

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MISE EN PLACE D'UNE TÂCHE D'ANALYSE DE COOCCURENCES
POUR GUIDER L'USAGE DES PRÉPOSITIONS
CHEZ DES ÉTUDIANTS ADULTES EN CONTEXTE DE FRANCISATION

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN DIDACTIQUE DES LANGUES

PAR
MATHIEU MÉNARD

DÉCEMBRE 2017

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.07-2011). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Merci à Simon Collin, mon directeur de recherche, pour ses suggestions éclairées et pour son accompagnement tout au long du processus de recherche et de rédaction. Merci à Véronique Fortier et à Tom Cobb pour leurs conseils et pour leur généreux partage de ressources; le prototype conçu à l'occasion de ce projet de recherche s'appuie sur plusieurs outils provenant du site Compleat Lexical Tutor, accessible à l'adresse <http://www.lextutor.ca>.

Merci à Melissa Dubuc pour son étroite collaboration, qui a permis le développement du prototype ainsi que la cueillette de données essentielle à cette recherche.

Finalement, merci à Marie Dupuis et à Michel Ménard pour leur appui, et un grand merci à Katy Boyer-Gaboriault pour son soutien durant la maîtrise et pour son encouragement durant les moments ardu.

TABLE DES MATIÈRES

REMERCIEMENTS.....	ii
LISTE DES FIGURES.....	v
LISTE DES TABLEAUX.....	vi
RÉSUMÉ.....	vii
INTRODUCTION.....	1
CHAPITRE I	
PROBLÉMATIQUE.....	3
1.1 Apports de la technologie en classe de langue.....	3
1.2 Apports des approches guidées par les données en classe de langue.....	5
1.3 Limites des approches guidées par les données en classe de langue	10
1.4 Limites de l'avancement des connaissances sur les approches guidées par les données en classe de langue.....	13
1.5 Introduction des prépositions dans l'apprentissage du français.....	16
1.6 Questions de recherche	19
CHAPITRE II	
CADRE THÉORIQUE	21
2.1 Corpus linguistique	21
2.2 Apprentissage guidé par les données	23
2.4 Synthèse des recherches sur l'apprentissage guidé par les données en lien avec la langue française et avec les apprenants de niveau débutant.....	25
2.5 Objectifs spécifiques	28
CHAPITRE III	
MÉTHODOLOGIE.....	30
3.1 Description des participants	30
3.2 Instrumentation	33
3.3 Développement des corpus	35
3.4 Comparaison du contenu des deux corpus.....	39
3.5 Développement de l'outil.....	41

3.6 Enregistrement des actions.....	47
3.7 Développement du questionnaire.....	50
3.8 Déroulement de la collecte de données.....	51
3.9 Considérations éthiques	55
3.10 Analyse des données	56
CHAPITRE IV	
RÉSULTATS	58
4.1 Profil des participants.....	58
4.2 Résultats de l'exercice sur les prépositions.....	63
4.3 Perceptions du dispositif.....	66
4.4 Enregistrement des actions dans l'exercice.....	71
CHAPITRE V	
DISCUSSION.....	73
5.1 Profil des participants.....	73
5.2 Résultats de l'exercice sur les prépositions.....	74
5.3 Perceptions des participants	74
5.4 Enregistrement des actions dans l'exercice.....	77
CHAPITRE VI	
CONCLUSION.....	79
6.1 Objectifs du projet de recherche	79
6.2 Limites envisagées	81
6.3 Pistes de recherche futures	82
APPENDICE A	
LETTRE ENVOYÉE LORS DE LA PREMIÈRE PRISE DE CONTACT	85
APPENDICE B	
FORMULAIRE DE CONSENTEMENT REMIS À CHAQUE PARTICIPANT	87
APPENDICE C	
QUESTIONNAIRE COMPLÉTÉ PAR LES PARTICIPANTS	93
APPENDICE D	
RÉSULTATS À L'EXERCICE SUR LES PRÉPOSITIONS	97
BIBLIOGRAPHIE	98

LISTE DES FIGURES

Figure		Page
3.1	Capture d'écran de la page d'accueil de la bibliothèque d'ouvrages en langue française qui sont disponibles sur Logos Library	35
3.2	Capture d'écran de la page d'accueil de l'encyclopédie Vikidia.....	36
3.3	Phrase trouée à compléter dans l'exercice sur les prépositions	41
3.4	Liste d'exemples pour le mot-clé « vacances », qui s'affiche à l'écran après avoir appuyé sur le bouton correspondant.....	44
3.5	Extrait du questionnaire bilingue présenté aux participants.....	46
3.6	Données de Google Analytics (affichage des listes d'exemples).....	49
4.1	Préférence pour l'apprentissage des règles.....	59
4.2	Préférence pour l'apprentissage de langue à l'ordinateur	61
4.3	Réponses correctes dans l'exercice sur les prépositions.....	65
4.4	Facilité de lecture des exemples.....	67
4.5	Intérêt exprimé pour les exemples choisis	68
4.6	Souhait d'utiliser le logiciel dans le futur.....	69
4.7	Recommandation de l'outil.....	70
4.8	Clics enregistrés pour afficher les exemples (deux séries d'exemples par question dans chaque exercice).....	72

LISTE DES TABLEAUX

Tableau		Page
3.1	Caractéristiques des deux groupes de participants.....	31
3.2	Caractéristiques des extraits de corpus.....	40
4.1	Utilisation de l'ordinateur dans différents milieux.....	62
4.2	Résultats globaux des deux groupes pour l'exercice.....	64
D.1	Résultats (moyenne) des deux groupes pour des éléments spécifiques ...	97

RÉSUMÉ

L'apprentissage guidé par les données (une approche qu'on pourrait identifier par l'acronyme AGD) et la consultation de corpus linguistique, en fournissant aux apprenants des extraits de textes authentiques mettant en évidence des régularités dans la langue, peuvent aider les apprenants à mieux comprendre le fonctionnement d'aspects de la langue cible, notamment pour ceux où la langue maternelle peut interférer. Pour les apprenants du français langue seconde, les études suggèrent que l'utilisation des prépositions s'avère un sujet propice à une démarche d'AGD. Tandis que plusieurs recherches impliquent des étudiants avancés en milieu universitaire, ce projet de recherche a recruté des participants adultes provenant d'un programme de francisation dans un centre de formation continue, afin d'évaluer si certains facteurs identifiés dans la littérature (choix du contenu de corpus à l'avance, simplification de l'interface de l'outil) peuvent contribuer à rendre cette démarche accessible aux apprenants de niveau débutant. Les participants, séparés en deux sous-groupes, ont complété une tâche où ils devaient inscrire des prépositions dans des phrases trouées, en s'appuyant sur des phrases données en exemple. Autant pour les phrases trouées que pour les exemples, chaque groupe avait accès à du matériel provenant d'un corpus différent, l'un rassemblant des textes littéraires libres de droit, l'autre étant composé d'articles provenant d'une encyclopédie libre en ligne destinée à des apprenants âgés de huit à treize ans. Dans une perspective exploratoire, ce deuxième corpus a été développé au stade de prototype, afin d'explorer le potentiel de sources libres pour créer d'autres ressources pour les apprenants du français. L'absence de différence significative entre les deux groupes pour la réalisation de la tâche, offre une perspective intéressante pour développer des ressources à partir d'encyclopédies libres. Par ailleurs, les données récoltées dans un questionnaire sur l'appréciation du dispositif, de même que les données d'utilisation du dispositif de phrases-exemples, suggèrent que d'autres recherches seraient nécessaires pour continuer à améliorer les outils d'AGD, et pour mieux les adapter aux besoins des apprenants de niveau débutant ou intermédiaire inférieur.

MOTS-CLÉS : apprentissage guidé par les données (*data-driven learning*), corpus linguistiques (*language corpora*), apprenants débutants ou intermédiaires inférieurs (*beginner or lower intermediate learners*), français langue seconde (*French as a second language*).

INTRODUCTION

Le corpus linguistique, en rassemblant à l'écrit des productions langagières réelles dans une base de données (Golonka *et al*, 2014), est à la fois un sujet de recherche linguistique et un outil pédagogique en langue seconde depuis la fin des années 1980 (Yoon, 2011). Il contribue à guider autant l'enseignement que l'apprentissage des langues. Pour l'enseignant, le corpus peut être employé comme une banque d'exemples authentiques (Braun, 2005) afin d'éviter l'aspect artificiel et les difficultés d'interprétation sémantique des exemples inventés (Römer, 2004). Les corpus linguistiques plus généralistes, conçus pour refléter l'usage d'une langue donnée à un moment précis, peuvent fournir des listes des mots les plus fréquemment employés et guider les enseignants dans le choix du vocabulaire et des expressions à enseigner (Braun, 2005; Yoon et Hirvela, 2004).

Pour l'apprenant, travailler avec des exemples authentiques peut être plus motivant (Gordani, 2013). À partir du matériel trouvé dans le corpus linguistique, l'apprenant peut découvrir des formulations différentes qui complètent ou contredisent les ouvrages de référence traditionnels (Aston, 1997), menant potentiellement à une compréhension plus fine de la langue, dans la mesure où les exemples trouvés permettent aux apprenants de réviser les erreurs qui se sont fossilisées à partir de leur expérience limitée avec la langue cible (Sun et Wang, 2003). L'exploration de corpus peut aussi aider à développer l'autonomie chez l'apprenant, en lui permettant de se concentrer sur des expressions et des types de texte qui l'intéressent (Gordani, 2013). En situation d'écriture, la consultation de corpus linguistique peut aider les apprenants à utiliser correctement certaines formules où la langue maternelle pourrait interférer, par exemple les cooccurrences verbe-préposition et les expressions figées (O'Sullivan et Chambers, 2006).

Étant donné le potentiel du corpus linguistique comme outil en classe de langue, et à la lumière des recherches effectuées dans le domaine de l'apprentissage guidé par les données (*data-driven learning*), il apparaît pertinent d'explorer certains des facteurs qui pourraient contribuer à rendre cet outil accessible à un plus grand nombre d'apprenants. Dans cette optique, ce projet de recherche s'intéresse à l'utilisation de deux corpus linguistiques dans un contexte d'apprentissage du français en tant que langue seconde, dans le cadre d'un programme de francisation pour nouveaux arrivants.

CHAPITRE I

PROBLÉMATIQUE

Dans un premier temps, un survol de certains apports récents de la technologie en classe de langue sera présenté. Il sera suivi d'une description des apports des approches guidées par les données en classe de langue, en lien avec l'utilisation de corpus linguistiques. En contrepoint à ce compte-rendu des apports de cette approche, certaines limites associées à l'utilisation de corpus linguistiques, ainsi que des pistes de recherche qui pourraient contribuer à l'avancement des connaissances dans ce domaine, seront détaillées. Par la suite, une section consacrée à l'introduction des prépositions dans l'apprentissage du français contribuera à mettre en contexte le choix de cet élément dans la question de recherche.

1.1 Apports de la technologie en classe de langue

L'intégration de différents logiciels et outils technologiques dans l'apprentissage reflète non seulement la place importante qu'occupe la technologie dans le milieu scolaire (Breton, 2016), mais également le fait que la maîtrise des technologies d'information et de communication fait partie des objectifs énumérés dans les programmes de formation au Québec (Ministère de l'Éducation, du Loisir et du Sport, 2007). Au fur et à mesure que différentes technologies ont été intégrées en éducation, des recherches ont contribué à évaluer leur impact et leur utilité.

L'étude de Golonka et collègues (2014), avec sa vue d'ensemble des effets de l'intégration de technologie en contexte d'apprentissage de langue étrangère, illustre

la diversification de logiciels et de technologies qui sont élaborés pour les apprenants. En général, l'intégration de technologie contribue à la motivation des étudiants : « les apprenants ont tendance à être davantage engagés dans le processus d'apprentissage, et à avoir une attitude plus positive vis-à-vis de l'apprentissage [notre traduction] » (Golonka *et al*, 2014, p.92). Certains de ces outils, comme les plates-formes de gestion de cours (*course/learning management system*) et le dictionnaire électronique, contribuent à personnaliser l'apprentissage et à développer l'autonomie chez les étudiants. D'autres outils, comme les systèmes de clavardage et les logiciels de reconnaissance orale et d'assistance à la prononciation, peuvent aider les étudiants à améliorer leurs compétences en communication écrite et orale dans un contexte moins anxiogène. Toutefois, certaines approches, comme l'utilisation de forums de discussion, de réseaux sociaux, ou de systèmes d'écriture collaborative, paraissent moins fructueuses, puisqu'elles tendent à impliquer les participants de façon inégale.

Selon Landure (2011), l'introduction de technologie en salle de classe « pourrait contribuer à changer en profondeur la façon de penser et de concevoir l'enseignement » (p.164). Comme les logiciels visent à développer l'autonomie des apprenants, le professeur adopte le rôle « de facilitateur, de guide ou de conseiller » (p.164), plutôt que de transmettre directement le savoir. Néanmoins, Landure rappelle que cette réorganisation des rôles implique également l'étudiant et « [qu'] une démarche raisonnée d'apprentissage autonome [...] ne se réalisera qu'avec l'acceptation de l'apprenant de changer de rôle » (p.176).

Pour résumer, certaines des récentes intégrations technologiques en apprentissage de langue tendent à changer la dynamique entre enseignants et apprenants; dans la mesure où ce dernier accepte de prendre un rôle plus actif dans son apprentissage, la technologie peut l'aider à développer son autonomie. Certains outils, comme le clavardage et la reconnaissance vocale, peuvent encourager le

développement de l'écriture et de l'expression orale. De façon générale, on remarque que l'introduction de technologie en salle de classe tend à hausser la motivation des étudiants, même si concrètement certaines de ces initiatives ne semblent pas avoir des effets aussi marqués sur leur performance.

1.2 Apports des approches guidées par les données en classe de langue

Depuis environ vingt ans, des chercheurs en didactique des langues (Boulton, 2009b; O'Sullivan et Chambers, 2006; Gaskell et Cobb, 2004; Chambers et O'Sullivan, 2004; Kennedy et Miceli, 2001; Cobb, 1997) rendent directement accessible aux apprenants des outils permettant d'investiguer le corpus linguistique, principalement des logiciels permettant d'énumérer les cooccurrences pour un mot-clef donné. Selon O'Sullivan et Chambers (2006), cette interface permet aux apprenants de remarquer les régularités dans les expressions et dans la grammaire afin d'améliorer leurs compétences langagières, « en mettant des patrons langagiers récurrents en évidence de façon contextuelle, et en aidant les apprenants à internaliser les schémas dont ils ont besoin pour connaître du succès comme apprenant et comme utilisateur de langues [notre traduction] » (p.52). Au fil des années, cette approche, combinant le contenu du corpus linguistique aux listes de cooccurrences élaborées à partir de l'analyse de ce dernier, a été employée afin de soutenir plusieurs aspects de l'apprentissage, dont le développement du vocabulaire (Gordani, 2013), la langue académique (Yoon, 2011; Yoon et Hirvela, 2004), la compréhension de la conjugaison (Hunston et Francis, 1998), et la révision des productions écrites (Chambers et O'Sullivan, 2004; Gaskell et Cobb, 2004; Todd, 2001).

Selon Boulton (2009a), ce processus « peut être motivateur et centré sur l'apprenant, [...] et potentiellement augmenter l'autonomie pour l'apprentissage

continu [notre traduction] » (paragr. 5). Dans le contexte de tâches de rédaction pour des scripteurs en langue seconde, la consultation de corpus contribuerait à rendre saillantes les expressions propres à un type de texte donné (Yoon, 2011), pour aider ces derniers à exprimer leurs idées sans obstacles linguistiques, en répondant à ce que Chambers (2010) identifie comme « les normes d'une certaine communauté discursive » (p.11). En exposant l'apprenant à plusieurs exemples pertinents, l'AGD cherche à « trouver un moyen d'accélérer le passage des connaissances acquises par la lecture à l'activité d'écriture » (Chambers, 2010, p. 10), dans une situation d'apprentissage multicontextuel (Cobb, 1997). Cette situation permet de « découvrir des patrons communs et typiques d'usage des mots dans des styles et genres variés [...] et d'identifier les facteurs contextuels qui peuvent influencer la variation de ces patrons [notre traduction] » (Chang et Sun, 2009, p. 284). Selon O'Sullivan et Chambers (2006), la mise en évidence de patrons langagiers récurrents contribuerait à « aider les apprenants langagiers à internaliser les schémas dont ils ont besoin [notre traduction] » (p.52), et l'accumulation d'exemples contribuerait à rendre le patron plus mémorable que s'il avait été présenté dans un dictionnaire ou dans une grammaire. D'après les recherches de Hunston et Francis (1998), l'analyse de patrons langagier récurrents (combinaisons verbe-nom, verbe-nom-groupe prépositionnel, etc.) permettrait de rassembler des verbes sémantiquement similaires, ce qui pourrait aider les apprenants à découvrir et à mémoriser le sens de ces verbes.

En comparaison au dictionnaire où à la grammaire, le corpus possède l'avantage de présenter les mots en contexte dans une plus grande variété d'exemples. Tandis que la signification du mot et les règles encadrant son utilisation sont explicitement détaillées dans les ouvrages de référence, dans le corpus ces informations sont implicites – l'apprenant doit faire appel à ses connaissances antérieures et au contexte des phrases extraites du corpus. Néanmoins, la consultation du corpus peut contribuer à ajuster le choix de mots, à condition de trouver des

exemples pertinents dans le corpus et de les interpréter correctement. Ainsi, lorsque l'étudiant n'a pas encore une connaissance approfondie de la langue étudiée, il est possible que sa langue maternelle interfère dans le choix de mots, comme le remarquent O'Sullivan et Chambers (2006) dans leur analyse de dissertations produites par des apprenants du français : « dans la catégorie du choix des mots, 17 des 28 erreurs semble provenir d'une interférence de la langue maternelle [notre traduction] » (p.60).

Chambers et O'Sullivan (2004) avaient aussi remarqué dans une recherche précédente que la consultation du corpus pouvait favoriser l'emploi approprié d'expressions idiomatiques, « que les étudiants auraient eu de la difficulté à produire en consultant un dictionnaire, un grammaire ou un manuel de cours [notre traduction] » (p.170). Cette tendance se remarque également dans la recherche de Landure et Boulton (2010), où les modifications positives suite à la consultation d'un corpus « port[aient] principalement sur les expressions idiomatiques et les unités lexicales composées qui sont difficilement repérables dans les dictionnaires ou traducteurs en ligne et que les étudiants ont souvent tendance à traduire littéralement » (p.20). De cette façon, l'utilisation du corpus peut évoluer au-delà de l'acquisition de vocabulaire et de compréhension de la grammaire, permettant de comprendre des expressions plus complexes, pour éventuellement « pousser les études vers le niveau discursif, en se concentrant sur le texte et sur les caractéristiques de genre, de position, etc. [notre traduction] » (Cobb et Boulton, 2015, p.495).

D'après Hadley (2002), l'AGD représente une piste de solution potentielle pour réintroduire des notions grammaticales dans un programme suivant l'approche communicative. Pour des étudiants qui préfèrent la communication avec des locuteurs natifs et qui tendent à trouver les règles grammaticales traditionnelles abstraites ou incomplètes, l'exploration d'un corpus, avec ses productions langagières authentiques,

peut s'avérer plus stimulante que l'apprentissage de phrases créées pour illustrer des règles grammaticales et parfois jugées comme artificielles. Par ailleurs, la consultation d'un corpus bilingue, comme dans la recherche de Johns et collègues (2008), peut contribuer à « libérer les étudiants de la traditionnelle traduction mot-à-mot [notre traduction] » (p.486). Le projet de Chujo et collègues (2006) a également fait appel à un corpus bilingue, afin de rendre le contenu plus accessible aux apprenants de niveau débutant. Dans ces deux cas, l'exploration de ce corpus peut aider les étudiants à prendre conscience des différences stylistiques et structurelles qui varient entre la langue maternelle et la langue cible, en étant exposé à plusieurs instances d'une même expression traduite ou d'un terme polysémique.

Selon Yoon (2011), l'utilisation de corpus peut s'avérer particulièrement utile dans une tâche d'écriture avec des apprenants en contexte de langue seconde, surtout si ledit corpus se spécialise dans la compilation d'un type de composition écrite : « les apprenants peuvent découvrir du vocabulaire, des combinaisons de mots et des patrons grammaticaux qui caractérisent ces genres et ces disciplines [notre traduction] » (p.131). Le corpus peut également devenir un ouvrage de référence dans des tâches de traduction ou de correction d'erreurs (Boulton, 2007a). Plus généralement, l'AGD semble pertinent dans les tâches qui développent la compréhension de la langue et la résolution de problèmes en écriture (Yoon, 2011). Elle permet à l'apprenant « de trouver de multiples exemples d'expressions dans un contexte pertinent » (Chambers, 2010, p. 17), pour s'assurer que sa rédaction répond aux normes discursives propres au type de texte rédigé. Selon les recherches de Chang et Sun (2009), la consultation guidée de corpus peut également contribuer à améliorer la révision chez les apprenants. Pour ces différentes activités, Yoon (2011) rappelle que plusieurs facteurs peuvent influencer la motivation des apprenants pour l'AGD ainsi que l'impact de cette approche, comme les préférences des participants en tant qu'apprenants (démarche indépendante ou activité dirigée par le professeur),

le temps consacré à la formation des participants pour utiliser l'outil d'exploration de corpus et la facilité d'utilisation de cet dernier, ainsi que l'intérêt des participants pour le contenu du corpus.

Au-delà des variations évoquées ci-dessus, comme le style d'apprentissage des étudiants ainsi que les approches employées pour familiariser ceux-ci à différents outils, l'apport des technologies d'information et de communication pour l'apprentissage linguistique est perçu de façon positive à plusieurs égards (Grgurovic, Chapelle et Shelley, 2013), pour hausser la motivation des apprenants et pour promouvoir le développement de l'autonomie, par exemple. À cet effet, Golonka et collègues (2014) indiquent dans leur revue de littérature que, selon les données compilées par plusieurs chercheurs en ayant recours à des questionnaires et à des entrevues, l'apport de la technologie au cours de langue étrangère tend à motiver les apprenants et à favoriser une attitude plus positive vis-à-vis du processus d'apprentissage en général. Gaskell et Cobb (2004) rappellent que même si la séquence générale d'apprentissage pour une langue et le niveau de l'apprenant sont connus, il n'est pas garanti qu'il y aura rétention des structures enseignées en classe. En proposant des logiciels où l'apprenant peut évoluer de façon autonome, les chercheurs en apprentissage linguistique soutenu par ordinateur espèrent offrir à l'élève un soutien plus personnalisé.

Dans cette optique d'utilisation de technologie pour l'apprentissage des langues, l'exploration de cooccurrences dans le corpus linguistique, que ce soit par l'analyse de phrases exemples ou de mots-clefs en contexte (« *keyword in context* », ou KWIC), semble une piste de recherche intéressante. Plusieurs auteurs (Braun, 2005; Chambers, 2005; Yoon et Hirvela, 2004; John, 2001) sont d'avis que la consultation de corpus linguistiques est compatible avec une approche favorisant l'autonomie et la découverte. Elle offre aux apprenants une banque d'exemples

naturels et authentiques, dont l'analyse permettrait de développer à la fois les connaissances lexicales et grammaticales. Pour John (2001), l'exploration de cooccurrences dans un corpus transformerait l'étudiant en chercheur linguistique, lui conférant ainsi un certain contrôle sur son apprentissage en le laissant développer sa compréhension de la langue dans ses propres mots. Par ailleurs, Yoon et Hirvela (2004) suggèrent que l'utilisation de corpus peut contribuer de façon significative au processus d'écriture en langue seconde. Dans une tâche d'écriture, le corpus « crée un environnement où cette ressource accompagne l'emploi du dictionnaire, permettant aux scripteurs de trouver de multiples exemples d'expressions dans un contexte pertinent » (Chambers 2010, p. 17). À partir de son survol de la littérature, Yoon (2011) soutient que d'autres applications doivent être développées et que les enseignants doivent recevoir plus de formations pour introduire cette approche en salle de classe. À la lumière des informations trouvées dans la littérature scientifique, ce projet s'inscrit dans la lignée des initiatives et des recherches en didactique qui visent à outiller les apprenants avec différents logiciels bonifiant l'apprentissage langagier, tout en s'assurant que ces outils soient faciles d'utilisation et pertinents au processus d'acquisition d'une langue seconde ou étrangère.

