



HAL
open science

Analyse des erreurs verbales dans un corpus d'apprenants sinophones du français

Jingrao Li

► **To cite this version:**

Jingrao Li. Analyse des erreurs verbales dans un corpus d'apprenants sinophones du français. Sciences de l'Homme et Société. 2020. dumas-03239432

HAL Id: dumas-03239432

<https://dumas.ccsd.cnrs.fr/dumas-03239432>

Submitted on 27 May 2021

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

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

Analyse des erreurs verbales dans un corpus d'apprenants sinophones du français

LI

Jingrao

Sous la direction d'Agnès Tutin

Laboratoire : LIDILEM

UFR LLASIC

Département Science du langage

Mémoire de master 2 - 20 crédits

Parcours : Linguistique

Année universitaire 2020-2021

Analyse des erreurs verbales dans un corpus d'apprenants sinophones du français

LI

Jingrao

Sous la direction d'Agnès Tutin

Laboratoire : LIDILEM

UFR LLASIC

Département Science du langage

Mémoire de master 2 - 20 crédits

Parcours : Linguistique

Année universitaire 2020-2021

Remerciements

Je tiens à remercier toutes les personnes qui m'ont aidée à réaliser mon mémoire.

Tout d'abord, je souhaite remercier ma directrice de mémoire Agnès Tutin. Le savoir scientifique et l'attitude rigoureuse qu'elle a déployés m'ont énormément impressionnée. Un grand merci tout particulier à Agnès Tutin de m'avoir offert l'opportunité d'effectuer un stage sur la constitution du corpus d'apprenants. Grâce à cette expérience, j'ai pu mieux connaître la linguistique de corpus et le langage XML.

Je tiens à remercier Madame Rui Yan et Madame Anne-Cécile Perre qui m'ont aidée à collecter les corpus auprès des apprenants sinophones au CUEF, et aussi Madame Tatiana Aleksandrova, grâce à qui la fiche sociobiographique a été élaborée. Un grand merci à tous les apprenants d'avoir permis la diffusion de leurs textes.

Pour l'annotation des erreurs, j'adresse mes remerciements les plus sincères pour Magdalena Augustyn de m'avoir initiée au traitement du corpus d'apprenants en XML. Grand merci pour sa patience et sa contribution dans la conception des codes.

Mes remerciements s'adressent également aux enseignantes de l'organisme Coup de Pouce, qui m'ont soutenue pendant l'élaboration de mon mémoire. Ce sont Madame Marie-Jo Baffert, enseignante de FLE et Madame Elisabeth Calaque, spécialiste en linguistique. Merci pour leur aide dans la vérification des erreurs verbales relevées dans les textes et leur relecture inlassable du mémoire. Je tiens à remercier le Coup de Pouce et les bénévoles pour le soutien qu'ils offrent aux étudiants étrangers. Je suis vraiment impressionnée par leur état d'esprit ouvert, leur attitude accueillante et les échanges culturels vivants.

Je remercie aussi toute équipe du laboratoire LIDILEM, surtout mon amie Yujing Ji, et mes camarades en linguistique, pour les nombreux d'échanges scientifiques et amicaux.

DÉCLARATION

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

NOM :

PRENOM :

DATE :

SIGNATURE :

Sommaire

Introduction.....	8
Chapitre I. Corpus d'apprenants.....	11
1.1 Le corpus d'apprenants et l'acquisition de la langue seconde.....	11
1.2 Le corpus d'apprenants et la didactique des langues.....	13
1.3 Les corpus d'apprenants dans les différentes langues.....	13
1.4 Les corpus d'apprenants pour le français.....	15
Chapitre II. Les erreurs dans les corpus d'apprenants.....	18
2.1 La notion d'« erreur ».....	18
2.2 L'intérêt d'étudier l'« erreur » et le système de l'interlangue.....	20
2.3 Les grilles d'erreurs des chercheurs.....	23
Chapitre III. Méthodologie : constitution du corpus et principes d'annotation..	31
3.1 La méthodologie de traitement des données.....	31
3.2 La constitution d'une grille d'erreurs.....	34
3.3 L'annotation systématique dans le corpus en XML.....	54
Chapitre IV. L'analyse et l'interprétation des données.....	64
4.1 L'analyse quantitative et qualitative des erreurs.....	64
4.2 La comparaison entre les textes écrits sur table et à la maison.....	85
4.3 L'interprétation des sources d'erreurs.....	89
4.4 Conclusion.....	103
Conclusion.....	106
Bibliographie.....	109
Sitographie.....	112
Glossaire.....	113
Sigles et abréviations utilisés.....	115
Table des illustrations.....	116
Table des annexes.....	117

Introduction

Le verbe est un élément fondamental et le noyau de la phrase, dont la fonction est d'organiser le discours, de se positionner, de présenter des informations essentielles. C'est aussi l'élément qui trahit le plus facilement le statut non natif. D'après la thèse de Ying Xiaohua (2013) consacrée aux erreurs commises par les apprenants sinophones à l'écrit, les erreurs sémantiques de TAM (temps, aspect, mode) et les erreurs concernant les conjugaisons arrivent en deuxième place parmi tous les types d'erreurs, ce qui représente une proportion de 15 %. On peut donc affirmer que les apprenants sinophones ne maîtrisent pas bien les verbes. Cela pourrait être expliqué par le fait que les verbes sont invariables en chinois, donc la maîtrise des verbes en français qui implique l'acquisition de la conjugaison, de l'accord et de TAM (temps, aspect, mode) s'avère particulièrement difficile pour les Chinois. Ce constat nous amène à proposer d'établir un diagnostic détaillé qui nous permettra de mieux comprendre leurs difficultés.

Nous pensons que l'analyse des erreurs verbales commises par les apprenants sinophones à l'écrit nous fournit une approche efficace pour envisager ce problème. Classifier les erreurs verbales des apprenants sinophones par catégorie devrait permettre de révéler la régularité des erreurs commises et d'entrevoir leurs principales difficultés dans l'apprentissage. Les données écrites ou orales produites par les apprenants sont une ressource centrale pour l'étude de l'acquisition des langues secondes, dans le but de mieux comprendre les mécanismes de l'apprentissage des langues. La constitution du corpus d'apprenants fournit des données authentiques pour étudier le système langagier des apprenants. Depuis les années 1980, les recherches dans ce domaine se sont développées dans le champ de l'acquisition de la langue seconde, de la linguistique de corpus et de la didactique des langues.

Plusieurs études ont été menées sur l'analyse des erreurs des apprenants en anglais langue seconde (Zheng & Park 2013, Omidipour 2014, Phuket 2015). Cependant, l'analyse des erreurs à l'écrit reste un sujet peu abordé pour le français. Il existe quelques recherches consacrées à l'analyse des erreurs lexicales fondée sur les corpus, comme le mémoire d'Anctil (2005) et sa thèse (2010), mais sur le français langue maternelle et non sur le français comme langue étrangère. Peu de travaux se sont concentrés sur l'analyse des erreurs verbales, en particulier pour les apprenants sinophones. La thèse de Yan (2017) fournit cependant une très bonne base en combinant l'observation des erreurs verbales et le statut d'apprenants sinophones. Parce que sa recherche s'inscrit dans le cadre de l'écrit scientifique, elle s'intéresse particulièrement au lexique verbal scientifique transdisciplinaire, qui ne représente qu'une facette de tous les verbes. Dans le cadre de notre mémoire, nous nous intéresserons de

façon plus large à tous les verbes apparaissant dans les corpus afin d'évaluer la proportion des verbes erronés et de les analyser.

Du point de vue méthodologique, dans les travaux précédents sur l'analyse des erreurs en français (Anctil 2010, Yan 2017), les auteurs ont extrait les erreurs du corpus en les classant par catégorie dans une typologie d'erreurs, mais un travail d'annotation systématique des erreurs sur corpus n'a pas été effectué pour le français, mis à part le travail de Granger (2007) avec le corpus Frida. Dans notre étude, outre la typologie d'erreurs, nous constituerons un corpus d'apprenants sinophones indiquant les erreurs verbales au format XML.

Nous avons mentionné l'importance d'étudier les difficultés des apprenants sinophones à maîtriser le système verbal du français. Il faut également souligner l'importance du corpus d'apprenants pour l'étude de l'acquisition de la L2, et le faible nombre de recherches sur l'analyse des erreurs fondées sur les corpus d'apprenants sinophones en français. Nous avons donc décidé de travailler sur les erreurs verbales commises par les apprenants sinophones en prenant en compte leurs spécificités, à l'aide d'une grille d'erreurs, et en effectuant un travail d'annotation systématique des erreurs dans le corpus, ce qui permettra de déboucher sur une analyse quantitative et qualitative.

Dans ce mémoire, en étudiant le corpus d'apprenants de niveau B1-B2, nous cherchons à répondre aux questions suivantes : quelles sont les principales erreurs verbales commises à l'écrit par les apprenants sinophones ? Comment peut-on expliquer ces erreurs ?

Parce que les verbes sont invariables en chinois, c'est-à-dire qu'ils n'ont pas de conjugaison ou d'accord en fonction des personnes, du temps, de l'aspect ou du mode, nous faisons l'hypothèse que les erreurs de conjugaison, d'accord et de TAM des verbes seront plus fréquentes que d'autres types d'erreurs. En ce qui concerne les sources d'erreurs, le système conceptuel du monde des apprenants chinois a été établi à partir de leur langue maternelle, à savoir le chinois, donc il y a lieu de croire qu'ils ne se débarrassent pas du système sémantique du chinois pendant l'apprentissage du français. De plus, la plupart des apprenants sinophones ont commencé à étudier l'anglais depuis l'école primaire, et son apprentissage est obligatoire jusqu'à l'éducation supérieure. Du fait que l'anglais ressemble morphologiquement au français, nous faisons l'hypothèse que l'influence de l'anglais se manifeste davantage au niveau morphologique.

Dans un premier temps, nous présenterons les recherches sur les corpus d'apprenants portant sur l'acquisition de la L2 et la didactique des langues. Dans un second temps, nous proposerons une synthèse des travaux consacrés à l'analyse des erreurs dans les corpus, avec une attention particulière aux typologies d'erreurs élaborées par les chercheurs. Nous

exposerons ensuite la méthodologie de recherche, avec la présentation du corpus, de la grille d'erreurs élaborée et de l'annotation semi-automatique des erreurs dans le corpus. Dans un dernier temps, nous interpréterons les résultats à travers une analyse quantitative et qualitative, pour observer la répartition des erreurs, faire ressortir les circonstances où les erreurs apparaissent fréquemment, et identifier les sources possibles d'erreurs.

Chapitre I. Corpus d'apprenants

Dans les années 1980s, des corpus d'apprenants d'envergure ont vu le jour grâce au développement de l'informatique. Les corpus d'apprenants sont définis comme « les collections électroniques des données naturelles ou presque naturelles produites par des apprenants étrangers ou de la langue seconde, et rassemblées selon des critères explicites » (Granger, Gilquin & Meunier, 2015, p.1). C'est un domaine de recherche interdisciplinaire qui relève à la fois de la linguistique de corpus, de l'acquisition des langues secondes, de la didactique des langues et du traitement automatique des langues.

Dans ce chapitre, nous commencerons par la présentation de ce domaine en lien avec l'acquisition de la langue seconde et la didactique des langues. Nous présenterons ensuite les corpus d'apprenants existants dans différentes langues étrangères, surtout pour le français langue étrangère, afin d'illustrer le développement de ce domaine.

1.1 Le corpus d'apprenants et l'acquisition de la langue seconde

Les recherches sur les corpus d'apprenants constituent une nouvelle approche dans l'étude de l'acquisition de la langue seconde (Lozano & Mendikoetxea 2013, De Cock & Tyne 2014, Li 2018).

L'acquisition de la L2 s'intéresse au langage des apprenants et à son développement. Dans l'étude de la L2, Lado (1957) propose le concept de « contrastive hypothesis » (hypothèse contrastive) : nous pouvons prédire les constructions susceptibles d'être mal employées par les apprenants, et celles qui ne poseront pas de difficultés dans leur apprentissage de la L2, à travers une comparaison entre le système langagier et culturel de la L1 et la L2 des apprenants. Les structures différentes dans les deux langues devraient être le plus souvent source d'erreurs. Mais cette idée est contredite par le fait que les apprenants ayant des langues maternelles différentes commettent souvent des erreurs identiques dans la L2, donc toutes les erreurs ne peuvent pas être expliquées par l'influence de la L1. En effet, seulement une minorité des erreurs trouvent leur origine dans la L1, mais la majorité des erreurs apparaissent avec l'apprentissage. Ce constat remet en cause l'approche contrastive dans la recherche de l'acquisition de la L2. C'est pourquoi cette approche contrastive évolue dans le sens inverse dans les années 1970s. Les chercheurs partent des productions authentiques des apprenants et travaillent sur les formes et les erreurs observées pour décrire le développement langagier des apprenants (Granger, Gilquin & Meunier, 2015). L'accent n'est mis non plus sur la comparaison entre deux langues, mais sur l'exploration des données

produites par les apprenants eux-mêmes : la structure interne des données, leur développement, et la façon dont les données d'apprenants sont dérivées de la langue cible.

Ainsi, les données écrites ou orales produites par les apprenants deviennent une ressource significative pour l'étude de l'acquisition de la L2. Cependant, ces données sont parfois considérées comme biaisées à cause du caractère contrôlé des tâches langagières, de sorte que les productions des apprenants ne représentent pas forcément celles qui apparaissent dans les contextes de communication plus naturels. Par ailleurs, les données recueillies sont de petite taille, impliquant un nombre d'apprenants limité. Ces limites remettent en cause la représentativité de ces données. Mais grâce au développement de l'informatique, il est plus facile de traiter un grand nombre de données en constituant un corpus d'apprenants d'envergure, ce qui augmente la représentativité des données authentiques. Outre l'augmentation de la représentativité, le format électronique des données permet également une analyse automatique avec les outils informatiques pour améliorer l'efficacité de l'analyse.

Toutefois, il existe une déficience potentielle dans l'étude du corpus d'apprenants pour observer le système langagier des apprenants : certains phénomènes linguistiques pourraient être surestimés ou sous-estimés à cause des stratégies adoptées par les apprenants dans leur production écrite, comme les stratégies d'évitement des constructions compliquées, surtout dans les situations formelles qui exigent de minimiser les erreurs pour obtenir de meilleurs résultats, comme pendant des examens. Il est incontestable que la méthode que l'apprenant adopte pour acquérir le vocabulaire, les expressions figées et les règles grammaticales se manifeste plus clairement dans les textes écrits qu'il produit que dans les tests à choix multiples, mais en même temps, l'écriture libre donne plus d'autonomie aux apprenants pour contrôler leur production. Ils ont ainsi tendance à éviter les expressions susceptibles d'être mal employées et les structures compliquées, minimisant ainsi les erreurs, surtout pendant les évaluations sur table. Ces stratégies d'évitement peuvent constituer un frein pour accéder à l'état authentique de leur apprentissage.

Malgré ces déficiences, le corpus d'apprenants occupe quand même une place importante dans la recherche sur l'acquisition de la L2. Il est particulièrement bien adapté pour l'étude de l'acquisition lexicale, car il permet une recherche automatique des lexèmes dans une grande masse de données.

Bien que la recherche sur le corpus d'apprenants s'inscrive dans la recherche sur l'acquisition de la L2 pour comprendre les mécanismes d'acquisition de la L2, elle est caractérisée par sa visée didactique.

1.2 Le corpus d'apprenants et la didactique des langues

Le corpus d'apprenants est caractérisé par sa visée didactique. L'élaboration d'un corpus d'apprenants est un pont qui relie les recherches de la linguistique de corpus avec la didactique des langues. L'utilisation du corpus d'apprenants représente de nombreux avantages pour la pédagogie. Tout d'abord, le corpus d'apprenants est une ressource d'exemples authentiques pour la rédaction des manuels, qui pourraient remplacer les exemples artificiels fabriqués par les éditeurs. Par ailleurs, les enseignants peuvent élaborer leur propre corpus d'apprenants auprès de leurs étudiants et identifier les erreurs dans une approche empirique, afin de mieux comprendre le développement de l'apprentissage de leurs étudiants. Les résultats serviraient à l'élaboration des programmes d'enseignement destinés à répondre aux besoins spécifiques des étudiants. Il est aussi possible de laisser les apprenants explorer le corpus et les corrections proposées. Dans ce processus, les apprenants ne sont plus des acteurs passifs qui reçoivent les informations de l'enseignant, mais des acteurs actifs qui contrôlent leur apprentissage sous la direction de l'enseignant. Ils peuvent ainsi bénéficier non seulement de la correction de leurs propres erreurs mais aussi de celle des erreurs de leurs camarades. De plus, le corpus d'apprenants peut servir à organiser les activités pédagogiques. Par exemple, dans l'étude de Bloch en 2009, les phrases issues du corpus sont données aux apprenants avec les verbes soulignés, et il est demandé aux apprenants de choisir ou de proposer des verbes plus appropriés que ceux qui sont utilisés (Granger, Gilquin & Meunier, 2015, p.449).

