terme *lexicographie* est ambigu. En effet, selon le *Trésor de la langue française informatisé* (TLFi), il peut à la fois désigner 'la technique de confection des dictionnaires' et 'l'analyse théorique de cette pratique'. Le premier de ces deux concepts, soit la production de dictionnaires en tant que telle, est également connu — particulièrement en France — sous le terme de *dictionnairique*, proposé par Bernard Quemada (1987). En anglais, on préfère le terme *dictionary-making* (Atkins, 2008b). Le deuxième concept se désigne souvent par le terme *métalexigraphie*, puisque qu'il réfère à des théories de production de dictionnaires pratiques.

Nous nous intéresserons particulièrement, dans ce mémoire, à la métalexigraphie, mais il nous faut d'abord présenter la distinction entre la métalexigraphie et la *lexicographie théorique*. Cette dernière vise la construction de dictionnaires théoriques (i.e. pour la recherche) et non de dictionnaires commerciaux. En effet, des travaux tels que le *Dictionnaire explicatif et combinatoire du français contemporain* (DECFC) (Mel'čuk et al., 1984, 1988, 1992, 1999) et celui de Wierzbicka (1987), *English Speech Act Verbs, A semantic Dictionary*, proposent des méthodes de description rigoureuses et parfois très formelles du lexique, dont l'application mène à un dictionnaire partiel idéalisé. De tels dictionnaires théoriques peuvent servir de modèles expérimentaux pour améliorer la qualité (complétude, cohérence, etc.) des dictionnaires grand public (Polguère, 2008) .

D'ailleurs, selon Atkins, lexicographe travaillant entre autres à un dictionnaire grand public bilingue, le Collins-Robert English-French Dictionary (CREFD), depuis plus de trente ans, ces travaux sont des exemples de ce à quoi la lexicographie pourrait ressembler dans un monde idéal :

The lexical analysis of Wierzbicka (1985, 1987, 1988) and of Asprejan, Mel'čuk, and those who work with them on immensely detailed lexicographical entries (...) shows us what lexicography could be like if we had the time (and the talent). (Atkins, 2008b, 42)

Ce commentaire illustre la situation particulière dans laquelle les lexicographes de dictionnaires grand public se trouvent : les travaux théoriques fournissent des critères pour la prise de décision concernant différents aspects de la conception d'un dictionnaire, cependant, les décisions éditoriales doivent prendre en compte des contraintes d'espace et de temps très rigides. Aussi, un dictionnaire étant une publication vaste, complexe et coûteuse, on ne peut pas aisément modifier sa structure (Atkins, 2008b). Pour ces raisons, les rédacteurs de dictionnaires grand public ne peuvent pas toujours adopter les procédures prescrites par les spécialistes de la lexicographie théorique. Ils suivent tout de même des procédures, plus ou moins propres à chaque maison d'édition, mais celles-ci, conscrites dans les feuilles de style, ne font pas l'objet de publications scientifiques régulièrement. Néanmoins, cet état de fait commence à changer, comme nous le verrons dans la prochaine section.

1.2 Métaléxicographie : état de la question

Auparavant, il était difficile d'accéder aux méthodes de travail utilisées par les lexicographes, puisque très peu de textes étaient publiés à ce sujet. Il a même été suggéré que cette discipline n'avait aucun fondement théorique :

It has often been said that lexicographers are people who work hard but who can never escape having guilty conscience, because lexicography has no theoretical foundations, and even the best lexicographers, when pressed, can never explain what they are doing, or why. (Wierzbicka, 1985, 5)

Cela étant dit, à l'heure actuelle, les lexicographes des dictionnaires grand public ont des méthodes de travail bien établies. Sur ce sujet, les principales sources d'informations dont nous disposons viennent de la lexicographie anglaise. Notamment, Atkins

s'est acquittée de la tâche de décrire la méthode de travail de son équipe, faisant mentir l'assertion de Wierzbicka ci-haut. Dans *Theoretical Lexicography and its Relation to Dictionary-making*, elle décortique le travail lexicographique et en présente les deux grandes phases, soit l'analyse et la synthèse (Atkins, 2008b). Le processus d'analyse correspond au moment où le lexicographe analyse le mot à décrire (en contexte), en essayant de découvrir le plus d'informations pertinentes sur son comportement, pour ensuite enregistrer et ordonner ces informations. L'analyse doit faire ressortir tous les éléments nécessaires à la synthèse, qui est la rédaction d'entrées à proprement parler.

La première étape de l'analyse décrite par Atkins dans un autre texte, *Then and Now : Competence and Performance in 35 Years of Lexicography* (Atkins, 2008a) consiste à consulter des corpus informatisés et la ressource FrameNet¹ pour examiner des occurrences de la lexie à décrire. Le projet FrameNet, développé à l'Université Berkeley, est une ressource en ligne sur l'anglais, basée sur la sémantique des *frames* de Fillmore (1976), qui s'appuie sur l'analyse de corpus. Le but de ce projet est de documenter les possibilités de combinatoire syntaxique et sémantique des unités lexicales (acceptions² spécifiques) en les associant à des « frames », sortes de scénarios conceptuels fédérant les unités du lexique. Cela se fait par une annotation informatique de phrases-exemples et une mise sous forme de table automatique des résultats d'annotations (Baker et al., 2003). La base lexicale FrameNet et les corpus permettent de régenter tous les faits nécessaires (sémantiques, syntaxiques, combinatoires, etc.) à la description lexicographique et de noter toutes les informations qui semblent pertinentes. Avec les données recueillies dans FrameNet, les rédacteurs du CREFD font ensuite une ébauche de l'entrée de dictionnaire, qui inclut une abondance d'informations, dont il faudra plus tard retrancher certains éléments moins importants. Ensuite, comme le CREFD est un dictionnaire bilingue, les rédacteurs doivent faire une recherche similaire pour les équivalents français, et utilisent un corpus parallèle (des mots avec leurs traductions les plus fréquentes, tirés de corpus).

¹FrameNet est disponible à l'adresse suivante : <http://framenet.icsi.berkeley.edu/>

²Voir (Polguère, 2008, 59) pour une présentation plus complète des notions d'acceptation et de vocable.

Cette description d'Atkins et d'autres ouvrages sur le travail lexicographique sont très informatifs, mais ils sont en général le résultat d'introspection, et se concentrent souvent sur un aspect particulier du travail, senti comme plus problématique. Par exemple, le thème de la désambiguïsation a été considérablement étudié, la notion même d'acceptation, ou *word sense*, étant mise en doute par plusieurs auteurs dont Hanks (2008), Kilgarriff (2008) et Atkins elle-même (citée par Kilgarriff (2008)) .

On doit constater qu'il n'y a pas, à l'heure actuelle, de modélisation complète et rigoureuse des connaissances et des compétences mises en jeu dans le travail lexicographique. C'est-à-dire qu'il n'y a pas encore à ce jour de description de l'ensemble des tâches impliquées en lexicographie, obtenue par l'observation effective de travail lexicographique en temps réel. Nous croyons qu'un modèle structuré, de type ontologique (Gomez-Pérez, 1999), des savoirs lexicographiques — c'est-à-dire des connaissances déclaratives et procédurales impliquées dans le travail lexicographique — serait nécessaire. Ce type de modèle nous permettrait de traiter l'activité lexicographique comme un cas particulier de résolution de problème (*problem solving*) (Polguère, 2007).

Une définition des concepts de *connaissances déclaratives* et de *connaissances procédurales*, tirée d'un ouvrage de pédagogie, permettra de clarifier notre propos :

Les connaissances déclaratives, comme les règles, les faits, les lois et les principes, sont des connaissances factuelles — des savoirs — qui sont statiques plutôt que dynamiques (...). Les connaissances procédurales concernent les étapes qu'il est nécessaire de franchir ou la démarche qu'il faut mener pour accomplir une action ; ces sont des connaissances dynamiques — des savoir-faire — qui sont acquises dans l'action, alors que l'apprenant réalise des tâches (...) (Boisvert, 1999)

En lexicographie, les connaissances déclaratives sont celles liées aux notions et règles linguistiques, tandis que les connaissances procédurales sont plutôt liées aux opérations impliquées dans le travail lexicographique. Les savoirs lexicographiques que nous voulons décrire sont donc de différentes natures. De plus, il y a différents types de connaissances déclaratives et procédurales. En effet, parmi les connaissances déclaratives rele-

vant de la lexicographie, on peut distinguer les 4 catégories suivantes, selon le descriptif du projet Lexitation (Polguère, 2007) :

1. concepts linguistiques fondamentaux structurant l'étude du lexique ;
2. concepts propres à la modélisation lexicale dans les dictionnaires et les bases de données ;
3. concepts liés à l'accès aux données linguistiques ;
4. concepts liés aux tâches descriptives (niveau de description, interdépendance des différents éléments descriptifs, critères et principes descriptifs).

Les connaissances procédurales peuvent aussi être de différentes natures :

1. ordonnancement des tâches impliquées dans la construction de descriptions lexicographiques ;
2. interdépendance de ces tâches ;
3. résolution des conflits lorsque plusieurs principes descriptifs ou types de données sont en contradiction.

Le projet Lexitation vise donc la construction d'un modèle ontologique des savoirs lexicographiques de natures diverses. C'est un vaste projet dont notre recherche constituera la première étape. En effet, notre présent travail de recherche entend démontrer, à travers la construction d'un échantillon, que la construction d'un tel modèle, selon une approche expérimentale, est possible. Mais d'abord, dans la prochaine section, nous entendons expliciter le lien entre la lexicographie et l'enseignement du vocabulaire et justifier la prémisse de notre projet selon laquelle un entraînement à la lexicographie rend plus aisé l'enseignement du vocabulaire.

1.3 Lien entre la lexicographie et l'enseignement du lexique

Rappelons que l'extraction des savoirs lexicographiques effectuée dans le cadre du projet Lexitation a pour but d'élaborer une ontologie des savoirs lexicographiques de laquelle seront dérivées des applications en enseignement-apprentissage du vocabulaire. En effet, Lexitation s'appuie sur le postulat suivant (Polguère, 2007) :

Les savoirs lexicographiques jouent un rôle crucial non seulement en lexicologie et lexicographie, mais aussi dans le contexte de l'enseignement-apprentissage des connaissances lexicales.

Ce postulat s'appuie sur l'observation que la maîtrise des savoirs lexicographiques augmente considérablement l'autonomie fonctionnelle des professionnels de la langue, et tout particulièrement celle des enseignants de langue. Cet état de fait s'explique par l'existence d'un lien entre la lexicographie et l'enseignement du vocabulaire, lien qui est perceptible à condition d'identifier ce que l'on veut enseigner à propos du lexique. En effet, ce qui est essentiel d'enseigner n'est pas limité à des listes de vocabulaire, comme l'a suggéré De Pietro (2003). S'il n'est pas souhaitable d'enseigner des listes d'associations signifiés-signifiants, il faut plutôt viser la maîtrise de connaissances lexicales qui, selon Polguère et Tremblay (2003), se trouvent sur deux axes : la **polysémie**, qui doit nous forcer à isoler les différentes acceptions d'un vocable pour les caractériser, et la **connexion lexicale**, qui fait qu'une unité lexicale n'existe que par (i) son positionnement dans le réseau lexical de la langue, — liens paradigmatiques — et (ii) ses propriétés individuelles, qui contrôlent la façon dont elle se combine aux autres unités lexicales dans la phrase — liens syntagmatiques — (Polguère et Tremblay, 2003, 27).

Pour rendre les notions de liens lexicaux paradigmatiques et syntagmatiques plus claires, nous donnerons ici une explication inspirée de celle de (Polguère, 2008, 106) :

Les liens lexicaux **paradigmatiques** connectent les lexies **à l'intérieur du lexique**. Par exemple, la lexie CHEVEUX est liée paradigmatiquement aux lexies CHEVELU, ÉCHELEVÉ, CHAUVÉ, COIFFEUR, etc.

Les liens lexicaux **syntagmatiques** connectent quant à eux les lexies **à l'intérieur de la phrase**. Par exemple, la lexie CHEVEUX s'emploie dans les expressions suivantes :

cheveux fins, épais, crépus, clairsemés, gras, sel et poivre ; se couper, se raser, se démêler...les cheveux, etc.

Les apprenants d'une langue maternelle, seconde ou étrangère doivent maîtriser les connaissances liées à la polysémie et à la connexion lexicale pour pouvoir utiliser correctement une lexie. Ces deux axes de connaissances sont également au cœur du travail lexicographique, car ce sont à partir d'eux que se décrivent, se comprennent, s'apprennent et s'utilisent les lexies. Autrement dit, les connaissances que les élèves doivent assimiler à propos des lexies sont celles que les lexicographes doivent modéliser dans les articles de dictionnaire. Dans leur travail quotidien, les enseignants doivent donc effectuer des tâches semblables à celles que font les lexicographes. Certains chercheurs en didactique ont d'ailleurs déjà proposé des méthodes d'enseignement du vocabulaire qui ont beaucoup en commun avec le travail lexicographique. Parmi les plus influentes, il y a celle du dictionnaire pédagogique anglais *Longman Language Activator* (Summers, 1993) et, pour le français, la méthode présentée par Jacqueline Picoche dans *Didactique du vocabulaire français* (Picoche, 1993), est également d'un grand intérêt. Dans cet ouvrage, Picoche propose une méthode didactique composée d'une suite d'activités qui sont similaires aux étapes de la phase du travail lexicographique qu'Atkins appelle *l'analyse* (voir Section 1.2). Il s'agit en gros d'utiliser la syntaxe comme révélateur de sens, d'ébaucher des définitions par genre prochain et différences spécifiques³ et de dégager la structure polysémique des vocables. En somme, Picoche propose trois types d'activités pédagogiques pour enseigner le vocabulaire, qui correspondent à une étude des lexies sous les facettes de la sémantique, de la syntaxe et de la combinatoire lexicale. Or, ce sont les trois facettes que l'on décrit lorsque l'on décrit une lexie dans un dictionnaire. En effet, selon les travaux en lexicologie explicative et combinatoire, un dictionnaire doit étudier les lexies sous les aspects sémantique, syntaxique et lexico-combinatoire (Mel'čuk et al., 1995). Il semble clair que la méthode de Picoche se rapproche de la lexicographie, en ce

³Le genre prochain est une *paraphrase approximative minimale du sens de la lexie définie*, et les différences spécifiques sont les *composantes sémantiques périphériques qui caractérisent le sens de la lexie par rapport 1) à son genre prochain et 2) au sens de toutes les autres lexies de la langue ayant le même genre prochain* (Polguère, 2008, 182). Polguère appelle les définitions en question *définitions analytiques*.

que sa méthode propose des activités qui permettent de traiter des aspects du lexique qui sont aussi au cœur des préoccupations de plusieurs ouvrages théoriques sur la lexicographie.

1.4 Enseignement du lexique : état de la question

Au Québec, comme ailleurs dans la francophonie, les chercheurs en didactique affirment que les compétences lexicales des élèves sont plus ou moins satisfaisantes, que les pratiques d'enseignement actuelles dans ce domaine sont inadéquates, et qu'il y a un déficit de recherches en didactique du lexique.

Au Québec, Claude Simard (1994) soulève le problème de l'enseignement du lexique, qui ne dispose pas d'une tradition pédagogique forte, contrairement à la grammaire, par exemple. D'ailleurs, Claude Simard, professeur au Département de didactique, psychopédagogie et technologie éducative de la Faculté des sciences de l'éducation de l'Université Laval, est l'un des rares chercheurs québécois qui s'intéressent à l'enseignement du vocabulaire. Parmi les autres, il y a Alain Polguère et Ophélie Tremblay, de l'Université de Montréal. Cette dernière, dans sa thèse de doctorat, qui propose un module de cours en didactique du lexique, expose les résultats d'une enquête qu'elle a menée pour connaître la part du lexique dans la formation des futurs enseignants du primaire. Elle a constaté que, dans les cursus de formations universitaires du système québécois actuel, il ne semble pas exister de cours en formation initiale des maîtres visant explicitement le développement des connaissances métalexicales pour l'enseignement du lexique au primaire en français langue maternelle (Tremblay, 2009). À l'exception de l'Université Laval, qui offre un cours intitulé *Didactique du lexique, de la grammaire et de l'orthographe* (DID-19984), aucun programme de formation initiale ne contient de cours explicitement consacré à la didactique du lexique, alors que l'on retrouve pourtant des cours thématiques, par exemple en didactique de l'oral ou de l'écrit. Nous supposons que l'enseignement du lexique est traité dans certains cours de didactique de l'écrit, sans qu'il en soit fait mention explicitement dans le titre du cours. Malgré cela, nous pouvons nous demander si les enseignants sont suffisamment préparés pour enseigner le vocabulaire

au primaire. Pour ce qui est des enseignants du secondaire, les principales universités québécoises leur offrent des cours d'introduction à la lexicologie et à la lexicographie, mais ne proposent pas à proprement parler de réflexion sur la manière de susciter le développement de connaissances lexicales chez les élèves. Donc, au secondaire autant qu'au primaire, la formation initiale des maîtres ne contient pas de cours explicitement consacré à la didactique du lexique. Y a-t-il un besoin pour une formation plus poussée en lexicologie et en didactique du lexique chez les futurs enseignants ? Si l'on en croit les spécialistes du domaine (d'ici et d'ailleurs), il semblerait que oui.

En effet, il y aurait une certaine « déficience » quant à la maîtrise du lexique chez les jeunes selon Simard (1994) (au Québec) et De Pietro (2003) (en Suisse), et le fait que le vocabulaire ne soit pas l'objet d'un enseignement adéquat à l'école n'aide pas à corriger cette situation.

En France, David et Grossmann (2003) remarquent que le lexique est négligé dans les apprentissages en langue, réduit à un enseignement fossilisé dans des routines pédagogiques, et que *l'élève est alors livré à lui-même face à son dictionnaire et mis en position d'accomplir seul les opérations de traitement lexical* (David et Grossmann, 2003, 3). Ainsi, il semble que ce qui se fait en ce moment dans la pratique de l'enseignement n'est pas efficace. Dans beaucoup de classes, l'apprentissage du vocabulaire prend la forme de mémorisation de listes lexicales, ce qui, pour certains, représente un risque de développer un *non-enseignement du vocabulaire fondé sur des notions aussi peu didactiques que l'accumulation, l'inculcation, le par cœur, bref, un enseignement qui favorise une certaine répulsion pour le français* (De Pietro, 2003). Il semblerait qu'en général, autant au Québec qu'en Europe, on espère que l'apprentissage du lexique se fera par simple imprégnation (Polguère, 2004). Or, selon De Pietro (2003), l'apprentissage par imprégnation ne constitue qu'une partie de ce qu'on observe lorsqu'on analyse la manière dont les locuteurs « non experts » traitent les unités lexicales dans les échanges quotidiens. De Pietro souligne aussi à raison qu'*on ne sait guère ce qui se fait en classe lorsque les élèves ne comprennent pas un mot ou s'avèrent incapables de produire le mot attendu, lorsque l'enseignant veut expliquer un mot nouveau, bref, pour que les élèves élargissent leur vocabulaire* (De Pietro, 2003, 13). Ces préoccupations sont aussi présentes dans le

contexte québécois, puisque, comme nous l'avons remarqué précédemment, il n'existe pas de véritable didactique du vocabulaire dans la culture scolaire québécoise.

À l'heure actuelle, les enseignants sont rarement outillés pour mener spontanément des interventions qui demandent de modéliser des phénomènes lexicaux, comme l'a remarqué Polguère (2004). Un des objectifs du projet *Lexitation*, et celui qui nous concerne, est donc d'accroître l'autonomie des enseignants — autant ceux du primaire que du secondaire — en élaborant un outil avec lequel ils pourraient s'initier à la lexicographie. Nous croyons en effet qu'un entraînement à la lexicographie leur permettrait de développer des méthodes d'analyse et des stratégies de réponse aux questions concernant le lexique. Plus exactement, nous voulons utiliser les savoirs lexicographiques recensés chez les lexicographes professionnels, savoirs que nous avons extraits dans les premières phases de notre projet et modélisés dans notre ontologie, pour élaborer des outils pédagogiques servant à cette initiation. Avant d'exposer comment des tels outils pourraient être construits, nous allons présenter les phases d'extraction des savoirs lexicographiques et de modélisation ontologique de ces savoirs. Le chapitre 5 sera consacré aux perspectives sur l'utilisation des connaissances lexicographiques dans un contexte d'enseignement.

CHAPITRE 2

PRÉSENTATION DE LA RECHERCHE : CONTEXTE ET OBJECTIFS

2.1 Projet Lexitation

2.1.1 Constat de départ et buts du projet Lexitation

Le présent mémoire de maîtrise s'inscrit dans le cadre d'un projet multidisciplinaire financé par le Conseil de recherches en sciences humaines du Canada (CRSH), intitulé *Modélisation ontologique des savoirs lexicographiques en vue de leur utilisation dans les disciplines de linguistique appliquée* — surnommé *Lexitation*. Selon son descriptif (Polguère, 2007), ce projet vise à faire sortir la recherche en lexicographie de son champ d'application traditionnel (la compilation de dictionnaires et de bases de données lexicales) pour lui donner un statut de champ de compétences général. En effet, il est fondé sur le postulat que les savoirs lexicographiques jouent un rôle crucial non seulement en lexicologie et en lexicographie, mais aussi dans le contexte de l'enseignement-acquisition des connaissances lexicales. Ce postulat s'appuie sur l'observation, faite par les membres de l'équipe de recherche, que *la maîtrise des savoirs lexicographiques augmente considérablement l'autonomie fonctionnelle des professionnels de la langue et qu'elle leur permet de fonctionner de façon optimale dans des situations où l'analyse des phénomènes linguistiques doit être effectuée instantanément* — par exemple, en enseignement de la langue, lors d'interactions en classe (Polguère, 2007). Autrement dit, nous voulons extraire les savoirs des lexicographes professionnels et les modéliser pour ensuite les utiliser dans le domaine de la formation à l'enseignement de la langue. Le projet se doit en outre de répondre aux questions suivantes, telles que présentées dans le descriptif du projet (Polguère, 2007) :

1. *Quelle peut être la modélisation formelle des connaissances déclaratives et procédurales qui sont présupposées par l'activité lexicographique ?*
2. *Comment ces deux types de savoirs interagissent-ils dans le processus d'ana-*

lyse des phénomènes lexicaux, processus qui est au cœur de l'activité lexicographique ?

3. *Comment intégrer l'accès aux données linguistiques volatiles que sont les phénomènes de la parole (langue en contexte d'emploi dans les corpus écrits et oraux ainsi qu'intuition linguistique de la personne pratiquant l'analyse des données) et leur utilisation dans un modèle général du savoir lexicographique ?*
4. *Comment évaluer l'influence que la maîtrise du savoir lexicographique peut avoir sur la capacité à travailler sur la langue, et notamment à l'enseigner ?*

2.1.2 Équipe de recherche

Lexitation est un projet multidisciplinaire : il regroupe des spécialistes des domaines de la lexicographie, de la modélisation ontologique des savoirs et de la didactique du français. Il est d'ailleurs le fruit d'une collaboration entre l'Observatoire de linguistique Sens-Texte (OLST) de l'Université de Montréal, le Département de didactique de l'Université de Montréal et le Laboratoire en informatique cognitive et environnements de formation (LICEF) de la Télé-université. L'équipe de recherche est composée d'Alain Polguère (OLST et Département de linguistique et de traduction de l'Université de Montréal), responsable du projet, d'Igor Mel'čuk (OLST et Département de linguistique et de traduction de l'Université de Montréal), de Pascale Lefrançois (OLST et Département de didactique de l'Université de Montréal), et de Jacqueline Bourdeau (LICEF). Les étudiants suivants, tous membres de l'OLST, sont aussi impliqués dans le projet, dans le cadre de leur recherche de maîtrise ou de doctorat : Anne-Laure Jousse, Sara-Anne Leblanc, Ophélie Tremblay et l'auteure du présent mémoire.

2.1.3 Moyens

Lexitation est une recherche originale de par les objectifs qu'elle poursuit, mais aussi de par sa méthodologie : nous appliquons les méthodes propres à l'analyse et à l'extraction des connaissances et compétences pour modéliser un savoir disciplinaire qui n'a jamais été étudié expérimentalement (voir Section 1.2). En effet, pour extraire les savoirs

lexicographiques et pour les représenter de façon formelle et explicite, nous utilisons les approches de modélisation ontologique généralement utilisées pour la construction d'environnements didactiques. Nous nous inspirons des recherches de Paquette (2002), de Psyché et al. (2003) et de Bourdeau et al. (2004), qui fournissent des modèles solides de modélisation des savoirs pour la construction d'environnements de formation. Avant de procéder à la modélisation des savoirs lexicographiques, nous devons d'abord extraire ces savoirs. Les travaux pour réaliser un modèle ontologique des savoirs lexicographiques se font donc sur deux axes méthodologiques :

1. expérimentation par simulation de situations de travail lexicographique et analyse de données ;
2. recours à des techniques et formalismes de modélisation propres à l'ingénierie des connaissances.

Pour réaliser nos expérimentations, dans lesquelles des lexicographes professionnels participent à des simulations de travail lexicographique, nous utilisons les ressources expérimentales et l'expertise disponibles au Laboratoire-Observatoire de recherche en Ingénierie du Téléapprentissage (LORIT) du centre de recherche LICEF de la Télé-université. Ensuite, nous analysons les données recueillies lors des expérimentations et nous identifions les savoirs à modéliser. La modélisation elle-même relève du deuxième axe méthodologique. Nous avons choisi un modèle d'encodage de type ontologique (voir Section 4.2) pour modéliser les savoirs lexicographiques. Enfin, dans une phase subséquente du projet, la modélisation ontologique sera testée pour explorer son utilisation en didactique de la langue, plus particulièrement, du français.

En tout, le projet Lexitation est composé de 5 grandes tâches (Polguère, 2007) :

1. explicitation des savoirs lexicographiques déjà présentés de façon semi-formelle dans notre modèle lexicologique de référence, la LEC (Mel'čuk et al., 1995) ;
2. exploration de la problématique de l'accès aux données linguistiques en lexicologie et en lexicographie, grâce à l'introduction des techniques de forage textuel

traditionnellement appliquées en recherche d'information (tâche initialement planifiée dont l'accomplissement a été écartée de Lexitation faute de ressources suffisantes) ;

3. extraction expérimentale des savoirs par simulation de l'activité lexicographique et analyse de données recueillies - utilisation de l'infrastructure du LORIT ;
4. modélisation ontologique formelle des connaissances et compétences lexicographiques complétant celles identifiées dans la première étape ; notamment pour ce qui est de l'accès aux données linguistiques ; mise en place de procédures de validation, test et révision du modèle ontologique ;
5. exploration du potentiel applicatif de l'enseignement des savoirs ainsi modélisés en didactique des langues.