1.3 Limites des approches guidées par les données en classe de langue

En général, l'analyse de corpus semble modérément bénéfique pour les apprenants (Golonka *et al*, 2014), notamment en raison de certaines limites qui ont été identifiées par différents auteurs dans la littérature scientifique. Pour un apprenant dont les compétences langagières et le vocabulaire sont limités, l'analyse d'un grand nombre de phrases afin d'en extraire une règle ou une formule utile nécessite un investissement de temps considérable et peut devenir source de frustration (Chambers, 2010; Chang et Sun, 2009; Gilmore, 2009; Yoon et Hirvela, 2004; Kennedy et Miceli, 2001). Même

dans le cas d'apprenants avec des connaissances langagières plus avancées, certains chercheurs sont d'avis que « les lignes discontinues des concordances ont le désavantage d'offrir trop peu de contexte » (Chambers, 2010, p. 14). À cet effet, Aston (1997) suggère que le décodage du contenu du corpus extrait de son contexte original nécessite une bonne connaissance non seulement linguistique, mais également culturelle, afin de saisir les nuances des phrases. Braun (2005) complète cette affirmation en rappelant que maîtriser une langue implique « le développement de connaissances et d'habiletés requises pour maîtriser les processus de production et de compréhension du discours [notre traduction] » (p.52), et que celles-ci sont difficiles à développer dans un extrait isolé de son contexte communicatif.

Au-delà des difficultés d'interprétation sémantique et grammaticale, d'autres obstacles rendent difficile l'exploration directe de corpus linguistiques par les apprenants. De fait, l'utilisateur doit être à l'aise avec l'ordinateur et la manipulation de données, et il doit comprendre le processus permettant de passer de l'observation de productions langagières à la formulation d'hypothèses sur celles-ci, pour réinvestir éventuellement ces hypothèses dans ses propres productions (Gordani, 2013; Boulton, 2010a). Aussi, bien que pour la plupart des étudiants il soit facile d'accéder à un ordinateur connecté à Internet, il est plutôt rare que ces derniers aient été initiés à une quelconque forme d'apprentissage linguistique assistée par ordinateur (Chang et Sun, 2009). Par ailleurs, même pour des étudiants apprenant une langue seconde ou étrangère depuis plusieurs années, effectuer une expansion de l'approche « communicative » et y incorporer une démarche « guidée par les données » nécessite un certain effort cognitif (Hadley, 2002). À cet effet, dans son expérimentation impliquant des apprenants de niveau universitaire, Chambers (2005) indique que « des habiletés lacunaires en consultation [de corpus] et en analyse [du contenu de corpus] étaient clairement un obstacle à l'expérience d'apprentissage [notre traduction] » (p.116). Pour tirer un profit maximal d'un logiciel de cooccurrences, il

faut savoir non seulement choisir le mot-clef important, mais comme certains outils traitent chaque forme d'un même mot indépendamment, il s'agit aussi de connaître la forme de ce mot qui sera susceptible de donner le plus de résultats probants (Kennedy et Miceli, 2001), ou de savoir incorporer les caractères spéciaux (« *wildcard characters* ») qui permettent d'étendre la recherche. Sans ces connaissances, l'utilisateur du logiciel peut se retrouver alors avec trop – ou trop peu – d'exemples, ce qui diminue l'utilité de l'outil (Yoon, 2011).

L'AGD présente également certaines faiblesses, à commencer par son traitement « mécanique » de la langue, qui tend à prioriser l'aspect linguistique sur les aspects discursif, sémantique, communicatif. Étant donné que les données sont isolées de la situation de communication qu'on peut imaginer à partir du texte complet, la composante discursive de la production langagière peut devenir difficile à interpréter (Braun, 2005). Dans la mesure où l'un des objectifs principaux de l'apprentissage d'une langue est la communication, retirer cette composante peut sembler contrintuitif. Dans un autre ordre d'idées, Boulton (2009b) rappelle que les étudiants qui sont habitués à une approche déductive auront besoin de plus d'accompagnement pour s'habituer à une approche inductive, cette dernière étant souvent favorisée dans les activités d'exploration de corpus. Dans un article subséquent (2010a), le même auteur suggère une introduction graduelle à l'AGD, en utilisant d'abord des activités plus contrôlées sur un support papier – désignée comme consultation indirecte par Chambers (2010) – avant d'introduire graduellement une exploration autonome du corpus à travers un extracteur d'exemples ou un logiciel de cooccurrences – aussi connue sous le nom de consultation directe (Chambers, 2010). Ainsi, pour introduire en douceur l'AGD, les activités peuvent être ajustées en fonction de l'autonomie de l'apprenant, en commençant avec des objectifs précis et des projets dirigés par l'enseignant avant de laisser l'étudiant explorer les ressources par lui-même (Boulton, 2010a). Ce point de vue est partagé par Oghigian et Chujo

(2010), qui suggèrent « de créer une version papier dans laquelle les chaînes plus complexes pourraient être retirées [notre traduction] » (p.203) afin de rendre l'activité plus facile d'approche pour les apprenants novices.

1.4 Limites de l'avancement des connaissances sur les approches guidées par les données en classe de langue

À partir des articles retenus pour ce projet, plusieurs avenues peuvent être envisagées pour contribuer à l'avancement des connaissances. En premier lieu, on remarque une certaine concentration des recherches dans le même milieu : tel que souligné par Landure et Boulton (2010), Braun (2005) et Chambers (2005), un grand nombre de projets concernant l'utilisation de corpus s'est limité au contexte universitaire et à l'apprentissage de l'anglais, avec les particularités et les limites propres à cet environnement et à cette langue. Et comme le rappelle Boulton (2010b), dans la mesure où l'un des objectifs du développement de la technologie est de rendre ces outils accessibles au plus grand nombre de gens possible, il serait pertinent d'évaluer comment ceux-ci sont utilisés dans une plus grande variété de contextes non seulement dans le système scolaire classique (au secondaire et au collégial, par exemple), mais aussi dans des institutions parallèles comme les écoles de langues. En travaillant avec des étudiants suivant un programme de francisation dans un centre de formation continue, ce projet vise à étudier la pertinence de l'AGD avec d'autres types d'apprenants, et à développer les connaissances sur cette approche lorsque le français est la langue cible.

En second lieu, plus de données semblent nécessaires pour identifier les facteurs clés qui peuvent rendre les corpus linguistiques accessibles aux apprenants de tous les niveaux. Plusieurs études s'intéressent aux paramètres qui peuvent guider

l'introduction de l'AGD en classe de langue; par exemple, Boulton (2007a) et O'Sullivan et Chambers (2006) constatent que l'AGD s'avère plus commun (et jugé plus pertinent) pour des apprenants avancés, car ceux-ci peuvent comprendre des formes lexicales et grammaticales plus sophistiquées, tandis que les apprenants de niveau débutant ou intermédiaire dépendent davantage du dictionnaire en raison de leur vocabulaire limité (John, 2001). Néanmoins, cette affirmation n'est pas nécessairement partagée par tous les auteurs. Par exemple, Yoon et Hirvela (2004) ont observé l'utilisation de corpus dans deux groupes (un de niveau intermédiaire, l'autre de niveau avancé) et constaté que les étudiants de niveau intermédiaire ont pu tirer des bénéfices de l'outil pour améliorer leurs productions écrites, notamment grâce à la formation et à l'accompagnement fourni par leur instructeur. Les observations de Landure et Boulton (2010) suggèrent que le corpus peut être utile de façons différentes selon le niveau de compétence langagière des apprenants : pour ceux qui sont plus avancés dans leur maîtrise de la langue cible, « la consultation du corpus leur permet simplement de confirmer leurs intuitions » (p.13), tandis que ceux de niveau plus modeste ont tendance à effectuer plus d'ajustements (autant positifs que négatifs). En observant les élèves analyser un corpus, Kennedy et Miceli (2001) constatent que ceux de niveau plus faible ont parfois plus de facilité à maîtriser cette approche que leurs collègues de niveau plus avancé. De plus, Boulton remarque dans une autre recherche (2009b) qu'avec la formation et l'accompagnement approprié, la formule « *keyword in context* » peut rendre les cooccurrences plus saillantes aux élèves de niveau inférieur en comparaison à des séries de phrases complètes. Le même auteur suggère aussi dans d'autres articles (2008, 2010a) qu'une mouture plus simple de l'AGD, soit en utilisant un support papier ou en concevant des activités guidées de façon plus stricte, pourrait profiter même aux apprenants de niveau plus faible. Ce projet de recherche implique des participants qui ont une compréhension rudimentaire de la langue cible avant le début de leur programme de francisation, et s'inspire des approches et des idées identifiées dans la littérature pour développer l'accessibilité de l'AGD pour ce profil d'apprenant. Il pourra ainsi contribuer à

l'avancement des connaissances sur les facteurs qui contribuent à rendre cette approche pertinente pour des étudiants de niveau débutant ou intermédiaire.

Troisièmement, selon certains articles identifiés dans la littérature, une plus grande variété de matériel devrait être élaborée pour répondre plus adéquatement aux besoins variés des apprenants (Yoon, 2011; Chambers, 2005). Braun (2005) remarque que « le nombre de corpus disponibles pour les différentes langues varie considérablement [notre traduction] » (p.48); de plus, les ressources disponibles ne correspondent pas nécessairement aux besoins des apprenants de langue seconde : « le type de corpus qui est facilement accessible n'est pas nécessairement le plus adapté à l'investigation des processus d'acquisition en langue seconde; ils sont presque toujours écrits, transversaux, et proviennent massivement d'apprenants avancés en anglais [notre traduction] » (Myles, 2005, p.388). À cet effet, Yoon (2011) rappelle qu'il est important de choisir un corpus qui s'accorde avec le niveau d'habileté des apprenants, et affirme que la création de corpus spécialisés serait bénéfique, en permettant aux apprenants de découvrir les expressions et le vocabulaire propre à un genre ou à une discipline donnée, tout en étant plus facile à manipuler et à interpréter, selon O'Sullivan et Chambers (2006). Bien que les corpus de différente taille aient le potentiel d'être complémentaires (Yoon, 2011), très peu d'articles trouvés dans la littérature impliquent plus d'une ressource. Par exemple, une recherche de Johns et collègues (2008) implique deux corpus (l'un bâti à partir d'une œuvre littéraire, et l'autre à partir de sa traduction chinoise), avec des résultats bénéfiques pour développer la connaissance du vocabulaire et de la grammaire et réduire l'aspect anxigène associé à la lecture de longs textes. On retrouve aussi un corpus bilingue dans la recherche de Chujo et collègues (2009), contribuant à mettre en lumière les différences grammaticales entre la langue maternelle et la langue-cible. Finalement, dans la recherche de Liu et Jiang (2009), l'accès à un ensemble de sous-

corpus peut permettre aux apprenants de mieux comprendre si un terme apparaît plus fréquemment dans la langue écrite ou dans la langue parlée.

Dans l'espoir de contribuer à la compréhension de l'influence des différents types de corpus sur les activités d'AGD, ce projet s'appuiera sur deux ressources différentes. Par la même occasion, ce projet visera à développer un prototype de corpus utilisant un autre type de texte, afin de diversifier les outils accessibles aux apprenants du français.

1.5 Introduction des prépositions dans l'apprentissage du français

La définition de la préposition tend à varier selon le système grammatical employé, mais selon Homma (2006), « l'idée cruciale, soutenue largement par de nombreux linguistes, est qu'une préposition est un relateur entre deux entités, à savoir X (repéré) et Y (repère) » (p.4). Dans son analyse critique de l'enseignement des prépositions « à » et « dans », l'auteure constate que les modèles trouvés dans les ouvrages de référence ne correspondent pas nécessairement à l'emploi des prépositions. Ainsi, les explications sémantiques opposant la notion de « point » pour « à » à la notion d'« interiorité » pour « dans », ou le « générique » et le « spécifique » pour ces mêmes prépositions, sont remises en question avec des exemples qui contredisent ces explications. Les formules grammaticales semblent aussi incomplètes; par exemple, pour ce même duo de prépositions, l'association entre « à » et articles définis, et entre « dans » et articles indéfinis, est remise en question avec d'autres exemples qui contredisent cette formule. À partir de cette étude de cas, l'auteure rappelle qu'une étude exhaustive des emplois d'une même préposition est essentielle à la bonne compréhension de son fonctionnement : « à

travers divers emplois d'une préposition, se manifeste un mécanisme générateur, autrement dit, l'ensemble des propriétés fonctionnelles de cette préposition » (p.11).

Les recherches de Sylvie Roy (1997, 1998), avec leur description des approches choisies dans quatre manuels pour enseigner les prépositions « à » et « de », contribuent à illustrer les défis propres à l'apprentissage des prépositions en contexte de français langue seconde, où les besoins communicatifs des apprenants requièrent une grande quantité de vocabulaire et d'expressions, parfois au détriment de notions de grammaire. Certaines des difficultés évoquées par Roy comprennent la représentation de règles grammaticales en formules, qui « demande une certaine capacité d'abstraction » (Roy, 1997, p.47) aux enseignants et aux étudiants, ou le fait que les définitions tendent à s'appuyer sur une nomenclature qui n'est pas nécessairement familière aux étudiants :

D'après les nombreuses descriptions et définitions fournies par les linguistes sur les prépositions A et DE, on peut déjà comprendre que leur enseignement et leur apprentissage ne sera pas chose aisée. De plus, il existe des problèmes de terminologie et de définition des fonctions grammaticales, car les descriptions évoquées [...] utilisent les termes de circonstanciel, de complément indirect et de groupe prépositionnel. (Roy, 1997, p. 79)

En plus des difficultés identifiées ci-dessus, Roy s'intéresse aux composantes associées à l'enseignement des prépositions, comme les verbes et les thèmes couramment utilisés. Elle remarque notamment l'emploi de verbes « à la fois très courants et proches des élèves » (Roy, 1997, p.189) pour enseigner « à » et « de », comme jouer, faire, aller et travailler. L'auteure remarque aussi que « des verbes qui ont un rapport avec les fonctions communicatives étudiées » (Roy, 1997, p.190) comme suggérer, proposer, ou recommander, peuvent servir à construire des exemples avec la préposition « de ». Finalement, Roy constate que « les aliments, les noms de pays et les passe-temps » (Roy, 1997, p.175) sont des thèmes communs pour

introduire l'emploi des prépositions, et suggère d'ajouter à cette liste des thématiques comme les vêtements et les loisirs. Dans l'ensemble, ces observations ont contribué à guider le choix des phrases à compléter et des exemples pour construire l'exercice complété par les participants, qui sera détaillé dans la méthodologie.

D'autres observations de l'auteure mettent en évidence des stratégies à considérer pour des apprenants de langue seconde, comme la simplification du métalangage en termes simples (Roy, 1997, p.172), la présentation d'expressions que les élèves pourront utiliser (Roy, 1997, p.156) et l'emploi de tableaux et de schémas accessibles (Roy, 1997, p.110). Dans l'ensemble, un constat qui ressort est que les prépositions comme « à » et « de » sont peu expliquées dans les manuels, entre autres parce que les explications dans les guides pédagogiques peuvent s'avérer lacunaires (Roy, 1998), et leur acquisition demeure difficile pour les étudiants.

Les recherches de Bruley-Meszaros (2010) soutiennent les observations ci-dessus : « les verbes de valence à et de, en tant qu'objectifs d'apprentissage, sont très peu nombreux dans les manuels et rarement abordés » (p.55); de plus, les règles présentées, « en plus d'être incomplètes, peuvent porter à confusion, voire être erronées » (p.56). Dans l'ensemble de méthodes et d'ouvrages de référence rassemblé par l'auteure, il semble que les « règles de grammaire [où sont introduits les verbes associés aux prépositions "à" et "de"] sont centrées sur les pronoms et ne présentent donc pas le fonctionnement de ces verbes dans leur ensemble. » (p.55). À travers le développement de la notion de « valence » pour expliquer l'utilisation des prépositions « à » et « de », l'article de Bruley-Meszaros permet de désigner d'autres verbes propices à l'enseignement de ces prépositions, comme parler, penser, ou rêver, qui peuvent être associés à l'une ou à l'autre (ou même aux deux prépositions simultanément), avec différentes nuances sémantiques selon la préposition employée.

En somme, l'introduction des prépositions dans l'enseignement du français relève d'un acte d'équilibre entre la nécessité d'offrir des exemples pertinents et des expressions utiles dans un contexte accessible aux apprenants, et l'importance de conserver les différentes nuances et complexités propres à l'emploi de ces prépositions. En retirant les emplois de préposition moins fréquents, il apparaît que « les auteurs des manuels de FL2 pour les étrangers simplifient le plus possible un système qui peut paraître complexe, ce qui peut engendrer chez les élèves de langue seconde quelques erreurs dans la compréhension de certains points grammaticaux » (Roy, 1997, p. 185-186). Le corpus linguistique, avec sa grande variété d'exemples provenant de textes authentiques et la possibilité de trier les exemples selon leur fréquence d'apparition, détient le potentiel pour faire le pont entre l'accessible et l'exhaustif, à condition que l'étudiant soit adéquatement accompagné dans son utilisation du corpus.

1.6 Questions de recherche

Bien que plusieurs chercheurs aient identifié des retombées intéressantes quant à l'emploi de corpus linguistiques en classe de langue, les initiatives développées spécifiquement pour les besoins des apprenants demeurent rares (Chambers, 2010). Dans la mesure où des recherches précédentes ont associé plusieurs bénéfices à l'utilisation de corpus linguistiques pour l'apprentissage de langues étrangères (hausse de la motivation et attitude positive, possibilité de personnaliser l'apprentissage, apport d'un outil de référence supplémentaire), ce projet de recherche vise à améliorer l'utilisation de cet outil, de façon à ce que des utilisateurs provenant de différents contextes (et de différents niveaux de compétence langagière) en tirent le meilleur avantage possible.

Dans une perspective de contribution à l'avancement des connaissances dans le domaine de l'utilisation de corpus linguistiques et de l'AGD, ce projet vise à s'inspirer des outils et des démarches trouvées dans les recherches énumérées dans la problématique, mais en recrutant des participants extérieurs au milieu universitaire, et qui ne sont pas encore à un niveau avancé de compétence langagière. De plus, comme le remarque Boulton dans un survol de recherches sur l'AGD (2007a), un plus grand nombre d'études se concentre sur l'anglais, non seulement en raison de la demande plus importante pour cette langue, mais également en raison de la plus grande disponibilité d'outils et de corpus (Braun, 2005). Dans cette optique, il s'avère important de continuer à développer des ressources et des outils pour les apprenants de d'autres langues, dont le français (Cobb, 2014a).

La question de recherche de ce mémoire est la suivante : dans une tâche où les étudiants insèrent des prépositions dans des phrases trouées en se basant sur les références fournies par deux corpus linguistiques différents, est-ce que le matériel du corpus assigné peut avoir une influence sur le résultat à la tâche? Dans la mesure où ces corpus proviennent de sources différentes, il est possible que des variations au niveau de leur contenu (comme le vocabulaire, la complexité syntaxique, ou les types de texte retenus) puissent entraîner une variation dans leur utilisation par les étudiants. Le cas échéant, on pourrait observer des variations dans d'autres aspects, comme les résultats à la tâche, les perceptions des participants vis-à-vis de l'outil (que ce soit la facilité de lecture, ou l'intérêt pour les références fournies par les corpus, par exemple), ou les interactions des participants avec le dispositif fourni..

CHAPITRE II

CADRE THÉORIQUE

Le cadre théorique se concentrera à priori sur les définitions du corpus linguistique et de l'apprentissage guidé par les données, suivi par un survol des recherches qui ont guidé le choix des prépositions comme objet langagier à développer dans l'exercice complété par les participants, ainsi qu'une synthèse des recherches connexes à ce projet de recherche. Ce chapitre sera complété par une énumération des objectifs spécifiques du projet.

2.1 Corpus linguistique

La définition retenue dans l'article rédigé par O'Sullivan et Chambers (2006) permet de résumer les caractéristiques essentielles du corpus linguistique : « un ensemble de textes, échantillonné de façon à être maximalelement représentatif d'une variété donnée d'une langue, et qui peut être entreposé et manipulé en utilisant un ordinateur [notre traduction] » (pp.49-50). Dans leur article, Golonka et collègues (2014) ajoutent plusieurs nuances à cette définition, en indiquant que le contenu du corpus peut être entreposé sous forme orale, écrite, ou dans un mélange des deux. Quant à la variété de langue représentée, on peut trouver à la fois des corpus généralistes et spécialisés, ces derniers étant particulièrement utilisés dans des tâches de traduction et d'interprétation où l'apprenant se spécialise dans un type de texte ou dans un domaine particulier (Aston 1997), ou pour les tâches de rédaction académique. Contrairement à ce qu'indique la définition retenue par O'Sullivan et Chambers ci-haut, Golonka et collègues (2014) affirment que le corpus peut être de

taille variable – dans le cas des corpus tirant leur contenu du web, par exemple (Acar *et al*, 2011). D'autres recherches ont même élaboré un corpus bilingue, en créant des associations de phrases entre une œuvre littéraire et sa traduction dans la langue cible (Johns *et al*, 2008), ou des corpus élaborés à partir des ouvrages de référence utilisés par les étudiants, avec la possibilité d'élaborer l'affichage « *keyword in context* » en consultant des exemples complets (Cambrensis et Gillway, 2006).

À travers la littérature, plusieurs autres types de corpus peuvent être identifiés : par exemple, il y a des corpus spécifiquement créés pour les apprenants (Chambers, 2005), ainsi que des corpus rassemblant à la fois des productions écrites d'apprenants et un système d'annotation des erreurs trouvées dans ces productions (Granger, 2003). L'exploration de ces bases de données linguistiques, qualifié de « recherche en corpus [notre traduction] » (Yoon et Hirvela, 2004, p.19) ou de « méthodes basées sur le corpus [notre traduction] » (O'Sullivan et Chambers, 2006, p.52), vise à « procurer l'accès à du matériel riche et authentique [notre traduction] » (Golonka *et al*, 2014, p.72), de façon à ce que les apprenants puissent étudier des formes spécifiques de production langagière, pour formuler et vérifier leurs hypothèses sur le fonctionnement de la langue (Yoon et Hirvela, 2004). Par l'énumération de plusieurs exemples authentiques de combinaisons de mots récurrentes, l'étude du corpus vise à « aider les apprenants langagiers à internaliser les schémas dont ils ont besoin [notre traduction] » (O'Sullivan et Chambers, 2006, p. 52). Selon le type de tâche, un corpus de taille différente sera à privilégier : pour accompagner la rédaction de textes, un corpus plus petit (mais spécialisé) serait préférable (Chambers, 2010), son utilité provenant du fait qu'il « offre une collection de performances d'experts dans des genres qui sont pertinents aux besoins et aux intérêts des apprenants [notre traduction] » (Kennedy et Miceli, 2001, p.38). De l'autre côté, un corpus généraliste offre l'avantage d'un plus grand nombre d'exemples (Yoon, 2011), ce qui pourrait contribuer à rehausser la compréhension de

la langue. Pour résumer, les corpus linguistiques sont des bases de données de productions langagières authentiques; l'exploration et l'analyse de ces données contribueraient à acquérir et à comprendre la langue étudiée, selon le processus d'AGD, défini dans la section suivante.