1.3 Les corpus d'apprenants dans les différentes langues

Bien que l'histoire du corpus d'apprenants ne soit pas très longue, ce domaine s'est bien développé ces dernières années, surtout les corpus oraux et écrits pour l'anglais langue seconde, comme *Louvain International Database of Spoken English Interlanguage* (Gilquin et al., 2010), *Longman Learner Corpus* (LLC, Samuel Johnson, 1755), *Cambridge Learner Corpus* (CLC) créé par l'université Cambridge. Dans le cadre de notre mémoire, nous ne présenterons que deux corpus emblématiques en anglais langue seconde : *International Corpus of Learner English (ICLE)* et *Written Corpus of Learner English (WriCLE)*.

Le corpus d'apprenants de l'anglais langue seconde le plus connu est sans doute *International Corpus of Learner English* (Granger et al., 2009). Ce corpus est constitué des textes argumentatifs rédigés par les apprenants de niveau intermédiaire à niveau avancé de différentes nationalités. L'élaboration de ce corpus a été réalisée grâce à la collaboration internationale de plusieurs universités.

La première version de ICLE a été publiée en 2002, comportant 2.5 millions de mots et concernant des apprenants de 11 langues maternelles différentes (bulgare, tchèque, néerlandais, finnois, français, allemand, italien, polonais, russe, espagnol et suédois.). La deuxième version, ICLEv2, comportant 4.5 millions de mots avec des apprenants de 16 nationalités différentes (bulgare, chinois, tchèque, néerlandais, finnois, français, allemand, italien, japonais, norvégien, polonais, russe, espagnol, suédois, tswana, turc), a été publiée en 2009. La deuxième version se différencie de la première par l'extension du corpus et des nationalités des apprenants, mais aussi par le développement de l'interface du corpus avec l'ajout de deux nouvelles fonctionnalités : le concordancier intégré et la répartition des résultats de recherche en fonction des variables enregistrées par ICLE. Le concordancier intégré permet aux utilisateurs de chercher directement les formes, les lemmes ou les balises de partie de discours, sans avoir à recourir à d'autres outils pour interroger les données. La répartition des résultats d'une requête linguistique en fonction de l'ensemble des variables enregistrées par ICLE (comme la langue maternelle, le sexe, l'âge, le type de tâche) met en évidence la relation entre les résultats de recherche et les variables d'apprenants ou de tâches.

Ce corpus a été élaboré dans le but de fournir une base empirique solide pour la théorie sur l'acquisition de la langue seconde, et d'autre part de contribuer à la conception des matériaux pédagogiques destinés aux apprenants avancés, ou à un groupe d'apprenants partageant la même langue maternelle (Granger et al., 2009).

Le corpus ICLE permet d'observer les différences entre la langue des apprenants avancés et celle des natifs en prenant en compte l'influence de la L1. Il a été très exploité et il fait l'objet d'études de nombreux travaux consacrés aux corpus d'apprenants (Stormbom 2018, Ringbom 1999, Virtanen 1996).

Il existe également des corpus d'anglais langue seconde plus spécifiques et destinés à un groupe d'apprenants partageant une même langue maternelle. Le corpus *Written Corpus of Learner English* (WriCLE, Mendikoetxea & O'Donnell, 2009) a été créé par l'université Autonome de Madrid. C'est un corpus constitué de 700 textes académiques en anglais rédigés par les apprenants hispanophones. Les apprenants sont de niveaux intermédiaire à avancé. Le corpus WriCLE se différencie du corpus ICLE par ses indications de niveaux des apprenants selon l'évaluation de Common European Framework of Reference for Languages (CEFR), une norme internationale pour évaluer la capacité langagière, de niveaux A1 à C2. Un sous-corpus d'écrits non académiques est en cours de constitution, principalement des blogs (WriCLEinf). Cela permet la comparaison entre différents registres, ainsi que l'étude de structures non formelles qui sont normalement absentes dans des rédactions académiques (Lozano & Mendikoetxea, 2013, p.71).

Outre le grand nombre de corpus d'apprenants pour l'anglais, les corpus pour d'autres langues étrangères ont également été développés. Par exemple, le corpus d'apprenants en espagnol *Corpus Escrito del Español L2* (CEDEL2) a été constitué par l'université Autonome de Madrid. Ce corpus s'inscrit dans le cadre du projet *Word Order in Second Language Acquisition Corpora* (WOSLAC) piloté par Cristóbal Lozano en 2009. Il comporte 750000 mots des textes écrits par différents niveaux d'apprenants anglophones de l'espagnol langue seconde : débutants, intermédiaires et avancés, et les textes sont répartis en fonction du niveau de l'apprenant. Ce corpus a été employé dans la recherche sur les collocations en espagnol (Alonso Ramos et al., 2010). Une étude contrastive fondée sur le corpus L1 anglais – L2 espagnol (CEDEL2) et le corpus L1 espagnol – L2 anglais (WriCLE) est envisagée pour mieux identifier l'origine des déficiences d'apprentissage (Lozano & Mendikoetxea, 2013).

1.4 Les corpus d'apprenants pour le français

Il existe également plusieurs corpus d'apprenants du français langue étrangère, par exemple les corpus écrits comme *The "Dire Autrement" corpus* (Hamel et Milicevic, 2010), *The Lund CEFLE Corpus* (Ågren, 2003), le *French Interlanguage Database* (Granger, 1998), et les corpus oraux comme *The COREIL corpus* (Delais-Roussarie et Yoo, 2010), *French Learner Language Oral Corpora* (Myles et Mitchell, 2001), *The InterFra corpus* (Bartning, 2016). Comme notre étude porte sur les corpus écrits, nous présenterons ici trois corpus écrits sur le français langue étrangère : *The "Dire Autrement" corpus*, *The Lund CEFLE Corpus* et le *French Interlanguage Database*.

Le corpus *Dire Autrement* est composé de 200 à 250 textes rédigés par une trentaine d'apprenants universitaires de différentes nationalités du français langue seconde, de niveau intermédiaire à avancé. Les textes se répartissent en quatre situations textuelles : informative, narrative, incitative et persuasive, et portent sur trois thèmes principaux : la santé, la bonne alimentation et le sport. Ce corpus s'inscrit dans le cadre du projet *Dire Autrement* qui vise à concevoir un dictionnaire électronique de reformulation pour les apprenants du français langue seconde. Le thème du projet porte sur les liens que les unités lexicales entretiennent au sein du lexique, comme le lien de collocation (comme *combler une lacune*), le lien de dérivation sémantique (synonymie, antonymie, nominalisation, par exemple, la nominalisation *essayage* du verbe *essayer* dans *essayer une robe*), et le lien de paraphrase (comme *analyser* et *effectuer une analyse*) (Hamel et Milicevic, 2007). Dans le but de produire un dictionnaire de reformulation pour favoriser l'acquisition de liens lexicaux et la réutilisation des paraphrases ou des dérivations sémantiques dans des situations de

communication semblables ou nouvelles, les erreurs ayant comme source une maîtrise insuffisante de liens lexicaux sont mises en avant dans l'analyse du corpus.

Le Corpus Écrit de Français Langue Étrangère (CEFLE) est un corpus d'apprenants suédois du français langue étrangère. Ce corpus est constitué de 400 textes écrits par des suédophones du français L2 au lycée en Suède et 60 textes rédigés par 30 scripteurs francophones en tant que groupe de contrôle. Les francophones natifs ont approximativement le même âge que les apprenants suédois. Les textes sont des rédactions libres de type descriptif et narratif. Le corpus est hautement homogène car les variables sont bien contrôlées. Les apprenants ont la même langue maternelle et approximativement le même âge. Les textes ont été écrits dans une même salle d'ordinateur, sous la direction d'une même expérimentatrice, dans un même temps limité et dans les mêmes conditions (les apprenants ont été prévenus que ce n'était pas un examen). Quand les autres variables sont contrôlés, la différence entre les rédactions des apprenants reflète mieux la différence entre leur compétence langagière.

Il est intéressant de noter que le corpus est à la fois longitudinal et transversal. 110 apprenants du français au lycée ont été suivis pendant une année scolaire, entre septembre 2003 et mai 2004. Ces apprenants ont entre 16 et 19 ans et ils suivent différents niveaux de cours de français : 1re, 2e, 4e et 5e année de français. La comparaison entre les rédactions d'un même groupe d'apprenants à différents moments d'apprentissage permet d'observer leur progression dans le temps, et la comparaison entre les rédactions des apprenants à des stades d'apprentissage différents permet de révéler les caractéristiques dans ce stade d'apprentissage.

Le corpus *French Interlanguage Database (Frida)* a été élaboré dans le cadre du projet Frida lancé par le « Centre for English Corpus Linguistics » de Sylviane Granger en 1998. Ce corpus contient actuellement 764 textes, soit 9466 phrases et 179642 mots (Biskri & Jebali, 2010, p.154). Il s'agit de textes rédigés par des apprenants de niveaux variables ayant des langues maternelles très diverses (néerlandais, allemand, japonais, anglais, etc.). Plusieurs types d'écrits figurent dans ce corpus : textes descriptifs, argumentatifs, narratifs, et ce sont des compositions libres : les textes sont réalisés sur un sujet imposé mais ouvert comme « raconter un fait divers » ou « qu'est-ce que le bonheur ? » (Ibid). Les erreurs dans ce corpus d'apprenants sont annotées en vue de constituer un outil d'apprentissage du français, et à long terme, de déboucher sur la conception d'un correcteur automatique destiné aux utilisateurs non francophones. Le corpus Frida a été très exploité dans les recherches sur l'acquisition du français langue seconde, comme celle de Bolly (2008) sur les unités phraséologiques autour des verbes *donner* et *prendre* en français écrit L1 et L2.

Comme nous pouvons le constater, les corpus d'apprenants ont été énormément développés ces dernières années. Les langues s'étendent de l'anglais à une grande variété de langues étrangères : l'espagnol, le français... Il existe des corpus comportant plusieurs nationalités d'apprenants, qui ont pour avantage de pouvoir comparer l'apprentissage des apprenants avec des langues maternelles différentes afin de mieux identifier l'influence de la L1, et aussi des corpus destinés aux apprenants partageant une même langue maternelle pour répondre aux besoins spécifiques de ce groupe d'apprenants.

Pour résumer, dans ce chapitre, nous voyons qu'à l'aide du traitement automatique des langues, les corpus d'apprenants constituent des ressources précieuses dans les recherches sur l'acquisition de la langue seconde et la didactique des langues : ils peuvent servir de bases empiriques pour les recherches sur l'acquisition de la langue seconde, mais aussi être appliqués dans l'élaboration des matériaux didactiques, dans la conception des correcteurs automatiques destinés aux apprenants étrangers. Malgré le développement remarquable des corpus d'apprenants, les corpus d'apprenants sinophones du français langue seconde restent peu développés. Nous sommes convaincue que le corpus d'apprenants sinophones que nous élaborerons dans le cadre de notre mémoire nous aidera à observer l'acquisition du verbe chez les apprenants sinophones et, à long terme, constituera une ressource applicable dans la didactique du français auprès des apprenants sinophones.

Chapitre II. Les erreurs dans les corpus d'apprenants

Pour construire une grille d'erreurs qui nous permette de mieux comprendre les difficultés des apprenants sinophones sur les verbes, nous avons besoin de connaître l'existant. Les erreurs ont été très étudiées dans les corpus d'apprenants en anglais (Zheng & Park 2013, Omidipour 2014, Na Phuket 2015), mais moins abordées en français.

Dans ce chapitre, nous commencerons par l'explication de la notion d'erreur. Nous montrerons ensuite l'intérêt d'étudier les erreurs avec la présentation du système d'interlangue. À la fin, nous nous concentrerons sur les caractéristiques des grilles d'erreurs élaborées par des chercheurs en anglais et en français, du point de vue sémantique et syntaxique.

2.1 La notion d'« erreur »

Les travaux sur l'analyse des erreurs se réfèrent à la notion d'« erreur » et celle de « norme ». Mais, la notion d'« erreur » apparaît complexe. C'est en effet un concept relatif qui implique un écart par rapport à une norme. Il faut également différencier l'erreur de la « maladresse » et de la « faute », qui sont aussi des déviations par rapport à une norme, mais qui possèdent un niveau d'acceptabilité ou un statut différents dans le processus d'apprentissage que l'erreur.

2.1.1 L'« erreur » et la « norme »

L'erreur est un concept relatif qui implique un écart par rapport à une norme. La définition de la norme est complexe. Nous présenterons trois types de normes dans le cadre de notre mémoire : la norme prescriptive, la norme descriptive et la norme fonctionnelle.

La norme prescriptive privilégie une certaine variation langagière et condamne les autres. C'est un choix souvent soumis à l'autorité politique pour une fonction normative, qui vise à former un système uniforme de la langue en éliminant les variations liées à l'usage. Ainsi selon la norme prescriptive, une erreur est un écart par rapport à une forme normative et privilégiée. La norme descriptive, quant à elle, se rapporte aux formes courantes utilisées par la majorité d'une communauté. L'usage devient ici le critère essentiel. Selon la norme descriptive, une erreur est donc un écart par rapport à l'usage de la majorité. Enfin, selon norme fonctionnelle, le bon usage doit être adapté aux diverses situations de communication et aux situations de la vie en société. Le contexte de communication et le genre deviennent essentiels, et une erreur apparaît quand la variété ou le registre de langue qu'on utilise ne correspond pas au contexte d'utilisation ou au genre du discours.

Nous allons illustrer les différents types de norme par des exemples concrets. Anctil (2012) adopte la norme prescriptive dans son étude. La définition de l'erreur dans son étude s'articule autour du concept de « locuteur expert » : une erreur correspond à une « forme linguistique ou combinaison de formes linguistiques qui diffère de ce qu'un locuteur expert aurait selon toute probabilité produit dans un même contexte de production. » (Anctil, 2012, p. 8). Selon lui, cette définition n'exclut pas le concept de norme, parce qu'un locuteur expert respecte généralement les prescriptions normatives (Anctil, 2012). Ainsi, cette définition de l'erreur s'inscrit dans une conception prescriptive de la norme. La recherche de Yan (2017), consacrée aux erreurs des constructions verbales scientifiques transdisciplinaires chez les apprenants sinophones dans l'écrit scientifique, adopte plutôt la norme fonctionnelle. Yan (2017) se base en effet sur l'usage dans l'écrit scientifique. Elle utilise des corpus d'experts Scientext-SHS comme corpus de référence pour délimiter les erreurs, corpus composé d'articles scientifiques rédigés par des experts. Plusieurs chercheurs comme Tran les utilisent également pour élaborer des séquences didactiques afin d'aider à la rédaction académique en français des apprenants (Yan, 2017). Ces corpus servent de « source d'exemples de l'usage des locuteurs natifs qui sont experts dans certains domaines, et non pas dans un sens prescriptif » (Chambers, cité par Yan, p. 93). L'usage dans les écrits scientifiques des experts représente donc la norme fonctionnelle dans le travail de Yan (2017). Yan (2017) définit l'erreur comme un élément linguistique qui ne respecte pas les règles grammaticales de la langue française ou les usages appropriés au genre de l'écrit scientifique. Il s'agit donc à la fois d'une norme prescriptive et d'une norme fonctionnelle. Pour les aspects morphologiques et grammaticaux, Yan (2017) adopte la norme prescriptive, tandis que pour les aspects sémantiques, elle se base plutôt sur la norme fonctionnelle.