2.2 Objectifs de la présente recherche

Lexitation est un projet de 3 ans, et certaines tâches déjà accomplies ou prévues ne seront pas couvertes par ce mémoire. Celui-ci concerne particulièrement les tâches 3, 4 et 5. Voici les objectifs spécifiques de la présente recherche, divisés en trois grandes tâches (3, 4 et 5) du projet Lexitation :

- Objectifs de la présente recherche relevant de la tâche 3 du projet Lexitation, *Extraction expérimentale des savoirs* :
 - élaboration d'une méthode d'extraction des savoirs lexicographiques ;
 - extraction des savoirs lexicographiques utilisés par les lexicographes lors de deux expérimentations.
- Objectifs de la présente recherche relevant de la tâche 4 du projet Lexitation, *Modélisation ontologique* :
 - identification de l'approche la plus appropriée pour le problème à l'étude ;

- construction du modèle ontologique ;
 - formalisation et encodage des savoirs lexicographiques extraits des deux expérimentations analysées.
- Objectifs de la présente recherche relevant de la tâche 5 du projet Lexitation, *Exploration du potentiel applicatif de l'enseignement des savoirs modélisés* :
 - explicitation du lien entre l'enseignement du vocabulaire et la lexicographie ;
 - identification des pratiques d'enseignement pour lesquelles un entraînement à la lexicographie pourrait être utile ;
 - évaluation sommaire des pratiques actuelles en didactique du lexique ;
 - réflexion sur la possibilité de dériver des outils pédagogiques à partir d'une ontologie des savoirs lexicographiques et présentation des caractéristiques de ces possibles outils pédagogiques ;
 - évaluation de la possibilité d'enseigner les savoirs lexicographiques modélisés dans le cadre des programmes de formation à l'enseignement.

En d'autres termes, le principal objectif de notre recherche était d'élaborer et de décrire une méthode appropriée pour extraire les savoirs lexicographiques mobilisés par les lexicographes durant nos expérimentations, qui consistent à simuler une activité lexicographique (voir Section 3.3). La particularité de cette tâche réside dans le fait que les savoirs lexicographiques doivent être extraits à partir de verbalisations faites par les lexicographes, qui doivent suivre le *Think Aloud Protocol* (Ericsson et Simon, 1993) (voir Section 3.2). Une fois la méthodologie d'analyse élaborée, nous avons comme objectif de procéder à l'analyse complète des données — sous forme de documents vidéo (voir Section 3.3.3) — d'une expérimentation complète (nous en avons effectué 5 jusqu'à maintenant, voir Section 2.4), et de tester le modèle d'analyse avec les données d'une autre expérimentation. Il a d'abord fallu déterminer un environnement approprié pour la modélisation ontologique des savoirs extraits, construire la charpente d'une base de données de type ontologique et élaborer un patron d'encodage des savoirs lexicographiques

dans cette base de données. Une fois la base de données construite, il s'agissait d'encoder les savoirs extraits des deux expérimentations analysées dans l'ontologie. Ce travail a été fait en collaboration avec le directeur de la présente recherche, Alain Polguère, qui est aussi le responsable du projet Lexitation. Nous avons comme dernier objectif d'amorcer une réflexion sur les possibles applications de notre modèle ontologique dans un contexte pédagogique : celui de l'enseignement du vocabulaire, entre autres par le biais de la formation des maîtres. C'est pourquoi nous avons fait appel aux services de Pascale Lefrançois, professeure au Département de didactique de l'Université de Montréal déjà impliquée dans le projet Lexitation, pour codiriger cette recherche.

Dans les prochains chapitres de ce mémoire, nous décrirons tout d'abord comment s'est organisé le contexte expérimental de la première série d'expérimentations du projet Lexitation (Chapitre 3). Ensuite, nous expliquerons comment nous procédons pour extraire des données expérimentales et pour modéliser les divers types de connaissances et stratégies utilisées par les lexicographes au cours d'activités lexicographiques (Chapitre 4). Finalement, nous expliquerons dans quelle mesure il nous semble possible d'exploiter le modèle ontologique des savoirs lexicographiques que nous avons commencé à construire dans un contexte pédagogique (Chapitre 5).

Notons que l'explicitation des savoirs lexicographiques, qui correspond à la première tâche du projet Lexitation, a été réalisée au préalable. On peut trouver une description de cette tâche dans la thèse de doctorat d'Ophélie Tremblay (2009, Chapitre 4). Aussi, la mise en place de procédures de validation du modèle ontologique reste à faire, et l'exploration du potentiel didactique du modèle sera aussi à approfondir, dans le cadre de travaux de plus longue haleine que le présent mémoire.

2.3 Cadre théorique : la lexicologie explicative et combinatoire

Le projet Lexitation vise la modélisation des concepts et des stratégies utilisés par les lexicographes quel que soit leur cadre théorique. En effet, comme nous avons vu, la troisième tâche du projet Lexitation consiste à extraire et à modéliser les savoirs de plusieurs lexicographes, issus de divers horizons théoriques. Or, même si on veut observer toutes

sortes de pratiques et d’approches lexicographiques, il est nécessaire de s’appuyer sur un cadre théorique unique. En effet, les lexicographes participant à nos expérimentations utilisent différentes terminologies, sous-tendant différentes théories, mais une ontologie des savoirs — quel que soit son domaine — doit être univoque, c’est-à-dire qu’il n’est pas souhaitable de voir des classes de savoirs équivalentes se multiplier dans une ontologie. Par exemple, si un participant utilise le terme *partie du discours*, un autre, *classe grammaticale*, et un autre, *catégorie syntaxique*, il faudra rassembler ces notions sous un seul terme (ou un terme de base, qui sera au besoin lié à des variantes), car on ne peut pas se permettre d’avoir un triplet de classes désignant une opération qui implique cette notion. Il est donc préférable d’avoir un réseau notionnel de référence comme base de travail. Or, l’approche de la LEC est particulièrement appropriée pour servir au développement d’une modélisation formelle de la compétence lexicographique. En effet, il existe déjà une littérature théorique (Mel’čuk et al. (1995) ; Polguère (2008)), qui correspond à ce que l’on pourrait appeler une *ontologie semi-formelle* de cette théorie du lexique : réseau notionnel, ensemble de critères et de méthodes, etc. Ces ouvrages de description sont présentés en langue naturelle, mais en suivant des contraintes d’expression et des contraintes terminologiques strictes.

En outre, l’approche de la LEC présente les avantages suivants, présentés dans le descriptif du projet Lexitation (Polguère, 2007) :

- *elle est formelle (notions descriptives et formalismes de représentation précisément définis) ;*
- *elle a une valeur universelle, ayant été appliquée sur une grande variété de langues très diverses ;*
- *elle vise spécifiquement la production de modélisations lexicales, appelées dictionnaires explicatifs et combinatoires (DEC) Mel’čuk et al. (1984, 1988, 1992, 1999) ;*
- *elle est appliquée à la construction de bases de données lexicales formelles, manipulables informatiquement, — cf. la base de données DiCo décrivant la com-*

binatoire lexicale d'un sous-ensemble représentatif du lexique français (Polguère, 2000);

- *son potentiel d'application à la didactique des langues — déjà examiné dans le passé (Leed et Nakhimovsky (1990); Steele (1990); Gentilhomme (1992)) — a été confirmé par des études récentes (Cauchon (2003); Tremblay (2003)) et intéresse de plus en plus les chercheurs en didactique;*
- *elle est utilisée avec succès dans d'autres domaines de linguistique appliquée, comme la terminologie monolingue et multilingue (Dancette et Réthoré (2000); L'Homme (2002));*
- *elle sert aussi à la construction de modèles vulgarisés adaptés à l'enseignement-apprentissage de la langue — cf. le Lexique Actif du Français (LAF) (Mel'čuk et Polguère, 2007).*

De plus, le réseau notionnel de base de la LEC est présenté dans *Lexicologie et sémantique lexicale, notions fondamentale* (Polguère, 2008), qui est une introduction à l'étude du lexique et de la sémantique lexicale, et qui se présente comme un réseau notionnel facilement « navigable ». Chaque chapitre introduit un ensemble de notions de base logiquement regroupés. Dans l'optique d'une utilisation pédagogique de l'ontologie, l'existence de cet ouvrage est particulièrement appréciable, les utilisateurs de l'ontologie pouvant s'y référer au besoin. Pour ces raisons, nous avons choisi d'utiliser la LEC comme réseau notionnel de base de notre ontologie.

2.4 Où en sommes-nous ?

Nous avons jusqu'à présent effectué cinq expérimentations. Les deux premières consistaient une sorte de phase de rodage, lors de laquelle des étudiantes initiées à la lexicographie, dont l'auteure du présent mémoire, ont servi de cobayes-lexicographes. Le lexicographe ayant participé à la troisième expérimentation était un lexicographe à l'emploi de *Druide informatique*, la compagnie qui commercialise le logiciel de correction *Anti-dote*. C'est principalement cette expérimentation qui a fait l'objet de l'analyse et de la

modélisation qui seront présentés dans les chapitres suivants. Pour la quatrième expérimentation, c'est un lexicographe du dictionnaire *Franqus*, un dictionnaire du français standard en usage au Québec développé à l'Université de Sherbrooke, qui s'est prêté au jeu. Finalement, un lexicographe du *Trésor de la langue française* (TLF) a accepté de participer à notre cinquième expérimentation. Nous prévoyons faire d'autres expérimentations, en lexicographie anglaise et éventuellement en lexicographie bilingue, mais celles-ci ne sont pas comprises dans le cadre du présent mémoire. Aussi, il n'est pas exclu que nous complétions notre ontologie avec des savoirs lexicographiques décrits dans des ouvrages spécialisés, tels que Fontenelle (2008).

CHAPITRE 3

APPROCHE EXPÉRIMENTALE DE L'EXTRACTION DES SAVOIRS LEXICOGRAPHIQUES

Rappelons que le but de nos expérimentations est d'extraire les connaissances déclaratives et procédurales impliquées en lexicographie. Pour ce faire, nous avons observé des lexicographes au travail. Pour chaque expérimentation, un ou une lexicographe a donc été invité(e) au Laboratoire-Observatoire de Recherche en Ingénierie du Téléapprentissage (LORIT), où il ou elle a accompli une activité lexicographique que nous avons enregistrée. Nous expliquerons dans ce chapitre comment a été implantée l'approche expérimentale et pourquoi une telle approche a été choisie.

3.1 Nature du travail lexicographique

Une approche expérimentale a été adoptée en raison de la nature même du travail lexicographique. Nous nous permettons de citer ici Comeau et Polguère (2008, 2) pour la description du travail lexicographique :

Le travail lexicographique est une activité dont on acquiert la maîtrise au cours d'un long apprentissage, fondé essentiellement sur la pratique. Il s'agit de ce que l'on appelle en anglais a craft. Comme tout apprentissage de ce type, celui du métier de lexicographe doit idéalement se faire sous la supervision d'experts, qui transmettent leur savoir d'une façon qui n'est généralement pas structurée linéairement, étape par étape. En ce sens, on pourrait envisager un enseignement de la lexicographie reposant sur des principes similaires à ceux du compagnonnage, pour l'apprentissage des métiers ouvriers traditionnels.

Nous savons qu'il faut beaucoup de temps pour maîtriser les techniques du métier de lexicographe. Pour devenir un expert, il faut aussi être exposé à un *input* linguistique

soutenu et s'exercer à la modélisation des données linguistiques auxquelles on est exposé :

Le lexicographe est un observateur qui se soumet à un bombardement continu de données linguistiques, qu'il perçoit et examine sans répit selon une perspective métalinguistique, et ce, même lorsqu'il n'est pas à sa table de travail. En fait, toute personne ayant pratiqué un tant soit peu la lexicographie a fait l'expérience de la véracité de cette règle : travaillez sur un mot donné, et vous pouvez être sûr de l'entendre (ou de le voir) employé dans les heures ou les jours qui suivent d'une façon qui viendra enrichir le travail de modélisation déjà effectué. (Comeau et Polguère, 2008, 2)

Alors comment modéliser les savoirs lexicographiques, si ceux-ci se définissent avant tout par des pratiques ? C'est un problème qui relève du domaine de l'extraction et de la formalisation des connaissances d'experts. Le travail à effectuer relève donc de l'ingénierie de la connaissance, en anglais *knowledge engineering*. Bien qu'il existe d'autres méthodes d'accès à l'expertise professionnelle (enquêtes auprès d'experts, extraction et formalisation des règles contenues explicitement et implicitement dans les textes des domaines concernés, etc.), une approche expérimentale, qui consiste à recréer le contexte de travail du lexicographe et à enregistrer la séance de travail (Comeau et Polguère, 2008), a été choisie. Il s'agit ensuite d'extraire les savoirs de la pratique lexicographique elle-même, à partir des données recueillies lors des expérimentations. Nous allons décrire plus précisément, dans les sections qui suivent, comment s'articule notre approche expérimentale de l'extraction des savoirs lexicographiques.

3.2 Utilisation du *Think Aloud Protocol*

Pour extraire les connaissances déclaratives et procédurales utilisées par un lexicographe, nous devons pouvoir accéder à celles-ci et, si possible, les observer directement. Or, lorsqu'un lexicographe travaille, il est généralement seul devant son ordinateur (nous supposons que toutes les ressources nécessaires à son travail sont disponibles sur support

informatique, comme cela est de plus en plus le cas). Conséquemment, les faits que nous pouvons observer sont les opérations qu'il effectue sur son ordinateur et qui se concrétisent à l'écran, et son travail final. Ces données observables sont utiles, mais elles ne sont que la manifestation du travail intellectuel effectué par le lexicographe. Les opérations mentales devront donc être inférées de ces données. Procéder ainsi nous expose à deux problèmes, identifiés dans (Comeau et Polguère, 2008, 3) :

Premièrement, on peut échouer dans l'identification de nombreuses connaissances inférables des données (soit par simple omission, soit par incapacité à comprendre le véritable modus operandi du lexicographe). Deuxièmement, on peut tout simplement mal interpréter ce que fait le lexicographe, par exemple confondre une succession quasi aléatoire d'opérations avec ce qui serait l'application méthodique d'une stratégie préconstruite.

Pour éviter ce problème, nous avons suivi un protocole verbal, le *Think Aloud Protocol* décrit par Ericsson et Simon (1993). Il s'agit, *grosso modo*, de faire en sorte que le lexicographe verbalise sa pensée lors de son travail. Cette verbalisation est enregistrée puis analysée en parallèle avec les autres données recueillies (enregistrement du travail effectué sur l'ordinateur).

Nous aurions pu choisir un autre type de protocole verbal, le *Think After Protocol*, par exemple. En effet, Ericsson et Simon (1984) présentent deux types de protocoles pour générer des données sur les processus cognitifs :

- les protocoles simultanés ou, en anglais, *concurrent protocols* (également connus sous le nom de *Think Alouds*) ;
- les protocoles rétrospectifs (également connus sous le nom de *Think Afters*).

Dans les protocoles simultanés, les participants doivent rapporter leur pensées et décrire ce qu'ils font pendant qu'ils accomplissent une certaine tâche, tandis que les protocoles rétrospectifs consistent à recueillir ces informations après coup, en faisant visionner au participant le vidéo de la tâche et en lui demandant de narrer l'enregistrement

en exposant ses opérations mentales. Les deux techniques peuvent être influencées par un « déplacement motivationnel » (*motivational shift*), qui risque toujours de se produire quand les participants d'une étude savent qu'ils sont observés. C'est-à-dire que ceux-ci agissent de la façon qu'ils croient être la façon attendue par l'expérimentateur au lieu d'agir comme ils le feraient normalement (Russo et al., 1989). Malgré ce possible biais, les protocoles verbaux, tant simultanés que rétrospectifs, fournissent une grande quantité de données sur les processus cognitifs, qui ne pourraient être obtenues autrement.

Il peut être affirmé que le *Think Aloud Protocol* est dérangent pour le participant, qui doit verbaliser ses actions tout en travaillant, ce qui peut alourdir la charge cognitive et ralentir la vitesse de travail. Cela étant dit, on l'a quand même préféré au *Think After Protocol*, parce que celui-ci est plus susceptible d'être influencé par l'oubli ou par l'invention, comme le montrent Branch (2000) et Ericsson et Simon (1984). De plus, nous avons remarqué, lors de nos premières expérimentations, que plus les participants sont expérimentés, moins ils éprouvent de difficulté à faire les verbalisations de type *Think Aloud*. Cette observation est cohérente avec les études d'Ericsson et Simon (1984) et de Branch (2000), qui ont indiqué que les *Think Alouds* sont moins utiles quand les participants effectuent une tâche nouvelle ou impliquant une charge cognitive importante. Il faut dire que, même quand le participant verbalise bien, il est possible qu'il oublie de verbaliser certaines opérations mentales, mais il semblerait que, plus il s'agit d'effectuer une tâche coutumière, plus le *Think Aloud Protocol* se fait naturellement. Nos participants étaient tous des lexicographes professionnels expérimentés (sauf pour la phase de test, où des étudiantes ont servi de cobayes) et ils n'ont aucunement eu l'air gênés par le *Think Aloud Protocol*. Voici les cinq raisons qui, selon nous, expliquent cet état de fait, telles que présentées dans (Comeau et Polguère, 2008, 3) :

1. *Nous nous concentrons sur le travail effectué par des lexicographes qui ont acquis une très grande maîtrise de leur domaine. On peut donc supposer qu'ils ont une vision assez claire de ce qu'ils font, même s'ils n'ont pas nécessairement une distanciation véritablement théorique vis-à-vis de leur discipline.*
2. *Si les lexicographes professionnels avec lesquels nous avons travaillé n'avaient*

aucun problème à verbaliser leur pensée, c'est peut-être parce que, d'une part, le matériau sur lequel ils travaillent est la langue elle-même et, d'autre part, leur pratique repose sur une terminologie bien établie (même si elle peut varier d'un contexte lexicographique à l'autre), terminologie qui supporte bien leur pensée et son expression.

3. *Un membre de l'équipe menant l'expérimentation — que nous avons appelé le pilote du lexicographe — est chargé d'« accompagner » le lexicographe dans son travail en stimulant sa verbalisation, au cas où celui-ci tendrait à oublier qu'il doit décrire les différentes tâches qu'il accomplit (voir Section 3.3.2).*
4. *Les autres membres de l'équipe assistent à l'expérimentation en tant que spectateurs cachés, dans un local séparé du laboratoire où elle se déroule (la salle de contrôle), et peuvent communiquer avec le pilote pour lui donner ponctuellement des instructions. Ils veillent à ce que l'ensemble de l'expérimentation se déroule sans accroc dans le temps alloué ; pour cette raison, nous les appelons les contrôleurs de l'expérimentation (voir Section 3.3.2).*
5. *Chaque expérimentation est structurée autour d'une activité lexicographique unique, préétablie, et est limitée dans le temps : une heure. Cela permet de garder à l'expérimentation une finalité pratique, dans un temps suffisamment long pour l'accomplissement d'un travail significatif, mais suffisamment court pour que la concentration du lexicographe soit maintenue (voir Section 3.3.1).*

Comme les premiers lexicographes avec lesquels nous avons travaillé n'avaient pas de difficulté à verbaliser leur pensée, nous avons décidé de continuer à utiliser ce protocole dans les expérimentations subséquentes. Cela nous permet d'éviter les problèmes de souvenirs incomplets qui pourraient arriver avec l'utilisation des *Think Afters*. En effet, étant donné que nos expérimentations sont relativement longues (une heure) et qu'elles comportent parfois des tâches relativement complexes, un protocole rétrospectif serait plus difficile à appliquer, et le risque d'oubli serait très élevé.

3.3 Organisation des expérimentations

3.3.1 Activités lexicographiques

Comme nous l'avons mentionné précédemment, lors de chaque expérimentation, le lexicographe effectue un travail particulier, que l'on appelle *activité lexicographique*. Il n'y a qu'une seule activité lexicographique par expérimentation, et celle-ci peut être choisie et élaborée par l'équipe d'expérimentateurs ou tirée du travail quotidien du lexicographe. Les activités lexicographiques peuvent donc être de différentes natures. Pour l'Expérimentation 3, l'activité avait été choisie par les expérimentateurs et consistait à dégager la structure polysémique d'un vocable. Le lexicographe devait identifier et décrire sommairement dans une fiche chaque acception qu'il identifiait (par introspection ou en consultant un corpus). Cette activité est intéressante parce qu'elle permet de mettre au jour le processus de conceptualisation même des lexies, qui est à la base du travail lexicographique. C'est cette même activité lexicographique qui avait été utilisée lors des deux premières expérimentations, en guise de test. Par contre, lors des Expérimentations 4 et 5 (voir Section 4.3 pour description de l'Expérimentation 5), nous avons demandé aux lexicographes de faire pendant une heure une activité tirée de leurs tâches quotidiennes, ou du moins, une activité représentative de leur travail habituel. Ils ont donc effectué respectivement de la révision d'articles de dictionnaire et l'ébauche d'une entrée de dictionnaire pour un néologisme. Les deux lexicographes travaillaient alors avec leur propres documents et utilisaient leurs propres outils. Pour les prochaines expérimentations, nous avons l'intention de continuer à procéder ainsi, c'est-à-dire de proposer aux participants des activités qui correspondent à celles qu'ils font dans leur travail habituel, afin de créer un contexte expérimental le plus naturel possible. Des activités pourront donc consister en différentes « phases » du travail lexicographique (ex : étude de la structure polysémique d'un vocable, rédaction de définitions, révision d'entrées de dictionnaire, etc.). Notons qu'une activité est au besoin effectuée récursivement au cours d'une même expérimentation. Par exemple, une activité de révision peut se faire pour une série d'articles de dictionnaire plutôt que pour un seul.

Dans le cas de l'Expérimentation 3 (et des deux premières), le lexicographe avait

reçu la consigne de dégager la structure polysémique du vocable **COMPTER**. Il avait à sa disposition un corpus informatisé, *Le Migou*, et il avait accès à Internet via le moteur de recherche Google. Il devait consigner chacune des acceptions qu'il identifiait dans une fiche du logiciel Filemaker, dont nous avons créé un modèle (voir la fiche modèle en Annexe 1). Il pouvait créer autant de fiches qu'il voulait. Chacune des fiches comportait 5 champs à remplir :

- **Lexie**, où le lexicographe écrivait le nom de la lexie et un numéro s'il le souhaitait (mais il n'avait pas reçu de consigne à cet égard) ;
- **Caractéristiques grammaticales**, où il pouvait écrire la partie du discours de la lexie, ou des caractéristiques qu'il jugeait pertinentes ;
- **Indice**, où il écrivait un indice qui lui permettait de distinguer la lexie d'un copolysème ¹, par exemple ;
- **Exemple**, où il devait écrire un exemple de phrase où apparaît la lexie en question. Cela pouvait être un exemple de son cru ou tiré d'un corpus ;
- **Remarque**, où il pouvait écrire toute remarque qu'il trouvait pertinente.

Nous avons demandé au lexicographe de travailler selon ses méthodes habituelles autant que possible, et d'utiliser le corpus ou Internet s'il en avait besoin, mais cela n'était pas obligatoire. Il avait jusqu'à une heure pour compléter l'activité, mais il pouvait arrêter avant s'il avait terminé.

3.3.2 Participants impliqués dans les expérimentations

Pour superviser la procédure du *Think Aloud Protocol* lors des expérimentations, un membre de l'équipe, appelé le « pilote », s'assied à côté du lexicographe pour l'encourager à penser à voix haute, en lui posant des questions telles que *Pourquoi avez-vous fait cette manipulation ?*, *Pouvez-vous me dire à quoi vous pensez en ce moment ?*, etc.

¹Deux lexies sont des *copolysèmes* si elles sont deux acceptions d'un même vocable (voir la page xiii pour la définition de *lexie* et la page 3 pour celle de *vocable*).

Le rôle du pilote est également d'aider le lexicographe en cas de problèmes techniques. Le lexicographe et le pilote sont dans une même pièce, la salle d'expérimentation (voir Section 3.3.3), à un poste de travail, tandis que les contrôleurs (les autres membres de l'équipe d'expérimentateurs) sont dans une pièce attenante, la salle de contrôle, équipée de moniteurs qui retransmettent ce qui se passe dans le laboratoire.

Voici un schéma de l'organisation matérielle des expérimentations. La double flèche indique une interaction orale bidirectionnelle entendue par tous les participants et la flèche pointillée une communication unidirectionnelle que le lexicographe ne perçoit pas.

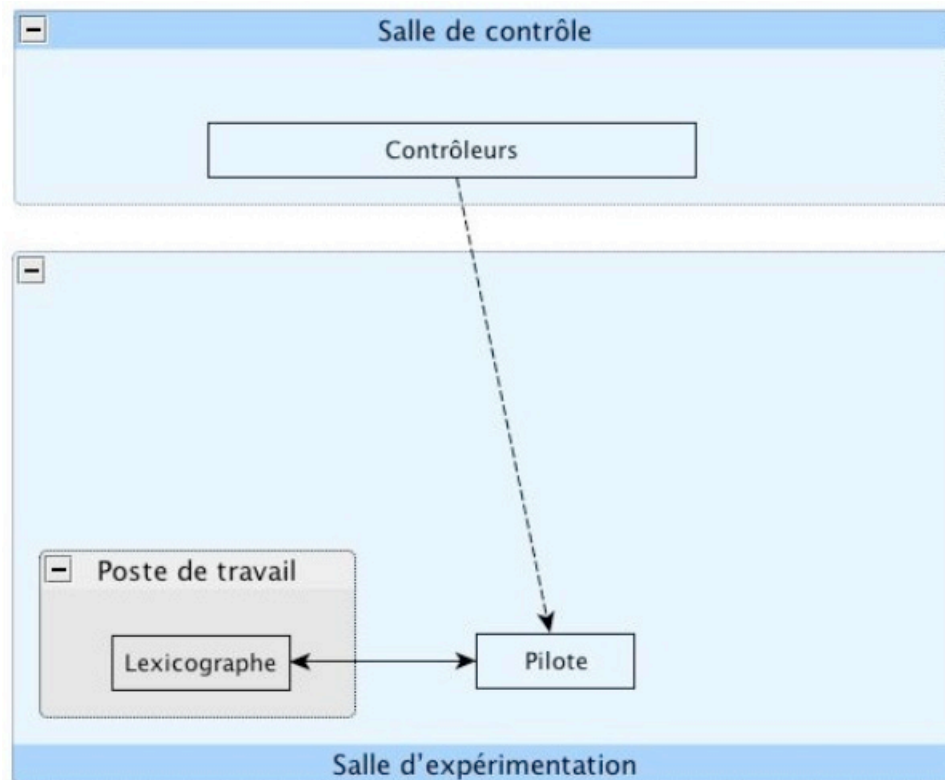


Figure 3.1 – Participants impliqués dans les expérimentations

Nous expliquerons dans la prochaine section comment le dispositif du LORIT permet une telle organisation.