2.2 Apprentissage guidé par les données

À la base, l'apprentissage guidé par les données (connu sous l'appellation de « *Data-Driven Learning* » en anglais, souvent résumée au sigle DDL, qui est traduit par AGD dans ce document) est conçu comme l'observation de base de données langagières, de façon à isoler des formulations communes émergeant d'exemples authentiques (Hadley, 2002; Landure, 2011). Dans Gordani (2013), on retrouve quatre principes à la base de cette approche : l'analyse empirique de « patrons langagiers » (*language patterns*) dans des textes authentiques, l'emploi d'un corpus à la base de cette analyse, l'utilisation d'ordinateurs pour diriger cette analyse (qu'elle soit automatisée ou interactive), et l'emploi de techniques d'analyses quantitative et qualitative pour réaliser la tâche (p.433). En somme, cette approche favorise les activités qui sensibilisent les apprenants aux patrons langagiers, à des régularités et à des tendances générales (Landure et Boulton, 2010), plutôt qu'un enseignement explicite de règles.

À partir de ces observations de cooccurrences, les apprenants développeraient leur compréhension et leur production de la langue non pas en suivant des règles grammaticales complexes, mais plutôt à partir de blocs de mots mémorisés en tant que « touts » (O'Sullivan et Chambers, 2006). Selon Hunston et Francis (1998), la « fréquence relative de cooccurrence [...], dit-on, faciliterait la tâche psychologique de traitement langagier [notre traduction] » (p. 66). D'après les observations de Todd

(2001), les types de mots où il est plus facile de déceler des patrons langagiers réguliers, comme les adjectifs et les verbes, seraient plus faciles à corriger pour les apprenants. Par ailleurs, ces blocs de mots permettraient de faire le trait d'union entre le vocabulaire et la syntaxe : les règles grammaticales seraient une expression abstraite de patrons langagiers s'appliquant à un ensemble lexical relativement vaste. Kennedy et Miceli (2001) abondent dans le même sens, suggérant que la formation de phrases est un assemblage de blocs de mots et que les apprenants maîtrisent la langue en prêtant attention à la façon dont ces assemblages s'agencent, mais aussi comment et dans quelles circonstances ceux-ci peuvent varier. Ainsi, l'étude de plusieurs exemples similaires serait compatible avec le processus cognitif de groupement d'information qui semble, selon certaines études psycholinguistiques (Ellis, 2003), favoriser la mémorisation.

Avec cette approche jugée compatible avec un style d'apprentissage inductif par certains chercheurs (Cobb et Boulton, 2015; Xue-hua et Ming-ying, 2013; Boulton, 2009a; Yoon et Hirvela, 2004), les apprenants construiraient leur propres « règles » selon leur compréhension des textes lus et à partir des tendances qu'ils observent dans les exemples, et raffinement graduellement ces règles au fur et à mesure que leur compréhension de la langue s'améliore. Selon Hadley (2002), cette démarche s'inscrit dans une lignée affirmant que l'enseignement de la grammaire « devrait consister en grande partie en activités de sensibilisation plutôt qu'en l'enseignement de règles [notre traduction] » (p.107). Todd (2001) est d'avis qu'il s'agit d'un processus « nécessitant peu d'effort de la part des apprenants, puisque les règles deviendront évidentes si suffisamment d'exemples appropriés sont à leur disposition [notre traduction] » (p.92). Boulton (2007b) favorise également l'approche guidée par les données en affirmant que des règles grammaticales enseignées explicitement seraient plutôt « des constructions artificielles qui nécessitent de grands efforts cognitifs pour être manipulées [notre traduction] » (p.2).

Selon Chambers (2005), l'utilisation de corpus serait particulièrement propice à l'élaboration de règles par les apprenants selon une démarche inductive. Bien qu'elle ait formé son groupe de quatorze étudiants aux approches déductive (commençant par la consultation d'une grammaire avant d'explorer le corpus) et inductive (commençant directement par l'exploration du corpus), « 9 ont choisi cette dernière méthode, suggérant qu'ils trouvaient peut-être l'approche inductive plus naturelle ou plus intéressante lorsqu'ils utilisaient des données du corpus [notre traduction] » (p.116). Sun et Wang (2003) apportent une nuance supplémentaire avec leur recherche : selon leur comparaison d'activités d'AGD de façon déductive (enseignement explicite de la règle et accès au corpus pour renforcer la compréhension de cette règle) et de façon inductive (élaboration d'une règle à partir de l'observation d'exemples tirés du corpus), chaque approche aurait sa raison d'être. Pour les cooccurrences jugées plus simples, une approche inductive pourrait donner des résultats supérieurs, mais un accompagnement s'appuyant sur une démarche déductive serait préférable pour certaines notions plus complexes.

2.4 Synthèse des recherches sur l'apprentissage guidé par les données en lien avec la langue française et avec les apprenants de niveau débutant

Comme ce projet implique des étudiants apprenant le français comme langue seconde, il est essentiel de se référer aux travaux publiés par Chambers et O'Sullivan (2006, 2004) pour prendre connaissance d'activités d'AGD avec des apprenants de cette langue. Dans l'article publié en 2004, Chambers et O'Sullivan détaillent la création d'un corpus spécialisé (avec des articles journalistiques sur la langue française, ainsi que des textes académiques sur le développement et l'histoire de cette

langue) que des étudiants universitaires aux cycles supérieurs en langue française ont utilisé comme outil de référence.

La méthodologie impliquait un processus à deux étapes, combinant un cours de rédaction à un cours sur les technologies d'information et de la communication (TIC). Premièrement, les étudiants étaient invités à écrire un commentaire sur un des enjeux du français, et les passages à travailler dans ce commentaire étaient soulignés. Deuxièmement, après une formation sur la consultation de corpus linguistiques, les étudiants devaient consulter ces ressources pour réviser les passages surlignés. Selon les analyses des auteures, la consultation de corpus a contribué à corriger des erreurs grammaticales (accord, utilisation de prépositions, négation), orthographiques et lexicogrammaticales (vocabulaire inapproprié, expressions figées). Dans l'ensemble, les réactions des participants vis-à-vis l'utilisation de cette méthode étaient généralement positives, avec quelques suggestions pour rendre l'outil plus facile d'accès.

Dans l'article publié par O'Sullivan et Chambers en 2006, une méthodologie semblable était appliquée (rédaction d'un court texte par les étudiants, surlignage des passages à réviser par les enseignantes-chercheuses, tentative de correction à l'aide du corpus) et le même type de corpus était utilisé, mais l'activité était complétée avec des étudiants au premier cycle universitaire. Selon les analyses, la consultation de corpus s'est avérée particulièrement pertinente pour ajuster l'emploi de prépositions et le choix de vocabulaire. Selon les données recueillies dans un questionnaire distribué aux participants, ceux-ci ont apprécié l'activité dans la plupart des cas, mais ont trouvé que la grande quantité de matériel produit par le corpus pouvait être difficile à filtrer et à interpréter correctement.

Pour trouver des pistes de solution aux difficultés évoquées dans les recherches de Chambers et O'Sullivan et intégrer des activités d'AGD auprès

d'étudiants de niveau moins avancé, les travaux de Boulton (2009b, 2010a) offrent plusieurs pistes de réflexion. Par exemple, pour éviter de présenter une trop grande quantité de contenu, il est possible de choisir à l'avance une série d'exemples probants à partir desquels les participants pourront former des hypothèses sur le fonctionnement de la langue. Pour les apprenants qui ne seraient pas à l'aise avec la formule « mot-clef en contexte », il est envisageable d'afficher plutôt des phrases complètes et de mettre le mot-clef en évidence autrement (surligné, en gras, etc.) dans un premier temps, avant de faire la transition vers un affichage de cooccurrences alignées. De la même manière, si la familiarisation avec le logiciel de consultation pose problème, une option à considérer serait de présenter le contenu sans la composante interactive (sur papier, ou à l'écran), afin qu'il soit plus facile de se concentrer sur les exemples fournis. Dans tous les cas, le principe est de morceler l'expérience d'AGD en composantes distinctes, et de choisir quelles composantes sont présentées avant de présenter l'expérience dans son ensemble.

Par ailleurs, Chujo et collègues (2009) rappellent certains principes pouvant guider l'introduction d'AGD avec des apprenants dont les compétences dans la langue cible sont limitées. Notamment, il apparaît préférable de choisir une tâche pour laquelle les exemples peuvent être interprétés de façon relativement superficielle pour élaborer une hypothèse simple et précise, en s'assurant que les exemples ne soient pas trop nombreux et qu'ils ne comprennent pas de variations hors-sujet. Dans leur article, Kennedy et Miceli (2001) suggèrent quelques autres principes directeurs pour guider les novices, dont : prêter attention aux mots de chaque côté des prépositions, ne pas oublier l'importance du lexique, et s'assurer que le problème à résoudre est suffisamment spécifique. En somme, plutôt que d'assumer que les étudiants adopteront sans heurts le rôle d'apprenant-chercheur, il est conseillé de contrôler les paramètres de l'exercice de façon à ce que celui-ci présente des objectifs clairs et réalisables, pour insérer graduellement une plus grande marge de manœuvre.

Finalemment, il est important de garder en considération la difficulté supplémentaire que peut représenter l'utilisation d'un outil informatique pour certains apprenants. Dans la littérature, cette difficulté est souvent nommée « surcharge cognitive », et elle « a lieu lorsque la quantité d'information à traiter [dans un environnement multimédia] dépasse la capacité cognitive de l'apprenant [notre traduction] » (Mayer et Moreno, 2003, p. 45). Bien que Mousavi et collègues (1995) suggèrent de varier et de combiner les modalités sensorielles pour faire appel aux différentes capacités cognitives, par exemple en offrant simultanément du texte écrit et une narration audio, cela pourrait affecter la capacité d'attention de l'apprenant. Aussi, le traitement d'information accessoire, c'est-à-dire « les processus cognitifs qui ne sont pas requis pour comprendre le matériel présenté, mais qui sont engendrés par la conception de la tâche [notre traduction] » (Mayer et Moreno, 2003, p. 45) peut mobiliser les ressources cognitives de l'étudiant. Pour que l'apprentissage soit significatif, l'apprenant doit « organiser mentalement le matériel présenté dans une structure cohérente, et intégrer le matériel présenté avec les connaissances antérieures [notre traduction] » (Mayer et Moreno, 2003, p. 44). À cet effet, selon les recherches de Chandler et Sweller (1996), il serait préférable de rassembler l'information dans un seul médium, plutôt que d'avoir une tâche à l'écran accompagnée d'instructions sur papier, par exemple. Ainsi, organiser le matériel de façon à faciliter son traitement (en réduisant les informations accessoires et en les concentrant dans un seul format, par exemple), peut aider à rendre la tâche plus accessible.

2.5 Objectifs spécifiques

Tandis que plusieurs recherches citées dans ce mémoire impliquent des apprenants de niveau avancé, ce projet vise à déterminer si des apprenants de niveau

intermédiaire inférieur peuvent bénéficier d'une approche d'AGD. Dans une perspective exploratoire, et considérant la quantité restreinte de ressources disponibles en français pour l'AGD, ce projet vise à explorer comment les résultats à une tâche guidant la manipulation d'un objet langagier complexe (dans ce cas, l'insertion de prépositions dans des phrases trouées) peuvent être influencés par le matériel de référence qui est fourni aux apprenants. Étant donné le nombre limité de corpus disponibles en français et accessibles aux apprenants débutants ou intermédiaires, nous avons élaboré un corpus inédit à partir d'une ressource destinée à ces apprenants. Ce corpus, qui pourrait présenter à la fois des similitudes et des contrastes avec un autre corpus de langue française déjà disponible, contribue à rendre possible l'exploration de comparaisons entre différentes sources de matériel de référence.

Par ailleurs, étant donné le nombre modeste de projets d'AGD effectués en français et avec des apprenants de niveau débutant ou intermédiaire inférieur, ainsi que les difficultés associées à cette approche, ce projet s'intéresse aux interventions qui contribuent à rendre ce type d'activité accessible à un plus grand bassin d'apprenants, et vise à évaluer comment les participants interagissent avec un outil simplifié. Dans cette optique, nous ajoutons deux autres objectifs spécifiques à celui mentionné ci-dessus : 1) documenter les perceptions des participants vis-à-vis du dispositif par l'intermédiaire d'un questionnaire; 2) documenter les interactions des participants avec le dispositif (plus spécifiquement, les clics effectués dans l'interface).

CHAPITRE III

MÉTHODOLOGIE

Tout d'abord, un portrait d'ensemble des participants à ce projet de recherche sera effectué; ensuite, l'instrumentation sera décrite en détail, en prêtant une attention particulière aux caractéristiques des deux corpus linguistiques utilisés ainsi qu'à l'objet langagier choisi pour l'activité complétée par les participants. Ces informations ouvriront la voie à une explication du fonctionnement des différentes composantes (exercice et questionnaire, enregistrement des actions). Finalement, ce chapitre sera complété par une description générale des deux rencontres réalisées avec les participants, des considérations éthiques ayant guidé la réalisation de ce projet, et de l'approche choisie pour récolter et pour analyser les données.

3.1 Description des participants

Les participants qui ont accepté de prendre part à ce projet de recherche sont dix-neuf étudiants inscrits dans un centre de formation continue d'une commission scolaire de la région métropolitaine de Montréal. Les participants sont de jeunes adultes dont l'âge varie entre 18 et 39 ans (moyenne = 23 ans), et qui proviennent de diverses régions de la Chine. Le fait que les participants proviennent de la même région n'était pas un tri intentionnel de la part du chercheur, mais simplement un reflet de la clientèle desservie par ce centre de formation continue.

Dans l'ensemble, on retrouve légèrement plus de femmes (onze) que d'hommes (huit). À l'occasion de ce projet, les participants sont séparés en deux groupes de dix et de neuf participants, à savoir un premier groupe de six femmes et de quatre hommes, ainsi qu'un second groupe de cinq femmes et de quatre hommes.

Tableau 3.1 Caractéristiques des deux groupes de participants

	Groupe 1	Groupe 2
Femmes	6	5
Hommes	4	4
Âge moyen (ans)	23,4 *	23,3

* Dans le groupe 1, une personne a choisi de ne pas dévoiler son âge; la moyenne a été ajustée en fonction du nombre de participants qui ont choisi de répondre.

Les résultats du test de Mann-Whitney ($p = 0,67$ pour l'âge, signification exacte unilatérale, et $p = 1,00$ pour le genre, signification exacte bilatérale) soutiennent l'absence de différence significative entre les deux groupes pour l'âge et pour le genre; cette vérification a été effectuée en amont pour s'assurer que les comparaisons subséquentes entre les groupes (résultats à l'activité, perceptions du dispositif, interactions avec le dispositif) ne seraient pas liées aux caractéristiques sociodémographiques.

En ce qui concerne le cheminement scolaire et professionnel des participants, la plupart de ces étudiants suivent parallèlement des formations professionnelles dans une variété de domaines (à titre d'exemples, on retrouve la soudure, la mécanique automobile, le dessin industriel, et le design d'intérieur) durant le jour, et un cours de francisation les soirs de semaine et la journée du samedi. Selon l'enseignante

responsable des cours de francisation pour ce groupe, la motivation de certains participants pour suivre ce cours est que ces heures d'études peuvent être reconnues pour indiquer un certain niveau de compétence langagière en français, ce qui peut devenir un atout important dans le cadre d'une éventuelle demande de résidence au Canada et de l'intégration au marché du travail dans la société d'accueil.

D'après l'enseignante responsable du cours de francisation, la plupart des participants ont un parcours assez semblable en ce qui a trait à leur apprentissage de la langue. De prime abord, ces participants ont une certaine capacité de communication orale et écrite en anglais; bien qu'ils soient débutants en ce qui a trait à la maîtrise du français, il est possible que les similitudes entre l'anglais et le français (par exemple, l'utilisation du même alphabet, de même que l'orthographe semblable pour plusieurs mots) aient contribué à faciliter leur transition vers l'apprentissage de la langue française. De plus, quelques-uns des étudiants de ce groupe ont eu accès à une formation de base pour se familiariser avec la langue française avant leur arrivée au Québec. Toutefois, leur apprentissage formel de la langue a commencé au printemps, en même temps que leur formation professionnelle. Le programme de francisation s'appuie sur des manuels destinés aux étudiants de niveau intermédiaire, comme la « Grammaire progressive du français » (Grégoire et Thiévenaz, 2003), des exercices originaux conçus par l'enseignante, ainsi que divers documents audiovisuels récents (comme des films populaires traduits en français, ainsi que des extraits d'émissions télévisuelles de Radio-Canada). Dans l'ensemble, ce programme a été élaboré de façon à ce que les étudiants soient prêts à compléter le Test de connaissance du français (TCF) pour le Québec, un des outils d'évaluation de la compétence en français reconnu par le Ministère de l'Immigration, de la Diversité et de l'Inclusion (Immigration, Diversité et Inclusion Québec, 2016) à la fin de la session d'automne.

Considérant le niveau de compétence limité en français des participants au début de leur formation, il fut décidé avec l'enseignante que les rencontres auraient plutôt lieu vers la fin du programme, à l'automne. À cette étape du parcours, les participants s'approcheraient d'un niveau intermédiaire inférieur, et seraient plus à l'aise avec la compréhension de phrases complètes et de messages complexes, ce qui permettrait l'évaluation d'aspects comme l'emploi de prépositions ou la construction de subordonnées relatives, pour lesquels l'utilisation de l'AGD est habituellement plus propice (Chang et Sun, 2009; O'Sullivan et Chambers, 2006; Chambers et O'Sullivan, 2004). Aussi, en attendant vers la fin du programme de francisation pour réaliser l'activité conçue dans le cadre de ce projet de recherche, celle-ci pouvait devenir un exercice de révision que l'enseignante pourrait utiliser pour réviser différentes notions enseignées durant les cours précédents.

Pour résumer, les participants qui ont pris part à ce projet de recherche apprennent le français comme langue seconde dans un contexte de francisation, tout en suivant parallèlement une formation professionnelle. Au moment de la rencontre, ces étudiants travaillaient à atteindre un niveau « intermédiaire » de compétence langagière en français, avec l'objectif de pouvoir réussir le TCF. Les prochaines sections détaillent les activités et les outils utilisés avec ces participants, l'objet langagier choisi pour ces activités, et le déroulement des rencontres.

3.2 Instrumentation

Pour recueillir une variété de données sur les participants, deux outils ont été développés : un exercice centré sur l'utilisation des prépositions, ainsi qu'un questionnaire pour obtenir les données sociodémographiques des participants, leurs

habitudes et leurs préférences comme apprenants, leur relation à la technologie, et leur appréciation de l'outil présenté avec l'exercice mentionné ci-dessus.

Les deux outils ont été créés en format HTML (*Hyper Text Markup Language*) en utilisant le gabarit gratuit « Foundation », créé par la firme de design ZURB pour créer du contenu web qui s'adapte automatiquement aux différents formats d'écran (ZURB, Inc., 2017), avec des fonctions interactives codées en JavaScript et en PHP (*PHP: Hypertext Preprocessor*), de façon à être accessibles à partir d'un navigateur web. Ce choix technologique présentait plusieurs avantages : d'abord, en passant par un navigateur, il n'était pas nécessaire d'obtenir les autorisations du centre de formation continue pour installer un logiciel supplémentaire sur les ordinateurs de la salle de classe. De plus, comme la majorité des participants étaient déjà familiers avec l'utilisation d'un navigateur web dans d'autres contextes (recherche d'information, communication par courriel et sur les réseaux sociaux, etc.), il était raisonnable de présumer que cette interface serait moins intimidante pour les participants qu'un logiciel présentant une interface moins familière, qui pourrait nécessiter une certaine période d'apprentissage et d'adaptation. Finalement, ce choix technologique ouvrait la voie à l'utilisation d'outils fournis par Google (Google Tag Manager et Google Analytics) pour intégrer plus aisément l'enregistrement des actions effectuées par les participants (clics, défilement du contenu) dans l'outil. Cet aspect est décrit de façon plus détaillée dans la section 3.7 de ce chapitre.

Dans l'éventualité où l'outil ne soit pas accessible au moment de rencontrer les participants (problème d'accès à Internet, serveur inaccessible, difficultés techniques avec les ordinateurs dans la classe), une version papier des outils a été préparée comme plan de rechange, avec essentiellement le même format que l'outil électronique, sans les aspects interactifs (par exemple, la possibilité d'afficher ou de cacher au besoin les listes d'exemples dans l'exercice sur les prépositions).

3.3 Développement des corpus

Pour ce projet de recherche, du contenu provenant de corpus francophones a été mis à la disposition des participants dans les exercices sur les prépositions. Premièrement, des phrases ont été extraites des corpus, puis transformées en phrases trouées en retirant les prépositions. Deuxièmement, des exemples illustrant le fonctionnement des différentes prépositions ont été extraits de ces corpus. Puis, dans la tâche, les participants devaient réintroduire une préposition dans la phrase trouée, en se référant aux exemples associés. Afin d'évaluer si le type de corpus pouvait avoir une incidence sur l'exercice (résultats et appréciation de l'activité par les participants), chacun des deux groupes a eu accès à du contenu provenant d'un corpus différent.

The screenshot shows the Logos Library website interface. At the top left is the 'logos' logo with the tagline 'non solo parole' and a turtle icon. At the top right is a logo for 'dona ora, sostieni gli ospedali di A'. The main content area is titled 'LOGOS LIBRARY - Word by word multilingual library'. On the left, there is a vertical navigation menu with links: 'about logos', 'logos dictionary', 'bilingual dictionaries', 'children's dictionary', 'volunteers', 'logos conjugator', 'learning conjugator', 'logos library', 'logos poetry', 'children's library', 'logos quotes', 'logos toolbars', and 'translation course'. The central search area is titled 'Search in the library' and contains the following fields: 'Sentence' (empty), 'Language' (dropdown menu set to 'French'), 'Title' (text input with 'A' entered), 'Author' (empty), and 'Publisher' (empty). Below the search fields are two circular icons and the text 'Title Index | Author Index | Publisher Index'. At the bottom, a status bar indicates 'French' and 'A : 144 documents found'. Below this, a list of search results is shown, including: '25 Annees De Lutttes Et De Victoires Sur La Voie Du Socialisme - Hoxha Enver', '32 Ans En Corrèze - Chapnikoff Roland', and '61 + 1 Aphorismes Sur La Parole Au Théâtre - Tordjman Charles'.

Figure 3.1 Capture d'écran de la page d'accueil de la bibliothèque d'ouvrages en langue française qui sont disponibles sur Logos Library

Le premier corpus utilisé est Corpatext (New et Wordtheque, 2006), un rassemblement de textes cumulant 15 millions de mots, créé en 2006 par Prof. Boris New à partir des ouvrages libres de droit disponibles sur la WordTheque (cette ressource étant désormais connue sous le nom de « Logos Library ») (Logos Group, 2017). Une version condensée de Corpatext (correspondant à approximativement 10% du corpus total), créée pour fonctionner avec les outils de Lextutor, a été gracieusement fournie par Prof. Tom Cobb pour les fins de ce projet de recherche (Cobb, 2014b).