Dans le cadre de notre mémoire, nous adopterons la norme prescriptive pour cerner les erreurs. Notre choix s'appuie sur trois arguments : tout d'abord, le corpus que nous utilisons pour effectuer la recherche est composé de textes rédigés par les apprenants dans le contexte universitaire, qui exige un français standard. De plus, la grille d'erreurs que nous élaborerons s'inscrit dans la didactique du français langue étrangère. La variété du français enseignée à l'université est censée correspondre à l'usage privilégié dans les ouvrages normatifs. Par ailleurs, quand nous délimitons les erreurs, la norme prescriptive nous permet de recourir à des ouvrages de référence comme *le dictionnaire Robert*¹. Cela nous aide à éliminer la subjectivité de notre part.

¹ Bibliothèques Universitaires Grenoble-Alpes. (2009-2018). Le Petit Robert de la langue française. Repéré à l'adresse : <https://pr-bvdep-com.sid2nomade-2.grenet.fr/robert.asp>

En somme, pour les corpus de niveau B1 et B2, nous décidons de faire référence à la norme définie par les grammaires et dictionnaires de référence en tant que critère d'évaluation des erreurs.

2.1.2 La « maladresse » et la « faute »

Anctil emploie le terme « maladresse » à côté du terme « erreur » dans son mémoire en 2005. En fait, il distingue deux types de problèmes lexicaux :

D'une part, les erreurs lexicales qui peuvent être condamnées catégoriquement par la norme prescriptive; d'autre part, les maladroites lexicales qui découlent d'un usage douteux du lexique, mais qui ne pourraient pas être de façon absolue jugées comme des erreurs. (Anctil, 2005, p. 67)

Il mentionne aussi que de telles maladroites sont très présentes dans l'écriture des apprenants non natifs. Les maladroites sont difficiles à traiter, car elles ne représentent pas une opposition totale entre un emploi adapté et un emploi erroné, mais un usage douteux. Les maladroites ne peuvent pas être rejetées totalement du point de vue strictement linguistique, mais elles ne sont pas toujours acceptées par la majorité de la communauté de discours. D'après Anctil (2005), une erreur est un problème plus grave qu'une maladroite.

Par ailleurs, dans sa thèse sur les erreurs des apprenants sinophones en français langue étrangère, Ying (2013) distingue les deux notions « faute » et « erreur » qui sont considérées comme synonymes dans la langue courante. Les erreurs ne peuvent pas être découvertes ou corrigées par l'apprenant lui-même et elles reflètent un défaut dans la connaissance sur la langue ou une confusion sur les règles fondamentales de la langue de l'apprenant. Les fautes sont commises à cause de l'inattention de l'apprenant. Elles peuvent être évitées ou corrigées par l'apprenant lui-même s'il fait plus d'attention. Les fautes ne signifient pas une compétence insuffisante de l'apprenant sur la langue.

Dans notre étude, l'objectif est de trouver les vraies « erreurs » sur les verbes et non des expressions maladroites. Étant donné que notre étude n'implique pas un entretien auprès des apprenants, il est difficile de distinguer une « faute » d'une « erreur ».

2.2 L'intérêt d'étudier l'« erreur » et le système de l'interlangue

Depuis longtemps, l'erreur est considérée comme un écart à éliminer, mais il faut rappeler le fait qu'étudier les erreurs est intéressant en didactique. Ying (2013) envisage l'erreur dans une perspective cognitive. D'après elle, « l'erreur appartient à l'interlangue de l'apprenant et reflète son état imparfait par rapport au système de la langue cible. » (Ying, 2013, p.29). L'interlangue, terme forgé en 1972 par Selinker selon l'approche cognitive à

l'étude de l'acquisition des L2, est un système intermédiaire entre la langue source et la langue cible existant dans la « tête » de l'apprenant, qui reflète l'état imparfait de son apprentissage. C'est un système autonome et fonctionnel qui est constitué d'une partie de la langue source (la langue acquise), d'une partie de la langue cible (la langue à apprendre) et d'une partie ne relevant d'aucune des deux, soit la partie imaginée par l'apprenant lui-même. Ce système intermédiaire comporte non seulement les formes correctes, mais aussi les formes incorrectes.

Le système de l'interlangue est caractérisé par les propriétés suivantes : l'instabilité, la perméabilité, la systématisme et la variabilité, la simplification et la complexification.

L'instabilité de l'interlangue se manifeste par les formes instables produites par les apprenants à un moment donné. Par exemple, nous pourrions observer dans un texte à la fois *ma grand-mère a adoré la cuisine du Sud* et *elle ne **cuir** pas autant aujourd'hui*. L'apprenant conjugue les verbes dans certains cas, mais ne le fait pas dans d'autres cas. Une variabilité concernant le genre est aussi perceptible chez les apprenants. Par exemple, un nom est parfois employé comme masculin, parfois comme féminin, sans que l'apprenant s'en rende compte.

La perméabilité se rapporte aux traces de contact avec d'autres langues du système de l'interlangue, y compris les langues acquises et la langue cible. Deux types d'erreurs sont identifiés en fonction de leur origine : les erreurs interlinguales et les erreurs intralinguales. Les erreurs interlinguales s'appellent aussi les erreurs de transfert, car elles proviennent du transfert négatif d'une autre langue sur la langue cible. Le transfert langagier est défini comme « l'influence s'écoulant des similarités et différences entre la langue cible et les langues acquises précédemment. » (Odli, 1989, cité par Granger, 2015, p.333). Les erreurs intralinguales sont liées au fait que l'apprenant applique ses connaissances sur la langue cible d'une façon inappropriée. Il existe deux modèles typiques des erreurs intralinguales : la surgénéralisation et la simplification. La surgénéralisation désigne le fait que l'apprenant surgénéralise des connaissances acquises antérieurement en langue cible à ses apprentissages ultérieurs. Par exemple, l'erreur d'auxiliaire dans la phrase *Il a venu* pourrait être considérée comme une surgénéralisation de la règle pour construire le passé composé avec l'auxiliaire *avoir*. La simplification désigne un oubli des règles ou une application incomplète des règles par l'apprenant.

La systématisme de l'interlangue renvoie à son état relativement stable à un moment donné. La stabilité des erreurs à un certain stade d'apprentissage pourrait illustrer cette propriété : les apprenants d'un même stade d'apprentissage commettent souvent des erreurs

identiques. Par exemple, les apprenants débutants commettent souvent l'erreur de ne pas conjuguer les verbes, mais avec l'apprentissage, ce type d'erreurs diminuera.

La variabilité de l'interlangue fait référence à son évolution dans le temps. Les apprenants remettent en question des principes de fonctionnement du système de l'interlangue, formulent sans cesse de nouvelles hypothèses sur la langue cible, et les évaluent dans leurs pratiques langagières. Dans ce processus, par le biais des erreurs, leur système d'interlangue se développe et se rapproche peu à peu du système de la langue cible. Ainsi, l'erreur incarne un processus mental et cognitif de l'apprentissage d'une langue seconde. L'erreur n'est plus un écart négatif que nous devons chercher à éliminer, mais un phénomène naturel, incontournable et nécessaire, qui constitue des indicateurs significatifs du processus d'apprentissage. La systématique et la variabilité de l'interlangue s'avèrent un peu paradoxales, mais il faut noter que les deux caractéristiques partent de différents points de vue : l'interlangue est systématique dans la mesure où elle est stable à un moment donné, mais elle est aussi variable car elle évolue sans cesse dans le temps.

Les dernières caractéristiques sont la simplification ou la complexification par rapport à la langue cible. La simplification est un phénomène fréquent au début de l'apprentissage. Par exemple, les productions de l'apprenant s'avèrent plus pauvres que celles des natifs, notamment sur le plan syntaxique et morphologique: certaines formes sont absentes ou évitées par les apprenants, comme le subjonctif, le plus-que-parfait, le conditionnel, la subordination. L'absence de déterminants, l'absence de conjugaisons sont aussi fréquentes chez les apprenants. Cependant, avec le développement de la compétence langagière, une complexification progressive est constatée, même une sur-complexification. Par exemple, nous avons observé que les apprenants sinophones utilisent beaucoup de phrases longues et lourdes avec des propositions relatives. Étant donné que les propositions relatives, très courantes en français, n'existent pas en chinois, on pourrait penser que les apprenants sinophones essaient d'imiter le modèle syntaxique du français en composant des phrases complexes, même si les phrases risquent d'être trop lourdes.

Pour conclure, les erreurs sont significatives pour l'étude de l'apprentissage de la langue seconde, car elles relèvent du système d'interlangue de l'apprenant et constituent des indicateurs cognitifs et mentaux du processus d'apprentissage. Par le biais des erreurs, le système d'interlangue de l'apprenant se développe et se rapproche peu à peu du système de la langue cible. Par ailleurs, la théorie de l'interlangue nous aide à analyser l'origine des erreurs : nous pouvons ainsi distinguer les erreurs interlinguales qui proviennent du transfert négatif des langues acquises sur la langue cible, et les erreurs intralinguales qui découlent de la surgénéralisation ou la simplification des règles de la langue cible.

Les enseignants, les apprenants et les chercheurs en traitement automatique des langues (TAL) peuvent tous bénéficier du traitement des erreurs. Pour les enseignants, le repérage des erreurs permet d'identifier les difficultés spécifiques d'un certain groupe d'apprenants et concevoir des matériaux didactiques correspondant à leurs besoins. Pour les apprenants, la mise en évidence des erreurs favorise la prise en conscience des difficultés ainsi que l'apprentissage par correction réciproque avec les camarades d'une façon collaborative. Pour les chercheurs en TAL, l'analyse des erreurs est indispensable en vue d'élaborer des outils automatiques pour la détection et la correction des erreurs (Biskri & Jebali, 2010, p.154).

Après avoir montré l'intérêt d'étudier les erreurs, nous passons maintenant à l'analyse d'un outil qui nous permet d'observer les erreurs : la grille d'erreurs. C'est une partie essentielle de notre étude, parce que les erreurs ne deviennent des indicateurs significatifs du processus d'apprentissage que quand nous les classons par catégorie et en révélons les régularités.

2.3 Les grilles d'erreurs des chercheurs

Dans cette partie, nous présenterons des grilles d'erreurs en anglais et en français fondées sur l'analyse des corpus.

2.3.1 La typologie d'erreurs de Dagneaux pour l'anglais

Pour les typologies en anglais, nous allons présenter celle de l'étude de Dagneaux, Denness et Granger en 1998 sur les erreurs commises par les apprenants francophones de l'anglais langue seconde. Cette étude a pour objectif d'observer la progression de la L2 chez les apprenants de niveau intermédiaire à avancé. Le corpus comporte, d'une part, un sous-corpus constitué des textes rédigés par des apprenants francophones de niveau avancé dans la base de données *International Corpus of Learner English (ICLE)*, et d'autre part, d'un corpus d'apprenants francophones de niveau intermédiaire, au total 150,000 mots. Une classification descriptive des erreurs a été proposée. Il existe 7 catégories principales : formel, grammatical, lexico-grammatical, lexical, registre, redondance / absence / l'ordre erroné des mots, et style, accompagnées de deux niveaux de sous-catégories. Par exemple, sous les « erreurs grammaticales », il y a les « erreurs grammaticales pour les verbes », « erreurs grammaticales pour les noms » et « erreurs grammaticales pour les adjectifs » ; et la sous-catégorie « erreurs grammaticales pour les verbes » est subdivisée en « erreurs de temps », « erreurs d'auxiliaire », « erreurs de voix », etc. La grille de Dagneaux (1998) comporte tous les types

d'erreurs commises par les apprenants, tandis que notre grille qui porte sur les erreurs verbales sera moins étendue, mais plus détaillée.

2.3.2 Les typologies d'erreurs pour le français

Pour le français, nous allons présenter trois grilles d'erreurs où les erreurs sont aussi classées selon les dimensions linguistiques : celle d'Anctil (2010), celle de Ying (2013) et celle de Yan (2017) et nous allons les comparer au niveau sémantique et syntaxique.

La grille d'Anctil (2010) s'intéresse aux erreurs lexicales commises par des francophones québécois natifs. Elle est structurée en trois grandes catégories : la forme, le sens et la combinatoire restreinte au niveau grammatical, lexical et pragmatique. Cette grille se caractérise par une analyse résolument descriptive, une organisation en fonction des dimensions de l'unité lexicale et une analyse fine des classes d'erreurs.

Comme la grille de Dagneaux (1998), la typologie de Ying (2013) s'intéresse à toutes les erreurs commises par les apprenants sinophones, les erreurs lexicales, mais aussi syntagmatiques et syntaxiques. Elle est composée de trois types d'erreurs : le type lexical, le type grammatical, le type sociopragmatique. Ying (2013) précise que sa grille est synthétique et peu détaillée afin de faire ressortir les problèmes principaux, ce qui sert mieux son objectif qui est d'améliorer d'enseignement du FLE en Chine. Nous pouvons constater que les catégories sont moins fines que celles d'Anctil, car il existe moins de sous-catégories.

La grille de Yan (2017) est inspirée de celle d'Anctil (2010) et se concentre sur les erreurs verbales dans l'écrit scientifique des apprenants sinophones. Yan (2017) classe les erreurs en quatre catégories : au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la grammaire et de la syntaxe et au niveau du genre et du contexte. Cette grille ne concerne que les erreurs liées aux verbes, et elle est donc moins étendue que celles d'Anctil (2010) et de Ying (2013). Nous adopterons ses quatre catégories pour construire notre grille d'erreurs.

Nous allons comparer les trois grilles d'erreurs pour le français au niveau sémantique et syntaxique.

2.3.2.1 Les caractéristiques sémantiques des grilles d'erreurs

Dans la grille d'Anctil 2010 fondée sur les erreurs des natifs, les problèmes sémantiques comportent cinq sous-catégories : l'utilisation inappropriée d'un quasi-synonyme, l'utilisation inappropriée d'une unité lexicale de sens proche, l'impropriété (le sens fictif et l'anglicisme), le non-respect des contraintes sémantiques pesant sur les actants d'une unité

lexicale et le pléonasme. Un quasi-synonyme désigne les synonymes substituables dans certains contextes. Les unités lexicales de sens proche font référence à celles qui partagent des composantes du sens et qui appartiennent à un même champ sémantique, mais qui ne sont pas substituables en contexte. Le sens fictif concerne la mauvaise attribution d'une lexie à un signifiant existant. Le non-respect des contraintes sémantiques pesant sur les actants d'une unité lexicale fait référence à la violation des contraintes sémantiques sur les participants de la situation dénotée par l'unité lexicale décrite. Par exemple, la combinatoire sémantique du verbe *manger* exige que son complément soit un objet mangeable, si le complément est *Pierre*, une erreur de non-respect des contraintes sémantiques apparaît.

Le cas de l'anglicisme dans la classe du sens nous paraît intéressant dans la grille d'Anctil (2010). Nous pouvons penser que cette catégorie est liée au contexte québécois de la recherche d'Anctil (2010), où l'anglais et le français sont deux langues qui co-existent dans la vie courante. L'anglicisme sémantique concerne l'utilisation erronée des unités lexicales en français qui possèdent la même forme en anglais, mais pas le même sens. Par exemple, l'utilisation de *définitivement* pour exprimer le sens de 'certainement' sous l'influence de l'anglais relève d'une erreur d'anglicisme sémantique. Même si l'unité lexicale *définitivement* existe aussi en français, elle signifie 'faire quelque chose de façon définitive, irrévocablement'. Elle ne remplit pas la fonction d'expression d'une certitude comme en anglais. Dans notre cas, la plupart des Chinois commencent à étudier l'anglais depuis l'école primaire, qui est donc leur première langue étrangère. Il se peut que les apprenants sinophones mélangent les deux langues étrangères dans l'apprentissage du français, tant au niveau morphologique qu'au niveau sémantique. Nous reviendrons sur ce point en analysant notre corpus pour observer si ce type d'erreurs est adapté à nos sujets d'étude.

Parmi les erreurs de sens d'Anctil (2010), le « pléonasme » réfère à la cooccurrence de deux termes dont le sens est inclus dans celui de l'autre. Par exemple, dans la locution *descendre en bas*, le verbe *descendre* est défini comme 'aller de haut en bas', qui implique déjà le sens 'vers le bas' exprimé par *en bas* qui suit le verbe *descendre*. Nous pouvons intégrer ce type d'erreurs dans notre analyse parce que nous pensons qu'il aura ce type d'erreurs chez les apprenants étrangers, découlant de leur maîtrise insuffisante des composantes sémantiques d'une unité lexicale.