3.3.3 Aménagement expérimental au Laboratoire-Observatoire de Recherche en Ingénierie du Téléapprentissage (LORIT)

Le LORIT est un laboratoire-observatoire de recherche en ingénierie du téléapprentissage. Il offre divers services pour effectuer des recherches en téléapprentissage ou en technologie de l'information, entre autres. Le dispositif technique du LORIT permet de collecter des données pour mener de telles recherches. C'est un environnement dont la configuration peut être aménagée pour réaliser des tests ou des observations. Le LORIT est formé d'un laboratoire (ou salle d'expérimentation), muni de postes de travail avec ordinateurs, et d'une régie (ou salle de contrôle). L'équipement audiovisuel et informatique permet de capturer les données d'une séance de travail faite dans le laboratoire. En effet, celui-ci est doté de trois caméras, de plusieurs micros et de dispositifs de capture d'écrans d'ordinateurs. Toutes les données ainsi recueillies peuvent être enregistrées et diffusées sur Internet en temps réel ou en différé. Les différents enregistrements (audio, vidéo et capture d'écran) peuvent être accompagnés d'un marquage temporel unique, qui permet la synchronisation des documents. Ainsi, les données recueillies peuvent toutes être mises en relation pour un segment d'activité lexicographique donné.

La régie, quant à elle, est équipée, entre autres, de moniteurs, qui peuvent diffuser les enregistrements du laboratoire, et de micros, par lesquels une personne située dans la régie pourra éventuellement communiquer avec quelqu'un qui se trouve dans le laboratoire.

Rappelons que, dans nos expérimentations, le lexicographe travaille dans le laboratoire, où il est filmé et où son écran d'ordinateur est capturé. Le pilote est également dans le laboratoire, auprès du lexicographe. Dans la régie, les autres membres de l'équipe de recherche, les contrôleurs, observent le lexicographe, grâce à des moniteurs, et veillent au bon déroulement de l'expérimentation.

Les enregistrements de la séance sont par la suite remis à l'équipe de recherche et sont analysés à l'OLST. Voici précisément toutes les données recueillies en cours d'expérimentation, présentées par ordre d'importance pour notre étude (Comeau et Polguère, 2008, 4) :

1. *enregistrement audio de la verbalisation effectuée par le lexicographe (Think Aloud Protocol) ainsi que des éventuelles interventions du pilote ;*
2. *enregistrement vidéo de la sortie d'écran du poste de travail du lexicographe ;*
3. *enregistrement audio des éventuelles interventions faites par les contrôleurs auprès du pilote ;*
4. *enregistrement vidéo de la gestuelle du pilote et du lexicographe (données d'une importance tout à fait secondaire pour notre projet).*

Une fois les données recueillies, il faut les analyser. Dans le prochain chapitre, nous présenterons la méthodologie d'analyse de ces données.

CHAPITRE 4

ANALYSE ET ENCODAGE DES DONNÉES

4.1 Analyse des données

4.1.1 Approche inductive

Lors de l'analyse des données recueillies dans les expérimentations, la « matière première » dont nous disposons pour extraire les connaissances déclaratives et procédurales du lexicographe sont les verbalisations que celui-ci effectue. Ainsi, nous n'analyserons pas le produit de l'activité lexicographique — i.e. les fiches lexicographiques produites —, car ce sont les processus en jeu lors du travail lexicographique qui nous intéressent. Or, ces derniers sont accessibles par les verbalisations du lexicographe, qui décrivent les démarches qu'il entreprend et dans lesquelles les connaissances qu'il mobilise sont en partie explicitées. Nous pouvons aussi extraire des connaissances à partir d'actions qui n'ont pas été verbalisées, mais qui ont laissé des traces à l'écrit. Nous expliquerons dans ce chapitre comment nous avons extrait ces savoirs et comment nous les avons encodés, c'est-à-dire comment nous les avons enregistrés à l'intérieur d'un modèle formel ontologique. Les explications qui seront données ici ont en partie déjà fait l'objet d'articles publiés, *Sharing the Knowledge of Lexicographers : Methodology for the Extraction of Lexicographic Abilities* (Comeau, 2009) et *Extraction et modélisation des connaissances lexicographiques en vue de leur transfert vers l'enseignement du vocabulaire* (Comeau et Polguère, 2008).

Dans ce dernier texte, le choix de la méthode d'analyse des verbalisations avait été expliqué en ces termes :

Ericsson et Simon (1984),(1993) décrivent plusieurs méthodes d'analyse de verbalisations effectuées selon le Think Aloud Protocol. Notre approche pour l'analyse des données entre dans la catégorie qu'ils appellent meaningful analysis of verbalizations. Dans ce type d'approche, il n'y a pas

d'accord préalable entre le sujet et l'expérimentateur sur les énoncés à utiliser en tant que verbalisations. Le participant utilise la parole libre plutôt que des formules figées, ce qui fait que le nombre de verbalisations possibles est illimité. L'expérimentateur doit, par la suite, analyser les verbalisations du participant et tenter d'en extraire le sens. Une des méthodes disponibles pour mener une telle analyse est de se baser sur une théorie, c'est-à-dire de choisir une « grille » de notions appartenant à une théorie et, lors de l'encodage, de faire correspondre à chaque verbalisation un des éléments de cette grille de notions. Nous n'avons pas utilisé cette méthode, parce qu'elle a le défaut de limiter l'encodage à certains éléments décrits par la théorie sur laquelle on s'appuie et, surtout, parce qu'il n'existe pas, à notre connaissance, de modèle structuré des savoirs impliqués en lexicographie. Comme nous l'avons mentionné en introduction, notre projet Lexitation vise justement la construction d'un modèle ontologique des savoirs lexicographiques. (Comeau et Polguère, 2008)

Bien sûr, il existe des théories lexicologiques et des approches établies en lexicographie. Les travaux de Zgusta (1971), de Béjoint (2000), de Landau (2001), et Mel'čuk et al. (1995), qui sont des références dans le domaine de la lexicographie, nous fournissent un cadre théorique pour ce qui est des notions statiques ou connaissances déclaratives mobilisées par les lexicographes. D'ailleurs, une ontologie des connaissances métalexicales (i.e. connaissances déclaratives sur le lexique), appelée le *Gros Tas de Notions* (GTN), dans laquelle des notions lexicologiques et lexicographiques sont modélisées, a été annexée à notre ontologie, et en est donc un module (voir Section 4.2). Néanmoins, il n'existe pas à l'heure actuelle de modèle des connaissances procédurales impliquées en lexicographie. Par conséquent, nous ne pouvions pas analyser les données à l'aide d'une liste de classes préétablies de connaissances lexicographiques déclaratives *et* procédurales et associer chaque verbalisation à une classe existante. Pour la première analyse d'expérimentation, nous avons plutôt procédé de façon inductive : nous avons observé les données brutes et tenté d'identifier quelles connaissances déclaratives et procédurales étaient utilisées par le lexicographe pour accomplir son activité, après

quoi nous avons construit des classes abstraites de connaissances procédurales correspondantes. Cette procédure est parfaitement légitime, si l'on se fie à Ericsson et Simon (1984, 6) :

In less formal kinds of analysis, the encoding scheme is not defined formally and a priori, but the search for interpretations proceeds in parallel with the search for an appropriate model or theory. We recognize clearly the need for and value of such interactive processes in the search for theories in new domains.

La lexicographie n'est pas ce qu'on peut appeler *a new domain* ; comme nous l'avons dit plus haut, les ouvrages de théorie lexicographique sont nombreux. Cependant, une théorie générale de la lexicographie, incluant les connaissances lexicographiques procédurales, reste à construire. Pour cette raison, l'analyse des données et la construction du modèle ont été effectuées en parallèle dans notre projet.

4.1.2 Notions de base

Avant d'aller plus loin dans la description de notre méthodologie, nous allons présenter trois notions de base sur lesquelles l'analyse des données repose. Ces trois notions ont par ailleurs été présentées dans (Comeau, 2009) :

- **Activité lexicographique** : Travail assigné au lexicographe pendant une expérimentation donnée. Il n'y a qu'une seule activité par expérimentation (voir Section 3.3.1).
- **Tâche lexicographique** : Segment de l'activité lexicographique associé à une période de temps dans lequel le lexicographe accomplit un travail clairement identifiable. Seule l'activité est prédéterminée ; les tâches sont improvisées par le lexicographe, c'est-à-dire qu'on ne lui dit pas de faire telle ou telle tâche. Ce qu'il faut remarquer ici est que la segmentation de l'activité en tâches est linéaire ; on attribue un numéro d'ordre à chaque tâche et on identifie son temps de début et de fin.

- **Opération lexicographique** : Travail lexicographique minimal effectué par le lexicographe à l'intérieur d'une tâche donnée, qui n'est pas associé à une séquence temporelle.

En raison de la complexité de la notion d'opération lexicographique, il sera nécessaire d'en donner une explication plus détaillée, accompagnée d'exemples (Section 4.1.2.2). Toutefois, il faut tout d'abord exposer l'architecture de base de l'ontologie, puisque des notions ontologiques seront nécessaires à la description de l'analyse des données.

4.1.2.1 Notions de base liées à l'ontologie

Comme nous l'avons mentionné plus tôt, l'analyse des données et la construction du modèle ont été réalisées en parallèle, l'une s'alimentant de l'autre. La forme des données et la structure de l'ontologie étant intrinsèquement liées, il est impossible de les traiter en vase clos, une à la suite de l'autre.

Voici donc quelques notions ontologiques qui seront nécessaires à l'exposé qui suivra. Disons d'abord ce qu'on entend par ontologie :

An ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology together with a set of individual instances of classes constitutes a knowledge base. (Noy et McGuinness, 2001, 3)

En fait, ce que nous appelons notre *ontologie* contient aussi des instances de classes, donc cela correspond à ce que Noy et McGuinness appellent une *base de connaissances*. Dans notre cas, le domaine du discours est, bien sûr, la lexicographie, et les classes (ou concepts) correspondent aux tâches et aux opérations lexicographiques (la Figure 4.1 présente les classes d'opérations). Les segments de l'enregistrement que nous identifions en tant que tâche ou opération sont des instances de ces classes.

L'éditeur d'ontologies que nous utilisons, Protégé, permet d'ouvrir une fenêtre pour chaque instance de tâche ou d'opération, dans laquelle plusieurs champs présentent les

attributs de ces instances. Le terme *attribut* réfère à l'information qu'on trouve dans les champs, et le terme *champ* (en anglais, *slot*) correspond à la case elle-même dans laquelle les attributs d'une instance sont décrits (la Figure 4.4 présente une instance de tâche).

Par exemple, la conceptualisation du sens d'une lexie est une classe de tâches dans notre ontologie. La première tâche de l'Expérimentation 3 en est une instance, parce que c'est ce type de tâche que le lexicographe accomplit à ce moment. Comme attributs de cette tâche, on peut nommer entre autres l'heure de début et l'heure de fin de cette instance particulière. Dans l'Expérimentation 3, il y a plusieurs instances de cette classe de tâches, c'est-à-dire qu'il arrive au lexicographe de conceptualiser le sens d'une lexie à plusieurs reprises au cours de l'activité lexicographique.

Remarquons que nous avons donné des noms anglais aux classes de notre ontologie. Chacune des classes sera prochainement reliée au terme français correspondant. En effet, dans l'ontologie, les notions (i.e. les classes de tâches et d'opérations) sont identifiées par un *concept* (nommé en anglais, en petites majuscules), qui lui-même peut être relié à un ou plusieurs *termes*. À un même concept peut donc correspondre un terme français et un terme anglais. Il sera donc possible de consulter l'interface française de l'ontologie quand nous auront relié les concepts aux termes français correspondant.

4.1.2.2 Nature et propriétés des opérations

Les notions ontologiques de base étant introduites, nous pouvons maintenant décrire plus précisément la nature des opérations lexicographiques et leurs propriétés.

Lors de l'analyse des données d'une expérimentation, les opérations lexicographiques effectuées par le lexicographe doivent parfois être inférées, puisque, même si les lexicographes ont en général bien suivi le *Think Aloud Protocol*, ils ont parfois omis de verbaliser certaines de leurs actions. Il peut donc arriver que nous inférons que le lexicographe exécute une opération, même s'il ne l'a pas mentionnée. Nous avons essayé d'inférer le moins possible, mais suffisamment pour obtenir une séquence d'opérations logique et complète. La plupart du temps, nous nous fions à un indice concret qui nous indique que le lexicographe a bien exécuté l'opération. Par exemple, dans la tâche 4 de

l'Expérimentation 3, le lexicographe écrit la partie du discours de la lexie, un verbe, tout en parlant d'autre chose. Même s'il n'a pas dit : *J'identifie la partie du discours de la lexie*, nous pouvons déduire qu'il l'a fait et ajouter une instance de la classe d'opérations POS_SPECIF (*part of speech specification*), dans la liste des opérations de la tâche en cours. Il est souvent impossible d'identifier le moment exact où les opérations inférées, donc « non tangibles », ont lieu. De plus, les opérations sont de très courte durée et peuvent souvent se chevaucher. Par exemple, dans cette même tâche 4, lorsque le lexicographe écrit la partie du discours de la lexie, il énonce à haute voix l'hypothèse qu'il s'agit bien d'une acception distincte d'une autre. Cette verbalisation constitue une opération qui consiste à évaluer si les sens de deux lexies sont équivalents ou non, instance de la classe SEM_EQUI_ASSESS. Les deux opérations sont donc simultanées. Pour cette raison, nous avons décidé de ne pas identifier la séquence temporelle de chaque opération, comme nous l'avions fait pour chaque tâche.

Toutefois, nous pouvons ordonner les opérations à l'intérieur des instances de tâches, « ordinalement » et logiquement. En effet, dans l'ontologie, dans la fenêtre de chaque instance de tâche, les instances d'opérations sont numérotées selon l'ordre logique dans lequel elles se sont vraisemblablement produites. Dans cette même fenêtre, nous indiquons par le champ *presupposed operations* (voir Section 4.2.) les opérations préalables à celle-ci.

Les opérations préalables à une autre opération sont celles qui l'ont précédée et sans lesquelles celle-ci n'aurait pas pu être réalisée. Toutes les opérations qui en précèdent une n'entrent donc pas dans cette catégorie. Par ailleurs, il est possible que deux opérations soient préalables à une autre, mais que ces deux opérations soient indépendantes l'une de l'autre. En outre, la relation de « prélabilité » est transitive : si l'opération 1 est préalable à l'opération 2, et que cette dernière est préalable à l'opération 3, alors 1 est préalable à 3 automatiquement. Pour ne pas alourdir la présentation des données inutilement, lors de l'encodage dans l'ontologie, nous n'indiquerons que l'opération 2 en tant qu'opération préalable à 3.

Les notions de *tâche* et d'*opération* introduites ci-dessus seront illustrées dans les sections suivantes. Il faut rappeler que l'analyse présentée ici comprend les données

d'une seule expérimentation (Expérimentation 3).

La Section 4.3 de ce chapitre portera sur l'analyse des données de la cinquième expérimentation. Nous avons décidé de présenter l'analyse de l'Expérimentation 5 séparément pour bien identifier les apports à l'ontologie qu'elle a permis, et pour montrer comment l'ontologie peut être utilisée pour l'encodage de nouvelles expérimentations. En effet, de multiples types de tâches et d'opérations différents se sont ajoutés à l'analyse de l'Expérimentation 5, mais beaucoup de classes ont été réutilisées. Notre ontologie est construite de façon à pouvoir être modifiée et enrichie facilement.

4.1.3 Division de l'activité en tâches

Dans chaque activité lexicographique, le lexicographe effectue des tâches particulières. Il est évident que le type de tâches effectuées dépend de la finalité de l'activité lexicographique en cours. Dans le cas de l'Expérimentation 3, l'activité consistait à dégager la structure polysémique du vocable *COMPTER*, c'est-à-dire de trouver et de décrire sommairement les différents sens de ce vocable. Notre travail consistait donc à observer les données (l'enregistrement de l'Expérimentation 3) et à identifier les différentes tâches que le lexicographe accomplit pendant l'activité.

Après avoir visionné plusieurs fois l'enregistrement de l'Expérimentation 3, nous avons premièrement divisé la séquence audio-vidéo en 40 grandes scènes, correspondant aux tâches effectuées par le lexicographe. Les tâches se distinguent par l'objectif que le lexicographe cherche à atteindre en les faisant. Une tâche se termine et une autre débute quand le lexicographe entreprend des opérations en ayant un objectif différent de celui qu'il poursuivait jusqu'alors. L'Expérimentation 3 ayant duré 55 minutes, la durée moyenne des tâches est de près d'une minute et demie. Nous avons numéroté ces tâches et identifié le moment de leur début et de leur fin. En tout, nous avons dégagé trois types de tâches. Dans notre ontologie, nous avons créé trois classes abstraites de tâches correspondantes, et chacune des 40 tâches de l'Expérimentation 3 est considérée comme une instance d'une de ces classes.

Voici les trois classes de tâches en question :

- **SEARCH_FOR_SENSE** : Le lexicographe cherche des acceptions auxquelles il n'a pas pensé, en utilisant le corpus ou par introspection. Il y a 18 instances de cette classe dans l'Expérimentation 3.
- **CONCEPTUALIZATION_OF_SENSE** : Le lexicographe conceptualise une acception et procède à une description lexicographique sommaire en remplissant une fiche lexicographique. Il y a 15 instances de cette classe dans l'Expérimentation 3.
- **MODIFICATION_OF_DESCRIPTION** : Le lexicographe effectue des modifications mineures dans les fiches lexicographiques qu'il a déjà rédigées. Il y a 7 instances de cette classe dans l'Expérimentation 3.

Cette dernière classe de tâches inclut toutes les modifications apportées aux fiches qui ne découlent pas de la conceptualisation même des acceptions. Parfois, la conceptualisation ou la reconceptualisation d'une acception peut amener le lexicographe à modifier une ou plusieurs fiches, mais nous avons choisi de n'inclure dans la catégorie **MODIFICATION_OF_DESCRIPTION** que les modifications qui n'impliquent pas de « remaniement » de la structure du vocable telle qu'élaborée par le lexicographe. En général, une modification de fiche vise à améliorer la description lexicographique, à la rendre plus complète ou conforme aux règles lexicographiques. Par exemple, dans la tâche 18 de l'Expérimentation 3, le lexicographe ajoute un exemple tiré du corpus dans sa fiche *COMPTER* -'dénombrer', afin que celle-ci comporte un exemple « réel », senti comme plus légitime que l'exemple de sa propre composition.

4.1.4 Division des tâches en opérations

À ce stade de l'analyse, nous avons obtenu des classes abstraites de tâches lexicographiques et des instances de ces classes, tirées de l'Expérimentation 3. L'étape suivante consistait à diviser les instances de tâches en actions plus simples, les opérations, et à créer des classes abstraites d'opérations, comme nous l'avons fait pour les tâches. Ces deux types de classes (tâches et opérations) ne sont pas subordonnés dans la hiérarchie (voir la Figure 4.1), mais y apparaissent de façon parallèle. En effet, les différentes classes d'opérations ne sont pas des sous-classes de tâches ; il s'agit

de segments d'activités de nature différente (voir Section 4.1.2). De plus, les classes d'opérations elles-mêmes sont indépendantes les unes des autres. Par exemple, dans la Figure 4.1 ci-dessous, on observe que les classes TESTING et LU_IDENTIFICATION sont dans une relation soeur-soeur plutôt que mère-fille (les définitions de ces classes sont dans la Section 4.1.5). Par contre, certaines classes peuvent avoir des sous-classes : BASIC_LU_IDENTIFICATION est une sous-classe (fille) de la classe LU_IDENTIFICATION. Notons que le même type d'opérations peut être utilisé dans différents types de tâches.

La Figure 4.1 présente les classes d'opérations identifiées dans l'Expérimentation 3 et leur nombre d'occurrences. Notons que la hiérarchie ne représente pas l'ordre dans lequel sont effectuées les opérations : dans chaque tâche, une ou plusieurs opérations sont effectuées, mais pas nécessairement dans l'ordre où elles apparaissent ici.

Notons les abréviations suivantes utilisées dans les noms d'opérations :

LU	=	<i>lexical unit</i> (unité lexicale) ;
QSYN	=	<i>quasi-synonym</i> (quasi-synonyme) ;
ASSESS	=	<i>assessment</i> (évaluation) ;
POS	=	<i>part of speech</i> (partie du discours) ;
EX	=	<i>example</i> (exemple) ;
IDENT	=	<i>identification</i> (identification) ;
SPECIF	=	<i>specification</i> (spécification).

4.1.5 Présentation des classes d'opérations identifiées pour l'Expérimentation 3

Après avoir divisé l'activité en 40 tâches, nous avons divisé chaque tâche en opérations. En moyenne, chaque tâche est constituée de 4 à 5 opérations. Les opérations peuvent être de différente nature. Nous présenterons ici toutes les classes d'opérations que nous avons identifiées dans l'Expérimentation 3, mais nous fournirons d'abord une explication sur la notion de *classe abstraite* de l'ontologie.

Nous avons dit que, dans une ontologie, il y a des classes de concepts et des instances de ces classes. En ce qui nous concerne, les classes sont des types de tâches ou d'opérations lexicographiques, telles que présentées dans la Figure 4.1, et les instances sont des segments de l'expérimentation correspondant à ces classes. Les classes de l'ontologie

- ▼ ● LEXICOGRAPHIC_OPERATION
 - ▼ ● LU_IDENTIFICATION
 - BASIC_LU_IDENTIFICATION
 - ▼ ● TESTING
 - ▼ ● TESTING_OF_COMBINATORICS
 - TESTING_OF_SYNT_COMBINATORICS
 - TESTING_OF_LEX_COMBINATORICS
 - TESTING_OF_GRAMM_COMBINATORICS
 - ▼ ● TESTING_OF_SEMANTIC_EQUIVALENCE
 - REPLACEMENT_BY_QSYN_IN_EX
 - PARAPHRASING
 - CORPUS_ANALYSIS
 - ▼ ● CHARACTERIZATION_OF_LU
 - ▼ ● SEMANTIC_CHARACTERIZATION_OF_LU
 - ▼ ● CLOSE_SENSE_IDENT
 - QSYN_IDENT
 - CLOSE_COPOLYSEME_IDENT
 - GENUS_IDENT
 - ▼ ● GRAMMATICAL_CHARACTERIZATION_OF_LU
 - POS_SPECIF
 - GOVERNMENT_PATTERN_SPECIF
 - RADICAL_IDENT
 - CANONIC_FORM_SPECIF
 - ▼ ● CONTEXTUALIZATION_OF_LU
 - EXAMPLE_CREATION
 - ▼ ● EXAMPLE_RETRIEVAL
 - EX_RETRIEVAL_FROM_MEMORY
 - EX_RETRIEVAL_FROM_CORPUS
 - EX_RETRIEVAL_FROM_DESCRIPTION
 - EXAMPLE_TRANSFER
 - EXAMPLE_MODIFICATION
 - ▼ ● REVIEW_OF_DESCRIPTION
 - REVIEW_OF_LU_DESCRIPTION
 - REVIEW_OF_POLYSEMIC_STRUCTURE
 - ▼ ● ASSESSMENT
 - ASSESSMENT_OF_FREQUENCY_OF_LU
 - ▼ ● SENSE_EQUALITY_ASSESSMENT
 - SENSE_EQUAL_ASSESS_THROUGH_COMBINATORICS
 - ▼ ● DESCRIPTION_ASSESSMENT
 - EXAMPLE_ASSESSMENT
 - PARAPHRASE_ASSESSMENT
 - QSYN_ASSESSMENT

Figure 4.1 – Hiérarchie des opérations lexicographiques

peuvent être soit abstraites, soit concrètes. Les explications qui suivent sont inspirées de celles de la thèse d’Ophélie Tremblay (Tremblay, 2009, Chapitre 3). Les classes abstraites fonctionnent comme des pointeurs de classes et servent plutôt à classer. Elles sont représentées par le symbole du cercle avec un point au centre, qui précède les classes dans la hiérarchie (voir la Figure 4.1). Les classes concrètes, elles, sont définies par ces classements, et, dans la hiérarchie, elles sont précédées d’un cercle plein. À ce type de concepts, on peut associer des instances directes, ce qui n’est pas le cas des classes abstraites.

Voici donc les classes d’opérations observées dans l’Expérimentation 3 et leur nombre d’occurrences, entre parenthèses. Rappelons que nous avons choisi de présenter les résultats de l’analyse de l’Expérimentation 5 séparément (dans la Section 4.3), pour bien mettre en évidence les ajouts que celle-ci a permis d’apporter à l’ontologie.

- LU_IDENTIFICATION (11 instances)

Le lexicographe identifie une lexie appartenant au vocable à décrire. Dans le cas de l’Expérimentation 3, le vocable à décrire était *COMPTER*, et parmi les lexies identifiées par le lexicographe, il y a *COMPTER* - ‘dénombrer’, *COMPTER* - ‘espérer’, *COMPTER* - ‘être important’, etc. Le lexicographe identifie les lexies soit en réfléchissant aux sens du verbe *COMPTER* qu’il connaît, soit en les identifiant dans des phrases qu’il examine dans le corpus.

- BASIC_LU_IDENTIFICATION (1 instance)

Le lexicographe identifie une acception du vocable comme étant l’acception de base de ce vocable. Dans le cas de *COMPTER*, l’acception de base, selon le lexicographe, est *COMPTER* - ‘dénombrer’.

- TESTING

Il s’agit d’une classe abstraite, comprenant les sous-classes suivantes :

- TESTING_OF_COMBINATORICS

Il s’agit d’une classe abstraite d’opérations consistant en des tests sur la combinatoire des lexies. Notons que ce sont les signes linguistiques qui ont des

propriétés de combinatoire (*contraintes qui limitent leur capacité de se combiner avec d'autres signes linguistiques, et qui ne peuvent être déduites ni de leur signifié ni de leur signifiant* (Polguère, 2008, 40)) à proprement parler. Les lexies ne sont pas des signes, mais des ensembles, des généralisations de signes. Par exemple, la lexie LAMPE est en fait un ensemble contenant deux signes différents, qui ne se distinguent que par la flexion, soit *lampe* et *lampes*. C'est par métonymie qu'on parle de la combinatoire des lexies.

* TESTING_OF_SYNT_COMBINATORICS (9 instances)

Opération qui consiste à tester des régimes potentiels pour la lexie à décrire (un verbe, dans le cas qui nous concerne). Dans l'Expérimentation 3, le but que le lexicographe poursuit en effectuant cette opération est de faire apparaître de nouvelles acceptions. Cela dit, ce test pourrait être effectué pour d'autres raisons.

* TESTING_OF_LEX_COMBINATORICS (1 instance)

Il s'agit d'opérations consistant à tester la combinatoire lexicale des lexies. Par exemple, dans l'opération 003-028-005, le lexicographe se demande si l'acception qu'il a en tête, COMPTER dans le sens d' 'espérer', ne s'emploie qu'avec l'adverbe BIEN, comme dans l'exemple *Il compte bien faire quelque chose*.

* TESTING_OF_GRAMM_COMBINATORICS (3 instances)

Il s'agit d'opérations consistant à tester la combinatoire grammaticale des lexies. Par exemple, le lexicographe teste la possibilité de mettre le verbe COMPTER à la forme pronominale (003-011-002) et à la voix passive (003-011-004).