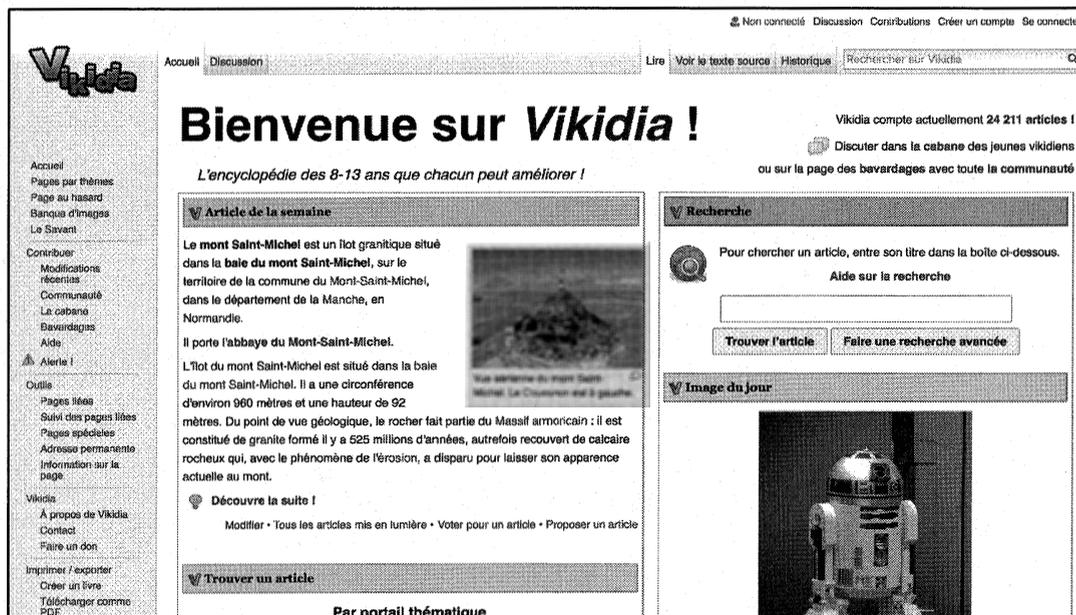


Figure 3.2 Capture d'écran de la page d'accueil de l'encyclopédie Vikidia

Le second corpus a été créé sur mesure à partir du contenu disponible sur l'encyclopédie en ligne Vikidia. Cette ressource s'inspire en grande partie de la structure et du fonctionnement de Wikipédia (notamment le contenu libre d'accès et de partage, ainsi que le fonctionnement communautaire et participatif de la rédaction), mais vise à offrir du contenu accessible à des utilisateurs dans la tranche d'âge entre

huit et treize ans (Association Vikidia, 2016). Outre son contenu libre d'accès, Vikidia présente également l'avantage d'être disponible en archive téléchargeable pouvant être utilisée hors connexion; une initiative développée pour rendre l'encyclopédie accessible dans des milieux scolaires ou éducatifs où l'accès à Internet peut être difficile ou sporadique (Association Vikidia, 2017). Cette archive téléchargeable permettait ainsi de faciliter l'extraction des données pour créer le second corpus de ce projet de recherche, sans avoir recours à un robot d'aspiration de site Web. À partir de cette archive, des scripts en format Perl ont permis d'automatiser l'extraction des articles, l'isolement du texte principal et leur conversion en texte brut, afin de pouvoir les manipuler plus aisément. Une fois extrait, le corpus de contenu de Vikidia a été soumis à l'outil « Vocabprofile » sur Lextutor (Cobb, s.d.), afin de vérifier la quantité de vocabulaire commun (appartenant aux 1000 lemmes les plus fréquents) par rapport aux termes moins communs. À partir de cette évaluation sommaire (sans effectuer de contrôle sur le traitement des noms propres, par exemple), le taux de vocabulaire commun dans Vikidia (environ 70%) semblait comparable à celui de Corpatext (environ 73%), suggérant que le contenu de Vikidia pouvait être suffisamment accessible pour être utilisé comme corpus.

Malgré des similitudes quant au pourcentage de vocabulaire fréquent, le contenu de ces deux corpus présente aussi des caractéristiques assez différentes. D'un côté, Corpatext rassemble des textes libres de droit en se concentrant sur des ouvrages littéraires qui ne sont plus protégés par le droit d'auteur, dont des ouvrages rédigés au XIX^e siècle ou au début du XX^e siècle. Par exemple, dans la version condensée disponible sur Lextutor (Cobb, 2014b), on retrouve notamment des livres de Balzac, de Dumas, de Maupassant, de Proust et de Zola, des œuvres théâtrales de Molière, des recueils de poésie et des ouvrages philosophiques (Descartes). Dans la mesure où ce corpus rassemble en grande partie des œuvres littéraires et poétiques, on peut s'attendre à ce que certaines figures de styles et métaphores puissent être difficiles à

interpréter pour des apprenants du français langue seconde. De plus, ce corpus rassemble en grande partie des textes de type narratif, alors que l'apprenant doit se familiariser avec plusieurs autres types de textes (informatif, argumentatif, etc.) dans le cadre du curriculum de francisation. De l'autre côté, étant donné que Wikidia a été créé en 2006 (Association Wikidia, 2016), son contenu a été écrit plus récemment que la plupart des ouvrages incorporés à Corpatext. De plus, dans la mesure où la mission de ce projet est de produire du texte accessible à un lectorat entre huit et treize ans, on peut présumer que les auteurs ont opté pour une syntaxe et du vocabulaire plus accessibles. Toutefois, en l'absence d'un guide stylistique précis du côté de Wikidia, la simplification du texte et l'accessibilité du contenu sont laissés à la discrétion des auteurs qui participent à la rédaction (Association Wikidia, 2015); il est donc possible que la complexité des textes (vocabulaire et syntaxe) puisse varier significativement d'un article à un autre, ou que certains textes présentent des emplois erronés ou non prescrits de la langue française. Malgré certaines incertitudes quant à la qualité et à l'uniformité des textes contribués à Wikidia, étant donné la visée encyclopédique du projet, on peut s'attendre à retrouver davantage de textes descriptifs, ce qui peut être avantageux pour offrir du contenu différent de ce qui est offert dans Corpatext.

Pour récapituler, les questions et les exemples utilisées dans l'exercice de ce projet de recherche sont extraits de deux corpus francophones : le premier, Corpatext, rassemble des livres libres de droit de plusieurs genres (littérature, philosophie, poésie) datant pour la plupart du début du XX^e siècle ou plus tôt, et pouvant inclure des expressions ou des figures de styles difficiles à comprendre pour des apprenants du français langue seconde; le second, créé sur mesure à partir de l'encyclopédie libre Wikidia, rassemble des textes descriptifs à propos de différents sujets et personnalités et s'adressant à un lectorat de huit à treize ans, mais dont le contenu créé bénévolement pourrait présenter des erreurs ainsi qu'une variation au niveau de la syntaxe et du vocabulaire jugé « accessible ».

3.4 Comparaison du contenu des deux corpus

Avant d'utiliser le contenu des deux corpus linguistiques (Vikidia et Corpatext) dans une tâche destinée aux participants, une comparaison de leur vocabulaire a été effectuée afin d'évaluer la possibilité qu'il y ait une différence significative entre ceux-ci. Plutôt que de comparer l'ensemble des deux corpus, seul le contenu utilisé dans l'exercice (phrases trouées et listes d'exemples, voir la section 3.6 ci-dessous pour plus d'information) a été retenu pour la comparaison. Ce contenu utilisé a d'abord été inséré dans l'outil Vocabprofile disponible sur Lextutor (Cobb, s.d.), qui permet de séparer les mots selon les vingt-cinq catégories de 1000 lemmes en ordre de fréquence d'apparition. Plutôt que d'opter pour cette catégorisation détaillée, les mots ont été séparés en trois catégories : mots appartenant aux 1000 lemmes les plus fréquents, mots appartenant aux autres lemmes de fréquence, et mots hors-liste (y compris les noms propres).

Dans le cas du contenu extrait du corpus Vikidia, l'outil Vocabprofile a déterminé que pour l'ensemble de 1338 mots, 1054 (78,8%) provenaient des 1000 lemmes les plus fréquents, 235 (17,6%) étaient issus des autres fréquences, et 49 (3,7%) s'avéraient des termes hors-liste. En ce qui concerne le contenu extrait du corpus Corpatext, l'outil Vocabprofile a déterminé que pour l'ensemble de 1163 mots, 993 (85,4%) provenaient des 1000 lemmes les plus fréquents, 152 (13,1%) étaient issus des autres fréquences, et 18 (1,5%) s'avéraient des termes hors-liste.

Tableau 3.2 Caractéristiques des extraits de corpus

	Vikidia	Corpatext
Nombre de mots	1338	1163
Pourcentage de mots appartenant aux 1000 lemmes les plus fréquents	78,8%	85,4%
Pourcentage de mots appartenant aux autres fréquences	17,6%	22,2%
Pourcentage de mots hors des listes de lemmes les plus fréquentes	3,7%	1,5%

À partir de ces données, des analyses statistiques inférentielles ont été effectuées pour vérifier la possibilité de différence significative entre ces deux extraits de corpus. Le test khi-carré effectué pour comparer ces deux groupes ($\chi^2 = 21,82$, $p < 0,001$) a permis de confirmer qu'il existe une différence entre les deux corpus en ce qui concerne le type de vocabulaire employé. Proportionnellement parlant, le contenu issu de Corpatext présente davantage de vocabulaire issu des lemmes des 1000 termes les plus communs, tandis que celui provenant de Vikidia présente légèrement plus de termes moins communs ou hors-liste. Toutefois, il est important de garder en tête que le nombre élevé d'observations valides (2501 mots au total, en combinant le contenu issu des deux corpus) contribue à expliquer le résultat de ce test; le résultat du V de Kramer ($V = 0,09$) suggère que l'intensité du lien entre les deux variables (type de corpus et distribution du vocabulaire) est relativement faible. Une piste d'explication possible est que, comme la mission de Vikidia est de fournir une encyclopédie en ligne et donc d'expliquer différentes notions et concepts d'une certaine complexité, on peut s'attendre à retrouver une certaine proportion de termes moins fréquents, que les contributeurs de Vikidia s'efforcent ensuite d'expliquer de façon aussi simple et accessible que possible. De l'autre côté, même si

le contenu de Corpatext semble contenir davantage de vocabulaire fréquent, les textes utilisés (soit des ouvrages littéraires libres de droit) pourraient néanmoins comprendre des passages plus difficiles à interpréter, notamment sur le plan syntaxique.

3.5 Développement de l'outil

Exercice #2

Écris un petit mot dans chaque .

Question #01
 Une colonie de vacances est un centre de vacances pour les enfants. Ce centre fonctionne les vacances scolaires, de juin à août.

Exemples
 Clique sur ou sur pour avoir des exemples.

Vous avez fini cette question? Cliquez sur !

Figure 3.3 Phrase trouée à compléter dans l'exercice sur les prépositions

Une fois l'objet langagier choisi, l'outil a été développé de façon à permettre aux apprenants d'analyser des cooccurrences qui contribueraient à guider l'usage des prépositions dans l'exercice. Ainsi, pour chaque phrase trouée à compléter avec une préposition appropriée, l'exercice mettait à la disposition des apprenants des phrases exemples associant des cooccurrences entre prépositions et mots-clefs (verbe, nom), à raison de deux mots-clefs et de quatre phrases exemples par mot-clef. L'analyse de ces cooccurrences aiderait à choisir la préposition appropriée pour compléter la phrase trouée.

Pour chaque version de l'exercice, il y a huit phrases à compléter, en s'appuyant sur les combinaisons de mots-clés et de prépositions pour lesquelles le chercheur a trouvé suffisamment de phrases appropriées dans le corpus (à la fois pour l'exercice et pour les exemples mis à la disposition des apprenants), selon les critères énoncés dans la section précédente (deux ensembles de quatre exemples pour chaque phrase trouée à compléter). Dans la mesure du possible, les phrases trouées employées dans ces deux exercices cherchent à couvrir les mêmes combinaisons de mots-clés et de prépositions (« jouer de » et « jouer à », « parler de » et « parler à », « travailler à » et « travailler pour », « depuis » et « pendant » avec un indicateur de temps, etc.), afin de permettre des analyses comparatives spécifiques entre les deux groupes, mais comme les combinaisons de mot-clé et de préposition choisies étaient souvent plus communes dans un corpus ou dans l'autre, certaines combinaisons apparaissent plus fréquemment dans un exercice ou dans l'autre.

Dans d'autres recherches consacrées à l'AGD, les auteurs soulignent certaines difficultés que les étudiants peuvent rencontrer. De prime abord, pour les apprenants de niveau débutant ou intermédiaire inférieur, le décodage d'une grande quantité de contenu (dans ce cas, les exemples tirés du corpus) pour trouver des similitudes peut nécessiter un grand effort mental (Yoon, 2011; Chambers, 2005; Kennedy et Miceli, 2001). De surcroît, plusieurs facteurs peuvent créer un état de surcharge cognitive : en plus de se familiariser avec une approche inédite (l'apprentissage guidé par les données), les participants doivent apprendre à utiliser l'interface d'un nouveau logiciel pour extraire le contenu du corpus, tout en réalisant d'autres tâches (rédaction de textes, correction d'erreurs, etc.) (Gordani, 2013; Boulton, 2010a). Comme ce projet de recherche s'adresse à des participants de niveau intermédiaire inférieur, quelques ajustements ont été effectués afin de réduire la possibilité que la complexité du contenu ou que la surcharge cognitive empêche les participants de profiter de l'outil et de l'activité en général.

En s'inspirant des travaux de Boulton (2008), qui avait travaillé avec des apprenants de niveau débutant en choisissant et en imprimant des listes de cooccurrences formatées à l'avance, et étant donné le temps très limité en compagnie des participants, il fut décidé de choisir en amont les exemples retenus pour l'exercice sur les prépositions, plutôt que de donner accès au corpus complet à travers des fonctions de recherche. Cette approche est également privilégiée par Warschauer et collègues (2017) : dans leur tâche de rétention de vocabulaire, les extraits de corpus sont choisis de façon à ce que les participants soient en mesure de saisir le sens du mot à partir des mots-clefs environnants, et les extraits contenant trop de mots inconnus sont laissés de côté. Cette stratégie de sélectionner des extraits de corpus à l'avance est également identifiée dans les travaux de Chambers (2010) comme une forme indirecte d'interaction avec le corpus, « permettant à l'apprenant d'analyser des concordances avant d'être obligé de maîtriser l'outil qu'est le logiciel de concordance » (p.15), tout en maintenant l'approche constructiviste et le développement de l'autonomie. Ainsi, afin de conserver les principes directeurs de l'AGD, ces séries d'exemples ont été choisis selon plusieurs paramètres – à partir d'un mot-clef, d'un verbe ou d'une préposition – afin que les apprenants doivent interpréter les séries d'exemples et trouver eux-mêmes les cooccurrences entre verbes et prépositions. Néanmoins, tel que suggéré par Oghigian et Chujo (2010), les phrases jugées trop difficiles pour les apprenants (par la présence de vocabulaire moins commun ou l'emploi de figures de style complexes), ou qui pourraient les induire en erreur (formulations erronées dans des articles de Wikidia ou emplois désuets dans certains textes plus anciens de Corpatext) ont été retirées afin de diminuer la charge cognitive liée à la lecture des exemples. De plus, comme les apprenants dans certaines recherches avaient de la difficulté à lire les phrases alignées par mot-clef (affichage dit « mot-clef en contexte », ou *keyword in context* en anglais), les exemples ont plutôt été affichés comme phrases complètes. Comme les mots-clefs n'étaient pas alignés, ceux-ci étaient plutôt affichés en gras dans les phrases exemples pour faciliter leur analyse par les apprenants.

Exemples
 Cliquez sur **fonctionne** ou sur **vacances** pour avoir des exemples.

Exemples: vacances
 Pendant le mois d'octobre, elle a été dans sa maison de **vacances** dans la région de l'Alentejo.

Pendant les vacances de printemps et les **vacances** d'été (début juillet et fin août), les élèves de CM1 et CM2 peuvent participer à des stages de remise à niveau. Ces stages proposent 3 heures d'enseignement supplémentaires pendant cinq jours pour aider les élèves qui auraient des difficultés plus importantes.

Il était parti en **vacances** et il n'avait pas pensé à s'inscrire.

Dans ce film, Bruce Willis s'appelle John McLain, il est un lieutenant de police en **vacances** pendant deux semaines.

Vous avez fini cette question? Cliquez sur **Suivant** !

Figure 3.4 Liste d'exemples pour le mot-clé « vacances », qui s'affiche à l'écran après avoir appuyé sur le bouton correspondant

Dans un même ordre d'idées, l'interface de l'outil a été développée en s'inspirant d'articles explorant différentes approches pour réduire la surcharge cognitive. Dans cette optique, en s'inspirant des recherches de Chandler et Sweller (1996) et de Mayer et Moreno (2003), il fut décidé de rassembler à l'écran les différentes composantes de l'exercice : les instructions, les phrases trouées à compléter, et les exemples tirés du corpus. Dans un premier temps, pour réduire la possibilité de confondre l'apprenant avec des informations accessoires, les instructions ont été ajustées pour préférer les mots communs au métalangage, par exemple en remplaçant « préposition » par « petit mot », en s'inspirant des recherches de Roy (1998, 1997). Ensuite, afin d'éviter que les apprenants soient déroutés par l'affichage simultané d'une grande quantité de texte à l'écran, les listes d'exemples

pour les différents mots clés ne s'affichent que lorsque l'utilisateur décide de les consulter en appuyant sur un bouton correspondant. Ainsi, les apprenants peuvent se concentrer sur la phrase trouée à compléter, en se référant au corpus au besoin.

Bien que certaines recherches aient priorisé des activités de correction d'erreurs à partir de consultation de corpus (O'Sullivan et Chambers, 2006; Chambers et O'Sullivan, 2004; Gaskell et Cobb, 2004; Sun et Wang, 2003; Todd, 2001), ce projet a plutôt opté pour une formule de phrases trouées dans lesquelles l'étudiant devait inscrire la préposition manquante. Cette approche de phrases trouées, qui avait déjà été utilisée dans d'autres recherches (Gordani, 2013; Landure, 2011; Boulton, 2010a; Boulton, 2009b), permettait de centrer sans ambiguïté l'activité sur l'utilisation des prépositions. Comme les participants sont des apprenants de niveau intermédiaire inférieur avec une connaissance limitée du métalangage, il est possible que ces derniers aient de la difficulté à identifier les prépositions, et à évaluer si celles-ci étaient appropriées ou non. Dans une tâche de correction, il y aurait également possibilité que les apprenants essaient de trouver d'autres erreurs (fautes d'orthographe ou de conjugaison, par exemple); la formule de la phrase trouée permettait donc de centrer l'exercice sur l'objet langagier choisi, et de maintenir un niveau de difficulté adapté aux participants.

Questionnaire

**J'aimerais mieux vous connaître !
I would like to know you better!**

<p>1) J'ai... I am...</p> <p><input type="text" value="40"/> ans years old</p>	<p>2) Je suis... I am...</p> <p><input type="radio"/> ...un homme ...a man <input type="radio"/> ...une femme ...a woman</p>	<p>3) Je suis arrivé au Québec le... I arrived in Quebec on...</p> <p><input type="text" value="5 juin 2014"/></p>
<p>4) J'étudie le français depuis... I have been studying French for...</p> <p><input type="text" value="5"/> ans years</p>	<p>5) J'ai étudié en... I studied in...</p> <p><input type="text" value="français"/></p>	<p>6) Je travaille (ou je veux travailler) comme... I work (or I plan to work) as a...</p> <p><input type="text" value="professeur"/></p>

**Quel est votre style d'étude ?
What is your studying style?**

7) J'aime apprendre les règles en premier, puis pratiquer avec des exemples en deuxième.
I like to learn the rules first, then practice with examples after.

Je ne suis pas d'accord / I do not agree 1 2 3 4 Je suis d'accord / I agree

Figure 3.5 Extrait du questionnaire bilingue présenté aux participants

Par ailleurs, le questionnaire a aussi été conçu de façon à être facile d'utilisation, en simplifiant le vocabulaire lorsque possible, par exemple en priorisant les affirmations à la première personne pour les questions démographiques (« Je suis un homme / une femme », « J'ai X ans », etc.). Comme le questionnaire ne visait pas à évaluer les compétences des participants en français, les questions étaient affichées simultanément en français et en anglais, pour favoriser la compréhension du contenu en offrant une traduction dans une langue plus familière à la plupart des participants.

Bref, à la lumière des obstacles potentiels à l'utilisation de l'AGD par des apprenants de niveau débutant ou intermédiaire notés dans la littérature, plusieurs ajustements ont été effectués pour tenter de diminuer la possibilité de surcharge cognitive chez les participants. Ainsi, l'exercice a priorisé la complétion de phrases trouées plutôt que la correction d'erreurs, les exemples ont été choisis à l'avance

plutôt qu'obtenus par des recherches manuelles dans les corpus, l'interface a été conçue de façon à offrir tout le contenu à partir d'un même médium, et le contenu du questionnaire a été simplifié et traduit en anglais. Toutes ces décisions visaient à faciliter la compréhension du contenu et à rendre l'outil accessible aux participants, malgré leur niveau limité de compétence en français.

3.6 Enregistrement des actions

Afin d'obtenir une autre source de données au-delà de l'exercice sur les prépositions et du questionnaire, les actions des participants lors de leur utilisation des outils ont été enregistrées, tel que suggéré par Perez-Paredes et collègues (2011). En plus des réponses au questionnaire, ces mesures directes contribueraient à évaluer l'appréciation et la pertinence de ces outils pour les participants. Cet ajout s'inspire de la recherche de Gaskell et Cobb (2004), où les accès à un logiciel de cooccurrences en ligne étaient enregistrés. Selon ces données, les étudiants accédaient davantage aux listes de cooccurrences et aux formulaires d'analyses d'erreur lorsqu'ils avaient accès à des liens pré-formatés vers ces ressources, « mais sans les liens, moins de la moitié des étudiants l'ont fait, avec toutefois de possibles indices d'un renversement de la tendance plus tard durant le cours, tandis que les apprenants devenaient plus confiants vis-à-vis de la technologie [notre traduction] » (p.311). En enregistrant les actions des participants, cela permettrait d'évaluer si l'utilisation de l'outil était la même tout au long de la tâche, ou si des variations seraient observées, par exemple, une baisse d'utilisation des exemples potentiellement associée à une diminution de l'intérêt des participants. On retrouve aussi cette approche de suivi des actions des participants dans la recherche de Warschauer et collègues (2017) : étant donné que « le traitement des lignes de cooccurrence comme information de glossaire requière un niveau relativement élevé de charge cognitive pour les apprenants [notre

traduction] » (p.36), il est possible que les apprenants ne soient pas autant motivés à consacrer les efforts nécessaires à la compréhension des cooccurrences et qu'ils favorisent plutôt les ressources nécessitant moins d'efforts cognitifs, comme la consultation de définitions. Les chercheurs ont donc intégré des outils permettant de mesurer le temps passé sur les différentes références lexicales accessibles (cooccurrences, définitions), et ajouté un système permettant aux apprenants d'indiquer quelles cooccurrences étaient jugées compréhensibles, afin de vérifier si les interactions avec les références lexicales variaient selon la difficulté des termes à définir. À la lumière de leurs observations, il semblerait que les apprenants se fient à leurs connaissances pour les questions qui leur paraissent évidentes, et qu'ils consultent davantage les ressources pour les éléments moins familiers.

Comme les outils de ce projet de recherche ont été développés pour être accessibles à partir d'un navigateur web, il était relativement simple d'intégrer les technologies de suivi de Google Analytics pour obtenir un aperçu des interactions effectuées par les participants. Notamment, avec les outils de « Google Tag Manager », il s'avère aisément possible d'enregistrer tous les clics effectués sur une page, à partir des identifiants CSS (*Cascading Style Sheets*) attribués aux différents éléments qui la composent. Comme les exemples tirés des corpus sont affichés lorsque l'utilisateur clique sur un bouton, il était pertinent d'enregistrer les clics sur les boutons affichant les listes d'exemples, en leur attribuant des identifiants CSS distincts. En enregistrant le nombre de clics sur ces boutons, cela permettrait de vérifier pour quelles questions les participants choisissent de consulter les exemples, ou s'il y a une tendance qui se développe à travers le temps (par exemple, si les exemples sont consultés au début de l'activité, puis consultés de moins en moins au fur et à mesure que l'activité progresse). Par la même occasion, le temps passé sur l'exercice et sur le questionnaire est également enregistré par Google Analytics, afin

de pouvoir vérifier s'il y avait des valeurs aberrantes (soit très peu, ou beaucoup de temps passé sur les outils).