Concernant la dimension sémantique, nous constatons une différence de classement entre la grille d'erreurs de Yan (2017) et celle de Ying (2013) : en ce qui concerne les erreurs du temps grammatical, Yan (2017) a classé les erreurs de conjugaison et les erreurs de temps/aspect/mode au niveau morphologique, tandis que Ying (2013) distingue deux cas : les

erreurs de conjugaison et les erreurs de choix de temps verbal. Pour Ying (2013), les erreurs de conjugaison relèvent du sous-type « erreurs morphologiques flexionnelles », tandis que les erreurs de choix de temps verbal appartiennent au sous-type sémantique. La classification de Ying (2013) nous paraît plus pertinente, parce que l'origine de ces deux types d'erreurs est différente. Les erreurs de conjugaison sont attribuables à la maîtrise insuffisante des règles de conjugaison, tandis que les erreurs de choix de temps verbal viennent de la confusion entre les valeurs sémantiques exprimées par les différents temps verbaux. Nous pouvons intégrer cette distinction dans notre travail.

Nous pouvons nous inspirer des sous-catégories « erreurs de collocations » et « erreurs de cooccurrences » au niveau sémantique et des cooccurrences lexicales de Yan (2017). La sous-catégorie « erreurs de collocations » concerne les collocations inappropriées de types Verbe-Nom. La collocation est une construction binaire constituée par la base et le collocatif. La base est choisie en fonction de son sens, tandis que le collocatif est choisi en fonction de la base et la signification prévue de la collocation (Hausmann, 1989; Mel'čuk, 1998). Par exemple, dans la collocation *peur bleue* qui signifie 'une grande peur', *peur* est la base, et *bleue* est le collocatif, dans la collocation *faire une hypothèse*, *hypothèse* est la base, et *faire* est le collocatif. Pour les collocations du type Verbe-Nom, le nom est souvent la base et le verbe est le collocatif. Cependant, Yan (2017) n'a pas bien précisé le type d'erreurs de collocation, par exemple, car il pourrait s'agir d'erreurs de collocatif, ou d'erreurs de base. Étant donné que les collocations sont importantes pour la maîtrise des verbes, nous souhaitons préciser ce sous-catégorie dans notre analyse. La sous-catégorie « erreurs de cooccurrences » prend en compte les adverbes inappropriés qui sont en cooccurrence avec les verbes. Par exemple, dans la phrase *Nous proposerons grossièrement une distinction de quatre grands courants : textuel, anaphorique, inactuel et aspectuel*, l'adverbe *grossièrement* est inadapté pour combiner avec le verbe *proposer*. Yan propose donc la correction *Nous proposerons sommairement une distinction de quatre grands courants : textuel, anaphorique, inactuel et aspectuel*. Les verbes ne doivent pas être étudiés isolément. Leurs cooccurrences avec les noms, les adverbes ou les compléments circonstanciels méritent aussi notre attention.

Pour résumer, du point de vue sémantique, nous pouvons nous inspirer de la classification sémantique des erreurs dans la grille d'Ancil (2010). L'« anglicisme sémantique » et le « pléonasme » dans sa grille nous paraissent intéressants et susceptibles d'être observés dans nos données. Nous retenons de la grille de Ying (2013) le classement des erreurs de conjugaison au niveau morphologique lors que les erreurs de temps/aspect/mode au niveau sémantique. Nous nous inspirons aussi des erreurs de

cooccurrences dans la grille de Yan (2017), mais préciserons dans notre étude les types d'erreurs de cooccurrences.

2.3.2.2 Les caractéristiques syntaxiques des grilles d'erreurs

Dans son travail, Anctil (2010) prend en compte non seulement la forme et le sens d'une unité lexicale, mais aussi ses propriétés combinatoires. Selon l'article d'Anctil en 2012, « la combinatoire restreinte d'un mot regroupe l'ensemble des propriétés qui régissent son "comportement" lorsqu'il apparaît en contexte de phrase et limitent les situations de parole dans lesquelles ce mot peut être utilisé ; elles (les propriétés) peuvent être d'ordre syntaxique, morphologique, pragmatique et collocationnel. » (Anctil, 2012, p.9) La recherche d'Anctil s'inscrit dans la Théorie Lexicologie Explicative et Combinatoire (LEC), qui est la partie lexicale de la théorie Sens-Texte, née en Russie dans les années 60, surtout développée par Mel'čuk et ses collègues à Montréal et à Paris (Mel'čuk et al., 1995). Selon la LEC, le signe linguistique comporte trois parties, le signifiant (la forme), le signifié (le sens) et les propriétés combinatoires. Les propriétés combinatoires d'un signe sont des contraintes propres à un signe qui limitent sa capacité à se combiner avec d'autres signes linguistiques. Elles ont deux aspects : les propriétés grammaticales, y compris les propriétés syntaxiques qui concernent la transitivité du verbe et le régime qui fait référence aux propositions régies par les lexiques ; les propriétés lexicales, soit la cooccurrence lexicale restreinte², comme la collocation³ (voir la partie 2.3.2.1).

Anctil (2010) insiste sur les propriétés combinatoires d'une unité lexicale. Ainsi, les erreurs lexicales incluent toutes les erreurs qui découlent de la méconnaissance d'une unité lexicale, même si elles se manifestent sur le plan syntaxique. Par exemple, Anctil (2010) étudie aussi l'utilisation d'une préposition erronée régie par le verbe : dans la phrase *L'horloger se dirigea à la caverne à grande vitesse*, il existe une erreur de préposition régie, puisque le verbe *se diriger* contrôle la préposition *vers* plutôt que *à*. Nous retenons de la grille d'Anctil (2010) qu'il est important d'étudier le verbe dans un sens étendu en prenant en compte toutes les erreurs causées par la transgression des propriétés combinatoires d'une

2 Cooccurrence lexicale : la cooccurrence lexicale est le simple fait pour deux ou plusieurs lexies d'apparaître ensemble dans un même énoncé (Anctil, 2005, p. 22). La cooccurrence lexicale restreinte est cette propriété qu'ont les unités lexicales d'apparaître ensemble de façon contrainte dans un même énoncé (Anctil, 2005, p. 38)

3 Collocation : L'expression AB (ou BA), formée des lexies A et B, est une collocation si, pour produire cette expression, le locuteur sélectionne A librement d'après son sens 'A', alors qu'il sélectionne B pour exprimer un sens 'C' en fonction de A. Nous dirons que la lexie choisie librement est la base de la collocation, alors que les expressions contraintes seront appelées collocatifs (Polguère, cité par Anctil, 2005, p. 25)

unité verbale, comme les erreurs de cooccurrences avec les noms et les adverbes, les pronoms régies par les verbes.

Dans le travail de Yan (2017), nous pouvons regretter que les pronoms, en tant que composantes régies par les verbes, n'aient pas reçu une attention suffisante. Les erreurs de pronoms n'apparaissent que sous les erreurs d'ordre des mots : erreurs qui concernent l'ordre pronoms - adverbes - verbes. Par exemple, dans la phrase *Tout au long du roman, l'auteur a nous décrit neuf voyages*, il y a une erreur verbale d'ordre des mots liée avec le pronom *nous*. La phrase devrait être *l'auteur nous a décrit neuf voyages*. Pronominaliser le complément indirect des verbes par *en* et *y* constitue sans doute une difficulté pour les apprenants sinophones, car le système de pronominalisation en chinois est plus simple qu'en français. Puisque notre recherche porte sur le verbe, en ce qui concerne les erreurs se rapportant aux pronoms, nous nous concentrons sur les erreurs de pronominalisation qui sont liées directement aux verbes, c'est-à-dire les pronominalisations erronées découlant d'une confusion sur les propriétés verbales. Par exemple, dans la phrase *Le concert est important, nous aimons bien le participer*, il y a aussi une erreur de pronominalisation. Le verbe *participer* est un verbe transitif indirect, qui est censé régir une préposition *à* pour introduire un complément objet indirect, donc nous devons pronominaliser le complément d'objet indirect *au concert* du verbe *participer* par le pronom *y* au lieu de *le*. Cette erreur trouve probablement son origine dans une mauvaise maîtrise de la transitivité du verbe *participer*. Cela fait donc partie de notre analyse des erreurs verbales. Nous reviendrons sur ce point dans la section 3.2.3.4.

De plus, les erreurs concernant la voix du verbe apparaissent dans la typologie de Dagneaux (1998) mais pas dans la grille de Yan (2017). Nous pensons que la voix mérite d'être traitée comme les autres dimensions du verbe : le temps, le mode et l'aspect. Selon Siouffi (2017), la voix définit « la structure agentive du procès-verbal, autrement dit précise ce qui fait, et ce qui est fait, dans le processus » (Siouffi, 2017, p. 159). Intuitivement, elle exprime la relation existant entre le sujet et le complément d'objet direct. Elle comporte la voix active, la voix passive et la voix moyenne. Selon la définition sémantique traditionnelle, à l'actif, le sujet grammatical assume l'action décrite par le procès-verbal, tandis que le complément d'objet direct la subit. Au passif, c'est le patient de l'action qui devient le sujet grammatical du verbe. Par ailleurs, l'expression du passif en français varie beaucoup selon les verbes et les situations en utilisant différentes tournures, comme la voix moyenne, à savoir les pronominaux de sens passif : *Les jupes se portent court*. La voix est une notion grammaticale complexe, surtout en ce qui concerne la voix moyenne, qui n'a pas de structure équivalente en

chinois. La voix pose donc potentiellement des problèmes aux apprenants sinophones. Il nous semble pertinent d'intégrer les erreurs de voix dans notre grille.

Pour résumer, du point de vue syntaxique, nous retenons des grilles d'Anctil (2010) qu'il est important d'étudier le verbe dans un sens large en prenant en compte les cooccurrences des verbes avec les noms et les adverbes, et les erreurs de pronoms liées aux propriétés du verbe. La grille d'erreurs de Yan (2017) nous fournit une bonne base pour analyser les erreurs verbales chez les apprenants sinophones. Nous adopterons ses quatre grandes catégories, en aménageant les erreurs de temps/aspect/mode au niveau sémantique et intégrant les erreurs de voix et les erreurs de pronoms dans notre grille.

2.3.2.3 La typologie d'erreurs pour le corpus Frida

Dans les trois grilles d'erreurs présentées ci-dessus, les erreurs sont classées selon les dimensions linguistiques. Nous exposerons ensuite une grille d'erreurs un peu différente, celle proposée pour l'annotation du corpus Frida (Granger, 2007). La grille est composée de deux parties : le domaine d'erreurs et la catégorie grammaticale. Il existe 8 domaines : forme graphique, morphologie, grammaire, lexique, syntaxe, registre, style et ponctuation, avec 36 catégories d'erreurs. Par exemple dans le domaine « syntaxe », il y a 4 catégories d'erreurs : incomplet, manque, ordre, redondance, dans le domaine « grammaire », il y a 9 catégories d'erreurs : classe, auxiliaire, genre, mode, nombre, personne, temps, voix, euphonie. En même temps, il existe 11 catégories grammaticales : adjectif, adverbe, article, conjonction, déterminant, nom, préposition, pronom, verbe, séquence, ponctuation, accompagnées de 55 sous-catégories grammaticales, par exemple, sous la catégorie « verbe », il y a : simple conjugué, participe simple, gérondif simple, infinitif simple, complexe conjugué, participe complexe, gérondif complexe, infinitif complexe. Cette grille se différencie des typologies précédentes par l'ajout de catégories grammaticales en parallèle avec les domaines grammaticaux, ce qui permet de mettre en évidence la catégorie grammaticale la plus fréquemment erronée. Nous nous inspirons des sous-catégories grammaticales de la catégorie grammaticale du verbe dans notre étude.

Dans le chapitre II, nous avons vu que les erreurs, relevant du système d'interlangue de l'apprenant, sont significatives pour observer le processus d'apprentissage de la L2. Mais l'identification des erreurs n'est pas toujours évidente. Dans notre étude, nous adopterons une norme prescriptive en recourant aux dictionnaires et grammaires de référence pour cerner les erreurs. Les erreurs ne deviennent des indicateurs du processus d'apprentissage que quand on les classe selon les catégories et en dévoile la régularité. Pour décrire les erreurs verbales, la classification des erreurs selon les dimensions linguistiques nous paraît pertinente : au niveau

morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la grammaire et de la syntaxe et au niveau du genre et du contexte.

Nous procéderons ensuite à la présentation de la méthodologie de notre étude, avec la présentation du corpus, de la grille d'erreurs élaborée et de l'annotation semi-automatique des verbes et des erreurs verbales dans le corpus.

Chapitre III. Méthodologie : constitution du corpus et principes d'annotation

Rappelons que notre étude consiste à analyser les erreurs verbales dans un corpus d'apprenants sinophones. Pour atteindre notre objectif, il faut d'abord accueillir les textes rédigés par les apprenants sinophones pour construire un corpus d'apprenants, et puis repérer les erreurs verbales dans le corpus et les classer selon les dimensions linguistiques pour élaborer une grille d'erreurs. Nous effectuerons ensuite un travail d'annotation systématique des verbes et des erreurs verbales dans le corpus avec le codage de la grille d'erreurs. Dans un premier temps, nous présenterons la méthodologie du traitement des données. Dans un second temps, nous illustrerons les catégories d'erreurs avec des exemples tirés du corpus. Pour finir, nous nous inspirerons de quelques propositions d'annotation des erreurs pour l'anglais et le français, afin de concevoir notre système d'annotation et de construire un corpus d'apprenants indiquant les verbes et les erreurs verbales. Nous exposerons aussi notre méthode de calcul automatique des erreurs.

3.1 La méthodologie de traitement des données

Pour élaborer un corpus d'apprenants, nous avons recueilli des rédactions auprès des apprenants sinophones effectuant leurs études à l'Université Grenoble-Alpes, par le biais d'enseignants⁴. Les apprenants sont de niveau B1 et B2 selon le CECRL⁵. Certains d'entre eux étudient au CUEF (Centre universitaire d'études françaises) ⁶, et les autres en troisième année de licence de sciences du langage. Le corpus est constitué de 95 textes, dont 86 sont rédigés par les apprenants de niveau B2 et 9 textes par les apprenants de niveau B1. Le corpus comporte 50,014 mots au total. Plusieurs types de textes figurent dans le corpus : des textes

⁴ Nous tenons à remercier les enseignants qui nous ont aidés dans la collection des données, en particulier Rui Yan et Anne-Cécile Perre.

⁵ CECRL : le cadre européen commun de référence pour les langues. Il s'agit d'une norme internationale qui définit des niveaux de maîtrise d'une langue étrangère et qui permet de décrire la compétence linguistique en langue étrangère des apprenants.

⁶ Le CUEF de l'université Grenoble-Alpes a su acquérir puis maintenir sa position d'établissement d'excellence pour l'enseignement du Français Langue Étrangère au fil des décennies. Il accueille chaque année environ 3 000 étudiants originaires d'une centaine de pays différents, y compris un grand nombre d'apprenants sinophones. Le CUEF est aussi un centre d'expérimentation de la recherche en didactique du FLE au sein de l'Université Grenoble-Alpes. Site du CUEF: <https://cuef.univ-grenoble-alpes.fr/menu-principal/presentation/>

argumentatifs, des textes descriptifs, des résumés, des lettres et des commentaires. Certains textes sont rédigés à la main pendant des examens sur table, et les autres sont tapés à l'ordinateur en tant que devoirs à la maison. Les textes sur table sont écrits en temps limité sans recours aux dictionnaires ou à d'autres ouvrages de référence. Étant donné que les devoirs à la maison n'excluent pas l'utilisation de correcteurs automatiques, nous ferons l'hypothèse que les erreurs morphologiques seront moins présentes dans les textes écrits à la maison.