– TESTING_OF_SEMANTIC_EQUIVALENCE

C'est une classe abstraite. Le lexicographe peut vouloir tester l'équivalence entre le sens d'une lexie et d'un quasi-synonyme, entre deux occurrences de la même lexie, ou entre deux copolysèmes (voir Section 3.3.1 pour la définition de *copolysème*). En général, si ces tests se révèlent négatifs, on est

en présence de deux lexies différentes. Cette considération peut amener le lexicographe à identifier une nouvelle acception et à créer une fiche lexicographique pour la décrire. À ce jour, nous avons pu observer des tests d'équivalence de deux types, correspondant aux sous-classes suivantes :

* PARAPHRASING (11 instances)

Il s'agit de paraphraser le sens d'une lexie ou d'une phrase dans laquelle se trouve une lexie pour tester l'équivalence entre cette lexie et une autre.

* REPLACEMENT_OF_QSYN_IN_EX (4 instances)

Cette classe d'opérations relève aussi du paraphrasage. Il s'agit de tests pour déterminer s'il y a équivalence sémantique entre deux lexies, consistant à énoncer une phrase comprenant la lexie à comparer, puis à remplacer cette lexie dans la phrase par un quasi-synonyme. Si la phrase obtenue a le même sens que la première, c'est qu'il y a équivalence sémantique entre les deux lexies.

● CORPUS_ANALYSIS (15 instances)

C'est une « métaopération » pour laquelle il faut spécifier les classes de *sous-opérations* impliquées. En effet, quand le lexicographe consulte le corpus, il parcourt les occurrences et applique une série de sous-opérations dont la composition est plus ou moins variable, selon l'objectif qu'il poursuit. Il peut, par exemple, pour chaque occurrence du corpus :

1. déterminer s'il s'agit bien du verbe **COMPTER** et non du nom **COMPTE** (sous-opération : POS_SPECIF, décrite plus bas) ;
2. déterminer s'il s'agit d'une occurrence d'une acception déjà recensée. Pour ce faire, le lexicographe peut effectuer des tests d'équivalence sémantique (Sous-opération : une des classes de TESTING_OF_SEMANTIC_EQUIVALENCE) ;
3. évaluer le résultat de ce test pour déterminer si oui ou non il s'agit d'une occurrence de cette acception (Sous-opération : SENSE_EQUALITY_ASSESSMENT).

Spécifions que ces sous-opérations sont des instances d'opérations qui peuvent être effectuées dans d'autres tâches. Dans ce cas, elles n'ont pas le statut de « sous-opération » mais sont des opérations à part entière. Le fait d'avoir une métaopération nous permet de faire l'économie de chacune des sous-opérations dans la liste des opérations de la tâche en question. De plus, généralement, une telle série de sous-opérations est appliquée récursivement sur le corpus. Le lexicographe peut appliquer en boucle des dizaines de fois la même série de sous-opérations, jusqu'à ce qu'il trouve quelque chose d'intéressant (une nouvelle acception à décrire, un exemple à ajouter dans une fiche, etc.). Quand il y a récursivité, nous pouvons l'indiquer en cochant la case *recursivity*, qui se trouve dans les instances de la métaopération CORPUS_ANALYSIS, et dont les valeurs possibles sont *yes* ou *no*.

- CHARACTERIZATION_OF_LU

Il s'agit d'une classe abstraite qui inclut plusieurs opérations visant la caractérisation des lexies :

- SEMANTIC_CHARACTERIZATION_OF_LU

C'est une classe abstraite qui comprend des opérations diverses de caractérisation sémantique de lexies :

- * CLOSE_SENSE_IDENT

Classe abstraite qui permet d'identifier des sens proches, de l'un des deux types suivants :

- QSYN_IDENT (18 instances)

Il s'agit d'identifier le quasi-synonyme d'une lexie.

- CLOSE_COPOLYSEME_IDENT (14 instances)

Deux lexies peuvent être des *copolysèmes* si elles sont des acceptions appartenant à un même vocable. Ici, il s'agit donc de déterminer à quelle acception appartient un usage.

- * GENUS_IDENT (1 instance)

Opération qui correspond à l'identification du genre prochain d'une lexie (voir Section 1.3 pour une définition de *genre prochain*).

– GRAMMATICAL_CHARACTERIZATION_OF_LU

Classe abstraite d'opérations qui consistent à caractériser une lexie en identifiant ses propriétés grammaticales :

* POS_SPECIF (12 instances)

Il s'agit d'identifier la partie du discours de la lexie (verbe, nom, adjectif, etc.)

* GOVERNMENT_PATTERN_SPECIF (4 instances)

Il s'agit d'identifier le régime d'une lexie, c'est-à-dire *la composante de sa combinatoire restreinte qui regroupe l'ensemble des contraintes que cette lexie impose sur l'expression syntaxique de ses actants* (Polguère, 2008, 138).

* RADICAL_IDENT (1 instance)

Il s'agit d'identifier le radical d'une lexie. Cette opération peut être effectuée lors de la recherche dans un corpus, par exemple, le lexicographe peut entrer le radical de la lexie à décrire dans le moteur de recherche du corpus, afin de trouver toutes les formes fléchies de la lexie.

* CANONICAL_FORM_SPECIF (1 instance)

Il s'agit d'identifier la forme canonique de la lexie à décrire (la forme infinitive, dans le cas d'un verbe). C'est aussi une opération qui peut être effectuée lors de recherches dans un corpus.

– CONTEXTUALIZATION_OF_LU

C'est une classe abstraite. La mise en contexte d'une lexie peut servir différents objectifs :

* EXAMPLE_CREATION (9 instances)

Le lexicographe crée un exemple contenant la lexie à décrire pour bien identifier le sens qu'il veut illustrer ou pour compléter la fiche lexicographique.

* EXAMPLE_RETRIEVAL

Il s'agit d'une classe abstraite comprenant des opérations dans lesquelles le lexicographe rappelle des exemples de la lexie qu'il veut décrire, pour alimenter la conceptualisation. Le rappel peut venir de diverses sources :

- EX_RETRIEVAL_FROM_MEMORY (2 instances)

Dans ce cas, l'exemple vient de la mémoire du lexicographe. Il peut s'agir par exemple d'une phrase qu'il a entendu dans les jours précédant l'expérimentation.

- EX_RETRIEVAL_FROM_CORPUS (15 instances)

Il s'agit de rappeler une phrase du corpus, pour la comparer à une autre, par exemple.

- EX_RETRIEVAL_FROM_DESCRIPTION (1 instance)

C'est une opération dans laquelle le lexicographe rappelle une phrase qu'il a lui-même écrite dans une autre fiche.

* EXAMPLE_TRANSFER (1 instance)

C'est le transfert d'un exemple du corpus vers une fiche lexicographique, ou d'une fiche à une autre.

* EXAMPLE_MODIFICATION (2 instances)

Opération dans laquelle le lexicographe modifie un exemple tiré du corpus pour l'ajouter dans une fiche.

● REVIEW_OF_DESCRIPTION

Il s'agit d'une classe abstraite d'opérations consistant à réviser différents aspects du travail lexicographique effectué.

- REVIEW_OF_LU_DESCRIPTION (9 instances)

Opération qui consiste à revoir une fiche lexicographique, par exemple pour la comparer à une nouvelle acception.

- REVIEW_OF_POLYSEMIC_STRUCTURE (5 instances)

Opération qui consiste à récapituler les acceptions déjà listées, pour en faire surgir d'autres ou pour les comparer entre elles, par exemple.

- ASSESSMENT

Classe abstraite regroupant des évaluations ou des jugements que le lexicographe peut porter sur différents aspects du travail lexicographique qu'il accomplit ou sur des caractéristiques des lexies elles-mêmes.

- ASSESSMENT_OF_FREQUENCY_OF_LU (2 instances)

Évaluation de la fréquence d'usage d'une lexie. Cela peut servir de critère pour déterminer si un usage particulier doit avoir une entrée dans le dictionnaire ou non.

- SENSE_EQUALITY_ASSESSMENT (13 instances)

Il s'agit de formuler un jugement d'équivalence sémantique, c'est-à-dire de déterminer si deux occurrences correspondent au même sens ou non.

- * SENS_EQUAL_ASSESS_THROUGH_COMBINATORICS (5 instances)

Cette opération s'applique quand le jugement d'équivalence sémantique se base sur la combinatoire de la lexie. Par exemple, le lexicographe peut déterminer que deux occurrences relèvent de deux lexies différentes puisqu'elles ne mettent pas en jeu le même régime.

- DESCRIPTION_ASSESSMENT (1 instance) Dans ce cas, l'opération consiste à faire une évaluation de l'ensemble de la description lexicographique.

- * EXAMPLE_ASSESSMENT (11 instances) Dans ce cas, le jugement porte sur une phrase que le lexicographe a mise en exemple ou qu'il veut mettre en exemple dans une fiche lexicographique.

- * PARAPHRASE_ASSESSMENT (1 instance) Jugement ou évaluation de la justesse d'une paraphrase du sens d'une lexie ou d'une phrase.

- * QSYN_ASSESSMENT (5 instances) Jugement ou évaluation de la justesse d'un quasi-synonyme attribué à une acception.

La Figure 4.1 récapitule toutes les classes et sous-classes d'opérations que nous avons observées dans une tâche ou plus de l'Expérimentation 3. La plupart des classes ont plusieurs instances, puisque durant l'activité lexicographique, le lexicographe a effectué 187 opérations, appartenant à 28 classes différentes.

Afin de rendre la méthodologie d'extraction plus claire, nous allons présenter dans la prochaine section la transcription d'une tâche de l'Expérimentation 3 et son analyse.

4.1.6 Transcription d'une tâche et son analyse

Voici la transcription de la tâche 8 de l'Expérimentation 3. L'italique est utilisé pour citer le lexicographe et les termes en petites majuscules sont les noms des classes d'opérations correspondantes que nous avons identifiées.

1. *Ça me fait penser, hier, je jouais avec ma nièce... elle a dit : « Ça, ça compte pas ».*

→ EX_RETRIEVAL_FROM_MEMORY

2. *Donc... C'est pas conforme au règlement.*

→ QSYN_IDENTIFICATION

3. *Ou... Il faut pas en tenir compte... Il faut pas prendre en considération. Ça ressemble un petit peu, mais c'est pas « compter avec », là.*

→ LU_IDENTIFICATION

Premièrement, dans cette opération, le lexicographe compare le sens de l'exemple dont il s'est souvenu, *Ça compte pas*, avec le sens d'une acception qu'il a déjà décrite et qu'il a nommée *COMPTER AVEC*- 'prendre en considération, tenir compte de'. Ensuite, il affirme qu'il s'agit d'une acception différente de *COMPTER AVEC* ; il crée donc une nouvelle fiche pour cette unité lexicale.

4. Il identifie et écrit la partie du discours de l'unité lexicale.

→ POS_SPECIF

5. Il écrit une expression quasi-synonymique, ‘être officiel’.

→ QSYN_IDENTIFICATION

6. Il écrit un autre quasi-synonyme, ‘[être] réglementaire’.

→ QSYN_IDENTIFICATION

7. Il identifie un autre quasi-synonyme, ‘[être] homologué’, mais ne l’écrit pas.

→ QSYN_IDENTIFICATION

8. Il invente une phrase, *Cette ronde de pratique¹ ne compte pas*, et l’écrit comme exemple.

→ EXAMPLE_CREATION

9. Il rappelle la phrase entendue la veille, *Ça, ça compte pas*, et l’entre comme exemple.

→ EX_RETRIEVAL_FROM_MEMORY

10. Il porte un jugement sur l’exemple : *C’est peut-être un peu trop relâché*.

→ EXAMPLE_ASSESSMENT

Il regarde sa fiche un moment, et quand la pilote lui demande à quoi il pense, il dit cette phrase, en pointant le quasi-synonyme ‘officiel’ : *Je ne suis pas entièrement satisfait de ma description... Il y aurait peut-être quelque chose de plus juste*.

→ QSYN_ASSESSMENT

Le lexicographe passe ensuite à autre chose, même si le quasi-synonyme ne lui semble pas tout à fait juste. La tâche se termine donc là. Son analyse n’a pas causé de problèmes particuliers, parce qu’il s’agit d’une tâche dans laquelle les opérations sont relativement faciles à identifier, ainsi que leur ordonnancement. Toutefois, certaines tâches ou suites de tâches font surgir des problèmes d’analyse, comme nous le verrons dans la prochaine section.

¹ Au Québec, *pratique* a aussi le sens ‘mise en jambes, échauffement’.

4.1.7 Cas difficiles

Durant l'analyse des données, nous avons effectivement rencontré quelques difficultés, de natures diverses, dont nous discuterons ici.

4.1.7.1 Cas concernant des opérations présupposées

- Premier cas : la recherche d'acceptions

Les tâches 005, 007 et 009 de l'Expérimentation 3 sont des instances de la classe SEARCH_FOR_SENSE. Dans ces trois tâches, le lexicographe cherche à trouver de nouvelles acceptions en combinant le verbe avec différentes prépositions. Ce type d'opérations est appelé TESTING_OF_SYNT_COMBINATORICS (voir Section 4.1.5). La difficulté de l'analyse, dans ce cas, est de déterminer si ces trois opérations sont liées par la relation de préalabilité. Les trois tâches en question sont entrecoupées par des tâches de conceptualisation d'acceptions.

Dans la tâche numéro 005, le lexicographe énonce les syntagmes *compter sur quelque chose*, *compter à* et *compter de*, qui ne lui inspirent pas d'acceptions nouvelles. Ensuite, il dit *compter avec*, qui l'amène à conceptualiser une acception qu'il nomme *COMPTER AVEC*. Cette conceptualisation est elle-même une tâche, la tâche 006, et appartient à la classe CONCEPTUALIZATION_OF_LU. Après avoir conceptualisé et décrit l'acception *COMPTER AVEC*, le lexicographe reprend où il s'était arrêté, en quelque sorte, dans sa « liste » de prépositions et continue à chercher de nouvelles acceptions. La tâche 007 est donc de la classe SEARCH_FOR_SENSE. Il reprend *compter avec*, et ensuite énonce *compter pour*, après quoi il se souvient soudainement d'une phrase entendue la veille : *Ça, ça compte pas*. À partir de cette phrase, il crée l'acception *COMPTER* - 'être officiel'. Commence alors la tâche 008, de la classe CONCEPTUALIZATION_OF_LU. Ensuite, dans la tâche 009, il reprend sa liste de prépositions : *Compter pour... compter pour 2 points...*, ce qui l'amène à créer la lexie qu'il nomme *COMPTER*- 'valoir'.

Dans le cas de ces trois tâches de type SEARCH_FOR_SENSE, dira-t-on que 005

est préalable à 007, et que 007 est préalable à 009 ? Nous croyons que oui, car le lexicographe entreprend ici un travail méthodique ; il passe en revue les prépositions une à une pour les combiner avec le verbe. Sa recherche est entrecoupée de tâches de conceptualisation d'acceptions, mais il y revient et continue sa « revue » où il l'avait arrêtée. D'ailleurs, si on isole ces trois tâches et si on les écoute une à la suite de l'autre, on constate qu'il y a une démarche unique et linéaire qui s'étale sur les trois tâches, la tâche 007 étant la suite logique de la tâche 005, et ainsi de suite.

- Deuxième cas : l'exemple retrouvé

Toujours dans la tâche 007, de la classe `SEARCH_FOR_SENSE`, le lexicographe en était à énoncer le syntagme *compter pour* (003-007-001) lorsqu'il s'est souvenu d'une phrase entendue la veille, *Ça, ça compte pas* (003-008-001). Le fait de se souvenir d'une phrase entendue est une opération en soi, de la classe `EX_RETRIEVAL_FROM_MEMORY`. Nous avons décidé de considérer cette opération comme la première opération de la tâche suivante (003-008), qui appartient à la classe `CONCEPTUALIZATION_OF_SENSE`. Cette décision sera discutée dans la section suivante. Ce qui nous intéresse pour l'instant est de savoir si le fait d'avoir énoncé le syntagme *compter pour* est une opération préalable à l'opération de se souvenir de l'exemple entendu la veille. C'est peut-être la ressemblance phonétique entre *compte pour* et *compte pas* qui a permis au lexicographe de se souvenir de l'exemple, mais cela n'est pas vraiment pertinent du point de vue lexicographique. Le souvenir aurait pu être provoqué par un autre déclencheur, il n'y a donc pas de lien de « prélabilité », selon notre définition, entre l'opération 003-007-001, `TESTING_OF_SYNT_COMBINATORICS`, et 003-008-001, `EX_RETRIEVAL_FROM_MEMORY`.

4.1.7.2 Appartenance d'une opération à une tâche

Il est parfois difficile de déterminer avec certitude à quelle tâche appartient une opération. C'est le cas de l'opération 003-008-001, `EX_RETRIEVAL_FROM_MEMORY`, qui,

selon nous, fait partie de la tâche 003-008, *CONCEPTUALIZATION_OF_SENSE*, et non de la tâche 003-007, *SEARCH_FOR_SENSE*. En effet, le lexicographe cherchait de nouvelles acceptions quand soudainement il s'est souvenu d'une phrase entendue la veille, *Ça, ça compte pas*. Or, dès qu'il a évoqué cet exemple, il a commencé à conceptualiser une acception, *COMPTER* - 'être réglementaire'. L'opération est donc vraiment la première des étapes constituant la conceptualisation de cette acception. Par conséquent, elle doit se trouver dans la liste des opérations de la tâche dont elle est le point de départ, soit 003-008, *CONCEPTUALIZATION_OF_SENSE*.

4.1.7.3 Tâches difficiles à classer

- Cas de la tâche 003-032 : modification de description ou conceptualisation de sens ?

La tâche 003-032 commence par ce qui semble être une conceptualisation de sens, mais finalement le lexicographe prend la décision de mettre l'exemple de l'acception en question dans une fiche déjà existante. En effet, en survolant les occurrences du corpus, il rencontre le syntagme *sans compter*. À prime abord, il considère que cet usage devrait être décrit dans une fiche à part. Ensuite, il se demande s'il s'agit du même sens que *COMPTER* - 'tenir compte de', prendre en considération'. Il consulte ses fiches pour les comparer, puis finalement il ajoute la phrase du corpus comme exemple dans la fiche *COMPTER* - 'tenir compte de', prendre en considération'. Il ajoute comme remarque qu'il s'agit peut-être de deux sens différents : *Éventuellement un sens pour les emplois négatifs*.

Si on regarde cette tâche du point de vue de la finalité, il n'y a pas eu conceptualisation d'une acception à proprement parler, mais plutôt une modification de description (il a ajouté un exemple dans une fiche). Par contre, si on prend le point de vue du lexicographe, de son intention, on ne peut pas dire que son but était ici d'enrichir une fiche, ou de la modifier. Le lexicographe a en fait conceptualisé un sens, qui finalement n'est peut-être pas assez différent d'un autre pour avoir sa propre fiche. Il écrit d'ailleurs la remarque qu'il faudra peut-être les séparer, il per-

çoit donc un certain écart entre les deux sens. Le terme *sens* est employé ici parce qu'il n'est pas encore question d'une « acception » à proprement parler. La limite entre mêmes sens et sens différents est parfois difficile à déterminer ; nous croyons qu'il faut considérer ici le cheminement lui-même et pas seulement la finalité de cette démarche, et conclure que l'opération en jeu ici est du type conceptualisation d'acception plutôt que modification de description.

4.2 Encodage des données dans l'ontologie

Dans cette section, nous présentons l'encodage des tâches et des opérations dans l'ontologie, élaborée avec l'éditeur d'ontologies Protégé. Rappelons que les notions de base liées à l'ontologie ont été présentées dans la Section 4.1.2.1.

L'éditeur d'ontologies Protégé permet d'ouvrir une fenêtre pour chaque instance de tâche ou d'opération, dans laquelle plusieurs champs présentent les attributs de ces instances. Le terme *attribut* réfère à l'information qu'on trouve dans les champs, et le terme *champ* (en anglais, *slot*) correspond à la case elle-même dans laquelle les attributs d'une instance sont décrits.

L'ontologie est construite de manière à ce que l'on trouve de l'information sur les expérimentations et sur les participants, en plus des informations sur les tâches et opérations lexicographiques. Nous avons donc créé une (méta)classe EXP-THING, dont les classes sont EXPERIMENTATION, PARTICIPANT, LEXICOGRAPHIC_TASK et LEXICOGRAPHIC_OPERATION. Chacune de ces classes a des instances, qui elles-mêmes ont des attributs particuliers, que l'on indique dans différents champs. Il est à noter que ces champs se retrouvent dans toutes les instances de la classe à laquelle ils appartiennent.

- La classe EXPERIMENTATION

Chacune des 5 expérimentations que nous avons menées est une instance de la classe EXPERIMENTATION. Chacune d'elle a donc sa propre « fenêtre », dont la Figure 4.2 est un exemple (dans lequel les noms de personnes et d'institutions ont été remplacés par des noms fictifs afin de conserver l'anonymat des participants), et où les informations (ou attributs) suivantes sont données par différents champs :

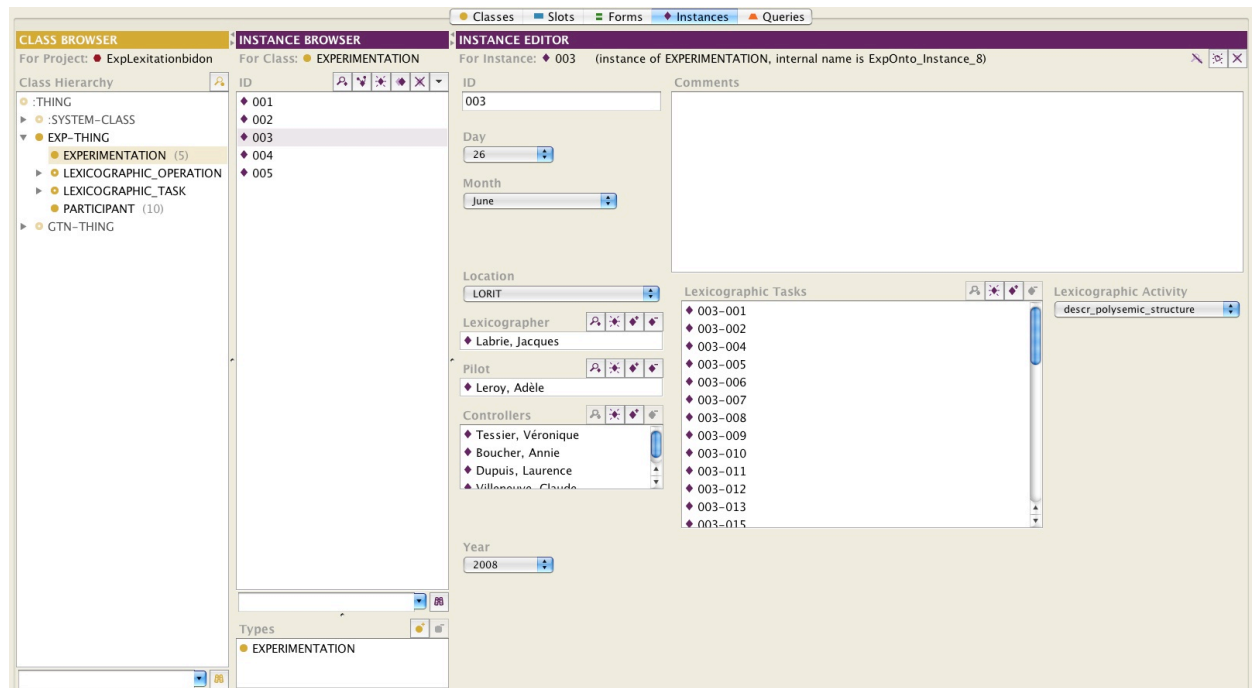


Figure 4.2 – Instance d’expérimentation

– *Id*

Le numéro d’identification de l’expérimentation. Par exemple, le numéro d’identification de la troisième expérimentation est 003.

– *day*

Le jour où a eu lieu l’expérimentation.

– *month*

Le mois où a eu lieu l’expérimentation.

– *location*

L’endroit où a eu lieu l’expérimentation. Jusqu’à maintenant, elles ont toutes eu lieu au LORIT (voir Section 3.3.3), mais nous pourrions éventuellement organiser des expérimentations ailleurs.

– *lexicographer*

Le nom du lexicographe participant à l’expérimentation.

- *pilot*

Le nom du pilote de l'expérimentation.

- *controllers*

Le nom des contrôleurs de l'expérimentation.

- *year*

L'année en cours lors de l'expérimentation.

- *comments*

Tout commentaire jugé pertinent.

- *lexicographic tasks*

Toutes les tâches qui ont été effectuées durant l'expérimentation (par leur code d'identification).

- *lexicographic activity*

L'activité lexicographique faite par le lexicographe durant l'expérimentation.

- La classe PARTICIPANT

La classe PARTICIPANT fournit des informations sur chaque participant des expérimentations. Ici, *participant* réfère à tous les gens impliqués dans les expérimentations, autant les membres de l'équipe de recherche que les lexicographes. Dans les fenêtres des instances de la classe PARTICIPANT, comme celle de la Figure 4.3, on trouve les champs suivants (ici aussi, les données nominatives sont fictives) :

- *participant name*

Le nom du participant.

- *participant background*

Ici, il s'agit d'une information sur le statut de la personne. Cela peut être *professional lexicographer, trained student, academic researcher in lexicography, academic researcher in education*.

- *comment*

Tout commentaire jugé pertinent.

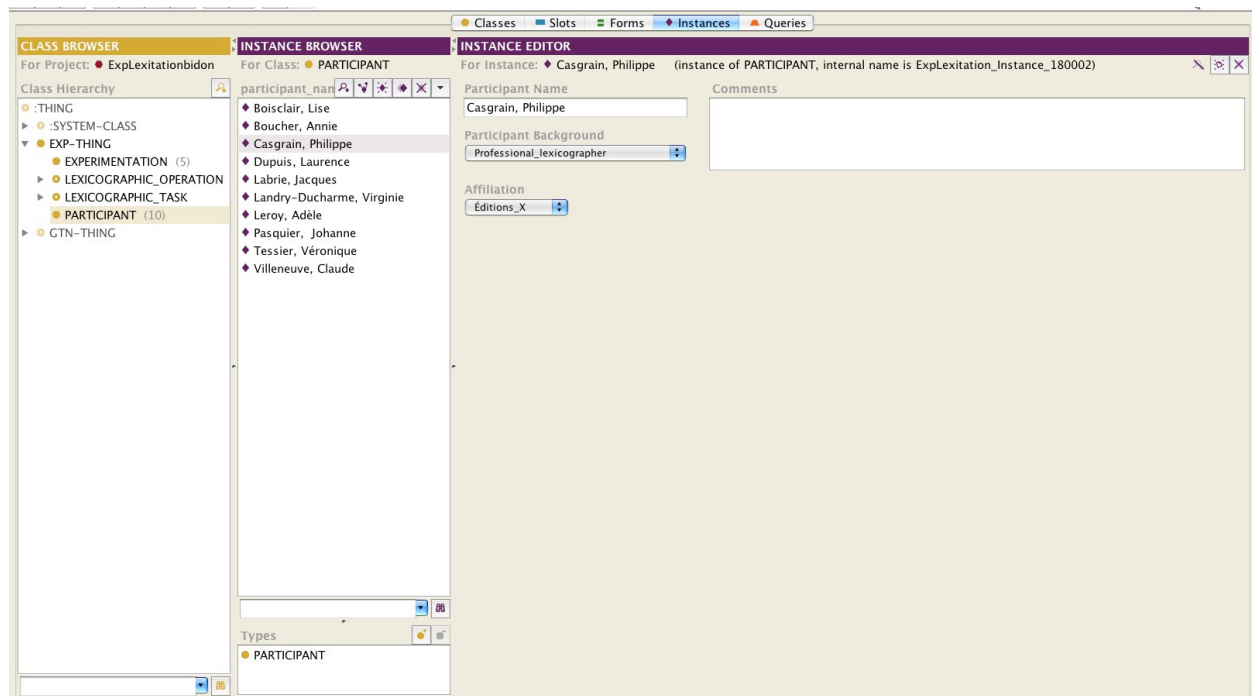


Figure 4.3 – Instance de participant

– *affiliation*

L’affiliation est soit l’université ou le groupe de recherche auquel appartient le participant, dans le cas d’un membre de l’équipe de recherche, ou l’organisme pour lequel il travaille, dans le cas d’un lexicographe.