Dimension principale : Libellé d'événement Autre ▾				
Dimension secondaire ▾ Type de tri : Paramètre par défaut ▾ small round button avancé				
Libellé d'événement	Nombre total d'événements	Événements uniques	Valeur de l'événement	Valeur moyenne
	183 % du total: 5,03 % (3 640)	100 % du total: 6,77 % (1 477)	0 % du total: 0,00 % (916)	0,00 Valeur moy. pour la vue: 0,25 (-100,00 %)
<input type="checkbox"/> 1. small round button Londres	20 (10,93 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 2. small round button aller	14 (7,65 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 3. small round button établi	13 (7,10 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 4. small round button Europe	13 (7,10 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 5. small round button mort	13 (7,10 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 6. small round button travailler	13 (7,10 %)	6 (6,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 7. small round button travaille	12 (6,56 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 8. small round button cinq ans	11 (6,01 %)	6 (6,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 9. small round button golf	11 (6,01 %)	7 (7,00 %)	0 (0,00 %)	0,00
<input type="checkbox"/> 10. small round button Paris	11 (6,01 %)	5 (5,00 %)	0 (0,00 %)	0,00

Afficher les lignes : 10 Accéder à : 1 1 à 10 sur 17

Figure 3.6 Données de Google Analytics (affichage des listes d'exemples)

En somme, comme l'exercice et le questionnaire ont été développés pour être accessibles par navigateur web, il était relativement aisé d'intégrer des technologies existantes (dans ce cas, la plate-forme de Google Analytics) pour documenter de façon détaillée comment ces outils étaient employés, de façon à fournir des données complémentaires aux renseignements déjà recueillis (soit les réponses dans l'exercice et dans le questionnaire).

3.7 Développement du questionnaire

Comme dans plusieurs autres recherches consacrées à l'AGD (Landure et Boulton, 2010; Chang et Sun, 2009; Chujo *et al*, 2006; Yoon et Hirvela, 2004), les participants devaient compléter un questionnaire après avoir complété l'activité. Au-delà des données extraites par le biais de l'activité (compétences langagières, utilisation de l'outil, variation possible selon le type de corpus offert), le questionnaire permet de récolter des données complémentaires, qui peuvent contribuer à expliquer les variations possibles entre les deux groupes de participants. Étant donné que l'activité sur les prépositions était déjà complétée à l'ordinateur, le questionnaire a suivi un format relativement semblable, et s'affichait immédiatement après avoir complété l'activité.

En s'inspirant des questionnaires trouvés dans d'autres articles (Boulton, 2009a; Chang et Sun, 2009; O'Sullivan et Chambers, 2006; Yoon et Hirvela, 2004), le questionnaire de ce projet de recherche (disponible en appendice C) comprend vingt-deux questions, séparées en quatre catégories : données sociodémographiques de base (comme l'âge et le genre), style d'apprentissage préféré (approche communicative, enseignement explicites des règles grammaticales, etc.), relation à la technologie (temps consacré à l'ordinateur, lieu d'utilisation, et tâches accomplies), et appréciation générale de l'activité (facilité des exemples, utilité perçue du logiciel, etc.). Les questions exigeaient soit des réponses courtes (comme l'âge, la date d'arrivée au Québec, le domaine d'études, l'emploi recherché) ou d'inscrire leur degré d'accord ou de désaccord sur une échelle de Likert (par exemple : facilité de lecture des exemples, utilité du logiciel).

Les données recueillies par ce questionnaire avaient l'objectif de compléter et d'approfondir l'analyse des données recueillies avec l'activité : dans un premier

temps, la comparaison des données permettrait de vérifier si les deux groupes étaient relativement similaires. Dans un second temps, dans l'éventualité où il y aurait une variation significative entre les résultats à l'exercice des deux groupes, certaines données recueillies dans le questionnaire, comme les préférences d'apprentissage ou la relation à la technologie, pourraient contribuer à expliquer si cette variation provenait de l'utilisation de deux corpus différents, ou de différences entre les deux groupes. En effet, certains auteurs (Boulton, 2009a; Gilmore, 2009) ont noté que ces aspects pouvaient influencer sur l'intérêt et sur la motivation des participants pour compléter une activité d'AGD.

3.8 Déroulement de la collecte de données

Étant donné le temps et les ressources limitées, ce projet a été restreint à deux rencontres avec les participants, chacune d'une durée d'approximativement deux heures.

La première rencontre a eu lieu durant l'après-midi du samedi 28 novembre 2015, un moment jugé propice par l'enseignante responsable du cours; alors que les cours donnés les soirs de semaine priorisent l'enseignement explicite s'appuyant sur un manuel et des activités, les cours donnés le samedi incorporent à l'occasion des activités spéciales comme la présentation de logiciels d'apprentissage de la langue (Rosetta Stone, par exemple) ou le visionnement de contenu vidéo (films traduits, extraits d'émissions télévisuelles, etc.). Il y avait onze participants à l'occasion de cette rencontre, soit environ la moitié du groupe total; selon l'enseignante, certains étudiants moins motivés ne sont pas aussi assidus pour les cours du samedi, et optent à l'occasion de s'absenter ou d'arriver en retard.

L'objectif principal de la première rencontre était de réaliser les actions qui permettraient d'effectuer la cueillette de données à l'occasion de la seconde rencontre : présenter le projet de recherche, distribuer les formulaires de consentement, et effectuer un pilotage d'un prototype de l'outil pour finaliser le contenu de la version finale de l'exercice. Dans un premier temps, les participants ont assisté à une courte présentation donnée par le chercheur pour découvrir les caractéristiques du projet de recherche auquel ils étaient invités, suivi d'une séance de questions-réponses. Par la suite, les formulaires de consentement (en français et en anglais) ont été distribués aux participants pour obtenir la signature de ceux-ci. Au besoin, le chercheur et l'enseignante responsable étaient disponibles pour répondre aux questions des étudiants quant au contenu du formulaire.

Dans un second temps, les étudiants se sont installés aux ordinateurs disponibles dans la salle de classe pour compléter le prototype d'exercice préparé par le chercheur, rassemblant des phrases à compléter extraites des deux corpus utilisés. L'objectif principal de cette première rencontre était de vérifier à quel point les prépositions choisies étaient déjà familières pour les participants – si certaines prépositions étaient parfaitement maîtrisées ou complètement inconnues, elles seraient alors retirées de l'exercice à l'occasion de la deuxième rencontre. Pour diminuer la possibilité que les étudiants puissent se consulter et échanger des réponses entre eux durant cette activité, les phrases à compléter étaient présentées dans des ordres différents aux étudiants qui étaient assis côte à côte. De plus, le chercheur indiquait aux étudiants que les exercices présentaient des questions différentes.

Une fois l'activité terminée, l'enseignante responsable du cours de francisation a réalisé une séance de rétroaction avec les étudiants, pour confirmer les prépositions appropriées pour chaque phrase utilisée dans le prototype d'exercice, et

fournir des explications supplémentaires pour les règles grammaticales guidant le choix de préposition.

Dans l'ensemble, les participants ont obtenu en moyenne 40% de bonnes réponses dans le prototype d'exercice, et l'ensemble de phrases à compléter a donné des résultats variés (entre 9% et 82% de bonnes réponses, selon la question). À partir de cette séance de pilotage de l'outil, il fut décidé d'utiliser approximativement le même ensemble de combinaisons de mots-clés et de prépositions, en fonction du nombre suffisant d'exemples pertinents qui pourraient être trouvés dans les deux corpus. Plus précisément, pour chaque phrase à compléter dans l'exercice, l'objectif était de trouver deux mots-clés pouvant guider les apprenants, et de trouver quatre phrases pouvant être utilisées comme exemples pour chaque mot-clé. Cette quantité d'exemples semblait appropriée pour donner une vue d'ensemble des possibilités de combinaisons de mot-clé et de préposition, tout en évitant de surcharger avec une grande quantité de contenu à analyser.

La deuxième rencontre a eu lieu l'après-midi du samedi 12 décembre 2015, qui correspondait au dernier cours du programme de francisation avant l'examen final. À cette occasion, dix-neuf étudiants étaient présents; selon l'enseignante responsable, ces derniers étaient motivés à prendre part aux activités de révision (dont la participation au projet de recherche, qui représentait l'occasion de revoir certaines formules grammaticales pratiquées durant le programme, et d'évaluer leur capacité à lire des extraits de textes authentiques), ce qui pourrait expliquer le plus grand nombre d'étudiants présents en comparaison à la première rencontre. Le jour de la rencontre était à la mi-décembre et des activités du temps des Fêtes avaient lieu à proximité; il est possible que le bruit provenant de ces festivités ait été une source de déconcentration pour certains des participants.

De fait, comme plusieurs étudiants étaient absents à l'occasion de la première rencontre, la rencontre a commencé avec la présentation du projet de recherche et la distribution de formulaires de consentement à ceux qui n'avaient pas eu l'occasion de le signer deux semaines plus tôt.

Étant donné que l'examen final avait lieu quelques jours plus tard, certains étudiants étaient anxieux à l'idée de compléter l'exercice sans accompagnement. Afin de réactiver les connaissances antérieures et de diminuer l'aspect anxiogène de l'activité, l'enseignante responsable a décidé d'effectuer d'abord une révision des unités du manuel consacrées à l'utilisation des prépositions, en employant une approche d'enseignement explicite et en révisant avec le groupe les règles et les exercices présents dans le manuel.

Après cette activation des connaissances antérieures, les participants se sont installés aux ordinateurs pour compléter l'exercice sur l'utilisation des prépositions, soit des phrases trouées à compléter avec les prépositions appropriées. À partir du groupe initial de participants pour ce projet de recherche (se référer à la section 3.1 pour plus d'information à ce sujet), deux sous-groupes ont été créés au moment de recueillir les données. Une fois installés à leurs postes informatiques, les participants étaient assignés aux deux groupes de façon systématique : en partant du premier participant, le chercheur a assigné chaque étudiant en alternance, soit au contenu associé au corpus de Vikidia, soit au contenu associé au corpus de Corpatext. De cette façon, les étudiants ne seraient pas en mesure de consulter leurs voisins immédiats. Avec cette approche, le groupe associé au contenu de Vikidia s'est retrouvé avec dix participants, tandis que le groupe associé au contenu de Corpatext s'est retrouvé avec neuf participants. Tout comme dans la première rencontre, les étudiants qui étaient côte à côte n'avaient pas accès au même contenu, pour éviter la possibilité d'échanger des réponses. Dans le cas de la deuxième rencontre, les participants avaient soit accès

au contenu (phrases trouées et exemples) provenant du corpus Corpatext, soit au contenu provenant du corpus Wikidia.

Après avoir complété les huit questions de l'exercice sur les prépositions, les participants étaient automatiquement redirigés vers le questionnaire à compléter. Comme les étudiants complétaient ces tâches à un rythme différent, ceux-ci étaient invités à prendre une pause après avoir terminé l'exercice et le questionnaire.

Une fois l'exercice et le questionnaire complétés par tous les participants, le contenu de l'exercice a été vérifié et révisé avec l'enseignante responsable du cours. Cela a permis aux étudiants de vérifier si leur réflexion pour l'utilisation des prépositions était juste, et quelles formules grammaticales devraient être révisées avant l'examen final. Dans l'ensemble, après avoir complété l'activité de la révision, les étudiants semblaient rassurés que leur niveau de compréhension du français était adéquat pour l'examen final.

3.9 Considérations éthiques

Tel qu'indiqué dans la lettre de contact et dans le formulaire de consentement (tous deux disponibles en appendice), une attention particulière a été accordée à la communication des paramètres du projet de recherche aux participants. Lors de la présentation du projet, il fut clairement indiqué que la participation au projet était volontaire, qu'elle n'impliquait aucune compensation ou avantage, et qu'elle n'aurait aucune incidence sur leurs résultats au cours de francisation. Dans cette optique, les rencontres ont eu lieu deux samedis en début d'après-midi, une plage horaire à laquelle l'enseignante avait l'habitude d'insérer des activités complémentaires (des visionnements de films, par exemple). Comme les participants étaient libres

d'omettre toute réponse ou d'arrêter l'activité à tout moment, l'exercice sur les prépositions et le questionnaire ont été conçus de façon à accepter des réponses incomplètes ou vides.

Afin de préserver la confidentialité des participants, les données ont été recueillies sans pouvoir être associées à des renseignements personnels qui pourraient permettre d'identifier les participants. Les informations entreposées sur le serveur (résultats de l'exercice et réponses au questionnaire) ont été conservées sur un dossier non-accessible au public, puis transférées dans des fichiers encryptés et protégés par mot de passe. À partir de ces informations, les données ont été agrégées pour représenter les deux groupes dans leur ensemble, plutôt que d'isoler des participants spécifiques.

Finalement, étant donné que chaque participant avait une relation différente vis-à-vis de la technologie et de l'utilisation d'un ordinateur au quotidien et dans les études, il y avait possibilité que l'utilisation de la technologie durant une activité d'apprentissage puisse être anxiogène pour certains participants moins habitués à travailler à l'ordinateur. Afin de diminuer le stress que cette activité pourrait engendrer, une démonstration était effectuée avant le début de l'activité, et le chercheur demeurait présent durant la complétion de l'exercice et du questionnaire pour assister les participants qui auraient des difficultés techniques.

3.10 Analyse des données

Pour la récolte des données tirées de l'exercice sur les prépositions et du questionnaire, étant donné l'envergure modeste du projet et le petit nombre de participants, aucun filtre n'a été effectué en amont sur les données enregistrées. De

prime abord, plusieurs prépositions étaient acceptables pour certaines phrases trouées; de plus, comme l'objectif était d'évaluer le choix de préposition, pour donner le plus de latitude possible aux participants, il fut décidé de ne pas prendre en compte les fautes de frappe ou d'orthographe, en autant que la préposition choisie demeurait claire. Au fur et à mesure que les participants complétaient l'exercice et le questionnaire, les données brutes étaient enregistrées dans un dossier protégé sur le serveur hébergeant ces activités, à raison d'un fichier par participant. À partir de ces données enregistrées sur le serveur, les réponses ont été vérifiées et formatées dans une feuille de calcul Excel par le chercheur, en s'appuyant sur le corrigé conçu en collaboration avec l'enseignante responsable du programme de francisation, selon les critères énoncés ci-dessus (plusieurs prépositions acceptées pour certaines phrases trouées, fautes de frappe ou d'orthographe ignorées si le choix de préposition restait clair). Par la suite, les réponses au questionnaire ont également été extraites du serveur et formatées dans une feuille de calcul Excel, et les données produites par Google Analytics ont également été exportées dans ce format. Une fois rassemblées, ces données ont permis d'effectuer des analyses inférentielles pour comparer les deux groupes : résultats selon le contenu du corpus, perceptions et interactions avec le dispositif.

CHAPITRE IV

RÉSULTATS

Premièrement, afin de mieux cerner le contexte de l'expérimentation, certaines caractéristiques des participants seront présentées à partir des éléments du questionnaire concernant leurs préférences en tant qu'apprenants. Deuxièmement, les résultats des deux groupes à l'exercice de phrases trouées seront présentés et comparés : d'abord à partir des moyennes des deux groupes, puis en se concentrant sur des éléments spécifiques qui montrent des similarités. Troisièmement, les perceptions des participants sur le dispositif utilisé dans l'exercice (listes d'exemples tirés de corpus linguistiques) seront détaillées. Finalement, en marge des résultats à l'exercice, les données d'utilisation de l'outil, récoltées à partir de Google Analytics, seront analysées pour vérifier s'il y avait une tendance pour la consultation des listes d'exemples au fur et à mesure que la tâche progressait. Comme le nombre de participants est inégal entre les deux groupes (dix dans le premier groupe et neuf dans le deuxième groupe), les résultats sont présentés sous forme de pourcentage, afin de faciliter la comparaison.

4.1 Profil des participants

Dans le questionnaire rempli après avoir complété l'exercice, certaines questions visaient à cerner les préférences des participants en tant qu'apprenants, afin de vérifier s'il y avait possiblement d'autres différences entre les deux groupes, au-delà des caractéristiques du genre et de l'âge présentées dans le tableau 1 de la section 3.1, pour lesquelles aucune différence significative n'a été décelée.

Parmi ces questions, les participants devaient indiquer par l'intermédiaire d'une échelle de Likert s'ils préféraient apprendre les règles en premier pour ensuite prendre connaissance des exemples, à l'opposé d'une approche où les apprenants devaient concevoir les règles à partir des exemples; en combinaison avec d'autres éléments du questionnaire, cette question tentait de situer les apprenants sur un continuum entre les approches déductives et inductives. Dans le premier groupe, 80% des participants accordaient la priorité à l'apprentissage des règles (60% « très en accord » et 20% « en accord ») contre 20% en désaccord, tandis que 66,7% étaient en accord (33,3% « très en accord » et 33,3% « en accord ») et 33,3% en désaccord dans le deuxième groupe.

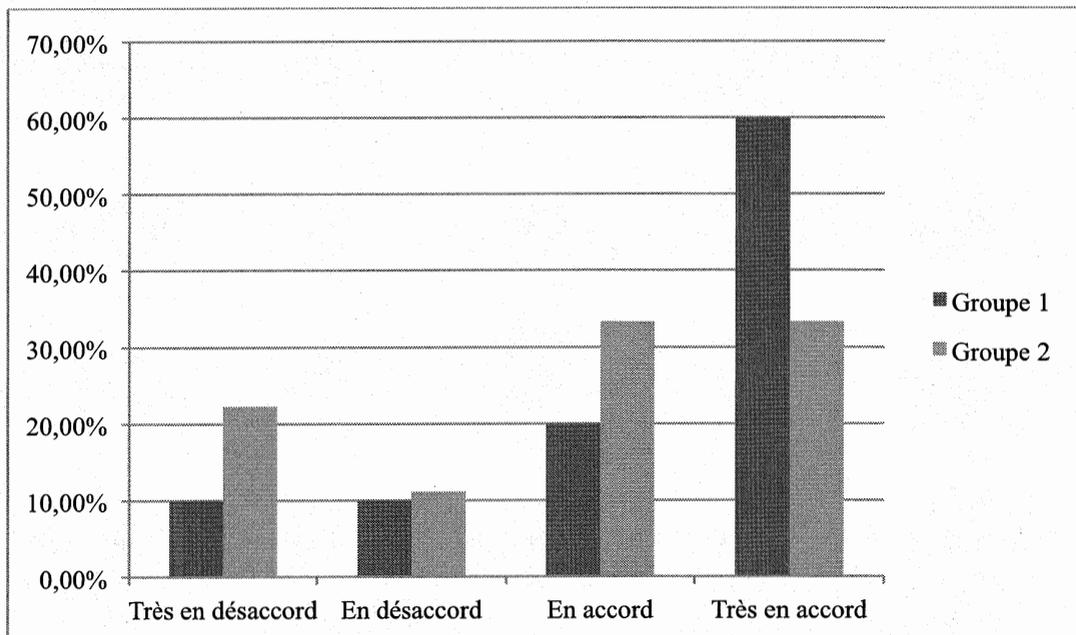


Figure 4.1 Préférence pour l'apprentissage des règles

Le test de Fisher a été utilisé pour vérifier si ces distributions permettaient de distinguer les deux groupes de participants; comme la signification exacte unilatérale

était de $p = 0,32$, ceux-ci ne présentent pas de différence significative pour cet aspect du questionnaire.

Dans un autre ordre d'idées, il est possible que la préférence pour une approche spécifique ou pour un média particulier (activités sur papier ou à l'ordinateur, par exemple) puisse être associée à un intérêt plus ou moins marqué pour une démarche d'AGD. Cet aspect était exploré de plusieurs façons dans le questionnaire, à commencer par une question demandant aux participants d'indiquer leur préférence pour l'apprentissage de langue à l'ordinateur à l'aide d'une échelle de Likert.

Pour cet élément, on remarque dans le premier groupe une distribution moitié-moitié entre ceux qui sont en accord et ceux qui sont en désaccord avec l'affirmation d'une préférence pour l'apprentissage de langue à l'ordinateur : 20% « très en accord », 30% « en accord », 30% « en désaccord » et 20% « très en désaccord ». Dans le deuxième groupe, 66,7% des participants sont en accord (11,1% « très en accord » et 55,5% « en accord »), contre 33% en désaccord (11,1% « en désaccord », 22,2% « très en désaccord »).

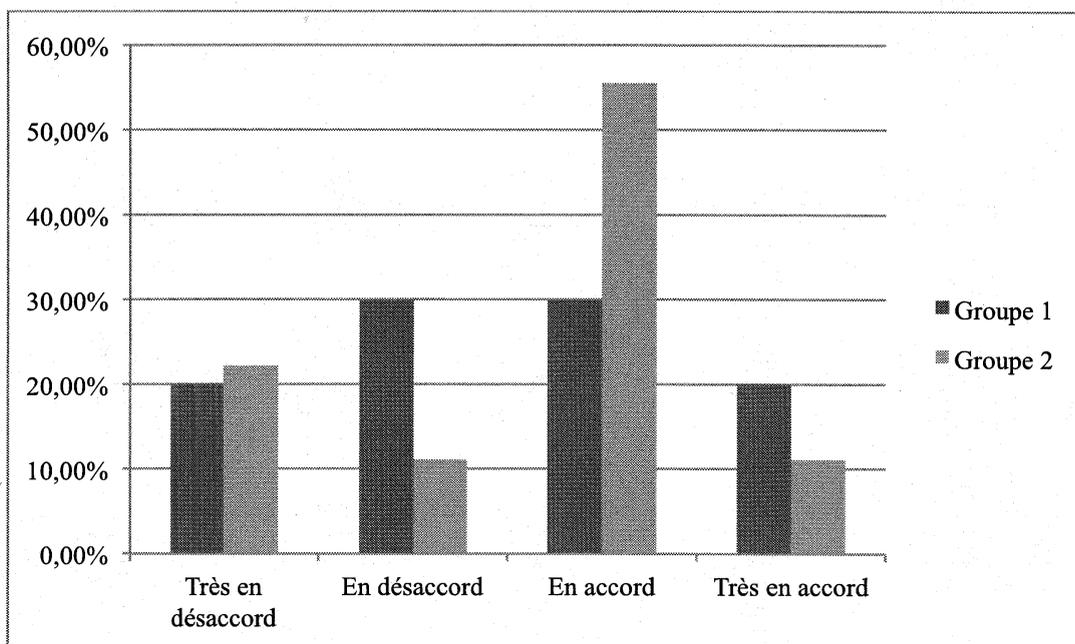


Figure 4.2 Préférence pour l'apprentissage de langue à l'ordinateur

Au-delà de la différence apparente, avec un plus grand pourcentage de participants exprimant leur préférence pour l'apprentissage de langue à l'ordinateur dans le groupe 2 (66,7%, contre 50% dans le groupe 1), le résultat du test de Fisher ($p = 0,91$ pour la signification exacte unilatérale) confirme l'absence de différence significative entre les deux groupes.

En plus de mesurer la préférence pour l'apprentissage de langue à l'ordinateur, le questionnaire demandait également aux participants dans quels milieux l'ordinateur était utilisé, avec trois possibilités : à la maison, à l'école et au travail. Dans l'ensemble, tous les participants ont indiqué avoir recours à l'ordinateur dans au moins un de ces milieux, dont 60% utilisant cet outil dans deux ou même trois des contextes énumérés.

Les réponses récoltées dans les deux groupes se sont avérées plutôt similaires pour l'utilisation à domicile (aux environs de 90%) et dans le cadre d'un emploi (aux environs de 20%). Seul le contexte scolaire semble présenter une différence, avec 60% des participants du groupe 1 répondant qu'ils utilisent l'ordinateur à l'école, contre 88,9% dans le groupe 2. À cet effet, bien que tous les étudiants inscrits au programme de francisation aient accès à des ordinateurs dans leur salle de classe avec des logiciels d'apprentissage de langues (Rosetta Stone, par exemple), l'utilisation de ces ressources était facultative et laissée à la discrétion des étudiants durant les périodes « autonomes » (c'est-à-dire sans activités dirigées par l'enseignante responsable). Ainsi, il est possible que certains participants aient opté pour d'autres activités, comme réviser le contenu des leçons effectuées dans le manuel, plutôt que d'utiliser les logiciels d'apprentissage de langues mis à leur disposition.