Nous récapitulons notre corpus dans le tableau ci-dessous :

Niveau du scripteur	Types de corpus	Types de textes	Taille	Contexte d'écriture
B2	L3 Linguistique	Résumés des récits d'expériences personnelles, présentations de la langue des signes, présentations d'une approche pour identifier l'auteur d'un texte.	20 textes (27430 mots)	À la maison
	DUEF B2	Résumés d'un texte	23 textes (8108 mots)	Sur table
	DUEF B2	Lettres au maire	20 textes (7593 mots)	Sur table
	DUEF B2	Commentaires des chiffres	23 textes (5160 mots)	Sur table
B1	DUEF B1	Descriptions d'une photo	9 textes (1723 mots)	Sur table

Tableau 3.1: Présentation du corpus

Pendant la collecte des données, les apprenants qui permettaient de diffuser leurs textes ont signé une fiche sociobiographique, où figurent des informations sur les scripteurs et sur les conditions de rédaction, comme l'âge, le niveau d'études, le nombre d'années d'étude du français, les langues acquises, les moyens d'aide à la rédaction. La fiche sociobiographique a été élaborée par Tatiana Aleksandrova⁷, chercheur à l'université Grenoble-Alpes. Pour les textes écrits à la main, nous avons fait la transcription en respectant la mise en page, le retour à la ligne et les ratures présentes dans les textes. Ensuite, nous avons repéré les erreurs

⁷ Nous tenons à remercier sincèrement Madame Tatiana Aleksandrova.

verbales dans les textes et les avons classées selon les quatre grandes catégories de la grille d'erreurs de Yan (2017) que nous avons présentée dans la partie 2.3.2 : au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la grammaire et de la syntaxe, au niveau du genre et du contexte. Selon les erreurs relevées dans le corpus, nous y avons ajouté des sous-types d'erreurs plus détaillées, ce qui a débouché sur une grille d'erreurs verbales précise. Suite à l'élaboration de la grille d'erreurs, nous avons transféré les textes au format XML et y avons intégré les métadonnées du scripteur et du texte afin de construire un corpus d'apprenants. La dernière étape a consisté à annoter semi-automatiquement les verbes et les erreurs verbales dans le corpus en indiquant le lemme du verbe, le type d'erreurs et le sous-type d'erreurs. Plusieurs annotateurs se sont engagés dans ce travail pour garantir la pertinence des annotations. À partir des textes numérisés et annotés, nous avons pu calculer automatiquement le nombre des verbes et des verbes employés de façon erronée, et observer la distribution des erreurs verbales. Nous avons également effectué une analyse qualitative des erreurs pour mettre en avant les circonstances où les erreurs apparaissent fréquemment et identifier leurs sources possibles.

À la différence de la grille d'erreurs d'Anctil en 2010, qui est élaborée à partir d'une approche théorique, nous avons procédé d'une façon empirique. La typologie d'erreurs élaborée en 2010 par Anctil avait pour objectif de couvrir l'ensemble des erreurs susceptibles d'être commises par un apprenant dans une production écrite, c'est-à-dire que les classes d'erreurs sont déterminées à l'avance en prévoyant tous les types de problèmes lexicaux possibles. La prévision précise et exhaustive a été assurée par la théorie de la Lexicologie Explicative et Combinatoire qui permettait d'envisager toutes les dimensions d'une unité lexicale afin de la maîtriser. C'est une typologie complète qui est réutilisable dans tous les genres de corpus, par tous les niveaux d'apprenants. De notre côté, en nous fondant sur le corpus d'apprenants sinophones de niveaux B1 et B2, nous avons adopté une démarche plus empirique en choisissant de construire une typologie à partir des erreurs verbales observées dans notre corpus, afin de déboucher sur une typologie plus adaptée aux apprenants sinophones de niveaux B1 et B2, qui pourrait apparaître extrêmement utile pour la didactique du français auprès ce public.

3.2 La constitution d'une grille d'erreurs

Pour construire notre typologie, nous réutiliserons les quatre grandes catégories de la grille de Yan (2017) présentée dans la section 2.3.2 : au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la grammaire et de la syntaxe et au niveau du genre et du contexte. Nous allons expliquer chaque catégorie en donnant des exemples concrets tirés de notre corpus.

3.2.1 Au niveau morphologique (MORPHO)

3.2.1.1 Erreurs de conjugaison et d'accord (CA)

Les erreurs de conjugaison renvoient aux verbes mal conjugués (1), et les erreurs d'accord se rapportent à l'accord du verbe avec le sujet (2), l'accord de l'auxiliaire ou du participe passé ayant trait aux temps composés du passé ou à la voix passive (3), l'accord du pronom dans les verbes pronominaux avec le sujet (4).

(1) *Le narrateur **choit** (correction proposée : choisit) de faire une explication supplémentaire qui, néanmoins, ne se consacre pas au déroulement.*

(2) *Les artistes de la région **obtient** (correction proposée : obtiennent) la chance de se présenter.*

(3) *Après la guerre, la jupe de Ubu Mère a été **déchiré** (correction proposée : déchirée)*

(4) *Nous **s'intéressons** (correction proposée : Nous nous intéressons) à la valeur culturelle du concert.*

(5) *Elle nous **regardes** (correction proposée : regarde) et demande "Qui es-tu ?".*

Nous constatons que les erreurs de conjugaison se manifestent au niveau morphologique (1), tandis que les erreurs d'accord se rapportent parfois à la structure syntaxique (4). Mais nous remarquons aussi que les erreurs de conjugaison et d'accord sont parfois difficiles à distinguer comme dans l'exemple 5. Nous pouvons dire qu'il s'agit d'une erreur de conjugaison, mais aussi d'une erreur d'accord du verbe avec le sujet, parce que le sujet est à la troisième personne, mais le verbe est à la deuxième personne. Pour éviter ce type de confusion et étant donné que les erreurs de verbe mal conjugué ne sont pas très fréquentes dans notre corpus, nous choisissons donc de ne pas séparer ces deux types d'erreurs et de les classer ensemble comme « erreurs de conjugaison et d'accord ».

3.2.1.2 Erreurs d'orthographe d'usage (ORT)

Les erreurs d'orthographe relevées dans notre corpus comportent deux cas : l'un concerne les erreurs de graphie qui ne nuisent pas beaucoup à la prononciation des unités verbales (6), alors que l'autre concerne les erreurs d'élimination (7). Par exemple :

(6) *Il y a beaucoup d'informations que nous ne pouvons pas utiliser directement, mais cela peut nous aider à mieux réfléchir (correction proposée : réfléchir) à notre sujet.*

(7) *C'est très utile de obtenir (correction proposée : d'obtenir) une bonne réputation de notre ville.*

3.2.1.3 Erreurs de verbes fictifs (VFIC)

Il s'agit de la création de formes verbales inexistantes en français. Par exemple :

(8) *Ce concert est une bonne chance de faire l'échange et se progressiver (correction proposée : progresser).*

(9) *Mais je sais que la ville est en train de facer (correction proposée : est confrontée) aux problèmes financiers, je pense qu'on peut gagner de l'argent par le concert quand même.*

3.2.2. Au niveau sémantique et des cooccurrences lexicales (SEMAN)

3.2.2.1 Erreurs de sens (SEN)

Rappelons la théorie Lexicologie Explicative et Combinatoire (Mel'čuk et al., 1995) que nous avons mentionnée dans le chapitre II, selon laquelle le signe linguistique est constitué de trois parties : le signifiant (la forme), le signifié (le sens) et les propriétés combinatoires (voir 2.3.2.1). Le signifié comporte trois dimensions : le genre prochain, qui se réfère au sens principal d'un signe ; les différences spécifiques, qui concernent les connotations différentes des lexies ; le typage des actants, qui renvoie à la combinatoire sémantique consistant à mettre des contraintes sémantiques sur les participants de la situation dénotée par l'unité lexicale décrite (voir 2.3.2.1). Par exemple, la combinatoire sémantique du verbe *manger* exige que son sujet soit un sujet animé.

En nous basant sur cette théorie et en nous inspirant de la grille d'Anctil (voir 2.3.2.1), nous proposons de distinguer deux sous-types d'erreurs de sens : les erreurs de synonymes, et les erreurs de sens inadapté. Les synonymes sont les lexies qui partagent les mêmes composantes de sens, et qui sont substituables dans certains contextes. L'utilisation inappropriée des synonymes concerne souvent le non-respect des différences spécifiques ou du typage des actants des lexies. Le sens inadapté se réfère à la mauvaise attribution d'une

acceptation à un signifiant existant. Dans ce cas, le sens principal du verbe employé ne convient pas au contexte et risque de rendre la phrase peu compréhensible. Nous allons illustrer ces deux types d'erreurs de sens avec des exemples concrets.

3.2.2.1.1 Erreurs de synonymes (SNM)

Nous allons illustrer l'utilisation inappropriée de synonymes par trois exemples :

(10) *Ce graphique **démontre** (correction proposée : montre) le nombre des temps utilisés par le texte mystérieux.*

(11) *Il y a beaucoup de techniques pour enrichir la syntaxe narrative mais nous **confinons** (correction proposée : limitons) nos études aux éléments évaluatifs qui concernent directement à la structure interne.*

(12) *Un lien faible ne nous conduit pas à maintenir une relation dente, mais il peut **amener** (correction proposée : apporter) toutes sortes d'informations à valeur importante.*

Dans l'exemple 10, d'après *le Petit Robert*⁸, le sens principal du verbe *démontrer* : 'faire voir quelque chose à quelqu'un' convient au contexte, mais par rapport au verbe *montrer*, *démontrer* a une connotation de 'faire connaître par raisonnement, prouver quelque chose', mais son sujet dans l'exemple, à savoir *le graphique*, ne peut pas raisonner. Le verbe *montrer* est donc mieux adapté.

Dans l'exemple 11, le verbe *confiner* possède le sens de 'ramener quelque chose à des limites plus étroites, tenir quelque chose dans d'étroites limites' exigé par le contexte, mais il s'emploie plutôt dans un sens concret. Par exemple, dans *Le père confine son enfant turbulent dans sa chambre*, *confiner* signifie 'enfermer quelqu'un ou quelque chose dans un lieu' (RL⁹). Cependant, dans l'exemple 11, le sens exprimé est que 'nous abordons seulement les éléments évaluatifs dans cette étude', il s'agit d'un usage plutôt abstrait. Nous proposons la correction *nous limitons nos études aux éléments évaluatifs*. Les verbes *confiner* et *limiter* partagent le sens de 'restreindre', mais avec des différences spécifiques : *confiner* ne peut être utilisé que dans un sens concret, mais *limiter* peut s'employer dans un sens concret ou abstrait.

Dans l'exemple 12, le verbe *amener* signifie 'faire venir quelqu'un avec soi, entraîner'. Il faut noter qu'il ne peut se combiner qu'avec des personnes, donc pour exprimer le sens de

⁸ *Le Petit Robert*, Paris, dictionnaires, (2009). Paris : Le Robert. Dans la suite du textes, les définitions sont reprises du *Petit Robert* (RP) si il n'y a pas d'indications du *Petit Larousse* (PL).

⁹ PL : *Le Petit Larousse illustré et en couleurs*, (2007). Paris : éditions Larousse.

‘porter avec soi quelque chose’ exigé par le contexte, il faudrait utiliser son synonyme *apporter*. C’est une erreur attribuable à la confusion sur le typage des actants du verbe *amener*.

3.2.2.1.2 Erreurs de sens inadapté (SIN)

Les erreurs de sens inadapté concernent un sens fictif. Par exemple :

(13) *Nous allons nous pencher sur la surdit  en mati re identitaire et culturelle, et **mettre en  uvre** (correction propos e : mettre en  vidence) dans quel contexte les langues des signes  voluent et comment l’identit  sourde se construit.*

(14) *Et d’apr s les colonnes dans le graphique, les phrases utilis es dans le texte myst re sont plut t courtes, ce qui nous **constate** (correction propos e : montre) que la fa on d’ criture du texte myst re est plus proche de celle de Zola.*

Dans l’exemple 13, la construction verbale *mettre en  uvre* signifie ‘faire usage de quelque chose pour obtenir un r sultat, appliquer quelque chose’. Cela ne correspond pas au sens de ‘montrer, expliquer’ exig  par le contexte. La construction *mettre en  vidence* est mieux adapt e.

Dans l’exemple 14, le verbe *constater* signifie ‘remarquer objectivement quelque chose, l’enregistrer comme vrai, r el’, le sujet doit  tre anim . Dans l’exemple, c’est *nous* qui *constatons*, le sujet *graphie* ne peut que nous *montrer* le fait.

En somme, les erreurs de sens regroupent toutes les erreurs perturbant la saisie du sens cible, soit   cause de l’utilisation inappropri e des synonymes qui ne respecte pas les propri t s sp cifiques ou le typage des actants des synonymes, soit   cause de la mauvaise attribution d’une acception   un signifiant existant qui ne correspond pas   son sens principal.

3.2.2.2 Erreurs de temps/d’aspect/de mode (TAM)

On divise en g n ral le temps en trois p riodes : le pass , le pr sent et le futur. L’aspect du verbe indique le fait que l’action traduite par le verbe est accomplie ou inaccomplie. Par exemple, la diff rence entre le pass  compos  et l’imparfait rel ve de l’aspect. Le pass  compos  exprime une action ponctuelle qui s’est d roul e et achev e dans le pass . L’imparfait pr sente une action du pass  dans son d roulement, soit une action en cours d’accomplissement. Le temps, l’aspect et le mode sont  troitement li s en fran ais et sont interpr t s par la conjugaison du verbe selon les TAM. Nous avons class  les erreurs concernant la conjugaison et l’accord du verbe au niveau morphologique (Voir 2.3.2.1), mais nous avons d cid  de mettre les erreurs de TAM au niveau s mantique en adoptant la

proposition de Ying (voir 2.3.2.1), parce que ce type d'erreurs trahit une confusion sur la signification véhiculée par les différents TAM du verbe. Par exemple :

(15) *Comme vous le saviez (correction proposée : savez), depuis l'existence de cette fête, un nouveau moyen de se reposer s'ajoute dans la vie des citoyens.*

(16) *En fin de compte, je suis sûre que l'événement soit (correction proposée : sera) meilleur qu'avant.*

(17) *Tout le monde, y compris les personnes qui sont chanteurs, étudiants, professeurs étudiés (correction proposée : étudiant) la musiques, les amateurs, vont au concert.*

Dans l'exemple 15, le verbe *savoir* est à l'imparfait, un temps du passé, mais selon le contexte de cette phrase, il vaut mieux utiliser le présent pour situer l'action de *savoir* dans le présent selon la réalité des faits.

Dans l'exemple 16, le subjonctif exprime un doute ou une hypothèse, et il ne correspond pas à la certitude exprimée par *je suis sûre*. Il faut donc utiliser l'indicatif.

Dans l'exemple 17, le participe passé *étudiés* est utilisé comme un complément du nom *professeurs*, mais *professeurs* est le sujet du verbe *étudier*, donc il faut utiliser le participe présent *classant* pour exprimer leur relation active.

En somme, les erreurs de temps/aspects/mode regroupent toutes les erreurs concernant l'utilisation inappropriée des différents TAM à cause de la confusion sur la valeur sémantique qu'ils véhiculent.

3.2.2.3 Erreurs de cooccurrences (COC)

La cooccurrence lexicale est le simple fait pour deux ou plusieurs lexies d'apparaître ensemble dans un même énoncé (Anctil, 2005, p. 22). La cooccurrence lexicale restreinte est cette propriété qu'ont les unités lexicales d'apparaître ensemble de façon contrainte dans un même énoncé (Anctil, 2005, p. 38). En nous inspirant de la grille de Yan (2017), nous prendrons en compte les erreurs verbales en cooccurrence avec les noms et avec les adverbes ou les compléments circonstanciels dans notre étude (voir 2.3.2.2).

3.2.2.3.1 Erreurs de cooccurrence avec les noms (CONO)

La détermination d'une cooccurrence lexicale restreinte est étroitement liée avec la notion de « sens restreint ». Selon Nesselhauf (2003), dans une combinaison lexicale du type Verbe+Nom, le sens d'un verbe ou d'un nom est restreint si un de ces deux critères suivants

s'applique : le premier critère est que le sens du verbe ou du nom est si spécifique qu'il se combine avec un petit nombre de noms ou de verbes; le deuxième critère est que le verbe ou le nom ne peut pas être utilisé dans ce même sens avec tous les noms ou les verbes syntaxiquement et sémantiquement possibles. Par exemple, le verbe *want* est de sens non restreint parce qu'il peut se combiner avec de nombreux noms, comme *toys, a child, a drink, a car, truth, etc*, mais le verbe *dial* est restreint car il ne peut se combiner qu'avec un seul nom ou très peu de noms, comme *number*; le verbe *perform* est aussi restreint selon le deuxième critère : même s'il peut se combiner avec plusieurs noms, certains noms semblent combinables d'un point de vue sémantique ne peuvent pas se combiner avec *perform*, par exemple on peut dire *perform a task*, mais on ne dit pas *perform a survey*. Pour une cooccurrence lexicale du type Verbe+Nom, si le verbe et le nom sont non restreints, c'est une combinaison libre : ils sont combinés selon leur sens ; s'ils sont tous restreints, c'est une expression figée ; si l'un de ces deux éléments est restreint et l'autre non restreint, il s'agit d'une collocation. La collocation est une construction binaire formée des lexies A et B où le locuteur sélectionne A librement d'après son sens 'A', alors qu'il sélectionne B pour exprimer un sens 'C' en fonction de A. La lexie choisie librement est la base de la collocation, alors que les expressions contraintes seront appelées collocatifs (Polguère, cité par Anctil, 2005, p. 25). Comme nous nous intéressons aux verbes, la majorité des collocations relevées dans notre corpus sont du type Verbe+Nom où le nom est la base et le verbe est le collocatif.