• La classe LEXICOGRAPHIC_TASK

Voici les champs que nous avons inclus dans les instances de la classe LEXICOGRAPHIC_TASK, dont la Figure 4.4 est un exemple :

– *id*

Ce champ contient le code par lequel une tâche est identifiée. Ce code est divisé en deux parties : la première partie est le numéro de l’expérimentation et le second est le numéro de la tâche. Par exemple, 003-001 est le code qui identifie la première tâche de la troisième expérimentation.

– *in experimentation*

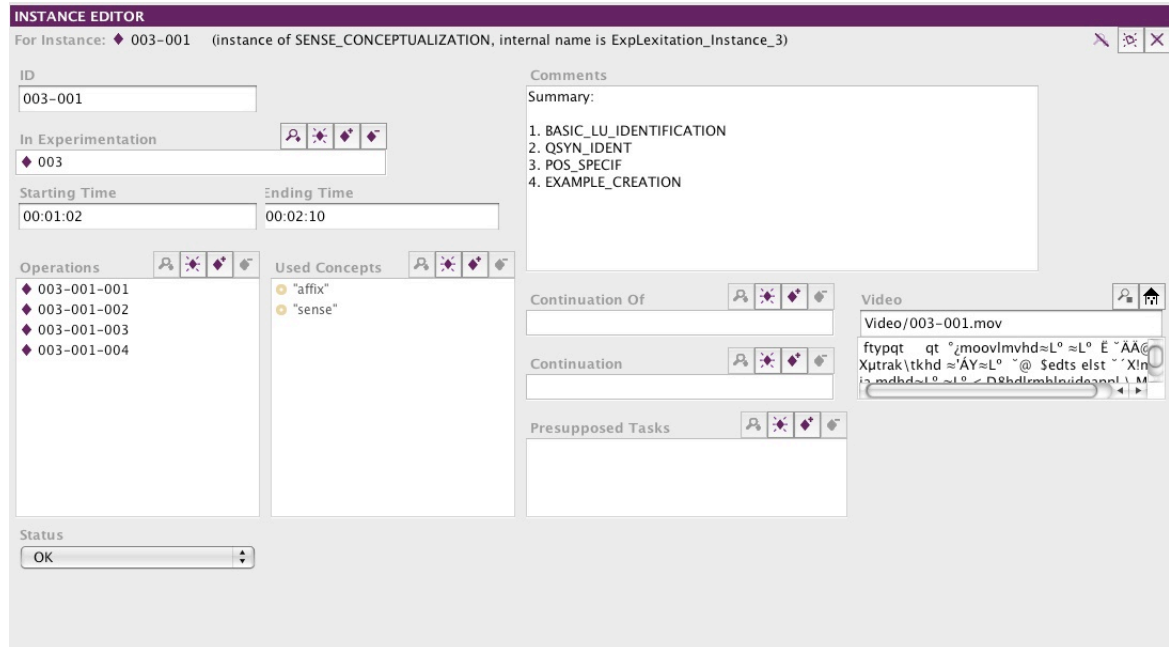


Figure 4.4 – Instance de tâche

Il s'agit d'un champ dans lequel on indique à quelle expérimentation appartient la tâche.

– *starting time et ending time*

On identifie l'heure de début et de fin de chaque tâche dans ces deux champs.

– *operations*

On indique ici toutes les opérations qui ont été effectuées dans la tâche.

– *used concepts*

Dans le cadre d'une ontologie des savoirs lexicographiques, il était important de présenter les notions lexicologiques et lexicographiques (connaissances déclaratives ou « statiques ») en jeu dans les tâches et opérations exécutées par les lexicographes. Comme une ontologie des notions lexicologiques et lexicographiques a déjà été construite, nous avons pensé qu'il serait intéressant d'utiliser ces notions directement. En effet, dans une phase antérieure du projet Lexitation, Alain Polguère et Ophélie Tremblay, une étudiante au doctorat participant au projet, ont construit une ontologie pour modéliser

les connaissances métalexicales, appelée le *Gros Tas de Notions* (GTN). Cette ontologie, qui s'inspire de la *Lexicologie explicative et combinatoire* (LEC)(Mel'čuk et al., 1995) (Polguère, 2008), nous fournit un cadre théorique pour ce qui est des connaissances déclaratives mobilisées par les lexicographes. Comme elle a également été développée avec Protégé, nous avons pu mettre les deux ontologies en relation. Ainsi, le GTN devient en quelque sorte un module de notre ontologie, par l'entremise du champ *used concepts*, présent dans chacune des instances d'opérations et de tâches. Dans ce champ, nous indiquons les connaissances lexicographiques déclaratives utilisées par le lexicographe dans la tâche donnée. Les notions que nous pouvons afficher dans ce champ sont des classes du GTN. Ce dernier est modifié et enrichi au besoin, au fur et à mesure que nos analyses avancent.

– *comments*

Tout commentaire jugé pertinent.

– *continuation of*

Ici, on indique si la tâche constitue la continuation d'une autre tâche.

– *continuation*

Inversement, si le travail accompli dans la tâche est continué dans une tâche subséquente, on l'indique ici.

– *presupposed tasks*

Les tâches préalables à une certaine tâche sont celles qui l'ont précédée et sans lesquelles celle-ci n'aurait pas pu être réalisée.

– *video*

Chaque instance de tâche est accompagnée de la séquence audio-vidéo qui lui correspond. Nous avons ainsi accès à la portion du vidéo qui correspond exactement à la tâche. Il n'est pas possible de faire la même chose avec les opérations, celles-ci ne se prêtant pas à l'exercice du découpage pour les raisons mentionnées précédemment. Il est à noter que, dans le champ *video* lui-même, on ne voit qu'un code incompréhensible : il faut cliquer sur

l'icône représentant une loupe (celui de gauche) pour faire apparaître la vidéo à l'écran.

– *status*

Ce champ ne donne pas une information sur la tâche en tant que telle ; il s'agit plutôt d'un moyen qui sert à la logistique de l'analyse des données, en indiquant si l'instance a été révisée. La valeur par défaut est donc *proposed* jusqu'à la validation, après quoi elle est remplacée par *ok*.

• La classe LEXICOGRAPHIC_OPERATION

Voici les champs présents dans instances d'opérations, dont la Figure 4.5 est un exemple :

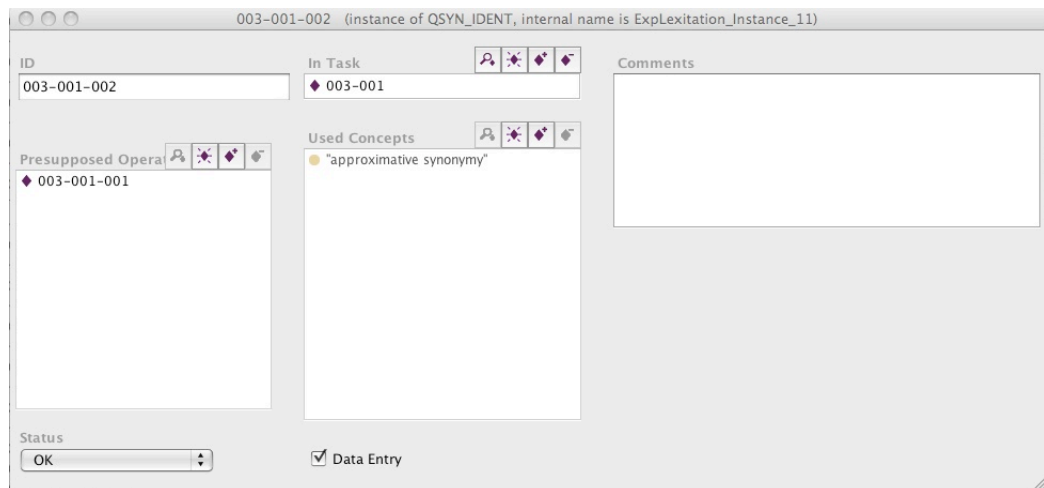


Figure 4.5 – Instance d'opération

– *id*

Le code d'identification des opérations comporte trois parties : le numéro de l'expérimentation, celui de la tâche puis celui de l'opération. Le code 003-002-001 identifie la première opération de la deuxième tâche de la troisième expérimentation.

– *in task*

Il s'agit d'un champ dans lequel on indique à quelle tâche appartient l'opération.

– *presuposed operations*

Les opérations préalables à une certaine opération sont celles qui l'ont précédée et sans lesquelles celle-ci n'aurait pas pu être réalisée.

– *used concepts*

Ce champ est également présent dans les instances de tâches (voir sa description dans la partie précédente).

– *comments*

Tout commentaire jugé pertinent.

– *status*

Cette comme nous l'avons dit plus tôt, il s'agit d'un moyen d'indiquer si l'instance a été révisée et validée. La valeur par défaut est donc *proposed* jusqu'à la validation, après quoi elle est remplacée par *ok*.

– *data entry*

Nous avons choisi d'inclure, dans chacune des instances d'opérations, un champ qui indique si l'opération en question a mené ou non à une entrée de données. En effet, plutôt que d'avoir une opération `DATA_ENTRY`, qui serait omniprésente, nous avons choisi de donner cette information sous la forme d'un champ à valeur booléenne, nommé *data entry*, dont les valeurs possibles sont *yes* ou *no*. Cela permet de garder le statut d'opération pour des activités plus purement lexicographiques.

Notons que les champs *starting time*, *ending time* et *video* ne peuvent figurer dans les instances d'opérations, étant donné la nature « intangible » de celles-ci (voir Section 4.1.2.2).

Voyons maintenant comment l'analyse de l'Expérimentation 5 a permis d'enrichir le modèle ontologique que nous avons développé.

4.3 Analyse de l'Expérimentation 5

4.3.1 Nouvelles classes de tâches et d'opérations

Après avoir analysé l'Expérimentation 3, nous avons analysé l'Expérimentation 5 plutôt que l'Expérimentation 4, parce qu'il y a eu des difficultés techniques lors de cette dernière (interruptions de l'enregistrement) et que, au moment d'effectuer les analyses, les données de l'Expérimentation 4 n'étaient pas encore disponibles. Nous avons déjà présenté les résultats de l'analyse de l'Expérimentation 3 (voir Section 4.1). Nous discuterons dans cette section des résultats de l'analyse de l'Expérimentation 5 et des différences que nous avons observées entre les deux expérimentations. Rappelons que les expérimentations 3 et 5 ont été menées avec des lexicographes distincts, qui ont effectué des activités lexicographiques différentes.

Il va sans dire que deux lexicographes n'effectueront pas exactement les mêmes tâches et les mêmes opérations pour réaliser une même activité lexicographique. La présence d'un nouveau lexicographe est donc en soi un facteur de variation dans les tâches et opérations. De plus, le choix de l'activité lexicographique conditionne directement le type de tâches et d'opérations qui sont effectuées, comme nous le verrons dans la section suivante.

4.3.1.1 Différences liées au type d'activité effectuée

Lors de l'Expérimentation 5, l'activité lexicographique choisie ne consistait pas à dégager la structure polysémique d'un vocable, comme ce fut le cas lors de l'Expérimentation 3. La lexicographe de l'Expérimentation 5, qui travaille sur la version informatisée du *Trésor de la langue française* (Imbs, 1971-1994)², a plutôt choisi de rédiger l'entrée de dictionnaire du vocable ADDICTION, un emprunt à l'anglais qui ne figure pas encore dans le TLF. Notons qu'il s'agissait d'un travail fait dans le cadre de la rédaction d'un supplément du TLF.

La polysémie beaucoup moins riche pour ADDICTION que pour COMPTER a eu une influence sur le type d'opérations effectuées pendant l'activité lexicographique. En

²Le *Trésor de la langue française informatisé* (TLFi) est disponible au <http://atilf.atilf.fr/tlf.htm>.

effet, le fait que l'activité de l'Expérimentation 5 ne consiste pas en un « recensement » de toutes les acceptions d'un vocable très polysémique, mais plutôt en la rédaction de l'entrée lexicographique complète d'un vocable peu polysémique a comme effet de réduire le nombre total de tâches accomplies. On compte en effet 40 tâches pour l'Expérimentation 3, contre seulement 19 pour l'Expérimentation 5. Bien entendu, s'il y a moins de tâches dans l'Expérimentation 5, celles-ci sont en revanche plus longues que celles de l'Expérimentation 3 : elles ont une durée moyenne d'environ 3 minutes, contre seulement 1 minute 30 secondes pour celles de l'Expérimentation 3.

Une autre différence remarquable entre les deux expérimentations est le nombre de tâches de la classe *SENSE_CONCEPTUALIZATION* effectuées. Il y a eu moins de tâches de conceptualisation de sens dans l'Expérimentation 5. Cette différence est en partie due au fait que, dans l'Expérimentation 5, le vocable à décrire n'a pas été choisi par l'équipe de recherche mais par la lexicographe elle-même. En fait, celle-ci avait déjà conceptualisé le vocable qu'elle allait décrire et elle avait déjà en tête la structure polysémique du vocable avant de se mettre au travail. La seule instance de tâche de la classe *SENSE_CONCEPTUALIZATION* de l'Expérimentation 5 est la première tâche de l'expérimentation (005-001), et ce n'est pas vraiment une tâche effectuée en direct ; les opérations sont évoquées dans une sorte de préambule de présentation de l'activité lexicographique. La démarche de la lexicographe est donc, dans un sens, inverse à celle du lexicographe de l'Expérimentation 3. En effet, elle a essayé de trouver des exemples qui correspondaient aux acceptions qu'elle avait en tête (ce qui ressemble plus à ce qu'on appelle, en psychologie, *top-down processing*), tandis que dans l'Expérimentation 3, le lexicographe a consulté le corpus pour trouver de nouvelles acceptions (ce qui ressemble plus à ce qui s'appelle *bottom-up processing* (Wolfe et al., 2003)).

4.3.1.2 Différences liées à des décisions éditoriales

Lors de l'Expérimentation 5, la lexicographe a travaillé en suivant les règles de rédaction du TLFi, le dictionnaire sur lequel elle travaille. Celles-ci ont influencé, dans une certaine mesure, le choix des tâches et opérations qui sont effectuées.

Toutefois, même s'il y a quelques règles à suivre particulières à chaque projet lexi-

cographique, cela ne veut pas dire que le travail des lexicographes consiste à suivre à la lettre un procédurier. Il y a en effet des décisions éditoriales qui régissent, par exemple, la présentation typographique, la présence et le choix des exemples, etc., et ces décisions peuvent conditionner le choix de certaines tâches ou opérations. Par contre, il n'y a pas de procédurier qui guide le lexicographe quand vient le temps d'analyser des faits de langue et de prendre des décisions ponctuelles quant au traitement de chaque lexie à décrire. Le lexique est trop vaste et trop complexe pour qu'un procédurier puisse couvrir toutes les situations auxquelles le lexicographe est confronté dans la pratique quotidienne de son travail.

Une anecdote racontée par Sue Atkins dans l'article *Then and Now : Competence and Performance in 35 Years of Lexicography*, sur ses débuts au CREFD, il y a une trentaine d'années, illustre d'ailleurs le fait que le lexicographe est bien souvent laissé à lui-même dans l'exercice de son travail :

I (...) went home with a list of words beginning with H, and the brief to write biligual entries for them all. No other instructions were given but I was encouraged to look at other dictionaries which I had to supply myself, and make my own decisions. (Atkins, 2008a, 248)

Bien sûr, les pratiques ont changé depuis ce temps, et les lexicographes ont maintenant certaines directives à suivre lorsqu'ils rédigent des entrées de dictionnaire. Toutefois, ce sont des règles qui concernent, le plus souvent, la structure générale du dictionnaire en construction ou le patron d'organisation des articles du dictionnaire.

Par exemple, dans le TLFi, chaque lexie doit être accompagnée d'un exemple, qui est généralement tiré d'un des corpus compilés par l'équipe du TLFi selon des critères spécifiques³. Les lexicographes du TLFi ne peuvent pas créer leurs propres exemples, ni modifier ceux qu'ils choisissent. Conséquemment, la lexicographe de l'Expérimentation 5 n'a pas effectué d'opérations de la classe MODIFICATION_OF_EXAMPLE. Elle aurait pu effectuer des opérations de la classe CREATION_OF_EXAMPLE à des fins de conceptualisation, mais elle ne l'a pas fait. Cette contrainte exige également de la lexicographe

³Voir la préface du TLF au <http://www.atilf.fr/>.

qu'elle effectue de longues séances de recherche dans les corpus, afin de trouver des exemples appropriés pour chacune des acceptions qu'elle veut décrire dans son entrée de dictionnaire. Nous avons donc créé une nouvelle classe de tâches correspondant à ces séances de recherche, `SEARCH_FOR_EXAMPLE`. Dans les instances de cette classe, la lexicographe ne cherche pas de nouvelles acceptions, mais des exemples pour illustrer les acceptions qu'elle veut décrire. C'est donc une classe de tâches différente de `SEARCH_FOR_SENSE`, qui était utilisée dans l'Expérimentation 3.

De plus, comme les exemples ont une grande importance dans le TFLi, ils sont choisis selon des critères particuliers. Pour identifier les opérations qui consistent à évaluer les exemples, nous avons utilisé la classe d'opérations `ASSESSMENT_OF_EXAMPLE`, que nous avons déjà dans l'Expérimentation 3. Nous avons également créé la classe `SOURCE_OF_EXAMPLE_ASSESSMENT` pour identifier les opérations qui consistent à juger de la qualité de la source d'un exemple, de la crédibilité de l'auteur de l'exemple et de la publication dans laquelle il a été publié.

4.3.1.3 Tâches et opérations relevant de la rédaction d'entrées lexicographiques

Une des particularités de l'Expérimentation 5 est que la lexicographe y rédige l'entrée lexicographique complète d'un vocable, `ADDICTION`, qui inclut une définition pour chaque acception. Rappelons que durant l'Expérimentation 3, le lexicographe n'avait pas à rédiger de définitions à proprement parler. Nous avons donc créé une classe de tâches qui s'appelle `DICTIONARY_ENTRY_WRITING`.

Les instances de cette classe de tâches sont souvent composées, entre autres, d'opérations relevant de la rédaction de définitions lexicographiques. Voici les classes d'opérations concernant la rédaction de définitions lexicographiques que nous avons identifiées dans l'Expérimentation 5 :

Il est à noter que l'opération `GENUS_IDENT` existait déjà, puisque le lexicographe de l'Expérimentation 3 l'avait utilisée, dans un but autre que celui de rédiger une définition lexicographique.

Normalement, une définition lexicographique est composée du genre prochain et des différences spécifiques (voir la Section 1.3) de la lexie définie. En fait, la définition

- ▶ ● CLOSE_SENSE_IDENT
- GENUS_IDENT
- DIFFERENTIA_IDENT
- SEM_ACTANT_CHARACTERIZATION
- WRITING_OF_DEFINITION

Figure 4.6 – Opérations concernant la rédaction de définitions

par genre prochain et différences spécifiques est le type de définitions que propose la plupart des dictionnaires de langue, et que (Polguère, 2008, 182) appelle *définition analytique*. Rappelons que les notions de genre prochain et de différences spécifiques ont été utilisées par Aristote dans *Les Topiques* (Aristote, 1997) pour fonder sa théorie de la définition des concepts. Toutefois, il est possible qu'un lexicographe identifie l'une ou l'autre de ces composantes (le genre prochain ou les différences spécifiques) dans un autre but que pour composer une définition ; par exemple, il peut identifier le genre prochain de deux emplois pour déterminer s'il s'agit de deux lexies différentes ou non. De plus, un lexicographe pourrait aussi rédiger une définition qui ne serait pas du type analytique, c'est pourquoi l'opération WRITING_OF_DEFINITION est nécessaire. En effet, on ne peut pas présumer que la spécification du genre prochain et des différences spécifiques équivaut à la composition d'une définition. L'opération WRITING_OF_DEFINITION s'applique donc aux opérations dans lesquelles le lexicographe rédige ce qu'il considère lui-même comme la définition de la lexie qu'il décrit.

Pour ce qui est de la nouvelle classe nommée SEM_ACTANT_CHARACTERIZATION, nous en avons un exemple dans l'opération 11 de la tâche 4 (005-004-011), lorsque la lexicographe note les types d'actants possibles pour l'une des l'acceptations d'ADDICTION, 'dépendance à une pratique'. Elle note *ordinateur* et *achat*.

Il est à noter que ces 3 nouvelles classes (DIFFERENTIA_IDENT, SEM_ACTANT_CHARACTERIZATION et WRITING_OF_DEFINITION) ont été ajoutées, en tant que sous-classes, à une classe qui existait déjà : SEMANTIC_CHARACTERIZATION_OF_LU. Nous pouvons donc voir que les classes d'opérations que nous ajoutons s'insèrent naturellement dans la hiérarchie existante, qu'elle viennent compléter.

Comme nous avons ajouté une opération de rédaction de définition, nous avons également ajouté une opération qui consiste en la révision de cette définition : la classe DEFINITION_ASSESSMENT. Celle-ci entre dans la classe générale ASSESSMENT, qui comprend des évaluations qui portent sur tous les aspects du travail lexicographique.

4.3.1.4 Opérations concernant le statut d'une lexie

En plus des opérations concernant la rédaction de définitions, il a été nécessaire d'ajouter des classes d'opérations à cause du fait que le vocable à décrire est un néologisme et un emprunt à l'anglais. En effet, la lexicographe a dû prendre une décision sur le statut d'ADDICTION, à savoir s'il doit entrer dans le dictionnaire ou non (opération 005-001-008). Nous avons donc ajouté la classe ASSESSMENT_OF_LEXICALIZATION, qui pourrait éventuellement servir à statuer sur la lexicalisation d'un usage français, par exemple à savoir si une métaphore est libre ou lexicalisée, auquel cas elle doit entrer dans le dictionnaire. Dans le cas de l'Expérimentation 5, pour prendre cette décision, la lexicographe a invoqué le critère de la fréquence de la lexie. Cette instance d'opération (005-001-005) entre dans la classe ASSESSMENT_OF_FREQUENCY_OF_LU, que nous avons déjà. Dans l'opération suivante (005-001-006), la lexicographe justifie sa décision de faire entrer ADDICTION dans le dictionnaire français en affirmant qu'il est utilisé dans les revues scientifiques :

(...) tout le monde préfère utiliser le mot anglais ADDICTION, qui est dans toutes les revues scientifiques.

Par ce commentaire, on voit que le registre de langue ou le domaine d'utilisation de la lexie est un critère qui influence sa décision. Le fait qu'ADDICTION soit utilisé par la communauté scientifique semble légitimer son insertion dans le dictionnaire français. Sous la classe CHARACTERIZATION_OF_LU, nous avons donc ajouté la classe PRAGMATIC_CHARACTERIZATION_OF_LU et, sous celle-ci, SPEECH_VARIETY_SPECIF :

De plus, dans l'opération suivante (005-001-007), la lexicographe invoque la polysémie du vocable ADDICTION comme preuve de son entrée dans le lexique français :

Le mot ADDICTION a l'air de s'implanter dans la langue, puisqu'on le trouve dans d'autres acceptions.

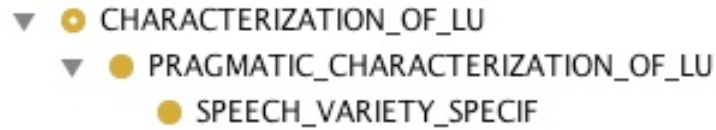


Figure 4.7 – Opérations concernant le statut d’une lexie

Pour rendre compte de cette opération, ainsi que d’autres dans lesquelles la lexicographe spécifie la structure polysémique du vocable, nous avons créé la classe POLYSEMIC_STRUCTURE_SPECIF, que nous avons placée sous la classe CHARACTERIZATION_OF_VOCABLE. Il est à noter que la classe POLYSEMIC_STRUCTURE_SPECIF inclut aussi une classe d’opérations qui consistent à spécifier la relation sémantique qui lie les acceptions d’un vocable. Dans l’opération 005-002-011, la lexicographe a spécifié que la deuxième acception de son vocable (ADDICTION - ‘dépendance à des produits alimentaires, à des pratiques, à des activités ou à des jeux’) était un sens « par extension » par rapport au sens de base, (‘dépendance à une drogue’). La classe SEM_RELATION_SPECIF a donc été ajoutée ici :

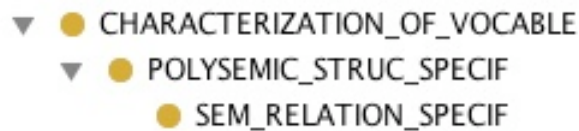


Figure 4.8 – Opérations concernant la structure polysémique d’un vocable

4.3.1.5 Consultation du dictionnaire

Une autre différence entre les deux expérimentations est liée aux consignes données aux lexicographes par l’équipe de recherche : dans l’Expérimentation 5, la lexicographe avait le droit de consulter des dictionnaires, mais ce n’était pas le cas pour l’Expérimentation 3. En fait, elle avait le droit de consulter tout le matériel qu’elle utilise normalement pour travailler. Elle a notamment utilisé le TLFi, pour y consulter la définition d’une lexie qui entrait dans la définition d’ADDICTION, qu’elle était en train de rédiger. Dans l’opération 005-005-001, elle cherche donc la définition de DÉPENDANCE, qui est le genre prochain de sa définition d’ADDICTION. Nous avons donc créé la classe

RETRIEVAL_OF_GENUS, sous DICTIONARY_ENTRY_CONSULTATION. Dans ce cas, il s'agit de la consultation du dictionnaire pour lequel la lexicographe travaille, mais la classe pourrait inclure éventuellement la consultation de dictionnaires « concurrents ». Plus tard, dans l'opération 005-006-003, elle cherche le mot INTOXICANT, qui figure dans l'un des exemples du corpus, et dont elle doute de la présence dans le TLFi. Après ses recherches, qui lui confirment qu'INTOXICANT ne figure pas dans le TLFi, elle prend la décision de ne pas utiliser cet exemple, puisqu'elle ne peut pas introduire dans une entrée une lexie qui n'est pas définie dans le TLFi : *Un dictionnaire doit être en boucle fermée, et on ne peut pas introduire de mots qui n'y sont pas.*

On voit donc que la consultation du dictionnaire peut être faite pour différentes raisons : pour s'assurer de la justesse d'une définition ou par soucis de cohérence interne du dictionnaire — pour s'assurer de ne pas utiliser dans une entrée un mot qui n'a pas d'entrée dans le dictionnaire — , etc.