Tableau 4.1 Utilisation de l'ordinateur dans différents milieux

	Groupe 1	Groupe 2
À la maison	90,0%	88,9%
À l'école	60,0%	88,9%
Au travail	20,0%	22,2%

Malgré l'écart observé entre les deux groupes vis-à-vis de l'utilisation de l'ordinateur en milieu scolaire, le résultat du test de Fisher ($p = 0,30$, signification exacte bilatérale) n'a pas permis de discerner de différence significative pour cet aspect. Comme les écarts étaient minimes pour les deux autres milieux (maison et travail) et que la différence entre les deux groupes n'était pas significative pour l'utilisation de l'ordinateur à l'école, des tests statistiques n'ont pas été effectués pour ces deux autres milieux.

Finalement, les participants devaient indiquer dans le questionnaire s'il s'agissait de la première fois qu'ils utilisaient une approche d'AGD. Dans le groupe 1, 80% utilisaient ce type d'outil pour la première fois; c'était également le cas pour 66,7% des participants du groupe 2. Aucune analyse statistique n'a été effectuée pour vérifier si cet écart était significatif; ces données sont présentées à titre descriptif.

4.2 Résultats de l'exercice sur les prépositions

Dans un premier temps, les résultats à l'exercice de phrases trouées ont été compilés pour les deux groupes. Pour le premier groupe (corpus Vikidia), le taux moyen de bonnes réponses obtenues par les participants est de 67.5%, avec une valeur minimale de 37.5% (trois phrases complétées avec succès) et une valeur maximale de 87.5% (sept phrases complétées avec succès), avec un écart-type de 17%. Du côté des phrases à compléter, le taux de succès varie entre 40% (c'est-à-dire quatre participants avec des bonnes réponses) et 90% (donc neuf participants avec des bonnes réponses).

Pour le second groupe (corpus Corpatext), le taux moyen de bonnes réponses obtenues par les participants est de 76,4%, avec une valeur minimale de 62,5% (cinq phrases complétées avec succès) et une valeur maximale de 100% (huit phrases complétées avec succès), avec un écart-type de 13,2%. Du côté des phrases à compléter, le taux de succès varie entre 22,2% (deux participants avec des bonnes réponses) et 100% (neuf participants avec des bonnes réponses).

Tableau 3.2 Résultats globaux des deux groupes pour l'exercice

	Groupe 1	Groupe 2
Moyenne des résultats à l'exercice	67,5%	76,4%
Résultat minimum	37,5%	62,5%
Résultat maximum	100%	100%
Écart-type des résultats	17%	13,2%

Entre les deux groupes, on remarque une différence approchant dix pour cent, pour laquelle des analyses supplémentaires s'avèrent nécessaires afin d'établir si cet écart est significatif. Bien que les distributions des deux groupes paraissent normales, étant donné la petite taille de l'échantillon, les tests non-paramétriques ont été utilisés pour vérifier s'il y avait une différence significative entre les deux groupes. Pour les résultats dans leur ensemble, il n'y a pas de différence entre les deux groupes selon le test de Mann-Whitney ($U = 30,50$, $p = 0,24$).

Par la suite, à la lumière des analyses effectuées pour les résultats dans leur ensemble, il semblait propice d'effectuer des comparaisons supplémentaires à partir de résultats obtenus pour des éléments distincts, afin de vérifier la possibilité de différences entre les deux groupes pour des combinaisons de mot-clé et de préposition spécifiques. Quatre combinaisons ont été retenues en s'appuyant sur les critères suivants : elles étaient présentes en nombre égal dans les deux exercices (ce qui faciliterait la comparaison des éléments), et plus de la moitié des participants de chaque groupe avaient inscrit une réponse appropriée pour cette combinaison. (Une liste complète des combinaisons et des résultats est présentée dans le tableau 5 de l'appendice D.)

Pour la première combinaison, soit le verbe travailler combiné à une préposition (dans / avec), le groupe 1 (corpus Vikidia) a obtenu une moyenne de 70%, tandis que le groupe 2 (corpus Corpatext) a obtenu une moyenne de 100%; il s'agit de la seule combinaison où l'un des groupes a obtenu une note parfaite. Pour la deuxième combinaison, soit le verbe parler combiné à une préposition (à / de), le groupe 1 a obtenu une moyenne de 70% et le groupe 2 a obtenu une moyenne de 88,9%. Pour la troisième combinaison, alliant une préposition (depuis) à une expression de temps, les deux groupes ont obtenu 70% et 77,8%, respectivement. Finalement, pour la combinaison de préposition (à / en) et de lieu avec un verbe d'état (qui rassemblait deux questions pour chaque groupe), le premier groupe a obtenu une moyenne de 90% et le deuxième une moyenne de 77,8%; des quatre combinaisons retenues, il s'agit de la seule où le groupe 1 a obtenu un résultat supérieur au groupe 2.

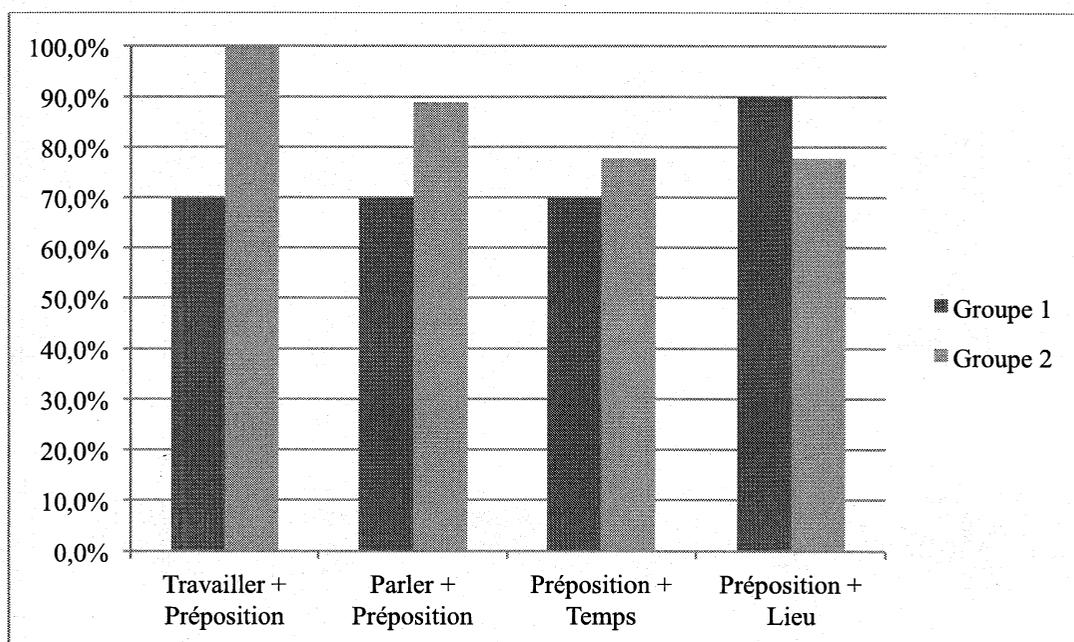


Figure 4.3 Réponses correctes dans l'exercice sur les prépositions

Afin de comparer les données croisées de chaque combinaison, le test de Fisher a été utilisé, étant donné que l'échantillon est petit et que certaines cellules ont un effectif théorique petit. Dans le cas de la première combinaison (travailler + préposition), la signification exacte bilatérale est de $p = 0,21$; pour la deuxième combinaison (parler + préposition), la signification exacte bilatérale est de $p = 0,58$; pour la troisième combinaison (préposition + temps), la signification exacte bilatérale est de $p = 1,00$; finalement, pour la quatrième combinaison (préposition + lieu), la signification exacte bilatérale est de $p = 0,35$. Dans tous les cas, les résultats du test de Fisher indiquent que les deux groupes ne sont pas significativement différents.

4.3 Perceptions du dispositif

En plus de récolter des données pour dresser un portrait des participants, une section du questionnaire a été conçue pour obtenir les perceptions de ceux-ci vis-à-vis du dispositif, notamment sur les listes d'exemples extraites des corpus, mais aussi sur l'interface en général (phrases à compléter, boutons avec mots-clés pour afficher les listes d'exemples). Comme les contraintes de temps n'ont pas permis d'entrevues individuelles avec les participants, ces réponses contribueront à établir l'appréciation des participants pour l'activité, tout en fournissant un point de comparaison où les deux groupes pourraient différer.

Une des questions retenues concerne la perception de facilité de lecture pour les exemples choisis. Étant donné que les démarches d'AGD requièrent de lire une quantité substantielle de contenu afin d'élaborer des hypothèses sur le fonctionnement de la langue-cible, la facilité de lecture des exemples (telle que perçue par les participants) pouvait devenir un autre facteur de variation entre les deux groupes. Dans cette activité, 60% du groupe 1 ont trouvé les exemples faciles à

lire (40% « en accord » et 20% « très en accord»), contre 40% qui n'étaient pas de cet avis (30% « en désaccord », 10% « très en désaccord »). Du côté du groupe 2, 55,5% des participants ont trouvé les exemples faciles à lire (11,1% « en accord » et 44,4% très en accord), contre 44,4% d'avis contraire (22% « en désaccord » et 22% « très en désaccord »).

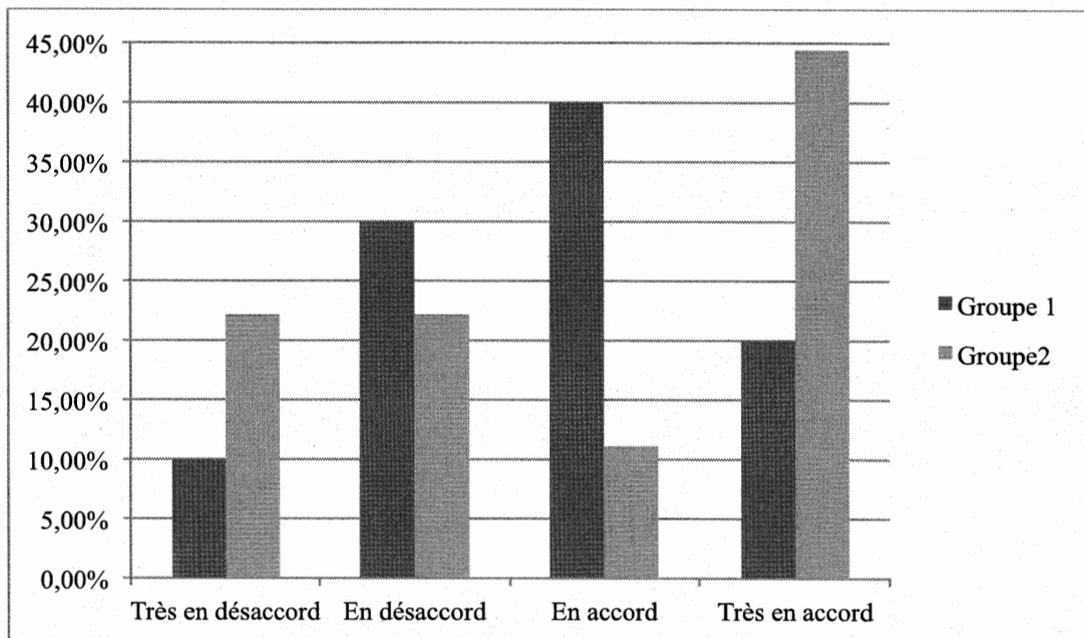


Figure 4.4 Facilité de lecture des exemples

Comme l'indique la figure 10 ci-dessus, les distributions des réponses des deux groupes semblent légèrement différentes, le groupe 1 favorisant les réponses modérées (« En désaccord » et « En accord »), tandis que le groupe 2 se situe davantage dans les extrêmes (« Très en désaccord » et « Très en accord »). Toutefois, selon le test de Fisher ($p = 0,84$ pour la signification exacte unilatérale), cette variation dans la distribution des réponses s'avère non-significative.

Finalement, trois questions provenant de la section conçue pour mesurer l'intérêt des participants envers l'outil ont été retenues : l'intérêt pour les exemples choisis, le souhait d'utiliser ce logiciel dans le futur, et la recommandation de cet outil à d'autres étudiants. En ce qui concerne l'intérêt exprimé par les participants pour les exemples choisis, les distributions des réponses pour les deux groupes paraissent très similaires entre les deux groupes : dans le premier groupe, 10% sont « très en accord », 40% « en accord » et 50% « en désaccord », contre 11,1% « très en accord », 33,3% « en accord », 44,4% « en désaccord » et 11,1% « très en désaccord » dans le deuxième groupe. En somme, pour les deux groupes, les réponses « en accord » et « en désaccord » semblent réparties de façon relativement égale. Comme les distributions paraissaient très similaires en apparence (hormis l'absence du groupe 1 dans la catégorie « très en désaccord »), aucune analyse statistique n'a été effectuée pour vérifier s'il y avait une différence significative.

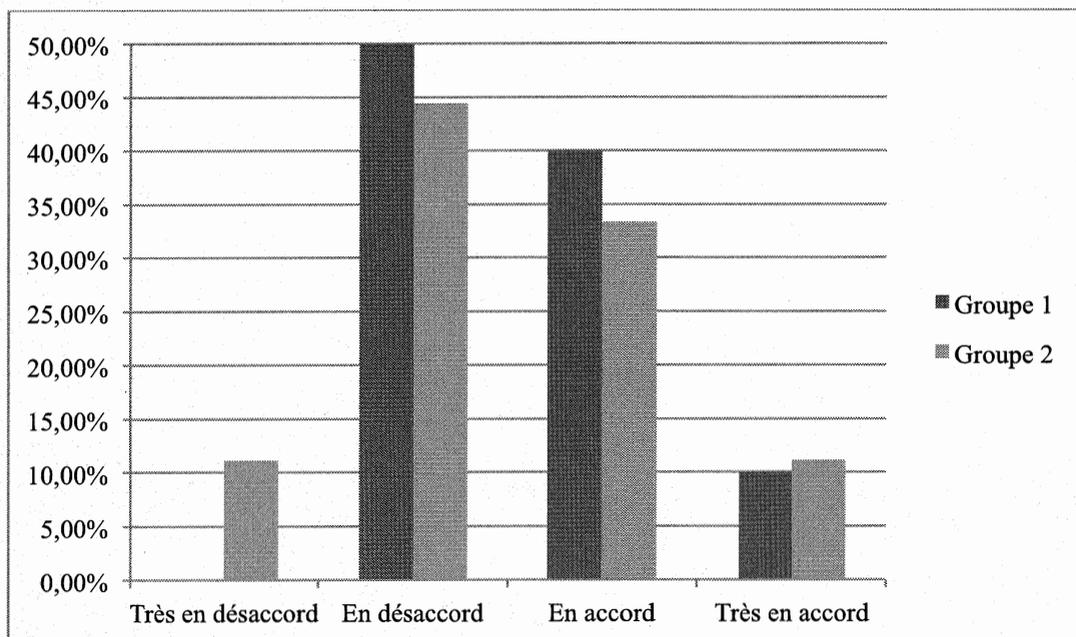


Figure 4.5 Intérêt exprimé pour les exemples choisis

Afin de mesurer l'appréciation des participants vis-à-vis du dispositif, une section du questionnaire demandait à ceux-ci s'ils étaient intéressés à utiliser cet outil dans le futur. Dans le groupe 1, 60% ont indiqué vouloir utiliser ce dispositif de nouveau (50% « en accord » et 10% « très en accord »), un pourcentage qui monte à 77,7% dans le groupe 2 (66,7% « en accord » et 11,1% « très en accord »). En contrepartie, 40% des participants du groupe 1 n'étaient pas intéressés à réutiliser l'outil (30% « en désaccord » et 10% « très en désaccord »), contre 22,2% « très en désaccord » dans le groupe 2.

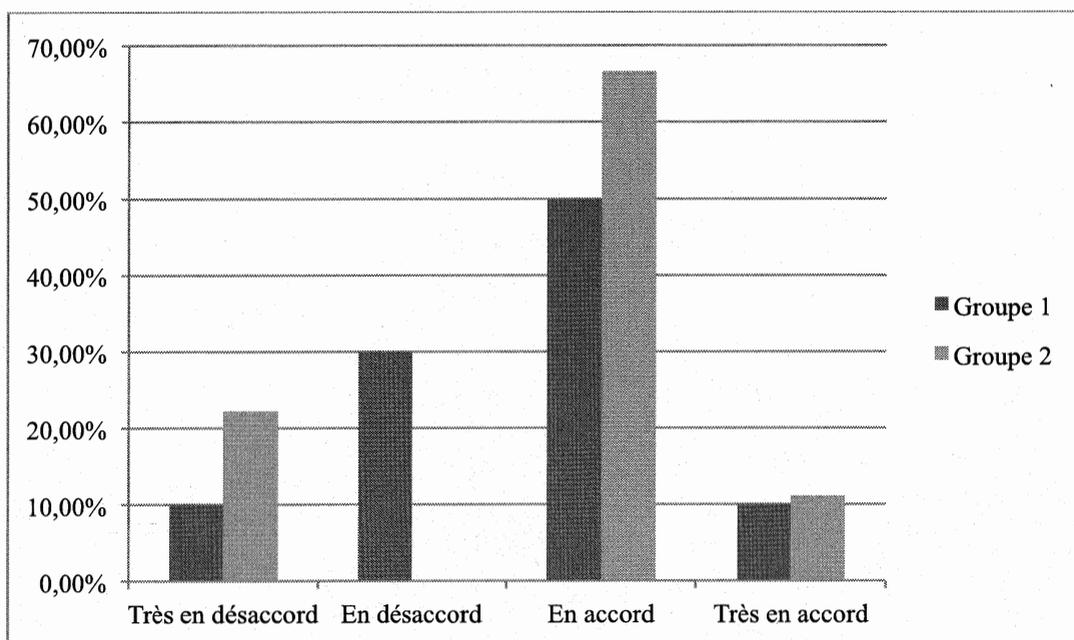


Figure 4.6 Souhait d'utiliser le logiciel dans le futur

En observant la figure 12 ci-dessus, les distributions des deux groupes semblent assez similaires dans l'ensemble, sauf l'absence de réponse « en désaccord » dans le Groupe 2; un test de Fisher (signification exacte unilatérale) a permis de confirmer ($p = 0,72$) qu'il n'y avait pas de différence significative entre les réponses des deux groupes pour cette question.

En s'inspirant de questionnaires trouvés dans la littérature, une deuxième question a été incorporée au questionnaire pour mesurer différemment l'appréciation des participants pour le dispositif, en demandant aux participants s'ils recommanderaient l'outil pour pratiquer le français. Dans le premier groupe, 100% des répondants ont répondu positivement, tandis que dans le deuxième groupe 77,7% ont répondu par l'affirmative.

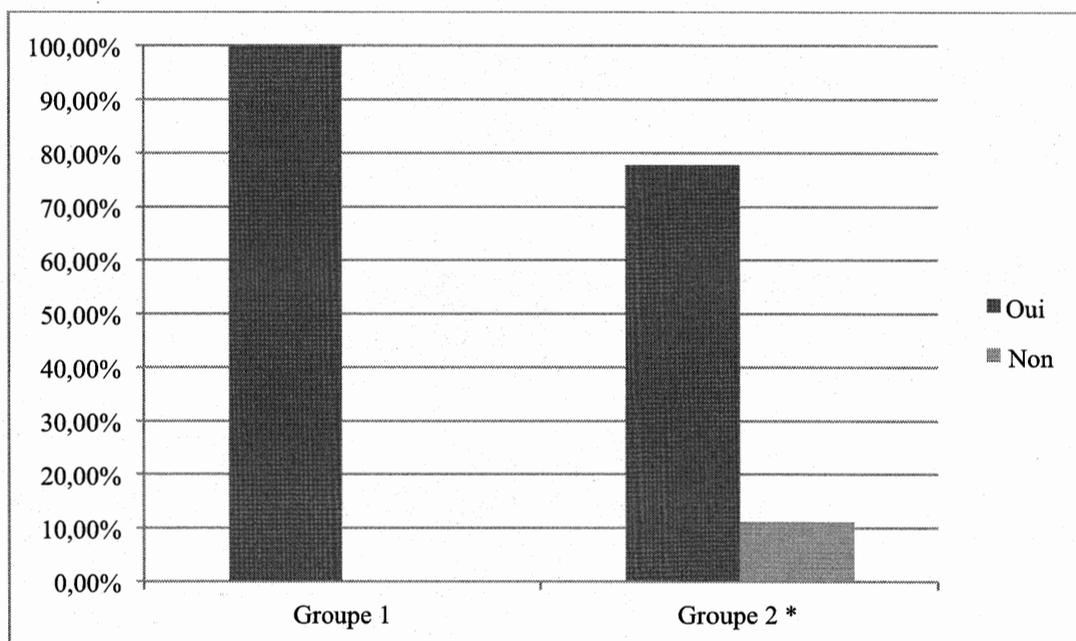


Figure 4.7 Recommandation de l'outil

* Dans le Groupe 2, une personne a choisi de ne pas répondre à cette question.

Dans la mesure où les données étaient partiellement incomplètes (avec une abstention dans le deuxième groupe), aucune analyse statistique n'a été effectuée pour cette question et les données sont présentées à titre descriptif.

4.4 Enregistrement des actions dans l'exercice

En parallèle à la question de recherche principale, un des objectifs secondaires de ce projet de recherche était de vérifier si une tendance pouvait être décelée quant à l'utilisation de l'outil fourni durant la complétion de l'exercice sur les prépositions. Par exemple, est-ce que les exemples étaient consultés plus fréquemment au début de l'exercice, puis de moins en moins au fur et à mesure que l'activité progressait? Dans la mesure où les participants à d'autres recherches sur l'AGD indiquaient avoir de la difficulté à décoder et à interpréter les exemples trouvés (Chambers, 2010), on pourrait s'attendre à ce que les participants laissent les exemples de côté au fur et à mesure que l'activité progresse, par perte de motivation ou d'intérêt.

Afin d'explorer cette hypothèse, le nombre de clics effectués sur chacun des boutons affichant les exemples a été enregistré au cours de l'activité. Étant donné que chaque question comprenait deux boutons menant à des listes distinctes de phrases exemples et que les exercices pour les deux groupes étaient différents, quatre nombres ont été enregistrés pour chaque question. Il est important de noter que ce nombre était enregistré indépendamment du nombre de participants; il était donc possible que le nombre de clics dépasse le nombre de participants, dans la mesure où certains d'entre eux ont cliqué plusieurs fois sur le même bouton pour afficher les exemples à plus d'une reprise.

Globalement, le nombre de clics paraît plus élevé pour la première question : dans le premier groupe (corpus Vikidia), 26 clics ont été enregistrés pour l'affichage de la première série d'exemples, et 14 clics pour la deuxième. Un comportement semblable émerge du deuxième groupe (corpus Corpatext), avec 14 et 20 clics respectivement pour les deux séries d'exemples de la première question de l'exercice.

Pour les autres questions de l'exercice, le nombre de clics pour afficher les listes d'exemples atteint une moyenne de 7,38 pour le groupe 1 (avec un minimum de 3 et un maximum de 15), contre une moyenne de 9,93 pour le groupe 2 (avec un minimum de 3 et un maximum de 13).