Nous distinguons deux sous-types d'erreurs de cooccurrence du type Verbe+Nom : les erreurs de verbes dans la cooccurrence du type *V+N* et les erreurs de noms dans la cooccurrence du type *V+N*.

Les erreurs de verbes dans la cooccurrence du type *V+N* renvoient au choix inapproprié du collocatif dans les collocations ou du verbe dans les expressions figées. Par exemple :

(18) *La deuxième manière est les comparateurs qui **désigne** (correction proposée : mesure) l'écart des événements narratifs pour envisager des possibilités non réalisées.*

(19) *Les adultes one déjà appris à bien s'exprimer et à **donner** (correction proposée : raconter) des vécus sous une structure plus complexe.*

(20) *Les amateurs de la musique, notamment ceux qui ne sont pas très habitués à assister à un concert, **sont** (correction proposée : ont) accès à divers genres de musiques grâce à la gratuité de cet événement.*

Pour les exemples 18 et 19, nous pouvons saisir facilement le sens cible du scripteur, mais ce n'est pas la bonne façon de l'exprimer car les verbes utilisés ne peuvent pas se combiner avec leurs compléments. Pour l'exemple 20, l'expression figée doit être *avoir accès à* au lieu d'*être accès à*.

Les erreurs de noms dans la cooccurrence du type *V+N* renvoient au choix inapproprié de la base dans les collocations ou du syntagme nominal dans les expressions figées, comme dans l'exemple 21, où on observe une erreur de la base dans la collocation *faire des compromis*.

(21) *Ce n'est pas un moyen pertinent de faire toujours des **compris** (correction proposée : *compromis*) avec les concurrents tels que Facebook.*

3.2.2.3.2 Erreurs de cooccurrence avec les adverbes ou les compléments circonstanciels (COA)

Nous avons déjà présenté la cooccurrence du type Verbe-Nom, il faut mentionner que les verbes peuvent aussi apparaître en cooccurrence avec les adverbes, ou les compléments circonstanciels. Par exemple :

(22) *Ce type peut se constituer **d'ordinaire** (correction proposée : *facilement*) mais manifeste profondément ce que le locuteur a évalué.*

(23) *Ce changement de la langue du signe conduit **exactement** (correction proposée : *inévitablement*) à l'incompréhension des interprétations.*

(24) *Ces parties se lient **avec des façons** (correction proposée : *d'une façon simple*) les plus simples.*

Dans l'exemple 22, le complément *d'ordinaire* du verbe *se constituer* est peu compréhensible en contexte. L'adverbe *d'ordinaire* signifie 'de façon habituelle', ce qui ne convient pas au contexte. Nous supposons que le sens cible du scripteur est *Ce type peut se constituer facilement mais manifeste profondément ce que le locuteur a évalué*. Dans l'exemple 23, *exactement* signifie 'avec une grande précision, une grande fidélité', mais pas le sens de 'forcément, inévitablement' exigé par le contexte. L'exemple 24 illustre une erreur du complément circonstanciel du verbe *se lier*.

3.2.3 Au niveau de la grammaire et de la syntaxe (GRAMSYN)

Six types d'erreurs figurent dans cette classe : les erreurs de voix des verbes, les erreurs de pronominaux, les erreurs d'ordre des mots, les erreurs de prépositions, les erreurs de pronoms et les structures syntaxiques erronées.

3.2.3.1 Erreurs de voix (VV)

En français, la voix comporte trois dimensions : la voix active, la voix passive et la voix moyenne (Riegel, Pellat & Rioul, 1994). La voix moyenne renvoie aux pronominaux de sens passif (voir la section 2.3.2.2). Par exemple : *Les livres se vendent bien*, le sujet *les livres* est le patient et le sujet du verbe *vendre*.

Les erreurs de voix concernent les erreurs de constructions passives à la place d'actives ou inversement, et les erreurs de pronominaux de sens passif. Par exemple :

(25) *Nous pouvons comparer les manières d'étude qui restent là ou qui ont été évoluées* (correction proposée : *qui ont évolué*) *avec les outils utilisés actuellement dans les musées de Grenoble.*

(26) *Sous le couvert de l'interrogation, les actes peuvent interpréter* (correction proposée : *peuvent être interprétés*) *comme demande d'information, demande d'action, défi, passage à l'acte.*

(27) *Ces événements présentent* (correction proposée : *se présentent*) *également au présent.*

Dans l'exemple 25, le verbe *évoluer* est un verbe intransitif, son usage est *quelque chose évoluée*, donc la forme passive de la phrase n'est pas correcte. Dans l'exemple 26, le sujet *les actes* est le complément d'objet direct du verbe *interpréter*, donc il faut utiliser la voix passive. Dans l'exemple 27, le verbe *présenter* est un verbe transitif direct, et le sujet *ces événements* est le patient qui subit l'action de *présenter*, donc il faut utiliser la forme passive. Puisque l'agent du verbe *présenter* n'est pas explicité, nous pouvons utiliser la voix moyenne *se présenter*.

3.2.3.2 Erreurs de pronominaux (EPN)

À côté des pronominaux de sens passif présentés dans la section 3.2.3.1, il existe d'autres types de pronominaux. Ce sont les pronominaux de sens réfléchi, de sens réciproque et de sens absolu. Les erreurs de pronominaux concernent l'utilisation erronée de ces trois types de pronominaux. Par exemple :

(28) Par exemple, on peut se rencontrer (correction proposée : *rencontrer*) les gens qui viennent des régions différentes et renforcer des amitiés.

(29) Elle vêt (correction proposée : *se vêt*) d'un t-shirt à manche court rose moulant.

(30) Il faut rendre compte (correction proposée : *se rendre compte*) que tous les verdicts ne sont pas sans appel car la cour d'appel et la cour de cassation sont là pour contester des décisions litigieuses et inopportunes.

Dans l'exemple 28, le verbe *rencontrer* est un verbe transitif direct, dont le complément d'objet direct est le nom *les gens*. La phrase *on peut rencontrer les gens* est déjà complète, donc le pronom *se* ne fait pas sens ici.

Dans l'exemple 29, le verbe *vêtir* est un verbe transitif qui signifie 'habiller quelqu'un de vêtements'. L'usage du verbe *vêtir* doit être *vêtir quelqu'un de quelque chose*, donc une erreur se produit à cause de l'absence de complément d'objet direct du verbe *vêtir* dans la phrase. Selon le contexte de la phrase, le sens doit être *Elle vêt elle-même d'un tee-shirt*, le sujet exerce une action physiquement sur lui-même, qui demande un verbe pronominal de sens réfléchi. Dans ce cas, nous proposons la correction *Elle se vêt d'un tee-shirt*, où le *se* est un pronom réfléchi désignant l'agent *elle*, qui est en même temps l'objet de l'action *vêtir*.

Dans l'exemple 30, la construction verbale *rendre compte* possède trois acceptions, la première est 'faire un rapport écrit ou oral', la deuxième est 'présenter quelque chose au public', la troisième est 's'exprimer et assumer les conséquences de quelque chose', mais aucune acception ne convient au contexte. En revanche, la construction pronominale *se rendre compte* qui signifie 'prendre pleinement conscience de quelque chose et la comprendre' semble adapté au contexte de la phrase. La construction *se rendre compte* est un pronominal de sens absolu car il n'existe que la forme pronominale dans le sens de 'prendre conscience'.

3.2.3.3 Erreurs d'ordre des mots (ORD)

Comme notre étude porte sur les verbes, nous nous intéressons particulièrement à l'ordre de cooccurrences des verbes, comme les adverbes, les noms et les pronoms. Ce type d'erreurs n'a pas apparu dans la grille d'erreurs d'Anctil (2010), ce qui pourrait s'expliquer par le fait que les natifs commettent peu ce type d'erreurs. Mais pour les apprenants non natifs, ce type d'erreurs est quand même fréquent. Par exemple :

(31) L'année dernière, plus de 1000 touristes sont y allés (correction proposée : *y sont allés*) et des artistes les plus célèbres se sont concentrés dans notre ville.

(32) *Par ailleurs, dans le monde scientifique, on est tenu d'être presque parfait pour ne pas **faire les autres souffrir** (correction proposée : *faire souffrir les autres*).*

Dans l'exemple 31, les pronoms doivent se placer devant le verbe ou l'auxiliaire. Dans l'exemple 32, le verbe *souffrir* est un verbe intransitif dans son sens de 'ressentir une vive souffrance physique ou morale, avoir des difficultés à faire quelque chose'. Selon la règle grammaticale, la construction doit être *faire+verbe intransitif à l'infinitif+sujet du verbe intransitif*, donc il faut être *faire souffrir les autres*.

3.2.3.4 Erreurs de prépositions (PREP)

Les prépositions constituent une difficulté importante pour les étudiants chinois, parce qu'il n'existe pas de système de prépositions équivalent en chinois et que les prépositions françaises sont compliquées. Vu le nombre important d'erreurs concernant les prépositions, nous avons décidé d'en préciser les sous-catégories : les prépositions surnuméraires, l'omission de prépositions et l'utilisation erronée de prépositions.

3.2.3.4.1 Prépositions surnuméraires (PS)

Les prépositions surnuméraires correspondent à l'emploi transitif indirect d'un verbe transitif direct dans la grille d'Anctil en 2010. Par exemple :

(33) ***J'attends de votre réponse** (correction proposée : *J'attends votre réponse*), M. le maire, et je vous remercie par avance.*

(34) *Chaque année, il y a beaucoup de français qui **viennent de tenter** (correction proposée : *viennent tenter*) leur chance.*

Dans l'exemple 33, le verbe *attendre* est un verbe transitif direct, qui régit un complément d'objet direct. Nous n'avons pas besoin de la préposition *de*. Dans l'exemple 34, le verbe *venir* s'emploie dans la construction *venir+verbe à l'infinitif* qui signifie 'se déplacer en direction d'un lieu pour faire quelque chose'. La préposition *de* est donc redondante ici.

3.2.3.4.2 Omission de prépositions (OP)

L'omission de prépositions correspond à l'absence des prépositions obligatoires. Par exemple :

(35) *La grande majorité de nos citoyens ont envie de **participer cet** (correction proposée : *participer à cet*) événement musical.*

(36) *Cette partie sert à s'arrêter le récit pour **faire l'auditeur mieux savoir** (correction proposée : *faire mieux savoir à l'auditeur*) où est l'intérêt de ce récit.*

(37) *La répétition est un genre où on répète les événements qu'on veut insister (correction proposée : *sur lesquels on veut insister*).*

Dans l'exemple 35, le verbe *participer* est un verbe transitif indirect qui régit une préposition *à* quand il est suivi d'un complément d'objet. Dans l'exemple 36, le verbe *savoir* est un verbe transitif. Son complément d'objet direct est la proposition *où est l'intérêt de ce récit*, son sujet est *l'auditeur*. Dans ce cas, selon la règle grammaticale, l'ordre de la construction doit être *faire+verbe transitif à l'infinitif+complément direct du verbe transitif+à+sujet du verbe transitif*. L'exemple 36 comporte donc à la fois une erreur d'ordre des mots et une erreur d'omission de la préposition *à*. Dans l'exemple 37, le verbe *insister* est un verbe transitif indirect qui régit la préposition *sur* quand il est suivi d'un complément. Nous pouvons restituer la phrase par *on veut insister sur les événements, donc on répète les événements sur lesquels on veut insister*.

3.2.3.4.3 Utilisation erronée de prépositions (EP)

Ce sous-type concerne un mauvais choix de prépositions. Par exemple :

(38) *On pourrait encourager notre peuple de (correction proposée : *à*) chanter.*

(39) *On met en avant les intérêts économiques et la productivité qui obligent aux employés de (correction proposée : *à*) sacrifier leur bien-être.*

Dans l'exemple 38, l'usage du verbe *encourager* est *encourager quelqu'un à faire quelque chose*. L'utilisation de la préposition *à* est arbitraire et on ne peut pas la substituer par d'autres prépositions. L'exemple 39 concerne la construction *obliger à quelqu'un à faire quelque chose*. Nous pouvons supposer que la superposition de deux *à* s'est avérée un peu étrange pour le scripteur, donc il a recouru à la préposition *de*. Il est aussi possible que le scripteur soit trop influencé par la construction connue *permettre à quelqu'un de faire quelque chose* et qu'il a effectué une surgénéralisation.

3.2.3.5 Erreurs de pronoms (PRO)

L'utilisation du pronom pourrait aussi concerner un problème de structure syntaxique, mais le pronom clitique étant assez particulier dans les langues romanes, cela vaut le coup d'en faire une erreur à part entière dans notre étude. Les erreurs classées dans cette sous-catégorie concernent le mauvais choix du pronom découlant d'une confusion sur les propriétés combinatoires syntaxiques du verbe. À la lumière de la grille de Yan (voir 2.3.2.2), nous avons décidé de classer les erreurs d'ordre de pronoms sous la catégorie « Erreurs

d'ordre des mots (ORD) » quand les pronoms sont bien choisis. Nous présenterons des exemples dans notre corpus :

(40) *Le concert est important, ne seulement pas les commerçant, mais aussi les artistes, peuvent **le bénéfici**er* (correction proposée : *en bénéfici*er).

(41) *Il est plutôt comme un déclencheur du texte, ce qui permet de donner envie aux lecteurs de savoir ce qui se passe en réalité et **leur pousse*** (correction proposée : *les pousse*) *à continuer de* (correction proposée : *à*) *lire le contenu suivant.*

(42) *Le narrateur s'arrête l'action d'événement pour attirer l'attention de l'auditeur et **les*** (correction proposée : *lui*) *faire mieux comprendre ~~que~~ ce qui s'est passé.*

Dans l'exemple 40, le verbe *bénéfici*er est un verbe transitif indirect, qui régit une préposition *de* quand il est suivi d'un complément d'objet. Il faut utiliser le pronom *en* pour pronominaliser un complément d'objet introduit par la préposition *de*. Dans la phrase erronée, le pronom choisi est un pronom direct *le*, ce qui prouve que le scripteur s'est trompé sur la transitivité du verbe *bénéfici*er. L'exemple 41 est presque le même cas. Le verbe *pouss*er est un verbe transitif direct dont l'usage est *pouss*er quelqu'un à faire quelque chose. L'utilisation du pronom complément d'objet indirect *leur* à la place du pronom complément d'objet direct *les* indique une confusion potentielle sur la transitivité du verbe *pouss*er. Dans l'exemple 42, le verbe *compr*endre est un verbe transitif direct, son complément d'objet direct est *ce qui s'est passé*, son sujet est *l'auditeur*, donc la construction doit être *faire mieux comprendre ce qui s'est passé à l'auditeur*. Pour remplacer le complément d'objet indirect à *l'auditeur*, il faut utiliser le pronom complément d'objet indirect *lui*.

3.2.3.6 Structures syntaxiques erronées (SYN)

3.2.3.6.1 Redondance syntaxique (RS)

Dans le cadre de notre mémoire, la redondance syntaxique fait référence à des phrases où figurent deux verbes principaux, alors qu'un seul est nécessaire, à des paraphrases autour du verbe, et à des formes pronominales inexistantes. Par exemple :

(43) *Je vous **prie** d'agr*er, *le maire de la ville, je vous **présente** mes respectueuses salutations* (correction proposée : *Je vous prie d'agr*er *mes respectueuses salutations*).