Nous avons donc créé la classe DICTIONARY_CONSULTATION, nom volontairement vague, afin que la classe puisse inclure des opérations ayant différents objectifs. Effectivement, la classe pourra inclure des opérations qui consistent à consulter des entrées de dictionnaire ou la nomenclature même de celui-ci, pour toutes sortes de raisons, même éventuellement pour vérifier l'encodage formel des entrées.

Voici la classe et les sous-classes qu'elle contient jusqu'à maintenant :

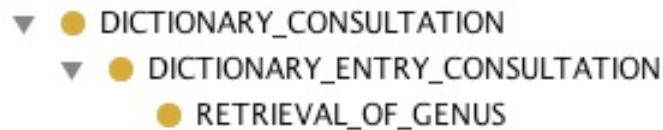


Figure 4.9 – Opérations concernant la consultation d'un dictionnaire

4.3.1.6 Nouvelles opérations concernant l'utilisation des corpus

De nouvelles classes d'opérations ont également dû être créées pour décrire la recherche dans les corpus faite dans l'Expérimentation 5. Il existait déjà des opérations de ce type : dans l'Expérimentation 3, le lexicographe soumettait le radical de la lexie

à décrire au moteur de recherche de son corpus pour retrouver les exemples contenant les formes (fléchies ou non) de la lexie en question. Nous avons donc créé l'opération `RADICAL_SPECIF` pour décrire cette opération. Or, dans l'Expérimentation 5, les moteurs de recherche des corpus utilisés par la lexicographe fonctionnent par chaînes de caractères, alors si la lexicographe soumet *addiction* au moteur de recherche, celui-ci ne lui fera correspondre que les phrases qui contiennent ce mot-forme. Par conséquent, ce que soumet la lexicographe au moteur de recherche n'est pas le radical de la lexie, mais bien sa forme au singulier, et elle effectuera plus tard une recherche avec la forme au pluriel. Nous avons donc créé les classes `WORD_FORM_SPECIF` et `ALTERNATE_WORD_FORM_SPECIF`, que nous avons placées sous `GRAMMATICAL_CHARACTERIZATION_OF_LU`, pour décrire ces opérations.

On voit donc que différents types d'outils demandent d'effectuer différentes opérations, mais que celles-ci viennent s'intégrer à notre hiérarchie naturellement. Elles viennent l'enrichir et, la plupart de temps, elles viennent se greffer à des classes qui existent déjà.

4.3.2 Classes de tâches et d'opérations réutilisées

Nous avons vu dans la Section 4.3.1.1 que seule la classe de tâches `CONCEPTUALIZATION_OF_LU` avait été réutilisée dans l'Expérimentation 5. Effectivement, la lexicographe n'ayant pas cherché d'acceptions (puisqu'elle avait déjà conceptualisé les acceptions qu'elle voulait décrire avant le début de l'activité) et n'ayant pas effectué de tâche consistant essentiellement à modifier la description, elle n'a pas effectué de tâches des classes `SEARCH_FOR_SENSE` et `MODIFICATION_OF_DESCRIPTION`. Or, si une seule classe de tâches a été réutilisée lors de l'Expérimentation 5, beaucoup de classes d'opérations ont été réutilisées. Voici la liste de classes d'opérations qui ont été réutilisées au moins une fois lors de l'Expérimentation 5 :

- `LU_IDENT`
- `BASIC_LU_IDENT`
- `QSYN_IDENT`

- CLOSE_COPOLYSEME_IDENT
- GENUS_IDENT
- POS_SPECIF
- CANONICAL_FORM_SPECIF
- EX_RETRIEVAL_FROM_CORPUS
- ASSESSMENT_OF_FREQUENCY_OF_LU
- SENS_EQUALITY_ASSESSMENT
- EXAMPLE_ASSESSMENT

Voici la hiérarchie des classes de tâches et celle des opérations maintenant que l'Expérimentation 5 a été analysée. Les nombres qui figurent entre parenthèses correspondent aux nombres d'instances de chaque classe au total (dans les deux expérimentations) :

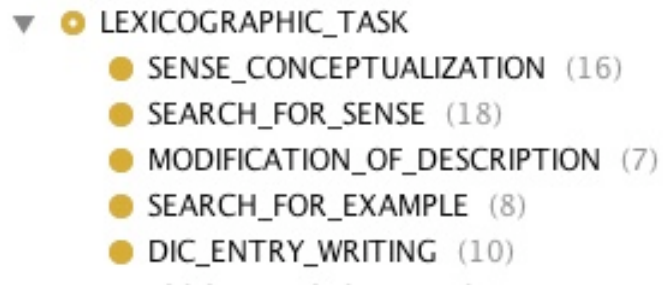


Figure 4.10 – Hiérarchie des tâches lexicographiques

- ▼ ● LEXICOGRAPHIC_OPERATION
 - ▼ ● LU_IDENTIFICATION (15)
 - BASIC_LU_IDENTIFICATION (2)
 - ▼ ● TESTING
 - ▼ ● TESTING_OF_COMBINATORICS
 - TESTING_OF_SYNT_COMBINATORICS (9)
 - TESTING_OF_LEX_COMBINATORICS (1)
 - TESTING_OF_GRAMM_COMBINATORICS (3)
 - ▼ ● TESTING_OF_SEMANTIC_EQUIVALENCE
 - REPLACEMENT_BY_QSYN_IN_EX (4)
 - PARAPHRASING (11)
 - CORPUS_ANALYSIS (15)
 - ▼ ● CHARACTERIZATION_OF_VOCABLE
 - ▼ ● POLYSEMIC_STRUC_SPECIF (1)
 - SEM_RELATION_SPECIF (1)
 - ▼ ● CHARACTERIZATION_OF_LU
 - ▼ ● SEMANTIC_CHARACTERIZATION_OF_LU
 - ▼ ● CLOSE_SENSE_IDENT
 - QSYN_IDENT (20)
 - CLOSE_COPOLYSEME_IDENT (17)
 - GENUS_IDENT (4)
 - SEM_ACTANT_CHARACTERIZATION (2)
 - DIFFERENTIA_IDENT (3)
 - WRITING_OF_DEFINITION (2)

Figure 4.11 – Hiérarchie des opérations lexicographiques

- ▼ ● GRAMMATICAL_CHARACTERIZATION_OF_LU
 - POS_SPECIF (13)
 - GOVERNMENT_PATTERN_SPECIF (4)
 - RADICAL_SPECIF (1)
 - CANONIC_FORM_SPECIF (2)
 - WORD_FORM_SPECIF (3)
 - ALTERNATE_WORD_FORM_SPECIF (2)
 - ▼ ● CONTEXTUALIZATION_OF_LU
 - EXAMPLE_CREATION (9)
 - ▼ ● EXAMPLE_RETRIEVAL
 - EX_RETRIEVAL_FROM_MEMORY (2)
 - ▼ ● EX_RETRIEVAL_FROM_CORPUS (23)
 - EX_RETRIEVAL_WITH_BROADER_CONTEXT (1)
 - EX_RETRIEVAL_FROM_DESCRIPTION (1)
 - EXAMPLE_TRANSFER (1)
 - EXAMPLE_MODIFICATION (2)
 - ▼ ● PRAGMATIC_CHARACTERIZATION_OF_LU
 - SPEECH_VARIETY_SPECIF (1)
 - ▼ ● REVIEW_OF_DESCRIPTION
 - REVIEW_OF_LU_DESCRIPTION (9)
 - REVIEW_OF_POLYSEMIC_STRUCTURE (5)
 - ▼ ● ASSESSMENT
 - ASSESSMENT_OF_FREQUENCY_OF_LU (4)
 - ▼ ● SENSE_EQUALITY_ASSESSMENT (16)
 - SENSE_EQUAL_ASSESS_THROUGH_COMBINATORICS (5)
 - ▼ ● DESCRIPTION_ASSESSMENT (1)
 - QSYN_ASSESSMENT (5)
 - DEF_ASSESSMENT (1)
 - ASSESSMENT_OF_LEXICALIZATION (1)
 - ▼ ● EXAMPLE_ASSESSMENT (19)
 - SOURCE_OF_EX_ASSESSMENT (1)
 - ▼ ● DICTIONARY_CONSULTATION
 - ▼ ● DICTIONARY_ENTRY_CONSULTATION (1)
 - RETRIEVAL_OF_GENUS (1)
-

Figure 4.12 – Hiérarchie des opérations lexicographiques (suite)

CHAPITRE 5

PERSPECTIVES SUR LES APPLICATIONS POSSIBLES EN DIDACTIQUE DES LANGUES

Nous avons mis sur pied le projet d'extraction des savoirs lexicographiques que nous avons présenté dans les chapitres précédents parce que nous croyons qu'il serait utile d'identifier les savoirs que les lexicographes ont acquis avec le temps, par leur expérience de travail, et de les « partager » avec un plus grand nombre de gens. Nous pensons en effet que des professionnels du langage et, parmi eux, les enseignants de langue, pourraient faire bon usage de ces savoirs. En effet, nous avons constaté qu'il y avait des améliorations à faire dans le domaine de l'enseignement/apprentissage du lexique (voir Chapitre 1). Nous croyons que la solution à cette lacune qui semble généralisée — dans les pays francophones, du moins —, réside en grande partie dans la formation des maîtres. Le lien entre la lexicographie et l'enseignement du vocabulaire ayant déjà été exposé dans la Section 1.3 du Chapitre 1, nous expliquerons dans ce chapitre pourquoi nous pensons qu'un entraînement à la lexicographie serait souhaitable pour les enseignants de langue et comment notre ontologie pourrait participer à cette formation en tant que base de connaissances de laquelle des outils pédagogiques pourraient être dérivés. Rappelons que l'objectif de la présente recherche n'est pas de développer de tels outils pédagogiques. Ceux-ci seront élaborés dans une phase ultérieure du projet Lexitation. Pour l'instant, dans la Section 5.1, nous nous contenterons de proposer des pistes de réflexion sur la nature et sur les caractéristiques de ces futurs outils pédagogiques. En dernier lieu, dans la section 5.2, cette réflexion concernera tout particulièrement leur possible intégration dans le modèle de formation scolaire québécois.

5.1 Entraînement à la lexicographie pour les maîtres

Pourquoi un entraînement à la lexicographie serait-il utile aux maîtres de langue ? Pour répondre à cette question, commençons par préciser le type de connaissances que

nous voulons que les maîtres acquièrent puis transmettent aux élèves.

5.1.1 Orientation qualitative plutôt que quantitative

Comme l'a remarqué Claude Simard (1994), l'enseignement du vocabulaire peut prendre deux grandes orientations ; l'une quantitative (quand il vise à augmenter le bagage lexical de l'élève) et l'autre, qualitative, quand il vise à *amener l'élève à mieux structurer son vocabulaire en lui faisant prendre conscience des régularités et des mécanismes de fonctionnement du lexique* (Simard, 1994, 30). Jusqu'à maintenant, selon Simard, c'est surtout l'orientation quantitative qui a été mise de l'avant dans les écoles, et malgré la volonté de nombreux enseignants d'aborder davantage le volet qualitatif, celui-ci est encore souvent mis de côté en classe. Il semble que le manque d'outils pédagogiques soit responsable de cet état de fait, de même que les limites des connaissances métalexicales, c'est-à-dire les notions théoriques et méthodologiques qui permettent de comprendre et de décrire l'organisation du lexique, des enseignants eux-mêmes. Remarquons que ces limites sont normales vu le peu de formation que les enseignants reçoivent en lexicologie et en lexicographie (voir Section 1.4).

Or, la compétence lexicale que nous voulons transmettre aux enseignants n'est pas quantitative, mais qualitative. En effet, ce n'est pas la taille du vocabulaire maîtrisé par les enseignants qui nous intéresse ici — bien que ce soit important —, mais leur compréhension des structures du lexique elles-mêmes.

Nous croyons en effet que la solution — partielle — au problème présenté plus tôt réside dans la formation des maîtres ; plus précisément dans un entraînement au travail lexicographique. Un tel entraînement leur permettrait de développer des capacités d'observation et d'analyse des phénomènes lexicaux qui sont nécessaires à l'acquisition du savoir lexical, selon Polguère :

Parce qu'il est impossible d'isoler un ensemble simple et clairement délimité de « règles lexicales » qu'il faudrait enseigner, on ignore tout simplement le lexique comme composante centrale de la connaissance linguistique et on passe à côté de la véritable solution : il faut apprendre à observer et à

décrire les phénomènes lexicaux pour mieux acquérir la connaissance lexicale. (Polguère, 2004, 1)

Il faut dire qu'à l'heure actuelle, les enseignants sont très peu formés à l'analyse des phénomènes lexicaux. En conséquence, ils manquent souvent de méthodes d'analyse et de stratégies pour répondre aux questions des élèves. Ils sont également peu à même de poser des diagnostics rapides et justes sur les erreurs de nature lexicale, comme l'a remarqué Polguère (2004) plus loin dans le même article.

Nous pensons donc que les maîtres devraient approfondir leurs connaissances métalexicales, et le travail lexicographique consiste justement en grande partie à observer, à comprendre puis à consigner des faits de langue relatifs au lexique. Or, nous croyons que, comme un sportif amateur peut s'inspirer des techniques d'entraînement des athlètes professionnels pour améliorer ses performances, même s'il n'a pas du tout l'intention de participer aux Olympiques, toute personne qui a besoin d'acquérir des connaissances métalexicales peut tout à fait s'entraîner au travail lexicographique pour les acquérir.

Bien entendu, il ne s'agit pas ici de faire des enseignants des professionnels de la lexicographie, mais bien de leur fournir quelques outils de base nécessaires à la pratique de l'enseignement de la langue et de son lexique (voir discussion à la Section 5.2.3).

Par exemple, avoir acquis une méthode de définition du sens des lexies serait un atout pour un enseignant de langue. Il faut souligner ici que ce qui aiderait l'enseignant dans son travail serait de connaître la **démarche** pour définir le sens des lexies, plutôt que d'avoir appris des définitions toutes faites. Cette démarche pourrait être utilisée pour définir n'importe quelle lexie (en autant que celle-ci soit connue de l'enseignant, bien sûr) rapidement et sans avoir recours à un dictionnaire. En plus de méthodes d'analyse du sens des lexies, l'entraînement à la lexicographie que nous souhaitons élaborer comprendrait des méthodes d'analyse des liens (paradigmatiques et syntagmatiques) que les lexies entretiennent entre elles.

Par ailleurs, des enseignants qui utiliseraient de telles méthodes en classe donneraient aux élèves un modèle d'observation et d'analyse des phénomènes lexicaux. Nous pensons que ces « bons réflexes » pourraient se transmettre aux élèves, et leur *donner*

progressivement les capacités d'apprentissage des connaissances lexicales, plus importantes que ces connaissances elles-mêmes (Polguère, 2004). Les apprenants pourront d'ailleurs faire usage immédiatement de ces capacités, si embryonnaires soient-elles, puisque *l'intelligibilité du message est facilitée par la propriété des termes et la discrimination des mots proches par le sens ou la forme* (Simard, 1994, 28). En d'autres termes, les capacités d'apprentissage des élèves leur permettront de décoder plus facilement les messages qui contiennent des mots inconnus, de faire des déductions, etc. De plus, elles leur permettront d'élargir leur vocabulaire plus rapidement. Cela veut dire qu'une amélioration des compétences dites qualitatives entraînera assurément une amélioration des compétences dites quantitatives.

Comment notre ontologie peut-elle intervenir dans un entraînement à la lexicographie pour les enseignants ? Dans la prochaine section, nous donnerons des pistes de réflexion pour répondre à cette question.

5.1.2 Le rôle de notre ontologie dans un entraînement à la lexicographie

Nous croyons qu'une ontologie comme la nôtre — mais plus complète que celle dont nous disposons à l'heure actuelle — peut servir de base pour élaborer un entraînement à la lexicographie. Notre raisonnement est le suivant :

Pour développer les compétences que les professionnels d'un domaine possèdent, il faut s'exercer à faire les tâches qu'ils font dans leur travail et qui leur ont permis de développer ces compétences. Nous proposons donc de faire faire aux enseignants des simulations — complètes ou partielles — d'activités lexicographiques qui ont été faites lors de nos expérimentations et qui sont décrites dans l'ontologie.

L'ontologie peut fournir des modèles de démarches à suivre pour effectuer différentes activités ou tâches lexicographiques (voir Section 4.1.2 pour les définitions de ces concepts). Par exemple, parmi les tâches lexicographiques dont l'ontologie peut fournir des modèles, il y a la conceptualisation (de sens) de lexies, qui fait partie du travail lexicographique, mais que tout individu qui apprend une langue doit également accomplir.

En effet, selon Simard (1994), des phénomènes qui entrent en jeu dans l'acquisition du vocabulaire, le plus important est la construction du sens. Il y a bien sûr aussi la mémorisation des formes, mais celle-ci est secondaire puisque *apprendre le vocabulaire ne se ramène pas à associer un signe à une chose pour accumuler des étiquettes* (Simard, 1994, 29). L'essentiel, donc, est la construction du sens, et celle-ci passe par un processus de conceptualisation qui aboutira à la constitution d'une définition mentale de la lexie :

La saisie du sens d'un mot passe par un processus de conceptualisation qui peut prendre diverses formes selon l'âge et les concepts à se représenter. L'apprenant crée une signification en abstrayant de la réalité un ensemble de traits devant former la définition du mot. (Simard, 1994, 29)

Ce processus est effectué intuitivement par les apprenants, mais il n'est pas l'objet d'un enseignement explicite. Nous pensons qu'il gagnerait à être révélé à la conscience des apprenants. Pour ce faire, on doit premièrement décortiquer ce processus et en expliciter les composantes, soit les opérations mentales qu'il implique. C'est, entre autres, ce que nous avons réalisé avec nos expérimentations. Celles-ci nous ont permis de mettre à jour les opérations mentales constitutives du processus de conceptualisation du sens des lexies chez des lexicographes. L'ontologie offre plusieurs instances de tâches de conceptualisation de sens, appelées SENS_CONCEPTUALIZATION.

Quand on a conceptualisé le sens d'une lexie, on peut normalement facilement composer la définition de cette lexie. Or, notre ontologie peut servir de base pour s'entraîner à composer des définitions de lexies, comme nous l'avons mentionné plus tôt. En effet, elle comporte une classe de tâches qui s'appelle DICTIONARY_ENTRY_WRITING. Les instances de cette classe peuvent servir de modèles de bonnes démarches à suivre pour composer une définition.

Par exemple, la tâche 003 de l'Expérimentation 5, de la classe DICTIONARY_ENTRY_WRITING, est composée des opérations suivantes :

1. CANONICAL_FORM_SPECIF
2. POS_SPECIF

3. BASIC_LU_IDENT
4. GENUS_IDENT
5. DIFERENTIA_IDENT
6. WRITING_OF_DEFINITION

En consultant l'ontologie, on peut avoir accès à la séquence même de la vidéo qui correspond à cette tâche, ainsi qu'à toutes les autres tâches de cette classe (il y a en une dizaine à l'heure actuelle). Rappelons que, pour participer à nos expérimentations, après une phase de tests, nous avons sélectionné des lexicographes professionnels. Pour l'instant, en analysant les données des expérimentations, nous décrivons toutes les tâches et opérations effectuées par les participants, sans juger de l'efficacité ou de la pertinence de chacune d'entre elles. Éventuellement, nous aurons un bassin de données assez vaste pour extraire des démarches-types servant à résoudre différentes sortes de « problèmes » lexicographiques, ou pour réaliser divers types de tâches.

Nous reconnaissons que l'ontologie, dans sa forme actuelle, est tout à fait embryonnaire, et qu'il faut absolument la développer pour atteindre une masse critique de classes de concepts permettant un véritable transfert pédagogique. Quand cette masse sera atteinte, les enseignants pourront non seulement apprendre par l'exemple en étudiant les démarches des lexicographes professionnels, mais ils pourront aussi s'initier à la pratique lexicographique, en ayant une bonne référence pour se guider et pour se corriger. Par exemple, ils pourront faire des activités semblables à celles des expérimentations : dégager la polysémie d'un vocable, composer la définition d'une lexie, etc. Ces activités pourraient se faire en consultant les démarches-types extraites de l'ontologie, pendant ou après. Les enseignants auront ainsi des modèles de bonne démarche à suivre pendant l'activité.

De plus, à partir des démarches-types de « résolution de problème » lexicographiques qui seront extraites de l'ontologie, des activités pédagogiques pourraient être dérivées (dans une phase ultérieure du projet Lexitation). Premièrement, des activités pourraient être élaborées pour les maîtres, puis elles pourraient être adaptées pour les élèves.

Cet entraînement à la lexicographie permettrait aux enseignants de développer les habiletés d'analyse et de modélisation de phénomènes lexicaux dont nous avons parlé plus tôt (Section 5.1.1), et qui seraient utiles dans deux types de pratiques didactiques.

En effet, non seulement des maîtres entraînés à la lexicographie pourraient-ils s'inspirer des démarches-types extraites de l'ontologie pour créer eux-mêmes des activités pour leurs élèves — en les adaptant à leur niveau—, mais ils pourraient aussi maximiser la pertinence de leurs interventions spontanées concernant le lexique.

Nous examinerons dans les prochaines sections ces deux types de pratiques d'enseignement, dans lesquels les connaissances lexicographiques acquises pourraient être mises à profit.

5.1.3 Les deux applications d'un entraînement à la lexicographie

Les enseignants peuvent se servir de compétences lexicographiques dans deux types de pratiques d'enseignement :

- Les analyses ponctuelles spontanées faites lors d'interactions en classe ;
- Les activités décrochées, planifiées ou improvisées, faites en classe avec les élèves.

5.1.3.1 Les analyses lexicales spontanées faites en classe

Premièrement, précisons ce que nous entendons par *analyses lexicales ponctuelles spontanées*. C'est le type d'enseignement qui consiste en interventions plus ou moins préparées, par exemple pour répondre aux questions des élèves sur telle ou telle lexie. Les enseignants pratiquent déjà, naturellement, ce type d'interventions, dont tous reconnaissent l'utilité. M.-J. Besson, cité par J.-F. De Pietro (2003, 13), rappelle que les interventions ponctuelles des enseignants concernant le lexique participent à l'enrichissement du vocabulaire des élèves :

En classe, des précisions de ce type sont fréquentes. Comme doivent être fréquentes les occasions de faire jouer les mots dans des contextes qui

en éclairent le sens. Tout cela dans une pratique naturelle et quotidienne de la langue visant enrichir le vocabulaire de chaque élève.

Les interventions ponctuelles spontanées ne sont peut-être pas suffisantes dans l'optique d'un enseignement du vocabulaire systématique, mais elles sont nécessaires, et nous croyons que le potentiel de notre ontologie se situe en grande partie ici. À l'heure actuelle, les enseignants ne sont pas toujours outillés pour mener efficacement ces interventions, comme nous l'avons remarqué plus haut. Une initiation à la lexicographie permettrait de les rendre plus autonomes quand ils sont confrontés à un nouveau phénomène à expliquer.

Par exemple, comme nous l'avons mentionné précédemment, les capacités d'analyse acquises les rendraient plus aptes à ébaucher des définitions précises « à la volée » — sans toutefois le faire aussi aisément que les lexicographes. Elles les aideraient aussi à poser des diagnostics plus rapidement sur des problèmes lexicaux, et à trouver plus facilement des réponses aux questions inattendues des élèves.

Par exemple, un professeur entraîné à la lexicographie est plus à l'aise pour composer une définition sur-le-champ pour répondre à une question telle que : *Que veut dire le mot addiction ?* En effet, il peut assez facilement composer une définition en déterminant le genre prochain ('dépendance') et les différences spécifiques ('à une drogue ou un médicament'). Il peut également valider la définition qu'il vient de former en appliquant des tests. Un professeur qui n'est pas entraîné à faire ce genre d'opérations sera plus enclin à chercher la définition d'ADDICTION dans un dictionnaire. Or, dans ce cas, la lexie en question ne figurerait probablement pas dans le dictionnaire consulté, et l'enseignant serait moins à même de répondre à la question de l'élève. Bien sûr, une situation comme celle-ci est peu susceptible de se produire, puisqu'il y a relativement peu de mots comme ADDICTION qui ne sont pas dans le dictionnaire. En fait, ce que l'on vise est que les enseignants soient capables de répondre à ce genre de question pour tous les mots qu'ils connaissent sans avoir à chercher dans un dictionnaire, et nous pensons qu'un entraînement à la lexicographie peut les aider à l'être.

Aussi, cet entraînement les rendrait plus conscients des phénomènes de collocation et, plus généralement, de phraséologisation, et donc davantage capables d'analyser et

d'expliquer ces phénomènes. Par exemple, si un élève demande pourquoi on dit *prendre une marche*, mais pas *prendre un pas*, le professeur initié à la lexicologie est en mesure non seulement de dire *On dit « faire un pas », et non « prendre un pas »*, mais il est également en mesure de connecter cette règle à un phénomène lexical général, celui de la collocation. C'est l'occasion de déclencher une discussion sur l'idiomaticité de l'expression en cause et le prétexte pour effectuer des activités impliquant l'analyse d'expressions de ce type, comme l'a suggéré Polguère (2004).

Bref, en outillant l'enseignant pour mener de façon optimale des interventions ponctuelles sur le vocabulaire, nous voulons renforcer et améliorer une pratique qui se fait naturellement et qui est nécessaire, même si elle n'est peut-être pas suffisante.

Nous verrons dans la prochaine section qu'un enseignement plus systématique et planifié du vocabulaire peut aussi être envisagé, dans l'exercice duquel des compétences lexicographiques pourraient également être mises à profit par les enseignants.

5.1.3.2 Activités planifiées

Plusieurs chercheurs en didactique du lexique souhaiteraient voir, à côté des interventions ponctuelles, un enseignement plus systématique du vocabulaire :

En plus des interventions ponctuelles que l'enseignant ou l'enseignante peut réaliser au fil des lectures et des autres activités scolaires, il y a sûrement place dans la classe de français pour des interventions plus méthodiques visant à mieux faire saisir à l'élève le fonctionnement des structures lexicales de sa langue. (Simard, 1994, 28)

Polguère (2004) avait remarqué que la maîtrise du savoir lexicographique donne à l'enseignant *la possibilité d'exploiter toutes les occasions d'improviser des activités d'exploration de la langue — à partir de questions, de faits linguistiques rencontrés dans les textes, etc.* (Polguère, 2004, 3) Or, nous pensons que ces savoirs permettront également à l'enseignant de *planifier* des activités structurées sur le vocabulaire en lien avec un projet de communication plus vaste. En fait, ces activités pourraient être de dif-

férentes natures : décrochées ou intégrées à un projet éducatif. Nous examinerons ces deux types d'activités dans la prochaine section.