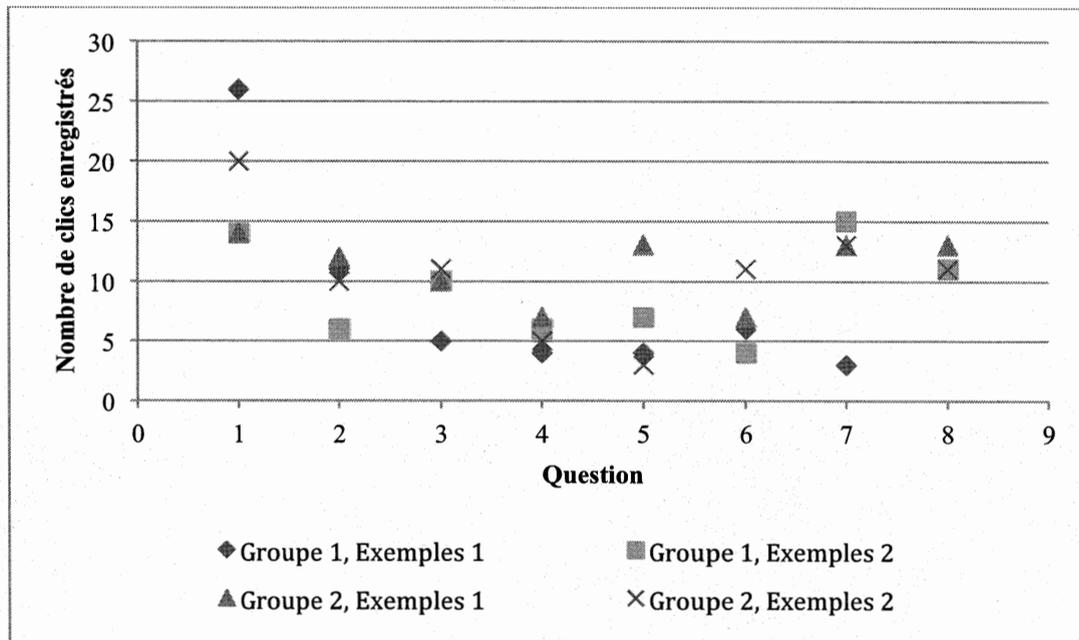


Figure 4.8 Clics enregistrés pour afficher les exemples (deux séries d'exemples par question dans chaque exercice)

À partir de l'enregistrement des clics déclenchant l'affichage des listes d'exemples, le graphique de nuages de points en figure 14 a été élaboré, en indiquant les questions sur l'axe des abscisses (de la première à la huitième question), et le nombre de clics effectués pour afficher les exemples sur l'axe des ordonnées. Les analyses statistiques (test de corrélation de Pearson) effectuées sur ces données indiquent que la corrélation observée entre la question et le nombre de clics pour afficher les exemples n'est pas significative ($r = -0,24$, $n = 32$, $p = 0,09$).

CHAPITRE V

DISCUSSION

5.1 Profil des participants

Le profil des participants, élaboré à partir des données qui ont été récoltées dans le questionnaire, rappelle l'importance de rendre les outils accessibles même si l'ordinateur fait partie du quotidien de la plupart des étudiants. Bien que plus de la moitié des dix-neuf participants ait indiqué un intérêt pour l'utilisation d'un ordinateur dans leur apprentissage du français et que tous aient répondu qu'ils utilisaient l'ordinateur dans au moins un volet de leur quotidien (soit à la maison, à l'école ou au travail), la majorité des participants découvraient une approche d'AGD pour la première fois à l'occasion de ce projet de recherche. À cet égard, le profil des participants s'apparente aux observations de Chang et Sun (2009) : dans leur recherche, les participants utilisaient déjà un ordinateur connecté à Internet, mais peu avaient précédemment utilisé un outil d'AGD. De plus, la majorité des participants ont exprimé une préférence pour une séquence d'enseignement où les règles sont présentées au début (voir la figure 7 du chapitre 4), tandis que les démarches d'AGD tendent à favoriser une approche inductive (Chambers, 2005; Sun et Wang, 2003). Devant ce changement, certains étudiants peuvent être décontenancés face à ce changement de la dynamique en classe (Landure, 2011), ou nécessiter plus d'accompagnement (Boulton, 2009b) pour tirer profit de cette démarche différente.

En somme, bien que plusieurs apprenants expriment un intérêt envers l'utilisation de technologie pour soutenir leur apprentissage, différents obstacles

peuvent les empêcher d'utiliser ces outils à leur plein potentiel (y compris leurs habitudes et leurs attentes en tant qu'apprenants), d'où l'importance de bien accompagner les apprenants dans les démarches d'AGD (Boulton, 2009b), et de rendre l'utilisation de ces outils aussi simple et intuitive que possible.

5.2 Résultats de l'exercice sur les prépositions

Bien que la performance du deuxième groupe (employant le corpus Corpatext) soit en apparence supérieure à celle du premier groupe (utilisant le corpus Vikidia), cette différence s'avère non significative. Étant donné l'absence de différence au niveau de la performance des deux groupes, il est possible que l'utilisation didactique des corpus par les participants soit indépendante de la nature des textes rassemblés dans ceux-ci; cet aspect demeure à confirmer, étant donné que la comparaison de ressources demeure peu présente dans les articles trouvés dans la littérature. Comme une partie importante des publications sur l'AGD se concentre sur l'enseignement de l'anglais (Boulton, 2009b; Chang et Sun, 2009; Gaskell et Cobb, 2004), la ressource principale est souvent le British National Corpus ou le Brown Corpus of Standard American English. Quelques-uns des articles retenus déploient plus d'une ressource (Liu et Jiang, 2009; Johns *et al*, 2008; Chujo *et al*, 2006), mais celles-ci sont utilisées de façon complémentaire plutôt que comparative. De fait, la comparaison de différentes ressources pourrait être développée dans des recherches futures.

5.3 Perceptions des participants

D'autres questions du questionnaire visaient à déterminer les perceptions des participants vis-à-vis du contenu des corpus. Les données présentées dans les figures

10 et 11 du chapitre quatre, respectivement sur la facilité de lecture des exemples et sur l'intérêt de ces derniers, rappellent qu'il est ardu de trouver un juste milieu dans la conception des activités en cours de langue. Étant donné les expériences et les intérêts variés des participants dans un même groupe, des contenus jugés faciles à lire par certains apprenants nécessiteront plus d'efforts pour d'autres. Malgré l'enthousiasme exprimé par plusieurs participants dans le questionnaire, les réponses nuancées vis-à-vis de l'intérêt et de la facilité de lecture rappellent que, en dépit des efforts consacrés à rendre les démarches d'AGD accessibles à un plus grand bassin d'apprenants, l'analyse d'exemples requière un effort cognitif important (Warschauer *et al*, 2017; Gordani, 2013; Chambers, 2010).

D'un autre côté, la distribution des réponses laisse tout de même présager la possibilité de donner une place plus importante à l'AGD avec des étudiants de niveau débutant ou intermédiaire. Ainsi, dans le groupe de participants, plus de la moitié d'entre eux ont répondu que les exemples choisis étaient faciles à lire. En lien avec les observations de O'Sullivan et Chambers (2006), le fait d'avoir concentré l'activité (et les listes d'exemples provenant des corpus) sur un objet langagier pour lequel la consultation de corpus a déjà fait ses preuves, c'est-à-dire le choix de préposition, semble avoir contribué au succès de l'activité. De plus, les résultats recueillis suggèrent que la sélection des extraits de corpus à l'avance, une approche inspirée par les travaux de Boulton (2008) et discutée dans les recherches de Chambers (2010), de Oghigian et Chujo (2010) et de Warschauer et collègues (2017), aide à rendre la démarche d'AGD plus accessible aux apprenants pour qui la recherche et l'analyse d'une grande quantité de contenu pourrait s'avérer difficile. Bien que cette approche semble avoir du potentiel chez les apprenants de niveau intermédiaire inférieur, dans une optique de développement de l'autonomie, il demeure important de réfléchir aux stratégies (formation, accompagnement, tutoriels) qui permettront éventuellement aux apprenants d'explorer le contenu des corpus par eux-mêmes.

Au-delà de la sélection des exemples, une attention particulière a été accordée à la conception de l'outil, en s'inspirant de suggestions trouvées dans la littérature (Yoon, 2011; Chambers, 2010; Boulton, 2010a; Chambers, 2005). Cette tentative de réduction de la surcharge cognitive semble avoir porté fruit : aucun participant n'a demandé de soutien à l'enseignante ou au chercheur pour maîtriser l'interface de l'exercice ou du questionnaire durant la deuxième rencontre. Indirectement, les réponses positives obtenues dans le questionnaire sur l'appréciation du dispositif, telles que détaillées dans les figures 12 et 13 du chapitre quatre, suggèrent que cette démarche de simplification pourrait avoir contribué à rendre l'exercice plus accessible.

D'autres éléments du questionnaire rempli par les participants visaient à déterminer leur appréciation du dispositif (c'est-à-dire l'exercice de phrases trouvées avec des listes d'exemples). Comme l'indique la figure 12 dans le chapitre quatre, plus de la moitié des participants ont indiqué vouloir utiliser de nouveau ce logiciel dans le futur. De plus, la quasi-totalité des participants recommanderaient l'utilisation de l'outil pour pratiquer le français (figure 13 du chapitre quatre). Dans l'ensemble, cet intérêt pour l'AGD évoque les conclusions partagées dans d'autres recherches énumérées dans le cadre théorique (Yoon, 2011; Chang et Sun, 2009; Gilmore, 2009; O'Sullivan et Chambers, 2006; Yoon et Hirvela, 2004). Les observations de Boulton (2009a) apportent une nuance supplémentaire : ce dernier a constaté un enthousiasme moins important chez des participants ayant pris part à plusieurs séances d'AGD étalées sur plusieurs mois. De fait, il est possible que l'intérêt des participants noté dans le questionnaire s'explique en partie par un effet de nouveauté, et que les réponses auraient été plus nuancées si l'exercice avait été décliné en plusieurs séances.

Étant donné l'envergure limitée du projet de recherche (petit groupe de participants, rencontre unique), il est difficile d'établir des liens concluants entre les

réponses au questionnaire et les préférences des participants en tant qu'apprenants. Dans la mesure où les expériences individuelles peuvent influencer l'utilisation de corpus (Yoon, 2011), il serait sans doute intéressant d'explorer davantage ces liens dans un projet de longue durée, avec un plus grand nombre de participants.

5.4 Enregistrement des actions dans l'exercice

D'après les données d'utilisation recueillies, la première question est celle qui a engendré le plus de clics pour afficher les listes d'exemples, autant pour l'exercice du groupe 1 que pour celui du groupe 2. Pour les autres questions, le nombre de clics pour afficher les exemples était moins important, comme l'illustre la figure 14 du chapitre quatre. Bien que les circonstances soient différentes, il est possible d'établir des parallèles avec les observations partagées par Gaskell et Cobb (2004) : durant la phase initiale, les participants utilisaient davantage les outils mis à leur disposition. Dans le cas de Gaskell et Cobb, l'utilisation était associée au fait que les participants avaient accès à des hyperliens qui facilitaient l'utilisation du corpus; dans le cas de cette recherche, il est possible qu'un plus grand nombre de clics ait été enregistré à la question initiale parce que les participants expérimentaient avec le dispositif. Pour les autres questions, l'analyse statistique suggère une absence de tendance au fur et à mesure que l'exercice progresse. Étant donné les ressources limitées de ce projet, il n'a pas été possible de réaliser des entrevues avec quelques participants pour mieux comprendre leur utilisation de l'outil ou d'effectuer d'analyses statistiques inférentielles plus complexes; toutefois, il est probable que d'autres facteurs puissent expliquer la fréquence d'utilisation du dispositif. Par exemple, comme dans les travaux de Warschauer et collègues (2017), il est possible que les participants aient choisi de ne pas consulter les outils pour les combinaisons de mot-clé et de préposition qui semblaient plus familières.

D'un point de vue global, les données d'utilisation d'outils d'AGD contribuent à mettre en évidence les éléments qui varient selon le profil des apprenants et les circonstances d'expérimentation, et à mieux comprendre les aspects qui doivent être travaillés pour rendre ces outils plus accessibles.

CHAPITRE VI

CONCLUSION

6.1 Objectifs du projet de recherche

Premièrement, un des objectifs de ce projet de recherche était d'évaluer si une activité d'AGD pourrait être potentiellement utilisée dans un contexte d'apprentissage de la langue avec des étudiants de niveau débutant ou intermédiaire inférieur. À cet égard, les résultats obtenus avec le groupe d'étudiants en francisation durant la deuxième rencontre paraissent prometteurs : en comparant le pourcentage de bonnes réponses à la première rencontre (40%) et celui à la deuxième rencontre (71,1%, en combinant les résultats des deux groupes), on note une amélioration des résultats à la deuxième rencontre, lorsque l'outil d'AGD était accessible aux participants. Par ailleurs, un peu plus de la moitié des participants à ce projet de recherche ont trouvé les exemples faciles à lire; bien que ces réponses soient encourageantes, elles rappellent par la même occasion que des efforts supplémentaires sont souhaitables pour rendre le contenu des corpus linguistiques accessible à un plus grand nombre d'apprenants.

Deuxièmement, tel que détaillé dans les questions de recherche, ce projet souhaitait explorer si la provenance du matériel de référence pouvait avoir une incidence sur différents paramètres de l'activité d'AGD (utilisation de l'outil, performance dans la tâche, perception du dispositif, etc.). À cette fin, un prototype de corpus s'appuyant sur le contenu de l'encyclopédie libre Wikidia a été créé, dans l'optique d'être comparé à Corpatext, un corpus de langue française librement accessible. Dans l'exercice de phrases à compléter, la performance obtenue par le

groupe assigné au contenu de Wikidia s'est avérée comparable à celle obtenue par l'autre groupe, assigné au contenu de Corpatext; autant pour les résultats dans leur ensemble que pour des éléments spécifiques, aucune différence significative n'a été détectée entre les deux groupes. En ce qui concerne les perceptions des participants, deux éléments ont été comparés (facilité de lecture des exemples et désir d'utiliser cet outil de nouveau dans le futur), mais aucune différence significative n'a été détectée entre les deux groupes. Considérant la nature exploratoire de ce projet de recherche, l'absence de différence significative entre les deux groupes, tant au niveau de la performance que des perceptions, suggère que l'encyclopédie Wikidia pourrait devenir une source viable de contenu écrit pour offrir un autre corpus linguistique francophone, avec des types de textes différents de ceux qu'on retrouve dans d'autres ressources comme Corpatext, afin d'offrir aux apprenants une plus grande diversité de contenu écrit. Étant donné la similitude des corpus vis-à-vis de leur vocabulaire, il est aussi possible que la nature du corpus importe peu pour le type d'activité mis en place dans ce projet de recherche.

Troisièmement, étant donné les objectifs secondaires visant à documenter les perceptions et les interactions des participants avec le dispositif, une attention particulière a été accordée à la conception de l'outil à la lumière des difficultés évoquées dans la littérature (Chambers, 2010; Chang et Sun, 2009; Gilmore, 2009; Yoon et Hirvela, 2004; Kennedy et Miceli, 2001). Dans l'ensemble, les données obtenues dans le questionnaire en lien avec l'accessibilité du contenu paraissent encourageantes, bien qu'imparfaites; si la sélection d'exemples à l'avance semble avoir du potentiel, il y a encore du chemin à parcourir pour concevoir des outils convenant à différents profils d'apprenants. Similairement, l'enregistrement des interactions suggère que l'AGD pourrait être utilisé par des apprenants débutants, mais qu'il y aurait encore place à amélioration pour hausser la consultation des ressources par ces derniers. Afin de mieux comprendre les données d'utilisation, une

analyse plus fine des circonstances (comme la familiarité ou la difficulté de chaque question, en lien avec les connaissances antérieures des apprenants) semble une voie intéressante pour mieux ajuster l'AGD aux apprenants novices ou intermédiaires inférieurs.

6.2 Limites envisagées

Dans un premier temps, bien que les résultats à l'exercice sur les prépositions aient été supérieurs lors de la deuxième rencontre, il est difficile de mesurer l'apport exact de l'outil d'introduction à l'AGD. Dans la mesure où cette augmentation des bonnes réponses pourrait s'expliquer par une variété de facteurs (notamment les apprentissages effectués durant les deux semaines qui séparaient les deux séances, la rétroaction effectuée avec certains participants après la première rencontre, ainsi que la séance d'enseignement explicite effectuée par l'enseignante lors de la seconde rencontre pour réviser et activer les connaissances antérieures des étudiants), les résultats doivent être considérés avec précaution. Sans considérer l'outil comme une panacée, on pourrait s'attendre à ce que sa contribution s'inscrive dans la lignée des observations notées dans la littérature : en général, l'AGD entraînerait des résultats neutres ou bénéfiques, selon l'échelle et la durée de l'intervention (Golonka *et al*, 2014; Yoon, 2011; Boulton, 2009a). Par ailleurs, il est possible que certaines circonstances entourant la dernière rencontre (tenue imminente de l'examen final, activités festives bruyantes à l'extérieur) aient pu devenir des sources de stress ou de distraction, et que les participants affectés n'aient pas investi autant d'efforts dans l'exercice.

Ensuite, les observations concernant la pertinence de Vikidia comme ressource doivent être nuancées, étant donné les limites de ce projet de recherche (dix-neuf participants séparés en deux groupes, prenant part à une seule séance avec

l'outil). Comme l'analyse du contenu utilisé pour créer les banques d'exemples a révélé une différence significative (mais faible) entre les deux corpus au niveau du pourcentage de vocabulaire fréquent, des analyses supplémentaires seraient essentielles pour mieux évaluer l'utilité potentielle de Vikidia comme outil dans une démarche d'AGD. Pour mieux jauger l'utilité de ce prototype de corpus, il serait donc préférable d'effectuer des recherches avec un plus grand nombre de participants et des projets étendus sur une plus longue durée, mais il faudrait aussi prévoir des activités où l'ensemble du corpus est mis à la disposition des apprenants avec une interface de saisie de mots-clefs et de recherche, sans qu'un tri soit effectué au préalable pour filtrer les exemples répétitifs ou plus difficiles à interpréter, pour se rapprocher d'une démarche plus typique d'AGD.

Au final, il est important de garder en tête la nature exploratoire de ce projet; étant donné le petit nombre de participants et la durée limitée de la rencontre avec ceux-ci, les données devraient être abordées comme un point de départ potentiel vers des investigations plus approfondies.

6.3 Pistes de recherche futures

Afin de mieux évaluer l'apport de différentes stratégies visant à réduire la surcharge cognitive associée aux outils d'AGD, des recherches supplémentaires permettant de comparer plusieurs variantes d'un même exercice (quelques possibilités à envisager : utilisation parallèle d'un outil informatique et d'un support papier, affichage des exemples tirés du corpus soit en phrases complètes ou avec les mots-clefs alignés, offrir l'accès au contenu du corpus soit à partir d'une sélection effectuée ou préalable ou à partir d'une interface de recherche) permettraient de mieux évaluer l'apport de ces différentes stratégies pour rendre cette approche accessible à différents

apprenants. Un autre aspect qui pourrait être développé avec un projet de plus grande envergure est l'apport de la durée : d'un côté, un plus grand nombre de rencontres aiderait à mieux évaluer l'apport de la consultation du corpus sur le développement des compétences langagières des participants; de l'autre côté, si la motivation des participants est associée à l'effet de nouveauté, une recherche plus longue donnerait une vision plus nuancée de la perception des participants envers l'AGD.

Dans un même ordre d'idées, l'enregistrement des actions effectuées dans les outils d'AGD, ainsi que l'enregistrement du temps passé dans les différentes composantes de ces outils, pourraient ouvrir la voie à des investigations supplémentaires. En lien avec les comportements observés dans les logiciels d'AGD, des entrevues effectuées avec les participants après les exercices contribueraient à illustrer comment les circonstances peuvent guider l'utilisation des différents outils. Par exemple, est-ce que les étudiants se réfèrent moins souvent aux exemples lorsqu'ils pensent bien comprendre les règles pour l'objet langagier spécifiquement ciblé par l'exercice? Similairement, est-ce que la difficulté de la question (perçue par l'étudiant ou évaluée par le chercheur) pourrait expliquer la fréquence d'utilisation de l'outil? Autrement, il est possible que les préférences des apprenants –notamment pour l'apprentissage explicite ou implicite, ainsi que leur intérêt pour l'utilisation de technologie – puissent avoir une incidence sur leur fréquence d'utilisation des outils d'AGD.

Enfin, l'élaboration d'un prototype de corpus à partir du contenu de Wikidia, l'encyclopédie libre destinée à un lectorat jeunesse, pourrait ouvrir la voie au développement d'un corpus de français simplifié, en s'inspirant de l'algorithme de simplification du français proposé par Brouwers et collègues (2012) et des principes identifiés par Collès (2002) pour expliquer la simplification de textes littéraires pour l'enseignement. D'un autre côté, certains auteurs (Ruiz de Zarobe, 2015; Oh, 2001)

suggèrent que l'élaboration, soit « la manipulation de l' "input" qui vise à clarifier le contenu du message par l'explication de la signification du texte » (Ruiz de Zarobe 2015, p. 471), serait préférable à la simplification pour favoriser la compréhension de l'apprenant tout en maintenant un aspect semblable aux textes authentiques. Ainsi, le développement d'un algorithme d'élaboration, permettant de remplacer des termes moins fréquents par des périphrases explicatives, pourrait être une autre avenue à explorer pour rendre l'AGD accessible à un plus grand nombre d'étudiants. Enfin, dans la mesure où les textes authentiques comme les textes modifiés peuvent donner des résultats nuancés en salle de classe (Gilmore, 2007), il apparaît nécessaire de poursuivre les recherches pour trouver les types de textes convenant le mieux aux différents profils d'apprenants, et aux différents contextes d'apprentissage de langue.

APPENDICE A

LETTRÉ ENVOYÉE LORS DE LA PREMIÈRE PRISE DE CONTACT

UQAM Département de didactique des langues

Objet : Participation à un projet de recherche

Bonjour !

Je m'appelle Mathieu Ménard et je suis étudiant à l'Université du Québec à Montréal. Comme votre professeur Melissa Dubuc, j'ai enseigné l'anglais en Asie pendant quelques années.

À l'université, je fais de la recherche sur l'étude de la langue aidé par ordinateur. Je veux créer des outils qui aident les étudiants à apprendre le français.

Je recherche des participants volontaires pour essayer un outil informatique. Cet outil aide à comprendre comment les prépositions (par exemple, « à », « de », « avec », « pour »...) sont utilisées en français. Si vous acceptez de participer, je vous demanderai aussi de remplir un petit formulaire pour mieux vous connaître comme étudiant. Vos informations personnelles, votre nom par exemple, ne seront pas utilisés.

La participation à ce projet est complètement volontaire et il n'y a aucun lien entre cette recherche et votre cours de francisation. En participant à ce projet, vous pourrez aider la recherche et vous pourrez découvrir un logiciel pour pratiquer votre français.

Merci de votre attention et au plaisir de vous rencontrer bientôt !

Mathieu Ménard

menard.mathieu.2@courrier.uqam.ca

Candidat à la maîtrise en didactique des langues
Université du Québec à Montréal

Subject: Participation in a Research Project

Hello!

My name is Mathieu Ménard and I study at Université du Québec à Montréal. Similar to your teacher Melissa Dubuc, I taught English in Asia for a few years.

I do research on computer assisted language learning. I want to create tools to help students learn French.

I am looking for voluntary participants to try a software language learning tool. This tool helps understand how prepositions (for example, small words like "à", "de", "avec", "pour"...) are used in French. If you agree to participate, I will also ask you to fill a small form to learn more about you as a student. Your personal information, such as your name, will not be used.

Your participation in this project is completely voluntary and there is no link between this research and your francisation class. By participating in this project, you can help research move forward and you can discover a new software to practice your French.

Thank you for your attention and looking forward to meeting you soon!

Mathieu Ménard

menard.mathieu.2@courrier.uqam.ca

Language Teaching Masters' Candidate

Université du Québec à Montréal

APPENDICE B

FORMULAIRE DE CONSENTEMENT REMIS À CHAQUE PARTICIPANT



«Mise en place d'une tâche d'analyse de cooccurrences pour corriger des erreurs grammaticales fréquentes chez des étudiants en contexte de francisation»

IDENTIFICATION

Chercheur responsable du projet : Mathieu Ménard
Programme d'enseignement : Maîtrise en didactique des langues
Adresse courriel : menard.mathieu.2@courrier.uqam.ca
Téléphone : 514-508-5462

BUT GÉNÉRAL DU PROJET ET DIRECTION

Vous êtes invité(e) à prendre part à ce projet visant à développer un logiciel aidant les apprenants à développer leur compréhension et leur maîtrise de la grammaire – plus précisément, à améliorer l'utilisation de prépositions et de la formation de groupes prépositionnels. Ce projet vise également à comprendre comment les apprenants de niveau débutant ou intermédiaire suivant un cours de francisation utilisent un logiciel d'apprentissage guidé par les données. Ce projet est réalisé dans le cadre d'un mémoire de maîtrise sous la direction de Simon Collin, professeur du département de didactique des langues de la Faculté des sciences de l'éducation. Il peut être joint au (514) 987-3000 poste 2446 ou par courriel à l'adresse : collin.simon@uqam.ca.