(44) *Chaque mot du modèle de phrase **est correspondant à*** (correction proposée : *correspond à*) *un signe afin que l'élève puisse s'exercer à utiliser le modèle de phrase après avoir appris le nouveau mot.*

Dans l'exemple 43, il existe deux verbes principaux : *prier* et *présenter*. Cependant, du fait que c'est une phrase simple sans conjonction de coordination, et qu'une phrase simple est

censée posséder un seul verbe principal, les deux verbes principaux dans cette phrase simple génèrent une erreur de redondance syntaxique.

Les exemples 44 illustrent les erreurs de paraphrase. La paraphrase renvoie à des expressions plus longues qui peuvent être reformulées par une autre expression moins lourde au contenu sémantique identique. En effet, l'expression *être correspondant à* est maladroite dans la phrase, le verbe *correspondre à* suffit pour exprimer le sens cible.

Il existe des redondances syntaxiques moins explicites associées aux propriétés inhérentes des verbes qui correspondent au pléonasma dans la grille d'Anctil (2010). Par exemple :

(45) *Je croie que ce concert gratuit peut **attirer** beaucoup de gens des autres villes **venir ici*** (correction proposée : *Je crois que ce concert gratuit peut attirer beaucoup de gens des autres villes*).

Dans l'exemple 45, le verbe *attirer* signifie 'faire venir'. Donc la partie *venir ici* dans la phrase est en quelque sorte une répétition du sens exprimé par le verbe. La phrase *Je crois que ce concert gratuit peut attirer beaucoup de gens des autres villes* est déjà complète.

Il faut mentionner un type de redondance syntaxique liée aux verbes pronominaux inexistantes. Il existe plusieurs cas où les apprenants ont inventé la forme pronominale des verbes. Par exemple :

(46) *Nous **nous déplorons** souvent **de** l'indisponibilité des médecins par manque du professionnel médical.*

Dans l'exemple 46, le verbe *déplorer* est un verbe transitif direct, il ne possède pas de forme pronominale. L'exemple comporte donc à la fois une erreur de redondance syntaxique et une erreur de prépositions surnuméraires.

Il faut souligner que les verbes pronominaux sont complexes. Nous avons donc décidé de les diviser en quatre cas : les erreurs d'accord du pronom avec le sujet dans les verbes pronominaux se classent sous type « Erreurs de conjugaison et d'accord » ; les erreurs concernant l'absence ou la redondance du pronom dans les pronominaux de sens passif relèvent de la sous-catégorie « Erreurs de voix » ; les pronominaux de sens réfléchi, réciproque ou absolu erronés sont classés dans la sous-catégorie « Erreurs de pronominaux » ; et les formes pronominales inexistantes sont considérées comme « Redondance syntaxique ».

3.2.3.6.2 Omission syntaxique (OS)

Contrairement à la redondance syntaxique, l'omission syntaxique désigne l'omission des composantes régies par les verbes qui sont indispensables, comme le sujet du verbe, le complément du verbe ou la conjonction, et l'omission du verbe lui-même dans une phrase.

Par exemple :

(47) Nous **ne trouvons pas supprimer** (correction proposée : *Nous ne trouvons pas que supprimer*) le concert est une bien (correction proposée : *bonne*) décision.

(48) Je sais que **cette décision à cause des** (correction proposée : *cette décision est prise à cause des*) raisons financières, mais je voudrais vous expliquer les avantages...

(49) Les musiciens auront la motivation à **créer** (correction proposée : *créer des musiques*).

L'absence de la conjonction *que* servant à introduire une proposition subordonnée est le cas le plus fréquent (47). Il existe également plusieurs erreurs concernant l'absence du verbe dans la phrase (48). L'absence de complément obligatoire dans l'exemple 49 renvoie à l'emploi intransitif d'un verbe transitif dans la grille d'Anctil (2010).

Nous avons aussi constaté l'absence de la reprise du sujet dans l'inversion sujet-verbe.

Par exemple :

(50) Pourquoi les jeunes adultes **prennent** (correction proposée : *prennent-ils*) une alimentation néfaste à la santé en sachant clairement l'importance de l'alimentation?

(51) Peut-être ce documentaire **veut** (correction proposée: *veut-il*) nous montrer ce qu'il faut savoir avant de plonger dans la Silicon Valley pour les entrepreneurs français.

Dans la forme interrogative (50), et dans la construction *peut-être+proposition* (51), l'inversion sujet-verbe et la reprise du sujet par un pronom personnel sont obligatoires.

3.2.3.6.3 Erreurs de constructions syntaxiques (CS)

Nous classons dans cette sous-catégorie toutes les erreurs syntaxiques liées aux verbes, en dehors de la redondance syntaxique et l'omission syntaxique, comme la fusion de plusieurs structures syntaxiques dans une phrase et les constructions verbales inexistantes. Nous illustrerons cette catégorie par des exemples ci-dessous :

(52) Plus profondément, cela va ouvrir une plus grande porte **pour nous d'aller plus loin** (correction proposée : *pour que nous puissions aller plus loin/pour aller plus loin*).

Dans l'exemple 52, il y a un chevauchement de deux structures. Nous pouvons dire que *cela va ouvrir une plus grande porte pour aller plus loin*. Dans ce cas, le sujet du verbe *aller* est le pronom indéfini *on* dans un sens général. Si nous voulons que le *nous* soit le sujet du verbe *aller*, il faut dire que *cela va ouvrir une plus grande porte pour que nous puissions aller plus loin*. Dans ce cas, le *nous* devient le sujet d'une proposition subordonnée introduit

par *que*. Les deux structures sont possibles, mais leur chevauchement *pour nous d'aller plus loin* n'est pas correct.

(53) *l'auteur nous **met en garde** (correction proposée : *met en garde en expliquant que*) que l'évaluation est parsemé (correction proposée : *présente*) dans tout le récit, non pas dans un seul endroit, bien qu'elle soit centrée dans le "foyer".*

Dans l'exemple 53, l'usage de la construction verbale *mettre en garde* est *mettre en garde quelqu'un sur/contre quelque chose*, qui signifie 'avertir quelqu'un de faire attention à quelque chose de négatif, comme un danger'. La construction *mettre en garde quelqu'un+préposition* n'existe pas. Nous proposons donc la correction *l'auteur nous met en garde en expliquant que l'évaluation est présente dans tout le récit*.

(54) ***Je semble*** (correction proposée : *il m'a semblé*) vraiment m'intégrer dans cet ensemble.

Dans l'exemple 54, le verbe *sembler* signifie 'présenter l'apparence de, donner l'impression'. Pour exprimer le sens de 'de mon avis, d'après moi', la construction est *il me semble que+proposition* ou *il me semble faire quelque chose*. L'usage *Je semble* n'existe pas dans ce sens.

(55) Enfin, il ***alerte les éditeurs de faire attention aux risques*** (correction proposée : *alerte les éditeurs sur les risques/alerte les éditeurs pour qu'ils fassent attention aux risques*) et les sensibilise à agir comme ce qu'a déjà fait le gros diffuseur Buzzfeed."

(56) ***Cela nous amène à réfléchir si on doit reconnaître le droit à l'erreur*** dans tous les cas ? (correction proposée : *Cela nous amène à réfléchir sur la reconnaissance du droit à l'erreur/Cela nous amène à nous demander si nous devons reconnaître le droit à l'erreur*).

Dans l'exemple 55, Le verbe *alerter* s'emploie dans les constructions *alerter quelqu'un sur quelque chose* ou *alerter quelqu'un pour que+proposition*. Il n'existe pas de construction *alerter quelqu'un de faire quelque chose*. La correction pourrait être *alerte les éditeurs sur les risques* ou *alerte les éditeurs pour qu'ils fassent attention aux risques*. Dans l'exemple 56, l'usage du verbe *réfléchir* est *réfléchir sur quelque chose*. La construction *réfléchir+préposition* n'existe pas. Nous proposons la correction *Cela nous amène à réfléchir sur la reconnaissance du droit à l'erreur* ou *Cela nous amène à nous demander si nous devons reconnaître le droit à l'erreur*.

3.2.4 Au niveau du genre textuel et du contexte : expression non conforme au registre de l'écrit au contexte universitaire (REG)

Le corpus collecté a été écrit en contexte universitaire, donc les textes sont censés respecter un registre formel. Conformément aux observations de Yan (2017), les erreurs

concernant le registre à l'écrit ne sont pas très présentes dans notre corpus. Parmi celles-ci, on relève les erreurs d'omission de *ne* dans la négation, et celles concernant l'utilisation des termes familiers. Par exemple :

(57) *On laisse tomber les travaux, les champs, les devoirs, juste pour écouter de la musique créée par nos musiciens locaux, on **se souvient plus** (correction proposée : *on ne se souvient plus*) de soucis, de pression.*

Dans l'exemple 57, la première partie de la négation *ne* qui est censée précéder le verbe est supprimée comme à l'oral. À l'écrit formel, il faut l'ajouter dans la phrase.

(58) *Remontons le fil de l'histoire, la génération de nos grands-parents revendique rarement le bonheur au travail, car on **bossait** (correction proposée : *travaillait*) juste pour subvenir à ses besoins.*

(59) *dans le monde scientifique, Edison a inventé l'ampoule après avoir tenté et **fait capoter** (correction proposée : *raté*) plein d'expériences.*

Dans les exemples 58 et 59, les verbes *bossier* et *faire capoter* relèvent du registre très familier et relèvent de l'oral. Ils sont considérés comme inadaptés dans un écrit universitaire.

3.2.5 Récapitulation de la grille d'erreurs

Après avoir expliqué d'une façon détaillée les classements des erreurs verbales, nous récapitulons ci-dessous notre grille d'erreurs :

Classes d'erreurs	Types d'erreurs	Exemples
	Erreurs de conjugaison et d'accord (CA)	<p>Ex : Le narrateur choit (choisit) de faire une explication supplémentaire qui, néanmoins, ne se consacre pas au déroulement.</p> <p>En tant qu'un concert gratuit, il nous permet de se retrouver (nous retrouver) et écouter la musique ensemble sous une atmosphère agréable.</p>

<p>Au niveau morphologique (MORPHO)</p>	<p>Erreurs d'orthographe d'usage (ORT)</p>		<p>Ex : Nous vous persuadons de ne pas supprimer (supprimer) cet évènement et de ne pas nous arrêter de nous plaire.</p> <p>C'est très utile de obtenir (d'obtenir) une bonne réputation de notre ville.</p>
	<p>Erreurs de verbes fictifs : néologismes (VFIC)</p>		<p>Ex : Également, ils sont capables d'exhibitionner (montrer) leur talent musical sur scène.</p> <p>Les artistes gagnent la réputation à cause du grand nombre de spectateurs et beaucoup d'opéra hors de la province les invitent aussi à performer (représenter leur musique).</p>
<p>Au niveau sémantique et des cooccurrences lexicales (SEMAN)</p>	<p>Erreurs de sens lexical (SEN)</p>	<p>L'utilisation inappropriée des synonymes (SNM)</p>	<p>Ex : Il y a beaucoup de techniques pour enrichir la syntaxe narrative mais nous confinons (limitons) nos études aux éléments évaluatifs qui concernent directement à la structure interne.</p> <p>Ce graphique démontre (montre) le nombre des temps utilisés par le texte mystérieux.</p>
		<p>Le sens inadapté (SIN)</p>	<p>Ex : Puisque vous êtes en face des raisons financières, j'ai trouvé quelque solutions pour rédigier (résoudre) ce problème.</p> <p>Je me permet de vous bien vouloir examiner la proposition consistant à faire endurer (durer) le festival musical à la fin de l'été.</p>

Au niveau de la	Erreurs de temps/d'aspect/ mode (TAM)	<p>Ex : Comme vous le saviez (savez), depuis l'existence de cette fête, un nouveau moyen de se reposer s'ajoute dans la vie des citoyens.</p> <p>Supposons qu'il y a (ait) un concert gratuits, des artistes pourraient être invités pour participer au concert.</p>
	Erreurs de cooccurrences (COC)	<p>Erreurs de cooccurrence avec un nom (CONO)</p> <p>Ex : Ce n'est pas un moyen pertinent de faire toujours des compris (compromis) avec les concurrents tels que Facebook.</p> <p>Les amateurs de la musique, notamment ceux qui ne sont pas très habitués à assister à un concert, sont (ont) accès à divers genres de musiques grâce à la gratuité de cet événement.</p>
		<p>Erreurs de cooccurrence avec l'adverbe ou le complément circonstanciel (COA)</p> <p>Ce changement de la langue du signe conduit exactement (inévitablement) à l'incompréhension des interprétations.</p> <p>Ce type peut se constituer d'ordinaire (facilement) mais manifeste profondément ce que le locuteur a évalué.</p>

grammaire et de la syntaxe (GRASYN)	Erreurs de voix (VV)	<p>Ex : Les gens qui passionnent à (qui sont passionnés par/qui se passionnent pour) la musique viennent ici, ils ont besoin des endroits pour dormir.</p> <p>Par exemple, on peut se rencontrer (rencontrer) les gens qui viennent des régions différentes et renforcer des amitiés.</p>
	Erreurs de pronominaux (EPN)	<p>Elle vêt (se vêt) d'un t-shirt à manche court rose moulant.</p> <p>Il faut rendre compte (se rendre compte) que tous les verdicts ne sont pas sans appel car la cour d'appel et la cour de cassation sont là pour contester des décisions litigieuses et inopportunes.</p>
	Erreurs d'ordre des mots (ORD)	<p>Ex : Il est significatif puis-que nous pouvons ainsi développer mieux (mieux développer) l'industrie.</p> <p>L'année dernière, plus de 1000 touristes sont y allés (y sont allés) et des artistes les plus célèbres se sont concentrés dans notre ville.</p>
	Erreurs de prépositions (PREP)	Prépositions surnuméraires (PS)
Omission de prépositions (OP)		La grande majorité de notre citoyens ont envie de participer (à) cet événement musical.
Utilisation erronée de		On pourrait encourager notre peuple de

		prépositions (EP)	(à) chanter.
	Erreurs de pronoms (PRO)		Ex : Le concert est important, ne seulement pas les commerçant, mais aussi les artistes, peuvent le bénéficié (en bénéficié).
Structures syntaxiques erronées (SYN)	Redondance syntaxique (RS)		Ex : Je vous prie d'agréer , le maire de la ville, je vous présente mes respectueuses salutations. (Je vous prie d'agréer mes respectueuses salutations) Je croie que ce concert gratuit peut attirer beaucoup de gens des autres villes venir ici (peut attirer beaucoup de gens des autres villes).
	Omission syntaxique (OS)		Ex : Je sais que cette décision (est prise) à cause des raisons financières, mais je voudrais vous expliquer les avantages que ce concert représente et vous peut-être changez votre avis. Nous ne trouvons pas (que) supprimer le concert est une bien décision.
	Erreurs de constructions syntaxiques (CS)		Ex : Plus profondément, cela va ouvrir une plus grande porte pour nous d'aller plus loin (pour aller plus loin/ pour que nous puissions aller plus loin) Je semble (il m'a semblé) vraiment m' intégrer dans cet ensemble.

<p>Au niveau du genre textuel et du contexte (CONT)</p>	<p>Expressions non conformes au registre de l'écrit au contexte scolaire (REG)</p>	<p>Ex : On laisse tomber les travaux, les champs, les devoirs, juste pour écouter de la musique créée par nos musiciens locaux, on se souvient plus (on ne se souvient plus) de soucis, de pression.</p> <p>Remontons le fil de l'histoire, la génération de nos grands-parents revendique rarement le bonheur au travail car on bossait (travaillait) juste pour subvenir à ses besoins.</p>
---	--	---

Tableau 3.1 : Grille d'erreurs élaborée pour l'analyse des erreurs verbales

3.3 L'annotation systématique dans le corpus en XML

À la différence des recherches précédentes comme celles d'Anctil (2010), notre méthodologie consiste à élaborer un corpus d'apprenants indiquant les erreurs comme celle de Granger (2007). Nous proposerons donc dans ce mémoire un travail d'annotation systématique des verbes et des erreurs verbales dans le corpus en XML.