5.1.4 Activités décrochées et intégrées sur le vocabulaire

Voyons d'abord ce que sont les activités *décrochées* :

Allal et al. (2001, 31-32) les utilisent dans le contexte de l'enseignement de l'orthographe. Voici la définition qu'ils en donnent :

En contexte de production textuelle, les activités décrochées proposent un traitement décontextualisé des savoirs orthographiques visés par la séquence pédagogique. Elles consistent en exercices centrés sur un objectif d'orthographe abordé en dehors de tout contexte de communication. Ce type d'activités s'oppose aux activités dites différées, qui gardent un lien fonctionnel avec la situation de production textuelle, et qui consistent en diverses exploitations de textes produits par les élèves.

Simard (1994), quant à lui, utilise le terme en enseignement du lexique. Il fait aussi une distinction entre deux modes d'insertion des activités sur le lexique dans la classe de langue maternelle : activités *intégrées* et activités *décrochées*. Les premières s'intègrent à un projet global de communication orale ou écrite. Les secondes se font surtout de façon planifiée : *l'enseignant peut décider de réaliser une leçon spécifique sur tel ou tel point du lexique sans s'obliger au départ à lier celle-ci à une pratique de discours complète.* (Simard, 1994, 32) On parle alors d'activités *décrochées*. Par ailleurs, *les connaissances développées durant ces activités d'analyse de la langue devront ultérieurement être réinvesties dans des activités globales de communication.* (Simard, 1994, 32)

Bref, l'enseignant pourrait élaborer des activités décrochées en s'inspirant des démarches extraites de l'ontologie et consignées dans une ressource élaborée pour les besoins spécifiques de l'enseignement. En effet, il est important de souligner que nous ne voulons pas que les enseignants manipulent l'ontologie dans sa forme actuelle. Il faudra la rendre plus « digeste » avant que les enseignants puissent s'en servir pour préparer des activités pédagogiques.

Par exemple, pour sensibiliser les élèves à la notion de *combinatoire* des lexies (combinatoire syntaxique, lexicale ou grammaticale), dans le cadre d'activités de découverte visant à enrichir les connaissances lexicales, un enseignant pourrait faire remarquer aux élèves des phénomènes relevant de la combinatoire des lexies, par exemple le fait que les différentes acceptions d'un verbe peuvent avoir des régimes différents, et donc s'utiliser avec des prépositions différentes. Pour ce faire, il pourrait s'inspirer des 9 instances d'opérations TESTING_OF_SYNT_COMBINATORICS de l'Expérience 3, dans lesquels le lexicographe cherche les différents sens du verbe COMPTER en testant sa combinatoire syntaxique. Bien sûr, cette stratégie pourrait être faite avec d'autres verbes.

Des activités de ce type pourraient être planifiées à l'avance, mais elles pourraient aussi se faire de façon spontanée au besoin, quand une lexie particulière est rencontrée en classe.

Une autre activité possible, pour rendre les élèves autonomes devant des questions concernant le lexique, est de leur montrer à utiliser des corpus (dont Internet est le plus facilement accessible) pour répondre à leurs questions. Le professeur peut s'inspirer des instances d'opérations de la classe CORPUS_ANALYSIS des expérimentations 3 et 5 pour créer des activités dans lesquelles les élèves apprendraient à se servir des corpus. En effet, ces instances offrent des modèles de techniques d'analyse de corpus, et montrent aussi que l'on peut consulter des corpus pour répondre à différents types de questions. Par exemple, pour vérifier si une lexie verbale ou adjectivale est toujours employée avec un même adverbe (question à laquelle un dictionnaire ne pourrait fournir de réponse claire), on peut effectuer une recherche dans un corpus pour confirmer ou infirmer l'hypothèse de départ. Par exemple, le lexicographe de l'Expérimentation 3 utilise le corpus pour vérifier son intuition selon laquelle COMPTER dans le sens d'«espérer» s'utilise toujours avec l'adverbe BIEN, ce qui se révèle faux. On voit aussi dans les instances de la classe CORPUS_ANALYSIS la méthode utilisée par le lexicographe : il parcourt les occurrences, et pour chacune, il vérifie d'abord s'il s'agit de la bonne lexie, c'est-à-dire du verbe COMPTER et non du nom commun COMPTE, en vérifiant par une analyse syntaxique la partie du discours de l'occurrence. Ensuite, il vérifie s'il s'agit bien d'une occurrence du sens recherché par une analyse sémantique. Les instances d'analyse de

corpus de l'ontologie permettent non seulement d'illustrer **comment** analyser un corpus, mais aussi **quand** et **pourquoi** le faire. Nous croyons que ces techniques d'analyse de corpus seraient utiles à tous, puisqu'elles rendent plus autonomes dans les apprentissages sur le sens et la combinatoire des lexies. Bref, on peut montrer aux élèves que les corpus sont des outils dont on peut se servir dans toutes sortes de situations, comme la rédaction ou la révision linguistique, et les encourager à le faire. Bien sûr, il faut user de jugement quand on scrute un corpus — et surtout Internet. En effet, il faut savoir faire la différence entre une « invention lexicale » individuelle et un usage répandu, mais les corpus sont tout de même une source d'information appréciable, surtout parce qu'ils présentent des informations que les dictionnaires de langue générale ne consignent normalement pas.

Nous avons donné plusieurs idées d'activités s'inspirant des instances de tâches et d'opérations lexicographiques de l'ontologie pouvant être faites en classe. Par ailleurs, nous croyons que des activités d'exploration du lexique peuvent être introduites aussi tôt qu'à l'école primaire. En effet, une expérience d'Ophélie Tremblay, qui visait à évaluer le degré de réceptivité des élèves face à un enseignement lexical et d'évaluer — informellement — les capacités et connaissances lexicales des élèves à travers la réalisation d'activités portant sur le lexique, a montré que des élèves du primaire sont capables *d'observer des données linguistiques de nature lexicale et sémantique, de réfléchir sur des questions faisant appel à leurs capacités métalinguistiques (plus spécifiquement à des compétences métasémantiques et métalexicales), de formuler des justifications pour expliquer certains phénomènes lexicaux et enfin de générer des exemples.* (Tremblay, 2004, 138). Tout porte à croire qu'un entraînement à l'observation et à la réflexion métalinguistique est possible dès l'école primaire. De plus, les activités utilisées par Tremblay mettaient en jeu les concepts suivants : la polysémie, les champs sémantiques, les prédicats sémantiques, les actants sémantiques, la dérivation sémantique, les définitions lexicographiques et les collocations. Ce sont des concepts qui sont liés au travail lexicographique et qui pourraient être abordés naturellement dans des activités dérivées de l'ontologie.

Ces dernières pourraient, d'autre part, être intégrées autant en classe de langue maternelle qu'en classe de langue seconde. En effet, des aptitudes à observer et à analyser des faits de langue serviraient à l'apprenant dans l'apprentissage de sa langue maternelle, et l'accompagneraient ensuite dans l'apprentissage d'une langue seconde ou étrangère.

Mais est-ce que les apprentissages visés et les activités que nous proposons ont leur place dans les programmes d'enseignement primaire et secondaire ? Nous en discuterons dans la section suivante.

5.2 Théorisation : où les apprentissages lexicographiques se situent-ils dans les programmes scolaires ?

5.2.1 Contenus enseignés

Claude Simard a divisé les contenus d'apprentissage du vocabulaire en deux classes : contenus linguistiques et contenus culturels. En effet, selon lui, *l'enseignement du vocabulaire comporte nécessairement un volet culturel, puisque les mots traduisent les valeurs et les modes de vie en vigueur dans une communauté à une époque donnée* (Simard, 1994, 30). En exemple de contenu culturel, il donne la question des régionalismes et des québécoisismes, qui devrait être abordée à l'école. Nous pensons aussi que ces notions doivent être abordées à l'école. Par contre, selon nous, ces notions sont également de nature linguistique, et pourraient donc être incluses dans la catégorie des contenus linguistiques. Cela étant dit, on trouve dans (Simard, 1994) une liste de contenus qui devraient faire partie de l'enseignement du lexique, liste qui est tout à fait digne d'intérêt, et que nous étudierons ici.

Les contenus y sont présentés par classes et sont parfois accompagnés d'une courte description (en italique) :

- La dénomination (*capacité de nommer les êtres, les choses et les concepts*)
- La formation des mots
 - La dérivation et la composition (*dérivation : formation d'un mot par l'ajout d'un affixe à une base ; composition : formation d'un mot par la jonction de plusieurs mots*)

- Les familles de mots (*ensemble de mots unis par la forme et le sens autour d'une base*)

- Le sens et les relations de sens
- Les mots en discours

Nous croyons que l'apport de notre ontologie se situerait surtout au plan des classes de contenus appelées *Le sens et les relations de sens* et *Les mots en discours*. Nous laisserons donc de côté les autres classes de contenus pour l'instant.

Par ailleurs, les classes *Le sens et les relations de sens* et *Les mots en discours* se divisent en sous-classes de contenus, toujours selon la classification de Simard (1994) :

- Le sens et les relations de sens :
 - Les traits de signification (*éléments de sens communs ou spécifiques entrant dans la signification d'un mot*)
 - La polysémie (*caractère d'un mot présentant, selon les contextes, plusieurs sens propres ou figurés*)

Remarque : la notion est décrite du point de vue de l'analyse, qui n'est pas la nôtre (voir (Polguère, 2008, 59) pour notre définition de la polysémie).
 - La synonymie et l'antonymie (*synonymes : mots ayant des sens proches et pouvant se substituer l'un à l'autre dans certains contextes ; antonymes : mots de sens opposés*)
 - L'hyponymie (*rapport entre un mot spécifique et le mot générique qui l'inclut*)
 - L'analogie (*association de mots appartenant à un même thème ou à un même domaine*)
- Les mots en discours :
 - Les constructions syntaxiques (*contraintes grammaticales régissant la combinaison des unités lexicales (type de sujet ou de complément, choix de la préposition pour tel verbe, etc.)*)

- Les suites lexicales (*combinaisons fréquentes de mots, figées ou libres*)
(Remarque : ces notions correspondent plus ou moins à ce que nous appelons *collocations* et *locutions* (voir (Polguère, 2008, 165 et 57) pour les définitions de ces notions).
- Les registres de langue (*usages variant selon les types de textes et les situations de communication*)

Nous croyons que toutes ces sous-classes de contenus pourraient être abordées avec des activités lexicographiques inspirées de l'ontologie, puisque ces notions sont également traitées dans les activités lexicographiques effectuées dans nos expérimentations (elles sont utilisées en lexicographie en général). Par exemple, le lexicographe de l'expérimentation 3, qui devait trouver les diverses acceptions d'un vocable, a effectué des opérations qui concernent la polysémie. Dans notre ontologie, nous avons donc une classe d'opérations qui s'appelle POLYSEMIC_STRUC_SPECIF. De même, les lexicographes ont souvent inclus des synonymes ou des quasi-synonymes (opération QSYN_IDENT) dans leurs descriptions lexicographiques. Les *traits de significations* renvoient en fait aux notions de *genre prochain* (opération GENUS_IDENT) et de *différences spécifiques* (opération DIFFERENTIA_IDENT), éléments constituant de la définition que le lexicographe met au jour quand il décrit une lexie.

L'hyponymie et l'hyperonymie sont deux relations lexicales mutuellement converses, correspondant à un cas particulier d'inclusion de sens (Polguère, 2008, 148), qu'il est difficile d'introduire séparément. Ces notions sont également couvertes par l'attribution d'un genre prochain, puisque le genre prochain d'une lexie est, en fait, un de ses hyperonymes.

Pour ce qui est des *mots en discours*, la notion de construction syntaxique est souvent abordée en lexicographie puisque le régime d'une lexie figure normalement dans sa description lexicographique, et c'est un aspect que les lexicographes peuvent prendre en compte pour déterminer si deux emplois relèvent de la même lexie. Dans la hiérarchie des opérations de l'ontologie, on trouve la classe TESTING_OF_COMBINATORICS, qui se rapporte à la combinatoire des lexies (donc aux « mots en discours »), et qui comprend les sous-classes suivantes : TESTING_OF_LEXICAL_COMBINATORICS, TES-

TING_OF_SYNTACTIC_COMBINATORICS et TESTING_OF_GRAMMATICAL_COMBINATORICS. Les *suites lexicales*, quant à elles, correspondent aux collocations et aux locutions. Les collocations contrôlées par une lexie sont idéalement mentionnées dans l'article lexicographique de celle-ci : c'est donc un aspect que les lexicographes doivent traiter dans les descriptions lexicales. Finalement, les *registres de langue* sont aussi pris en compte dans les descriptions lexicographiques, nous avons d'ailleurs une classe d'opération SPEECH_VARIETY_SPECIF dans l'ontologie.

Nous voyons donc que des classes de l'ontologie correspondent à des contenus d'enseignement proposés par Claude Simard. Or, ce qui est à retenir, c'est que l'ontologie, avec les **instances** de ces classes, nous permettra d'extraire des modèles de « bonnes pratiques » lexicographiques, et que ce sont ces modèles qui seront à la base de l'entraînement lexicographique que nous voulons développer pour les enseignants.

Mais d'abord, les contenus que nous venons d'énumérer sont-ils inclus dans le programme de formation scolaire québécois ? Nous verrons dans la prochaine section que, malgré que leurs descriptions soient brèves, il y a bien des connaissances lexicales au programme scolaire québécois.

5.2.2 Lien avec le programme de formation scolaire

Les connaissances métalexicales que nous voulons que les élèves acquièrent font partie des savoirs essentiels nécessaires au développement des compétences *lire des textes variés*, *écrire des textes variés* et *communiquer oralement*, compétences énoncées dans le *Programme de formation de l'école québécoise*, publié par le ministère de l'Éducation du Québec. Aussi, certaines notions lexicales sont comprises dans les *connaissances liées au texte* que les élèves doivent acquérir durant leur formation scolaire. Nous verrons premièrement dans cette section la place occupée par le lexique dans la formation primaire, puis dans la formation secondaire. En dernier lieu, nous ferons quelques remarques sur le programme lui-même.

5.2.2.1 Le lexique dans la formation primaire

S'il y a peu de notions lexicales ou métalexicales au programme de formation primaire québécois, il y a tout de même quelques éléments répartis dans la section des *Connaissances liées au texte* (Ministère de l'Éducation, 2006a, 88-90). Cette section en comprend une autre, *Prise en compte des éléments de la situation de communication* (Ministère de l'Éducation, 2006a, 89), qui prescrit l'enseignement des connaissances suivantes :

- Registre de langue
 - Distinction des registres familier et standard
 - Formes du registre standard, d'usage courant, reconnues comme correctes et valorisées socialement (ex : prononciation, finales de mots, formes verbales, vocabulaires, formulation)

Dans la section *Prise en compte d'éléments de cohérence*, on retrouve la connaissance suivante :

- Reprise de l'information en utilisant des termes substituts
 - pronoms
 - synonymes
 - groupes de mots

Dans la section qui s'appelle *Exploration et utilisation du vocabulaire en contexte* (Ministère de l'Éducation, 2006a, 90), il y a les connaissances suivantes :

- Mots variés, corrects, précis, évocateurs liés aux thèmes abordés en français et dans les autres disciplines
- Termes utilisés pour consulter des outils de référence (ex : index, table des matières, mots-clés, bibliographie, fichier)
- Termes liés au monde du livre et de la littérature (ex : auteur, illustrateur, livre, recueil, chapitre, page de garde, couverture, quatrième de couverture, collection, éditeur, année de publication, dédicace)

- Noms des lettres de l'alphabet et des signes orthographiques
- Vocabulaire visuel constitué des mots fréquents et utiles
- Familles de mots
- Formation des mots (base, préfixe, suffixe)
- Formation des temps de verbes (radical+terminaisons)
- Principe alphabétique et combinatoire (règles d'assemblage des relations lettres-sons)
- Expressions figées, expressions régionales, expressions courantes, sens commun et sens figuré des mots
- Termes liés à la construction des concepts grammaticaux et à utiliser en situation de travail sur la langue :
 - Nom (commun/propre)
 - Pronom, verbe, adjectif
 - adverbe, préposition

En août dernier, le ministère de l'Éducation du Québec a publié un complément au programme actuel, appelé *Progression des apprentissages, français langue d'enseignement*. Ce document apporte des précisions sur les connaissances que les élèves doivent acquérir au cours de chacune des années du primaire. Les connaissances sont groupées autour de trois compétences, dont *écrire des textes variés*, qui demande d'acquérir des connaissances lexicales :

À la fin du primaire, l'élève emploiera dans ses textes des mots précis et justes appartenant généralement au registre standard et variera son vocabulaire lorsque le contexte de communication s'y prêtera. Elle ou il aura atteint cet objectif par : la découverte de divers procédés de formation des mots, dont la dérivation et la composition ; l'exploration des différentes relations de sens entre les mots ; la découverte de certaines propriétés des mots ; le développement de sa capacité à recourir aux dictionnaires. Des activités proposées en classe amèneront l'élève à s'intéresser au sens des

mots, à leur forme, à leurs propriétés et au contexte de communication dans lesquels ils sont employés. Le travail avec le dictionnaire lui apprendra à utiliser efficacement cet ouvrage de référence pour y trouver des informations de divers ordres, dépassant la simple fonction de vérification orthographique. Des pratiques régulières d'écriture donneront à l'élève l'occasion d'utiliser ses connaissances sur le lexique et d'améliorer sa capacité à formuler précisément et correctement ses idées. (Ministère de l'Éducation, 2009a, 5)

Ce document fait davantage de place aux connaissances sur la structure du lexique, sur les liens que les lexies entretiennent entre elles, bref, aux connaissances qualitatives que nous suggérons d'enseigner aux élèves dans la Section 5.1.1. Cela vient également appuyer notre proposition selon laquelle les enseignants ont besoin de bonnes connaissances lexicales et métalexicales.

Il y a d'ailleurs dans ce document des précisions intéressantes sur les connaissances lexicales à développer, notamment sur la polysémie. Dans la connaissance qui s'appelle *Le sens des mots*, on dit que l'élève doit pouvoir :

- observer qu'un mot exprime au moins un sens
 - il a au moins un sens propre (ex. : le premier sens de *dévoré* est « manger avidement, gloutonnement »)
 - ses différents sens sont proches les uns des autres (ex. : *carte* peut être un « petit carton servant dans différents jeux » ou un « carton illustré sur une face et utilisé pour exprimer des souhaits »)
 - il a un ou plusieurs sens figurés (ex. : un autre sens de *dévoré* est « lire avec avidité ») (Ministère de l'Éducation, 2009a, 6-7).

La découverte du phénomène de la phraséologisation est aussi proposée dans le document *Progression des apprentissages*. En effet, on y dit que l'élève doit pouvoir :

- observer qu'une locution a un sens qui n'est pas l'addition des sens des éléments qui la constituent (ex. : la locution *tirer les vers du nez*

signifie « faire parler » et non « prendre des petits animaux cylindriques et mous dans la partie du visage qui abrite l'organe de l'odorat »)

- connaître le sens de locutions courantes rencontrées dans des textes
- connaître le sens figuré de mots courants qui figurent dans les textes lus par les élèves. (Ministère de l'Éducation, 2009a, 7).

La notion de collocation est aussi mentionnée dans le document, qui stipule que l'élève doit pouvoir :

- observer que certains mots se combinent à d'autres de manière privilégiée (ex. : on peut dire une *peur bleue*, une *colère noire* mais non *une *peur noire*, *une *colère bleue*.) (Ministère de l'Éducation, 2009a, 7)

Le document inclut même une section sur les définitions et la structure des articles lexicographiques, *Le regroupement des mots de la langue dans le dictionnaire*, qui stipule que l'élève doit pouvoir :

- Observer différentes sortes de dictionnaires, imprimés ou électroniques
 - thématiques (ex. : visuel)
 - descriptifs, encyclopédiques
 - autres sortes de dictionnaires (ex. : dictionnaires de synonymes, de rimes)
- Trouver un mot dans le dictionnaire en se référant à l'ordre alphabétique
- Connaître les principales abréviations utilisées dans les dictionnaires (ex. : n.m. pour nom masculin, fam. pour familier ; inv. pour invariable)
- Observer la construction d'une entrée dans les dictionnaires descriptifs et dans les dictionnaires encyclopédiques

- la classe de mots et le genre des noms et des pronoms
 - la forme
 - * féminine des adjectifs qui changent de forme selon le genre
 - * plurielle des noms et des adjectifs quand elle est irrégulière
 - la présence d'autant de définitions que le mot a de sens
 - les deux parties d'une définition suivies d'exemples
 - * un terme
 - générique (ex. : un *tabouret* et un *fauteuil* peuvent être définis comme un « siège »)
 - appartenant généralement à la même classe de mots que le mot défini (ex. : la définition d'un nom commence par un nom, celle d'un verbe par un verbe)
 - * des différences spécifiques, à la suite du terme générique, pour distinguer deux mots définis par le même terme générique
 - le registre quand il n'est pas standard
 - la présence de mots appartenant à la même famille, de synonymes, d'antonymes, de locutions ou de mots qui peuvent se combiner avec le mot décrit
- Observer que des homophones s'écrivant de la même façon font l'objet de deux entrées dans le dictionnaire quand leurs sens sont très éloignés (ex. : il y a deux entrées pour *mineur*)
(Ministère de l'Éducation, 2009a, 7-8)

Nous saluons l'arrivée de ces précisions au programme, qui font véritablement une place à l'analyse des phénomènes lexicaux en classe, et qui, nous le croyons, affirment la nécessité, pour les enseignants, de posséder de bonnes connaissances lexicales et métalexicales, ainsi que quelques techniques d'analyse et de modélisation de phénomènes lexicaux. Nous avons vu que certaines notions lexicales et métalexicales sont déjà au programme scolaire du primaire. Voyons maintenant ce qu'il en est pour le secondaire.

5.2.2.2 Le lexique dans la formation secondaire

Le programme scolaire du premier cycle du secondaire consacre une section entière à l'enseignement du lexique (Ministère de l'Éducation, 2006b, 137-139). Il y est notamment question des concepts suivants :

- Formation des mots
 - Dérivation (préfixes, suffixes)
 - Composition
 - Étymologie
 - Famille de mots
- Sens des mots
 - Polysémie et sens contextuel
 - Homonymie, homophonie, paronymie
 - Sens propre et sens figuré
 - Comparaison, métaphore

Parmi les indications pédagogiques accompagnant ces concepts, on dit que les élèves du premier cycle du secondaire doivent pouvoir :

- Attribuer aux mots un sens approprié en se référant au contexte et à un dictionnaire
 - Reconnaître les différents types de renseignements fournis par les dictionnaires usuels, en accordant une attention particulière à la distinction des définitions et des exemples
 - Se référer au contexte pour donner un sens approprié aux marqueurs de relation ainsi qu'aux homonymes, notamment à ceux utilisés dans plusieurs disciplines
 - Se référer au contexte pour reconnaître et utiliser des mots présentant des possibilités de confusion homophonique (ex : *ver*, *vert*, *vers*, *verre* et *vair*) ou paronymique (ex : *imminent* et *éminent*)

- Reconnaître et utiliser les sens propre et figuré des mots en tenant compte de l'effet produit ou à produire
 - Établir une relation entre la valeur expressive (connotative) ou neutre (dénominative) des mots et le point de vue adopté.
- Reconnaître et utiliser un vocabulaire permettant l'expression du temps, du lieu, du but, de la cause et de la comparaison

Les autres notions présentées sont les suivantes :

- Relations entre les mots
 - Inclusion (mots génériques, mots spécifiques)
 - Analogie
 - Synonymie
 - Antonymie
 - Champs lexicaux
 - Combinaison de mots (expressions figées)

De plus, parmi les indications pédagogiques accompagnant ces concepts, on dit que les élèves doivent pouvoir :

- Reconnaître et utiliser des mots génériques ou spécifiques pour classer, hiérarchiser et préciser l'information, de même que pour assurer la continuité et la progression dans le texte
- Utiliser un dictionnaire analogique et un dictionnaire des synonymes en distinguant les nuances ou les connotations particulières aux synonymes et aux autres substituts
- Assurer la continuité entre les phrases de son texte et éliminer les répétitions abusives en recourant à des synonymes
- Reconnaître et utiliser des antonymes en accordant une attention particulière au contexte
- Reconnaître et utiliser les champs lexicaux pour mettre en évidence les liens existant entre des éléments du texte et établir des repères d'interprétation ou d'appréciation

- Observer en contexte certaines expressions courantes que l'usage a figées (locutions, clichés, dictons, proverbes) et qui font partie de la mémoire collective des francophones

La notion de variété de langue est traitée plus loin (Ministère de l'Éducation, 2006b, 140) et se divise en plusieurs points :

- Variation linguistique
- Langue standard (soignée)
- Langue familière

Au deuxième cycle du secondaire, toutes ces notions sont encore au programme (Ministère de l'Éducation, 2009b, 132-138), en plus des suivantes :

- Formation des mots
 - Emprunts (anglicismes)
 - Néologisme
 - Archaïsme
- Sens des mots
 - Sens contextuel
 - Paronymie
 - Connotation
 - Figure de style/procédé stylistique (comparaison, métaphore, métonymie, répétition, gradation, accumulation)

L'objectif du présent mémoire n'étant pas d'évaluer le programme de formation scolaire, nous n'en ferons pas ici une analyse approfondie, mais nous nous permettrons cependant de remarquer que les lignes directrices du programme sont très générales, et qu'il ne semble pas y avoir de structure logique. De plus, l'approche privilégiée est celle du décodage — i.e. de l'analyse —, qui vise à aider l'élève à interpréter le contenu lexical des énoncés linguistiques. Or, nous pensons qu'un programme d'enseignement de la langue devrait également inclure une perspective et des outils d'**encodage**, c'est-à-dire visant à aider l'élève à encoder linguistiquement sa pensée.

Cela étant dit, nous constatons tout de même avec plaisir que, tant au primaire qu'au secondaire, il y a des notions lexicales et métalexicales au programme d'enseignement scolaire. Nous pensons que toutes ces notions pourraient être abordées dans des activités inspirées des démarches de « résolution de problèmes » extraites de l'ontologie. Ces activités pourraient par exemple consister à faire l'ébauche d'une entrée lexicographique pour une lexie (possiblement une de celles qui ont été traitées dans les expérimentations), en suivant la démarche-type de rédaction d'entrée lexicographique extraite de l'ontologie, et en comparant sa démarche avec celle du lexicographe. Par ailleurs, nous croyons que, pour superviser ces activités, les enseignants devront faire intervenir plusieurs de leurs compétences professionnelles, comme nous le verrons dans la prochaine section.