PROCÉDURE(S) OU TÂCHES DEMANDÉES AU PARTICIPANT

Votre participation consiste à compléter deux tâches de correction d'erreurs grammaticales au cours de deux rencontres différentes espacées d'environ deux semaines. Lors de la deuxième tâche, vous aurez accès à un outil informatique contribuant à guider vos corrections. Vos réponses à ces deux exercices de correction, de même que l'utilisation que vous faites de l'outil informatique (accès aux différentes fonctions et temps d'utilisation) seront enregistrées. Cette participation sera complétée avec un formulaire vous posant des questions sur votre appréciation de l'outil d'informatique, sur votre style d'apprentissage et sur votre motivation à continuer votre étude de la langue. Le questionnaire comprend aussi quelques questions démographiques (comme votre âge, votre sexe, le nombre d'années d'étude du français), mais les données recueillies (réponses aux deux tâches, données d'utilisation de l'outil informatique et réponses au questionnaire) ne permettront pas de vous identifier.

AVANTAGES ET RISQUES

Votre participation contribuera à l'avancement des connaissances par une meilleure compréhension des facteurs permettant aux apprenants de niveau débutant ou intermédiaire de profiter pleinement des outils informatiques pour l'apprentissage de la langue. Il n'y a pas de risque d'inconfort important associé à votre participation à cette rencontre. Vous devez cependant prendre conscience que les difficultés liées à l'apprentissage d'un nouveau logiciel ainsi que la limite de temps pour compléter les deux tâches de correction d'erreurs grammaticales pourraient engendrer un certain niveau de stress ou d'inconfort. Vous demeurez libre de demander du soutien technique au chercheur durant les activités, et de ne pas répondre à une question du questionnaire que vous estimez embarrassante sans avoir à vous justifier. Une ressource d'aide appropriée pourra vous être proposée si vous souhaitez discuter de votre situation. Il est de la responsabilité du chercheur de suspendre ou de mettre fin aux activités s'il estime que votre bien-être est menacé.

ANONYMAT ET CONFIDENTIALITÉ

Il est entendu que les renseignements recueillis lors de l'entrevue sont confidentiels et que seuls, le responsable du projet et son directeur de recherche, Simon Collin, auront accès aux données des tâches de correction, de l'utilisation de l'outil informatique et des réponses au questionnaire. Ce matériel de recherche (l'enregistrement numérique des réponses aux tâches de correction et des réponses au questionnaire ainsi que les données d'utilisation de l'outil informatique) ainsi que votre formulaire de consentement seront conservés séparément sous clé par le responsable du projet pour la durée totale du projet. Les données ainsi que les formulaires de consentement seront détruits 5 ans après les dernières publications.

PARTICIPATION VOLONTAIRE

Votre participation à ce projet est volontaire. Cela signifie que vous acceptez de participer au projet sans aucune contrainte ou pression extérieure, et que par ailleurs vous êtes libre de mettre fin à votre participation en tout temps au cours de cette recherche. Dans ce cas, les renseignements vous concernant seront détruits. Votre accord à participer implique également que vous acceptez que le responsable du projet puisse utiliser aux fins de la présente recherche (articles, mémoire, essai ou thèse, conférences et communications scientifiques) les renseignements recueillis à la condition qu'aucune information permettant de vous identifier ne soit divulguée publiquement à moins d'un consentement explicite de votre part.

COMPENSATION FINANCIÈRE

Votre participation à ce projet est offerte gratuitement. Un résumé des résultats de recherche vous sera transmis au terme du projet.

DES QUESTIONS SUR LE PROJET OU SUR VOS DROITS?

Vous pouvez contacter le responsable du projet au numéro (514) 508-5462 pour des questions additionnelles sur le projet. Vous pouvez également discuter avec le directeur de recherche Simon Collin au (514)-987-3000 #2446 des conditions dans lesquelles se déroule votre participation et de vos droits en tant que participant de recherche.

Le projet auquel vous allez participer a été approuvé au plan de l'éthique de la recherche avec des êtres humains. Pour toute question ne pouvant être adressée au directeur de recherche ou pour formuler une plainte ou des commentaires, vous pouvez contacter le Président du Comité d'éthique de la recherche pour étudiants (CÉRPE), par l'intermédiaire de son secrétariat au numéro (514)-987-3000 # 1646 ou par courriel à : (savard.josee@uqam.ca).

REMERCIEMENTS

Votre collaboration est importante à la réalisation de ce projet et nous tenons à vous en remercier.

SIGNATURES :

Je reconnais avoir lu le présent formulaire de consentement et consens volontairement à participer à ce projet de recherche. Je reconnais aussi que le responsable du projet a répondu à mes questions de manière satisfaisante et que j'ai disposé suffisamment de temps pour réfléchir à ma décision de participer. Je comprends que ma participation à cette recherche est totalement volontaire et que je peux y mettre fin en tout temps, sans pénalité d'aucune forme, ni justification à donner. Il me suffit d'en informer le responsable du projet.

Signature du participant :

Date :

Nom (lettres moulées) et coordonnées :

Je déclare avoir expliqué le but, la nature, les avantages, les risques du projet et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature du responsable du projet :

Date :

Nom (lettres moulées) et coordonnées :



" Introducing a concordance analysis task
to help correct francisation learners' frequent grammatical errors"

IDENTIFICATION

Researcher: Mathieu Ménard
 Program: Masters in Language Education (Maitrise en didactique des langues)
 Email address: menard.mathieu.2@courrier.uqam.ca
 Phone: 514-508-5462

PROJECT OVERVIEW

You are invited to participate in a project aiming to create a software program to help learners develop their understanding and mastery of French grammar – specifically, to improve their use of prepositions and their construction of prepositional phrases. This project also aims to gain a better understanding of how beginner and intermediate learners in a francisation program use data-driven learning software. This research is being completed, as part of a master's thesis, under the supervision of Drs. Simon Collin, professor at the department of language teaching in the Faculty of Education. Drs. Collin can be reached at (514) 987-3000 extension 2446 or by email at: collin.simon@uqam.ca.

WHAT YOU WILL BE ASKED TO DO

As a participant in this study, you will be asked to complete two grammatical error correction tasks during two different meetings approximately two weeks apart. During the second error correction task, you will have access to a software tool to guide your corrections. Your answers to these two grammatical error correction tasks, as well as how you use the software (which functions are accessed, and for how long) will be recorded. You will also be asked to complete a form asking you about your appreciation of the software tool, your learning style and your motivation to continue language learning in the future. This questionnaire will also ask demographic questions (such as your age, gender, number of years studying French), but the collected data (answers in the two tasks, software usage data and answers to the questionnaire) will be completely confidential: your name will not be included or in any other way associated with the data collected in this study.

PERSONAL BENEFITS AND RISKS TO PARTICIPATION IN THE STUDY

Your participation will contribute to the advancement of knowledge by improving the understanding of factors that allow beginner and intermediate level learners to fully benefit from language learning software tools. There are no major risks or side effects associated with participation in this project. However, you should be aware that the challenges of learning how new software works, along with the time limit to complete the grammatical error correction task, could induce some stress or discomfort. You remain free to ask the researcher for technical support during the activities, and you may decide not to answer any question if you feel uncomfortable doing so, without having to justify yourself. An appropriate support resource will be provided if you wish to discuss your situation. It is the researcher's responsibility to suspend or end the activities if he feels that your well-being is threatened.

CONFIDENTIALITY

All information you provide is confidential and only the researcher and his director, Drs. Simon Collin, will have access to the answers collected from the correction tasks, software usage data and questionnaire answers. This research material (digitally recorded answers to the correction tasks and questionnaire answers, as well as software usage data), in addition to your consent form, will be separately stored in a secure location only accessible by the researcher and his director for the duration of this project. The collected data and the consent forms will be kept for a period of 5 years following publication of the research, after which they will be shredded and/or permanently deleted.

VOLUNTARY PARTICIPATION

Participation in this research project is voluntary. This means that you accept to take part in the project without any constraint nor external pressure, and that you are free to withdraw from this research at any time. In that case, your data will be shredded and/or permanently deleted. Your consent to participate in this project confirms that you allow the researcher to use the collected data in his research (articles, essays or theses, conferences and scientific communications) on the condition that no information that allows your identification be publicly divulged without your explicit consent.

FINANCIAL COMPENSATION

No compensation will be given for participation in this study. A summary of the results will be sent to you upon the completion of this research project.

QUESTIONS ON THE PROJECT OR ON YOUR RIGHTS?

You can contact the researcher by calling (514) 508-5462 with any additional question on this project. You can also contact research director Simon Collin by calling (514) 987-3000 #2446 to discuss the conditions under which research is conducted and your rights as a research participant.

The project in which you will participate has been granted clearance according to the policy for the ethical review of research involving human participants. For any question that cannot be addressed to the research director, for a complaint or for comments, you may contact the chair of the Research Ethics Board for Students (Comité d'éthique de la recherche pour étudiants (CÉRPE)) by calling (514) 987-3000 #1646 or emailing (savard.josee@uqam.ca).

ACKNOWLEDGMENTS

Your collaboration is important to the completion of this project and we wish to thank you for your support.

SIGNATURES :

Your signature below indicates that you have read this letter of information and consent form and voluntarily agree to take part in this research project. You acknowledge that the researcher has answered your questions to your satisfaction and that you have had sufficient time to consider whether to participate in this study. You understand that your participation in this research is completely voluntary and that you can withdraw at any time, without any penalty, simply by advising the researcher.

Signature of Participant:

Date:

Print name and contact information:

I have explained the goals, the type of research, the advantages, and the risks of the project and answered, to the best of my abilities, the questions I was asked.

Signature of Researcher:

Date:

Print name and contact information:

APPENDICE C

QUESTIONNAIRE COMPLÉTÉ PAR LES PARTICIPANTS

Questionnaire

J'aimerais mieux vous connaître !
I would like to know you better!

1) J'ai...

I am...

40

ans

2) Je suis...

years old

I am...

...un homme ...une femme

...a man ...a woman

3) Je suis arrivé au Québec le...

I arrived in Quebec on...

6 juin 2014

4) J'étudie le français depuis...

I have been studying French for...

5

ans

5) J'ai étudié en...

years

I studied in...

français

6) Je travaille (ou je veux travailler) comme...

I work (or I plan to work) as a...

professeur

Quel est votre style d'étude ?

What is your studying style?

7) J'aime apprendre les règles en premier, puis pratiquer avec des exemples en deuxième.

I like to learn the rules first, then practice with examples after.

Je ne suis pas d'accord // *I do not agree* 1 2 3 4 Je suis d'accord // *I agree*

8) J'aime apprendre le français avec un manuel et un cahier d'exercices.

I like to study French with a textbook and a workbook.

Je ne suis pas d'accord // *I do not agree* 1 2 3 4 Je suis d'accord // *I agree*

9) J'aime apprendre le français en parlant avec les autres.

I like to study French by talking with others.

Je ne suis pas d'accord // *I do not agree* 1 2 3 4 Je suis d'accord // *I agree*

10) J'aime apprendre les règles en observant des exemples.

I like to learn rules by looking at examples.

Je ne suis pas d'accord // *I do not agree* 1 2 3 4 Je suis d'accord // *I agree*

11) J'aime utiliser un ordinateur pour apprendre le français.

I like to use a computer to study French.

Je ne suis pas d'accord // *I do not agree* 1 2 3 4 Je suis d'accord // *I agree*

12) J'aime utiliser un téléphone intelligent (iPhone, Android) ou une tablette (iPad, Kindle) pour apprendre le français.

I like to use a smartphone (iPhone, Android) or a tablet (iPad, Kindle) to study French.

Je ne suis pas d'accord // *I do not agree* 1 2 3 4 Je suis d'accord // *I agree*

13) Pour pratiquer mon français, j'aime...

To practice French, I like to...

- ...lire les nouvelles / *read the news* ...lire des livres / *read books*
 ...écouter de la musique / *listen to music* ...regarder des films / *watch movies*

Comment utilisez-vous la technologie ?

How do you use technology?

14) J'utilise un ordinateur...

I use a computer...

- ...à la maison / *at home* ...à l'école / *at school* ...au travail / *at work*

15) Chaque semaine, j'utilise l'ordinateur...

Every week, I use a computer

10

heures

hours

16) Avec mon ordinateur...

With my computer...

- ...je fais mes devoirs (Office) / *I do my homework*
...j'utilise un correcteur (Antidote) / *I use a spell check or proofing tool (Antidote)*
 ...je pratique avec Rosetta Stone / *I practice with Rosetta Stone*
...je communique (email, Facebook) / *I communication (email, Facebook)*

Que pensez-vous de l'activité ?

What did you think of the activity?

17) Les exemples sont faciles à lire.

The examples are easy to read

Je ne suis pas d'accord / *I do not agree* 1 2 3 4 Je suis d'accord / *I agree*

18) Le logiciel est utile pour pratiquer mon français.

This software is useful to practice French

Je ne suis pas d'accord / I do not agree 1 2 3 4 Je suis d'accord / I agree

19) Les exemples sont intéressants.

The examples are interesting

Je ne suis pas d'accord / I do not agree 1 2 3 4 Je suis d'accord / I agree

20) Je veux utiliser ce type de logiciel dans le futur.

I want to use this kind of software in the future.

Je ne suis pas d'accord / I do not agree 1 2 3 4 Je suis d'accord / I agree

21) J'utilise ce type de logiciel pour la première fois.

I used this type of software for the first time

Oui / Yes Non / No

22) Je recommande ce logiciel pour pratiquer le français.

I recommend this software to practice French

Oui / Yes Non / No

Vous avez fini? Merci de cliquer sur Envoyer !

APPENDICE D

RÉSULTATS À L'EXERCICE SUR LES PRÉPOSITIONS

Tableau D.1 Résultats (moyenne) des deux groupes pour des éléments spécifiques

	Groupe 1	Groupe 2
Travailler + Préposition (dans / avec)	70%	100%
Parler + Préposition (à / de)	70%	88,9%
Préposition + Temps (depuis)	70%	77,8%
Préposition + Lieu ¹ (à / en)	90%	77,8%
Préposition + Durée (pendant)	40%	22,2%
Jouer + Préposition ² (à / de)	55%	77,8%
Aller + Préposition ³	-	88,9%

1 Chaque groupe avait deux phrases à compléter avec la combinaison « Préposition + Lieu »; pour chaque groupe, les performances à ces deux phrases (90% et 90% pour le groupe 1, 100% et 55,6% pour le groupe 2) ont été combinées en un résultat unique.

2 Le groupe 1 avait deux phrases à compléter avec la combinaison « Jouer + Préposition », tandis que le groupe 2 avait une seule phrase avec cette combinaison.

3 La combinaison « Aller + Préposition » était seulement présente dans l'exercice du groupe 2.

BIBLIOGRAPHIE

- Acar, A., Geluso, J., & Shiki, T. (2011). How can search engines improve your writing. *CALL-EJ*, 12(1), 1-10.
- Association Vikidia. (2015). Vikidia: Règles. Récupéré le 20 mars 2017 de <https://fr.wikidia.org/wiki/Vikidia:Règles>
- Association Vikidia. (2016). Vikidia:À propos. Récupéré le 20 mars 2017 de https://fr.wikidia.org/wiki/Vikidia:À_propos
- Association Vikidia. (2017). Vikidia:Vikidia hors-connexion. Récupéré le 20 mars 2017 de https://fr.wikidia.org/wiki/Vikidia:Vikidia_hors-connexion
- Aston, G. (1997). Small and large corpora in language learning. Dans B. Lewandowska-Tomaszczyk & P.J. Melia (eds), *PALC 97: Practical applications in language corpora*, 51-62.
- Boulton, A. (2007a). But where's the proof? The need for empirical evidence for data-driven learning. *Proceedings of the BAAL Annual Conference 2007*, 13-16.
- Boulton, A. (2007b). DDL is in the details... and in the big themes. *Proceedings of the Corpus Linguistics Conference: CL2007*, 1-13.
- Boulton, A. (2008). Looking (for) empirical evidence of data-driven learning at lower levels. *Corpus Linguistics, Computer Tools, and Applications: State of the Art*, 581-598.
- Boulton, A. (2009a). Corpora for all? Learning styles and data-driven learning. *5th Corpus Linguistics Conference*.
- Boulton, A. (2009b). Testing the limits of data-driven learning: language proficiency and training. *ReCALL*, 21(1), 37-54.
- Boulton, A. (2010a). Data-driven learning: Taking the computer out of the equation. *Language Learning*, 60(3), 534-572.
- Boulton, A. (2010b). Learning outcomes from corpus consultation. *Exploring New Paths in Language Pedagogy: Lexis and Corpus-Based Language Teaching*, 129-144.

- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(01), 47–64.
- Breton, Pascale. (2016, 15 août). Ne tabletons pas la technologie. *La Presse+*, section DÉBATS, écran 2. Récupéré le 20 mars 2017 de http://plus.lapresse.ca/screens/2b3247a5-edea-48f1-ab2c-4280933c8a31%7C_0.html
- Brouwers, L., Bernhard, D., Ligozat, A.-L., & François, T. (2012). Simplification syntaxique de phrases pour le français. *Actes de JEP-TALN-RECITAL*, 211–224.
- Bruley-Meszaros, C. (2010). Quel enseignement des verbes en didactique du français langue étrangère? *Synergies France*, 6, 51–59.
- Cambrensis, J., & Gillway, M. (2006). The study of selected vocabulary in context—using technology to motivate learners. *UGRU Journal*, 3, 1-10.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2), 111–125.
- Chambers, A. (2010). L'apprentissage de l'écriture en langue seconde à l'aide d'un corpus spécialisé. *Revue Française de Linguistique Appliquée*, 15(2), 9–20.
- Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL: the Journal of EUROCALL*, 16(01), 158–172.
- Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology*, 10(2), 151–170.
- Chang, W.-L., & Sun, Y.-C. (2009). Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning*, 22(4), 283–302.
- Chujo, K., Anthony, L., & Oghigian, K. (2009). DDL for the EFL classroom: Effective uses of a Japanese-English parallel corpus and the development of a learner-friendly, online parallel concordancer. *Proceedings of 5th Corpus Linguistics Conference*.
- Chujo, K., Utiyama, M., & Miura, S. (2006). Using a Japanese-English parallel corpus for teaching English vocabulary to beginning-level students. *English Corpus Studies*, 13, 153–172.

- Collès, L. (2002). Les textes en français simplifié: une voie pour l'enseignement de la littérature? *Praxis Des Neusprachlichen Unterrichts*, 49, 402–409.
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301–315.
- Cobb, T. (2014a). A resource wish-list for data-driven learning in French. *French through corpora: Ecological and data-driven perspectives in French language studies*. Newcastle, Cambridge Scholars, 255-290.
- Cobb, T. (2014b). Where do Lextutor FRENCH Corpora come from? Récupéré le 20 mars 2017 de http://www.lexutor.ca/conc/fr/corpus_descriptions_french.html
- Cobb, T. [s.d.]. VOCABPROFILE (COMPLEAT). Récupéré le 12 mai 2016 de <http://www.lexutor.ca/vp/comp/>
- Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. *Cambridge Handbook of English Corpus Linguistics*, 478-497.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *The handbook of second language acquisition*, 63-103.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301–319.
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97–118.
- Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363–372.
- Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27(1), 70–105.
- Gordani, Y. (2013). The effect of the integration of corpora in reading comprehension classrooms on English as a Foreign Language learners' vocabulary development. *Computer Assisted Language Learning*, 26(5), 430–445.
- Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*, 20(3), 465–480.

- Grégoire, M., & Thiévenaz, O. (2003). *Grammaire progressive du français*. Paris: Clé International
- Grgurović, M., Chapelle, C. A., & Shelley, M. C. (2013). A meta-analysis of effectiveness studies on computer technology-supported language learning. *ReCALL*, 25(02), 165–198.
- Hadley, G. (2002). An introduction to data-driven learning. *RELC Journal*, 33(2), 99–124.
- Homma, Y. (2006). Analyse critique et révision de quelques point de vue théoriques sur l'alternance entre A et DANS en vue d'une problématique de l'enseignement des prépositions françaises en FLE. *Colloque international "Recherches en acquisition et en didactique des langues étrangères et secondes"*, p.1-13.
- Hunston, S., & Francis, G. (1998). Verbs Observed: A Corpus-driven Pedagogic Grammar. *Applied Linguistics*, 19(1), 45–72.
- Immigration, Diversité et Inclusion Québec. (2016). Connaissances en français et en anglais pour les candidats du Programme régulier des travailleurs qualifiés. Récupéré le 20 mars 2017 de <https://www.immigration-quebec.gouv.qc.ca/fr/immigrer-installer/travailleurs-permanents/conditions-requises/connaissances-linguistiques.html>
- John, E. S. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. *Language Learning & Technology*, 5(3), 185-203.
- Johns, T. F., Hsingchin, L., & Lixun, W. (2008). Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning*, 21(5), 483–506.
- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77–90.
- Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25(1), 105-121.
- Landure, C. (2011). Data-driven learning : Apprendre et enseigner à contre-courant. *Mélanges CRAPEL*, 32(spécial), 163-178.

- Landure, C., & Boulton, A. (2010). Corpus et autocorrection pour l'apprentissage des langues. *ASp. La Revue Du GERAS*, (57), 11–30.
- Liu, D., & Jiang, P. (2009). Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts. *The Modern Language Journal*, 93(1), 61–78.
- Logos Group. (2017). Logos Library. Récupéré de <http://www.logoslibrary.eu>
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1), 43-52.
- Ministère de l'Éducation, du Loisir et du Sport. (2007). Programme de formation de l'école québécoise: Enseignement secondaire, deuxième cycle. Récupéré le 20 mars 2017 de <http://www1.education.gouv.qc.ca/sections/programmeFormation/secondaire2/>
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology*, 87(2), 319.
- Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4), 373–391.
- New, B. et Wordtheque. (2006). Corpatext 1.01 Récupéré le 20 mars 2017 de <http://www.lexique.org/telAutresBases.php>
- Oghigian, K., & Chujo, K. (2010). An effective way to use corpus exercises to learn grammar basics in English. *Language Education in Asia*, 1(1), 200–214.
- Oh, S.-Y. (2001). Two Types of Input Modification and EFL Reading Comprehension: Simplification Versus Elaboration. *TESOL Quarterly*, 35(1), 69–96.
- O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49–68.
- Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, J. M., & Jiménez, P. A. (2011). Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation. *Computer Assisted Language Learning*, 24(3), 233–253.

- Römer, U. (2004). Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching. *Corpora and Language Learners*, 151–168.
- Roy, S. (1997). Le traitement des prépositions à et de dans quatre manuels scolaires de français langue seconde. Montréal : Université du Québec à Montréal.
- Roy, S. (1998). A l'attaque contre les prépositions!. *The Canadian Modern Language Review*, 55(2), 283–286.
- Ruiz de Zarobe, L. (2015). Aménager les textes pour mieux les comprendre en Français Langue Étrangère: le type «élaboration» comme proposition. *Çedille. Revista de Estudios Franceses*, (11), 469-483.
- Sun, Y.-C., & Wang, L.-Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16(1), 83–94.
- Todd, R. W. (2001). Induction from self-selected concordances and self-correction. *System*, 29(1), 91–102.
- Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, 21(2), 32–51.
- Xue-hua, A., & Ming-ying, X. (2013). An Empirical Research on DDL in L2 Writing. *US-China Education Review*, 3(9), 693–700.
- Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10(3), 130–139.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257–283.
- ZURB, Inc. (2017). Foundation | The most advanced responsive front-end framework in the world. Récupéré le 20 mars 2017 de <http://foundation.zurb.com>