5.2.3 Compétences professionnelles mobilisées

Nous avons vu que des activités inspirées de l'ontologie pourraient s'inscrire dans le programme de formation scolaire québécois, et nous allons voir maintenant que ces activités mobiliseraient des compétences professionnelles que le ministère de l'Éducation a déjà identifiées comme devant être maîtrisées par les enseignants.

En 2001, le ministère de l'Éducation du Québec a publié un document d'orientation qui s'appelle *La formation à l'enseignement — Les orientations — Les compétences professionnelles* (Ministère de l'Éducation, 2001), connu également sous le nom de *Référentiel de compétences professionnelles de la profession enseignante*, qui a pour but de décrire les compétences professionnelles que les enseignants doivent acquérir au cours de leur formation.

Toutes les compétences identifiées dans ce document ne seraient pas nécessairement mobilisées par les activités que nous proposons. Néanmoins, plusieurs d'entre elles, que nous présenterons ici, le seraient.

La deuxième compétence présentée dans *Le Référentiel de compétences professionnelles de la profession enseignante* est la suivante :

Communiquer clairement et correctement dans la langue d'enseignement, à l'oral et à l'écrit, dans les divers contextes liés à la profession en-

seignante.(Ministère de l'Éducation, 2001, 69)

La compétence comprend les composantes suivantes :

- Employer une variété de langage oral appropriée dans ses interventions auprès des élèves, des parents et des pairs.
- Respecter les règles de la langue écrite dans les productions destinées aux élèves, aux parents et aux pairs.
- Pouvoir prendre position, soutenir ses idées et argumenter à leur sujet de manière cohérente, efficace, constructive et respectueuse lors de discussions.
- Communiquer ses idées de manière rigoureuse en employant un vocabulaire précis et une syntaxe correcte.
- Corriger les erreurs commises par les élèves dans leurs communications orales et écrites.

Voici la description de cette dernière composante :

Tous les enseignants et les enseignantes doivent être soucieux de la qualité de la langue de leurs élèves, et ce, tant dans la langue parlée que dans la langue écrite. Sur ce point, le maître doit cependant faire preuve de discernement. Comme le souligne Brent (1999 : 125) le maître québécois est devant un dilemme : d'une part, l'application d'une norme pédagogique de la langue basée sur celle des classes moyennes ou supérieures, définie ailleurs, et trop éloignée de l'usage québécois et, d'autre part, une norme pédagogique basée sur l'usage populaire au Québec. Selon lui, un compromis entre ces deux extrêmes serait plus acceptable. Car il ne faut pas oublier que l'élève, de son côté, fait aussi face un dilemme de normes langagières : celles qui correspondent l'usage linguistique de son milieu d'appartenance et de ses pairs, et celles, externes, qui représentent la variété linguistique officielle de l'école qui peut parfois s'aligner sur un usage puriste de la langue. Selon Brent (1999), c'est la présence d'importantes réactions d'évaluation négative de la part des maîtres, à l'égard des façons

de parler et d'écrire d'un grand nombre de leurs élèves, qui provoque ou augmente l'insécurité linguistique, voire l'insécurité cognitive des élèves visés, ce qui n'est pas, selon lui, sans rapport avec le taux élevé d'abandon scolaire au Québec. C'est pourquoi, s'il semble essentiel de corriger les erreurs commises par les élèves dans leurs communications orales et écrites, il convient cependant pour l'enseignante ou l'enseignant de faire preuve de jugement. (Ministère de l'Éducation, 2001, 72-73).

Il va sans dire que, pour corriger adéquatement les erreurs lexicales des élèves, les enseignants doivent avoir de bonnes connaissances lexicales, leur permettant tout d'abord de repérer ces erreurs. Nous croyons de plus que des connaissances métalexicales plus approfondies, acquises en étudiant des méthodes de travail utilisées en lexicographie, pourraient aider les enseignants à corriger ces erreurs de façon plus juste et avec plus de discernement, en accord avec ce que prescrit le *Référentiel*.

Grâce à leurs capacités d'analyse métalexicale, les enseignants pourraient effectivement exposer clairement les diagnostics qu'ils posent sur les erreurs lexicales de leurs élèves, c'est-à-dire expliquer, quand l'élève n'utilise pas le mot juste, pourquoi ce n'est pas le mot juste.

À titre d'exemple, prenons la situation évoquée précédemment, mettant en jeu des collocations : pour expliquer l'erreur d'un élève qui dirait *prendre un pas* au lieu de *faire un pas*, l'enseignant pourrait expliquer le phénomène de la collocation et ne pas se contenter de dire : *On ne dit pas « prendre un pas » mais « faire un pas »*. Ce type d'intervention aiderait les élèves à comprendre leurs erreurs, puisqu'ils pourraient découvrir les structures lexicales en cause. Cela leur donnerait un sentiment de contrôle et, du coup, aurait un effet positif sur leur confiance linguistique, puisqu'ils recevraient des corrections compréhensibles et raisonnées, au lieu d'explications frustrantes telles que *Cela ne se dit pas*. L'enseignant pourrait aussi encourager l'élève à avoir recours à l'analyse de corpus — pour savoir ce qui se dit et s'écrit —, démarche dont l'ontologie fournit des instances.

Aussi, pour corriger les erreurs lexicales qu'on appelle *impropriétés*, l'enseignant pourrait s'inspirer des instances de l'ontologie et combiner des opérations lexicogra-

phiques pour élaborer une démarche d'analyse d'erreur lexicale.

En effet, il pourrait démontrer qu'une lexie n'est pas utilisée correctement en donnant premièrement une paraphrase du véritable sens de cette lexie et en remplaçant la lexie par cette paraphrase dans une phrase. Pour illustrer cette démarche, prenons l'exemple de cette erreur lexicale fictive commise par un élève dont la langue maternelle est l'espagnol : *Je suis un gars amusant qui aime beaucoup la diversion*. Dans cet exemple, la lexie DIVERSION n'est pas bien employée, et l'erreur est due à l'interférence de la langue espagnole. En effet, en espagnol, DIVERSIÓN veut dire 'plaisir, distraction'. L'élève a pensé que la lexie française DIVERSION pouvait aussi être utilisée dans ce sens.

Pour expliquer l'erreur de l'élève, l'enseignant doit premièrement donner le véritable sens de la lexie DIVERSION, en utilisant la méthode de définition par genre prochain et différences spécifiques (qui constitue une paraphrase du sens), ou en donnant un quasi-synonyme de la lexie, afin que l'élève en conceptualise le sens. Remarquons qu'on trouve dans l'ontologie plusieurs instances de ces classes d'opérations (GENUS_IDENT, DIFFERENTIA_IDENT, QSYN_IDENT), dont l'enseignant pourrait s'inspirer.

Dans le cas de DIVERSION, l'enseignant devrait donner une définition comme la suivante : 'action de détourner l'attention de quelqu'un'. Après avoir donné le véritable sens de DIVERSION, le professeur pourrait faire la substitution dans la phrase de l'élève, ce qui donnerait : *Je suis un gars amusant qui aime (l'action de) détourner l'attention (de quelqu'un)*. Grâce à cette paraphrase, l'élève se rendrait compte que DIVERSION n'est pas le terme qui convient, puisque ce qu'il veut dire, c'est qu'il aime s'amuser, se distraire, avoir du plaisir, etc. L'enseignant pourra alors en profiter pour lui fait découvrir les lexies PLAISIR, DISTRACTION, AMUSEMENT, DIVERTISSEMENT, etc.

Ce test de substitution peut être utilisé chaque fois que l'on veut vérifier si l'on utilise un bon synonyme (par exemple, lorsqu'on veut éviter la répétition dans un texte). Il y a d'ailleurs plusieurs instances de la classe REPLACEMENT_BY_QSYN_IN_EX dans l'ontologie.

Bref, pour analyser et corriger des erreurs lexicales ou pour expliquer les effets de

style que des inventions lexicales peuvent provoquer, il faut savoir identifier les différents constituants du sens des lexies, et plus on s'exerce à le faire, plus on devient habile. Voilà pourquoi nous pensons qu'un entraînement à la lexicographie pourrait aider les enseignants à accomplir la tâche de corriger les erreurs commises par les élèves dans leurs communications orales et écrites.

La dernière composante de la deuxième compétence décrite par le *Référentiel* est la suivante :

Chercher constamment à améliorer son expression orale et écrite.

Voici la description de cette composante :

(...) il existe bon nombre d'ouvrages de référence comme les grammaires et les dictionnaires qui constituent des ressources essentielles au personnel enseignant. Cependant, pour s'y référer et en faire un usage pertinent, l'enseignante ou l'enseignant doit avoir une maîtrise suffisante de sa langue. Un maître peut avoir oublié une règle ou une exception et peut se servir de ces outils pour éviter une erreur que reproduiraient les élèves. Toutefois, on ne peut douter de ce que l'on ne connaît pas. C'est pourquoi le maître doit avoir une connaissance suffisante de la langue pour faire un usage approprié de ces ressources essentielles. Le développement de cette compétence linguistique n'est cependant jamais achevé et fait partie d'un processus de formation continue. (Ministère de l'Éducation, 2001, 73).

Nous considérons que la connaissance lexicale est au cœur de la connaissance d'une langue. En effet, *la langue est un système sémiotique complexe constitué de signes qui sont, dans leur immense majorité, de nature lexicale.* (Polguère, 2004) Nous considérons également que la compétence métalinguistique fait partie de la compétence linguistique, et que, dès lors, les connaissances métalexicales que les enseignants pourront acquérir avec notre ontologie (dans sa version adaptée) et les activités qui en seront dérivées, peuvent tout à fait s'inscrire dans les compétences professionnelles que les enseignants doivent chercher à approfondir, dans l'optique d'une formation continue.

De plus, l'enseignant se doit d'être un modèle, selon le *Référentiel* :

Si la compétence linguistique s'avère un élément essentiel à bien des occupations, elle concerne tout particulièrement l'enseignement. En effet, le maître, quelle que soit sa discipline, emploie la langue dans l'ensemble des activités en vue de faciliter l'apprentissage des élèves. Il est constamment un modèle pour des apprentis. Il est donc nécessaire que les enseignantes et les enseignants, dès leur entrée dans la profession, manifestent une bonne qualité de langue écrite et orale. (Ministère de l'Éducation, 2001, 74).

Selon nous, le maître peut être un modèle de par le français qu'il utilise, mais aussi de par son attitude envers la langue. S'il adopte une attitude « investigatrice » et qu'il a l'habitude d'analyser des phénomènes linguistiques en classe pour répondre à des questions ou pour résoudre des problèmes, nous croyons qu'il donnera un bon exemple aux élèves et qu'il favorisera leur intérêt. En effet, en montrant aux élèves que n'importe quel locuteur du français peut analyser et expliquer des phénomènes de sa propre langue, grâce à une stratégie d'exploration et de modélisation, il leur donnera confiance en eux. De plus, les élèves utiliseront à leur tour ces stratégies d'analyse du lexique, et ils deviendront de plus en plus autonomes dans leur apprentissage.

Par exemple, le professeur pourrait se servir de ses connaissances lexicographiques pour analyser des effets de style ou des inventions littéraires tirés de lectures faites en classe. Des exemples d'analyses de ce type sont données dans (Polguère, 2008) et dans (Picoche, 1993). Ces analyses mobiliseraient des connaissances lexicographiques que les enseignants — surtout ceux du secondaire — pourraient acquérir directement en étudiant des ouvrages lexicographiques comme celui de Polguère (2008), bien sûr. D'ailleurs nous n'excluons pas la possibilité d'enrichir notre ontologie avec des connaissances lexicographiques extraites de textes spécialisés dans une prochaine phase du projet Lexitation. Toutefois, nous pensons qu'une adaptation de l'ontologie, consignait des instances de « bonnes démarches » de résolution de problèmes lexicographiques pourrait être particulièrement précieuse pour la formation des maîtres de langue, en vertu du caractère très détaillé de ces instances, et de la grande variété de tâches et d'opérations pouvant y être répertoriées.

Nous rappelons que notre but n'est pas de faire de l'enseignant un lexicographe pro-

fessionnel. Mais l'enseignant se doit tout de même de dépasser la connaissance « naïve » d'un locuteur qui utilise la langue sans réfléchir à son fonctionnement. Plus tôt, dans la section « Sens de la compétence », le *Référentiel* affirme que les connaissances linguistiques des enseignants ne doivent pas être « calquées » sur celles des spécialistes tels que les linguistes ou les lexicographes, mais qu'elle doit dépasser celle du locuteur moyen vu la mission d'éducation que l'enseignant accomplit :

Précisons que, dans le cas d'une enseignante ou d'un enseignant, un usage professionnel de la langue ne signifie pas la même chose que la connaissance manifestée par un professionnel de la langue, comme celle de l'écrivain, du spécialiste de la linguistique ou de la grammaire. La compétence linguistique de l'enseignante ou de l'enseignant n'a pas à être calquée sur celle de ces spécialistes, mais elle doit dépasser celle qui est communément partagée par tous les citoyens et comporter des usages précis qui la distinguent de celle des autres professions. L'initiation des jeunes à la langue et à la culture donne au travail enseignant un caractère particulier et spécialisé à cet égard. (Ministère de l'Éducation, 2001, 70).

Nous sommes tout à fait d'accord avec cette affirmation, et, comme nous l'avons mentionné plus tôt, il n'est pas question de faire des enseignants des spécialistes de la lexicographie. Notre objectif est plutôt de leur donner les connaissances métalinguistiques nécessaires et suffisantes à l'enseignement efficace de la langue, le lexique devant retrouver la part qui lui revient dans l'enseignement primaire et secondaire.

Finalement, une dernière compétence sera mobilisée par notre projet : celle de savoir créer des situations d'apprentissage. Voici l'énoncé de cette compétence :

Concevoir des situations d'enseignement-apprentissage pour les contenus à faire apprendre, et ce, en fonction des élèves concernés et du développement des compétences visées dans le programme de formation. (Ministère de l'Éducation, 2001, 75).

Les enseignants pourront effectivement se servir de leurs compétences nouvellement acquises en lexicographie pour créer des situations d'apprentissage à l'aide de la version

« vulgarisée » de l'ontologie. Cela n'est pas une demande irréaliste ou excessive : créer des situations d'apprentissage est l'une des compétences que tout enseignant doit avoir, tel que le stipule le *Référentiel* :

La conception de situations d'enseignement-apprentissage constitue l'un de (...) ces éléments les plus fondamentaux (du métier d'enseignant). (Ministère de l'Éducation, 2001, 75).

Le type d'activités que l'enseignant pourrait élaborer à l'aide de l'ontologie, comme faire faire du travail de type lexicographique aux élèves (par exemple, trouver eux-mêmes les différentes acceptions d'un vocable à l'aide d'un corpus) est parfaitement adapté à l'approche par compétences proposée par le ministère de l'Éducation du Québec :

L'approche par compétences valorise le processus menant aux apprentissages, l'intégration des apprentissages ainsi que le développement d'habiletés intellectuelles complexes. Cela sous-entend que, à la différence de situations où c'est le maître qui est actif et qui traite avec les objets d'apprentissage, les situations d'apprentissage significantes et ouvertes engagent les élèves dans des démarches qui mobilisent leurs représentations et leurs acquis antérieurs, leur en font percevoir les limites et les obligent à manipuler et à maîtriser de nouvelles connaissances pour repérer et résoudre un problème, accomplir une tâche ou réaliser une production ou un projet. (Ministère de l'Éducation, 2001, 75).

Bref, nous pensons que les activités qu'il sera possible d'inventer avec l'ontologie permettront de voir des notions qui sont déjà au programme scolaire, et que ces activités s'intégreront bien dans le modèle pédagogique actuel, dans lequel l'élève est actif dans son apprentissage.

5.3 Conclusion

Nous terminerons ce mémoire en rappelant d'abord l'importance d'un enseignement adéquat du vocabulaire dans la formation scolaire. Celui-ci est primordial, ne serait-ce

que par l'impact indirect mais non négligeable qu'il a sur l'ensemble de la formation scolaire. En effet, le vocabulaire est la base de tout apprentissage subséquent, quelle que soit la matière, et *l'on sait l'étendue des connaissances lexicales nécessaires, par exemple, pour la compréhension des tâches scolaires* (De Pietro, 2003).

Au final, nous espérons avoir montré, dans un premier temps, qu'une formation plus poussée en lexicographie serait utile pour les enseignants, parce qu'elle les rendrait plus autonomes, et, dans un deuxième temps, que le développement d'une ontologie des savoirs lexicographiques est possible, mais que ce développement n'est pas encore terminé. De la même façon, nous espérons avoir montré qu'on peut exploiter le potentiel d'une telle ontologie en didactique des langues. En effet, nous pensons qu'une ontologie des savoirs lexicographiques peut servir à développer des activités pédagogiques qui pourront aider les enseignants à acquérir des compétences métalexicales, mais aussi des activités qui seront élaborées expressément pour les élèves, activités qui traiteront de notions qui sont au programme scolaire, mais que les enseignants, à l'heure actuelle, ne savent pas toujours comment intégrer à leur enseignement, et qui susciteront chez les élèves de l'intérêt pour le lexique grâce à l'approche investigatrice et active qu'elles permettent, qui rend tout domaine du savoir — fût-il aussi vaste que le lexique d'une langue— accessible.

BIBLIOGRAPHIE

- L. Allal, D. Béatrix Köhler, L. Rieben, Y. Rouiller, M. Saada-Robert et E. Wegmuller. *Apprendre l'orthographe en produisant des textes*. Fribourg : Éditions universitaires, 2001.
- Aristote. *Organon V. Les Topiques*. Paris : Librairie J. Vrin, 1997.
- B. T. S. Atkins. Then and now : Competence and performance in 35 years of lexicography. Dans Thierry Fontenelle, dir., *Practical Lexicography, a reader*, pages 247–272. Oxford : Oxford University press, 2008a.
- B. T. S. Atkins. Theoretical lexicography and its relation to dictionary-making. Dans Thierry Fontenelle, dir., *Practical Lexicography, a reader*, pages 31–50. Oxford : Oxford University press, 2008b.
- Collin F. Baker, Charles J. Fillmore et Beau Cronin. The structure of the framenet database. *International Journal of Lexicography*, 16.3:281–296, 2003.
- H. Béjoint. *Modern Lexicography : an Introduction*. Oxford : Oxford University Press, 2000.
- J. Boisvert. *La formation de la pensée critique : théorie et pratique*. École en mouvement. Saint-Laurent : Erpi, 1999.
- J. Bourdeau, R. Mizoguchi, V. Psyché et R. Nkambou. Selecting theories in an ontology-based its authoring environment. Dans J. Lester, R. Vicari et F. Paraguaçu, dir., *Intelligent Tutoring Systems, Lecture Notes in Computer Science*, pages 150–161. Heidelberg : Springer, 2004.
- J. L. Branch. Investigating the information-seeking processes of adolescents : The value of using think alouds and think afters. *Library and Information Science Research*, 22 (4):371–392, 2000.

- J. Cauchon. La structure prédicative du verbe dans l'apprentissage de la grammaire de la phrase et du texte. Mémoire de maîtrise, Université de Montréal, 2003.
- S. Comeau. Sharing the knowledge of lexicographers : Methodology for the extraction of lexicographic abilities. Dans *Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'09)*. MTT, 2009.
- S. Comeau et A. Polguère. An experimental approach to the extraction of lexicographic abilities. Dans *Actes du 32e Colloque annuel de l'Association de linguistique des provinces atlantiques*. ALPA, 2008.
- J. Dancette et C. Réthoré. *Dictionnaire analytique de la distribution*. Montréal : Presses de l'Université de Montréal, 2000.
- J. David et F. Grossmann. Enseigner-apprendre le lexique, présentation du dossier. *Lettre de l'AIRDF*, 33:3–5, 2003.
- J.-F. De Pietro. L'enseignement du lexique en suisse ou comment en finir avec les listes à mémoriser ? *Lettre de l'AIRDF*, 33:12–18, 2003.
- A. K. Ericsson et H. A. Simon. *Protocol Analysis : Verbal Reports As Data*. Cambridge : MIT Press, 1984.
- A. K. Ericsson et H. A. Simon. *Protocol Analysis : Verbal Reports As Data*. Cambridge : MIT Press, 1993.
- C. J. Fillmore. Frame semantics and the nature of language. Dans *Annals of the New York Academy of Sciences : Conference on the Origin and Development of Language and Speech*, 1976.
- T. Fontenelle, dir. *Practical lexicography, a reader*. Oxford : Oxford University press, 2008.
- Y. Gentilhomme. Panorama sur le *Dictionnaire Explicatif et Combinatoire* : retombées pédagogiques. Dans Igor Mel'čuk *et al.*, dir., *DEC III*, pages 95–120. Montréal : Presses de l'Université de Montréal, 1992.

- A. Gomez-Pérez. Développements récents en matière de conception, de maintenance et d'utilisation des ontologies. *Terminologies nouvelles*, 19:9–20, 1999.
- P. Hanks. Do word meanings exist ? Dans Thierry Fontenelle, dir., *Practical Lexicography, a reader*, pages 125–134. Oxford University press, 2008.
- Paul Imbs. *Trésor de la langue française : dictionnaire de la langue du XIXe et du XXe siècle (1789-1960)*. Paris : Éditions du Centre national de la recherche scientifique, 1971-1994.
- A. Kilgarriff. I don't believe in word senses. Dans Thierry Fontenelle, dir., *Practical Lexicography, a reader*, pages 135–152. Oxford University press, 2008.
- S. L. Landau. *Dictionaries, The Art and Craft of Lexicography, Second Edition*. Cambridge : Cambridge University Press, 2001.
- R. Leed et A. Nakhimovsky. Meaning-text theory : Linguistics, lexicography, and implications. Chapitre Lexical Functions and Language Learning., pages 365–374. University of Ottawa Press, 1990.
- M.-C. L'Homme. Fonctions lexicales pour représenter les relations sémantiques entre termes. *Traitement automatique de la langue*, 43(1):19–41, 2002.
- I. Mel'čuk, A. Clas et A. Polguère. *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot, 1995.
- I. Mel'čuk, A. Clas, A. Polguère, S. Mantha et N. Arbatchewsky. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques (Volumes I à IV)*. Presses de l'Université de Montréal, 1984, 1988, 1992, 1999.
- I. Mel'čuk et A. Polguère. *Lexique Actif du Français*. Bruxelles : De Boeck et Larcier, 2007.
- Ministère de l'Éducation. *La formation à l'enseignement, les orientations, les compétences professionnelles*. Gouvernement du Québec, 2001.

- Ministère de l'Éducation. *Programme de formation de l'école québécoise. Enseignement primaire*. Gouvernement du Québec, 2006a.
- Ministère de l'Éducation. *Programme de formation de l'école québécoise. Enseignement secondaire, premier cycle*. Gouvernement du Québec, 2006b.
- Ministère de l'Éducation. *Programme de formation de l'école québécois ; Progression des apprentissages au primaire*. Gouvernement du Québec, août 2009a.
- Ministère de l'Éducation. *Programme de formation de l'école québécoise. Enseignement secondaire, deuxième cycle*. Gouvernement du Québec, 2009b.
- N. F. Noy et D. L. McGuinness. *Ontology development 101 : A guide to creating your first ontology*. Knowledge Systems Laboratory Technical Report and Medical Informatics Technical Report KSL-01-05 and SMI-2001-0880, Stanford University, 2001.
- G. Paquette. *Modélisation des connaissances et des compétences*. Montréal : Presses de l'Université du Québec, 2002.
- J. Picoche. *Didactique du vocabulaire français*. Paris : Nathan, 1993.
- A. Polguère. Une base de données lexicales du français et ses applications possibles en didactique. *Revue de Linguistique et de Didactique des Langues (LIDIL)*, 21:75–97, 2000.
- A. Polguère. Savoir consulter un dictionnaire, c'est bien... savoir comment on le construit, c'est mieux. *Québec français*, 134:65–67, 2004.
- A. Polguère. *Descriptif du projet lexication*. Non publié, 2007.
- A. Polguère. *Lexicologie et sémantique lexicale. Notions fondamentales*. Montréal : Presses de l'Université de Montréal, 2008.
- A. Polguère et O. Tremblay. Qu'y a-t-il à l'intérieur de noix ? ou comment décortiquer les unités lexicales. *Lettre de l'AIRDF*, 33:27–30, 2003.

- V. Psyché, O. Mendes et J. Bourdeau. Apport de l'ingénierie ontologique aux environnements de formation à distance. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 11:89–126, 2003.
- B. Quemada. Notes sur lexicographie et dictionnaire. *Cahiers de lexicologie*, 51 (2): 229–242, 1987.
- J. E. Russo, E. J. Johnson et D. L. Stephens. The validity of verbal protocols. *Memory and Cognition*, 17(6):759–769, 1989.
- C. Simard. Pour un enseignement plus systématique du lexique. *Québec français*, 92: 65–67, 1994.
- J. Steele. Meaning-text theory and first-language learning : Some implications. Dans J. Steele, dir., *Meaning-Text Theory : Linguistics, Lexicography, and Implications*, pages 377–389. Ottawa/London/Paris : University of Ottawa Press, 1990.
- Della Summers, dir. *Longman Language Activator*. Longman, Harlow, 1993.
- O. Tremblay. Une approche structurée de l'enseignement/apprentissage du lexique en français langue maternelle basée sur la lexicologie explicative et combinatoire. Mémoire de maîtrise, Université de Montréal, 2003.
- O. Tremblay. Pour une approche structurée de l'enseignement-apprentissage du lexique. Dans E. Calaque et J. David, dir., *Didactique du lexique : contextes, démarches, supports*. Bruxelles : De Boeck Université, 2004.
- O. Tremblay. *Une ontologie des savoirs lexicologiques pour l'élaboration d'un module de cours en didactique du lexique*. Thèse de doctorat, Université de Montréal, 2009.
- A. Wierzbicka. *Lexicography and Conceptual Analysis*. Ann Arbor : Karoma, 1985.
- A. Wierzbicka. *English Speech Act Verbs, A semantic Dictionary*. Sydney : Academic Press, 1987.

J. M. Wolfe, S. J. Butcher, C. Lee et M. Hyle. Changing your mind : On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology : Human Perception and Performance*, 29(2):483–502, 2003.

L. Zgusta. *Manual of Lexicography*. The Hague : Mouton, 1971.

Annexe I

Modèle de fiche lexicographique

The image shows a screenshot of a FileMaker Pro database form titled "Structure polysémique". The form is displayed in a window titled "StructPolyExp3". The interface includes a menu bar (File, Edit, View, Insert, Format, Records, Scripts, Window, Help) and a toolbar with icons for New, Delete, and Find. The form itself has a header section with "View Research Notes Report" and a sub-header with "New", "Delete", and "Find" buttons. The main content area contains several text input fields: "Lexie :", "Carac. gram. :", "Indices :", "Exemples :", and "Remarques :". At the bottom of the form, there is a "Date" field with the value "6/26/2008". On the left side, there is a sidebar with "Browse" and "Layout" sections, and a "Record" section showing "Total 2" and "Unsorted".

Figure I.1 – Modèle de fiche lexicographique utilisé lors de l'Expérimentation 3