L'annotation du corpus renvoie à la pratique qui consiste à ajouter les informations interprétatives aux corpus sans en modifier le contenu d'origine (Granger et al. 2015). L'annotation dans le corpus représente nombre d'avantages. Tout d'abord, cela nous permet de situer les verbes et les erreurs verbales dans le co-texte. Du fait que l'annotation représente seulement une interprétation possible des données de la part des annotateurs, il nous paraît plus pertinent de montrer aux lecteurs le co-texte des erreurs verbales que de les extraire du texte. Dans un second temps, les annotations assurent des fonctions importantes : par exemple, les annotations morphosyntaxiques permettent de désambiguïser les formes avec différentes fonctions, par exemple, la forme *analyse* peut être un nom ou un verbe conjugué. Par ailleurs, l'annotation XML nous permet d'intégrer toutes les informations utiles au sein du texte numérique. Un point important est qu'on peut aussi intégrer les métadonnées qui comportent des informations sur les scripteurs et sur les conditions de rédaction : l'âge, le nombre d'années d'études du français, les langues acquises et d'autres informations complémentaires,

à partir des informations indiquées dans les fiches sociobiographiques des apprenants. C'est un enrichissement des données qui nous aide à identifier l'origine des erreurs. De plus, le classement des erreurs verbales est abstrait et demande une analyse concrète. Les annotations dans le corpus explicitent cette modélisation et permettent de valider le modèle. En fin de compte, les textes numériques en XML sont interprétables par les machines, et les étiquettes homogènes des annotations favorisent une analyse quantitative automatique. Par exemple, nous pouvons calculer automatiquement le nombre d'un type d'erreurs donné dans le corpus. Les résultats de recherche seront affichés sous forme de concordance, ce qui facilite la vérification manuelle.

Le langage XML (eXtensible Markup Language) est un sous-ensemble de SGML (Standard Generalized Markup Language : un langage informatique normalisé de définition et d'échange de documents structurés). Il facilite l'intégration des ressources diverses par les balisages qui sont une manière formelle de caractériser les composantes d'une structure textuelle. Les annotations sont marquées en différente couleur que le texte original au format XML, elles sont donc mieux visualisées et le texte original est récupérable et indépendant des annotations. On peut aussi utiliser des feuilles de style pour mieux visualiser les annotations. Tous les corpus annotés à l'heure actuelle utilisent XML.

Avant d'exposer notre méthode d'annotation, nous allons présenter tout d'abord quelques propositions d'annotation des erreurs lexicales en anglais et en français.

3.3.1 Les propositions d'annotation des erreurs en anglais et en français

En anglais, dans le travail de Dagneaux et al. (1998) sur le corpus d'apprenants francophones de l'anglais langue seconde, il existe 7 catégories principales, comme nous l'avons vu précédemment : formel, grammatical, lexico-grammatical, lexical, registre, redondance / absence / l'ordre erroné des mots, et style (voir la section 2.3.1). Ces catégories sont encodées par des abréviations. Par exemple, G représente « erreurs grammaticales ». Le système d'annotation est hiérarchique : à la suite du code de catégories principales, il existe des codes de sous-catégories. Par exemple, suite à la catégorie G représentant les erreurs grammaticales, il y a GV représentant erreurs grammaticales pour les verbes, qui est subdivisée en GVT (tense errors : erreurs de temps), GVAUX (auxiliary errors : erreurs d'auxiliaire), GVV (voice errors : erreurs de voix), etc. Prenons un exemple d'annotation

dans le corpus : *I closed my eyes and saw the troops observing (FS) eachother \$each others\$ with hostility*. La catégorie d'erreurs (FS) est indiquée avant le mot erroné, et la correction proposée se trouve entre les deux signes \$. Dans le cas d'exemple, il s'agit donc d'une erreur formelle de nombre. Les codes d'annotations hiérarchiques ont pour avantage de pouvoir indiquer la catégorie et la sous-catégorie en même temps, et de rendre le système d'encodage flexible, c'est-à-dire que les annotateurs pourraient ajouter ou supprimer les codes de sous-catégories en fonction de leurs corpus (Dagneaux et al, 1998). Toutefois, étant donné que notre grille portant sur les erreurs verbales sera plus détaillée que la grille de Dagneaux (1998) et aura plus de sous-catégories, la superposition des codes risque de rendre le codage long et lourd. Nous choisissons donc d'attribuer des codes indépendants à des catégories et des sous-catégories d'erreurs, mais d'utiliser deux attributs « type » et « subtype » pour indiquer leur relation hiérarchique.

En français, nous allons présenter l'annotation des erreurs dans le corpus Frida. L'annotation a inspiré du système d'étiquetage anglais (Dagneaux et al, 1998) en adoptant l'encodage des catégories d'erreurs par des abréviations, mais avec un format différent. Les données sont standardisées au format XML dans le corpus Frida. Il faut mentionner que le corpus Frida possède un double système d'annotation des erreurs. Pour le premier niveau de balisage, chaque erreur se voit attribuer trois balises : domaine d'erreurs, catégorie d'erreurs et catégorie grammaticale (voir la section 2.3.2.3). Le second niveau de balisage est un étiquetage morphosyntaxique qui indique les informations sur le lemme, la catégorie grammaticale, le genre et le nombre. Les deux niveaux de balisage ont été imbriqués au format XML. Prenons un exemple (Biskri & Jebali, 2010, p.155) :

Le manager des relations

```
<err ide= "2" dom1= "G" cer1= "NBR" cgr= "ADJ">
  <ini>
    <tok lemme= "public" cat= "adj"
      tag= "fem sing">publique</tok>
  </ini>
  <cor>
    <tok lemme= "public" cat= "adj"
      tag= "fem plu">publiques</tok>
  </cor>
</err>
```

Chaque erreur se voit attribuer un identifiant (ide). Les balises de domaine d'erreurs (dom) et de catégorie d'erreurs (cer) sont numérotées pour pouvoir contenir plusieurs erreurs relevant d'une même balise d'un même token (par exemple, si l'adjectif employé est au masculin singulier, lorsque l'adjectif cible devrait être au féminin pluriel, il comporte deux catégories d'erreurs : une erreur de genre et une erreur de nombre, donc cer1= "GRE" cer2= "NBR"). L'état initial d'une erreur se trouve entre les balises <ini> </ini>, et sa correction entre <cor> </cor>. Pour l'état initial d'une erreur, comme pour sa correction, le lemme, la catégorie grammaticale ainsi que le nombre et le genre sont indiqués.

Nous pouvons constater une déficience dans ce système d'annotation, qui est la répétition de la catégorie grammaticale dans le premier niveau de balisage « cgr » et dans le deuxième niveau de balisage « cat ». En revanche, nous pouvons nous inspirer de la numérotation des domaines d'erreurs et des catégories d'erreurs pour contenir plusieurs erreurs d'un même token, car il existe de nombreux verbes qui comportent plusieurs types d'erreurs dans notre corpus. Puisque notre étude se concentre sur les verbes, nous indiquerons également le lemme du verbe, la catégorie d'erreurs et les sous-catégories d'erreurs. Cependant, nous ne proposerons pas de correction dans notre annotation, parce que pour le mauvais choix du verbe, il existe potentiellement plusieurs corrections possibles.

Après avoir regardé quelques propositions d'annotation des erreurs, nous allons présenter notre schéma d'annotation.

3.3.2 Le balisage d'annotation des erreurs verbales

Nous avons intégré un ensemble d'attributs dans notre balisage d'annotation, en suivant les recommandations de la Text Encoding Initiative, une communauté académique internationale du champ « humanités numériques » qui vise à fournir des recommandations pour l'encodage des données textuelles numériques (Burnard, 2014).

Le balisage que nous proposons a été conçu en collaboration avec Magdalena Augustyn¹⁰, chercheur à l'université Grenoble-Alpes. Le balisage fonctionne de la façon suivante. L'élément <teiHeader> entoure les métadonnées du fichier, y compris

¹⁰ Nous adressons nos sincères remerciements à Madame Magdalena Augustyn pour sa contribution dans la conception du schéma d'annotations.

l'établissement d'édition, le responsable du projet, l'adresse de l'organisme du projet, les informations sur les scripteurs. Les paragraphes sont marqués par l'élément <p>. Nous avons utilisé l'élément <w> qui signifie *word* pour entourer un mot. « pos » indique la catégorie du discours des éléments entourés, dans notre cas, c'est « pos= "v" » pour annoter les verbes. « lemma » indique le lemme des verbes entourés. Nous avons annoté tous les verbes dans la perspective de calculer la proportion des verbes erronés dans tous les verbes. L'annotation du lemme nous aide à mettre en lumière les verbes les plus utilisés. Pour les verbes erronés, outre les deux valeurs précédentes, nous intégrons la valeur « type » pour marquer la catégorie d'erreurs et la valeur « subtype » pour montrer le type ou le sous-type d'erreurs. Les quatre grandes catégories entrent dans la balise « type », et les types et sous-types se placent dans la balise « subtype ». Comme nous l'avons vu, il existe deux niveaux de sous-types dans notre grille d'erreurs, donc dans la balise « subtype », nous proposons de marquer les deux niveaux de sous-types par le signe « / ». Nous illustrons notre balisage par l'exemple suivant :

Nous <w pos="v" lemma="aller">allons</w><w pos="v" lemma="se pencher" type="MORPHO" subtype="CA">se pencher</w> sur la surdité en matière identitaire et culturelle, et <w pos="v" lemma="mettre en oeuvre" type="SEMAN" subtype="SEN/SIN">mettre en œuvre</w> dans quel contexte les langues des signes <w pos="v" lemma="évoluer">évoluent</w> et comment l'identité sourde <w pos="v" lemma="se construire">se construit</w>.

Pour les verbes corrects, comme *allons*, *évoluent* et *se construit*, la balise « pos » indique le statut verbal, et la balise « lemma » présente le lemme du verbe. Pour les verbes erronés, comme *mettre en œuvre*, à part le statut verbal et le lemme, il y a « SEMAN » dans la balise « type » signifiant « erreurs au niveau sémantiques et des cooccurrences lexicales », et « SEN/SIN » dans la balise « subtype » indiquant le type d'erreurs « sens lexical » et le sous-type d'erreurs « sens inadapté ».

Quand un verbe comporte plusieurs types d'erreurs, nous décidons d'ajouter une balise indiquant la deuxième catégorie d'erreurs et le type d'erreurs. L'exemple suivant qui comporte à la fois une erreur de redondance syntaxique et une erreur de prépositions surnuméraires a été annoté comme ci-dessous :

Nous <w pos="v" lemma="déplorer" type="GRAMSYN" subtype="SYN/RS"><w type="GRAMSYN" subtype="PERP/PS">nous déplorons</w></w> souvent de l'indisponibilité des médecins par manque du professionnel médical.

Dans la balise de la deuxième catégorie d'erreurs, nous avons supprimé les codes « pos » et « lemma », car d'une part, ce n'est pas la peine de répéter les mêmes informations, et d'autre part, cela sera plus pratique pour compter le nombre des verbes par l'étiquette « lemma ».

Avec le développement du traitement automatique des langues, il est possible de faire certaines annotations d'une manière automatique, surtout pour les annotations morphosyntaxiques. Cependant, il faut souligner que les outils destinés à annoter automatiquement les corpus sont souvent testés sur la base des données standard des natifs, ou sur la base des propriétés linguistiques conventionnelles. C'est pourquoi les erreurs dans les corpus d'apprenants, surtout celles qui sont liées à l'orthographe et à la conjugaison, perturbent sans doute l'identification du statut morphosyntaxique des données. Une vérification manuelle par des annotateurs pour décider d'accepter ou de rejeter l'annotation proposée par l'informatique paraît donc indispensable dans ce type de corpus. Il s'agit donc plutôt d'une annotation semi-automatique dans les corpus d'apprenants.

Passons maintenant à la présentation de la méthode d'annotation semi-automatique des erreurs dans notre étude.

3.3.3 L'annotation semi-automatique des verbes avec NooJ

Pour l'annotation des verbes dans notre corpus, nous avons aussi recouru à NooJ¹¹ (Silberztein, 2015), un outil du traitement automatique des langues, pour repérer automatiquement les verbes. NooJ est capable d'effectuer automatiquement des analyses linguistiques des textes.

Nous avons remarqué qu'il existe des déficiences du repérage automatique, comme les verbes aux temps composés sont annotés en tant que deux verbes, le pronom est exclu du balisage dans les verbes pronominaux, seulement les participes présents sont entourés dans les gérondifs. Il existe du silence et du bruit dans les données extraites. Le silence désigne les données adaptées à notre requête de recherche, mais qui n'ont pas pu être récupérées, et le bruit renvoie aux données inadaptées à notre besoin, mais qui sont affichées. Dans notre cas, le silence ou le bruit sont souvent attribuables à la nature de notre corpus, à savoir un corpus

¹¹ NooJ : <http://www.nooj4nlp.net/pages/nooj.html>

d'apprenants non-natifs, où les erreurs d'orthographe empêchent sans doute l'identification des verbes. Par exemple, dans la phrase *Lorsque les professeurs enseignent des mots ou des phrases anglais, ils présentent d'abord en langue écrite en anglais, puis joignent une langue écrite en chinois*, *ils* est annoté automatiquement comme un verbe, mais il est évident que c'est le pronom personnel *ils* qui a été mal orthographié. Certains éléments du silence ou du bruit sont dus au fait que les verbes conjugués possèdent une forme identique aux noms, par exemple dans la phrase *La notion de distance statistique mesure l'écart entre deux données*, le verbe *mesure* n'a pas été annoté automatiquement, car *mesure* peut aussi être un nom. Les déficiences mentionnées au-dessus constituent des arguments pour une vérification manuelle rigoureuse.

En dehors du fait d'examiner la pertinence des données extraites automatiquement, il faut aussi faire attention à vérifier les points ci-dessous : premièrement, pour les erreurs d'omission du verbe dans une phrase, nous devons compléter manuellement le verbe absent, comme ce que montre l'exemple ci-dessous :

Je `<w pos="v" lemma="savoir">sais</w>` que cette décision `<w pos="v" lemma="prendre" type="GRAMSYN" subtype="SYN/OS"></w>` à cause des raisons financières, mais je `<w pos="v" lemma="vouloir">voudrais</w>`vous `<w pos="v" lemma="expliquer">expliquer</w>` les avantages que ce concert `<w pos="v" lemma="représenter">représente</w>` et vous peut-être `<w pos="v" lemma="changer" type="GRAMSYN" subtype="ORD">changez</w>` votre avis.

Le deuxième point consiste à désambiguïser les verbes. La forme *être+participe passé* constitue le cas typique, il faut décider l'annotation du lemme du verbe. Par exemple, *La journée mondiale des sourds (JMS), créée en 1958, est une journée internationale destinée à sensibiliser la population sur la surdité et la présentation de la culture sourde*, nous devons annoter *est* comme verbe *être* et *destinée* comme verbe *destiner*. Tandis que dans la phrase *Le niveau est simplifié, les apprentissages sont réduits à l'essentiel, et dans la plupart du temps, les enfants ne peuvent apprendre que les notions de base*, nous devons annoter *est simplifié* comme verbe *simplifier* et *sont réduits* comme verbe *réduire*.

L'une des difficultés que nous avons rencontrées pendant l'annotation des verbes est la détermination du statut des participes présents ou participes passés utilisés comme adjectif. Par exemple, dans la phrase *Mais les problème se résident dans les programmes scolaires à l'école des sourds qui sont plutôt adaptés pour les sourds et le décalage entre l'école ordinaire et l'école spéciale se creuse*, la construction *sont adaptés* peut être considérée

comme la construction passive du verbe *adapter* ou la construction *être+adjectif*, son lemme peut être *adapter* dans le premier cas, ou *être* dans le deuxième cas. Cela dépend de l'interprétation de l'annotateur. Dans notre méthodologie, si les participes passés ou les participes présents peuvent être reformulés par une proposition relative, ils sont considérés comme des verbes. Par exemple : pour la phrase *Il existe des techniques **consacrés** à enregistrer le plus grand nombre possible de discours*, on peut la reformuler par *Il existe des techniques **qui sont consacrées** à enregistrer le plus grand nombre possible de discours*. Le participe passé *consacrées* est donc annoté comme un verbe car il peut être considéré comme une forme courte du passif. Pour la phrase *Il existe quatre outils principaux **consistant** à l'enrichissement de la syntaxe narrative*, on peut aussi la reformuler par *Il existe quatre outils principaux **qui consistent** à enrichir la syntaxe narrative*. Le participe présent *consistant* a aussi reçu un statut verbal. Il existe aussi des cas plus évidents, comme *la langue parlée, résultats escomptés, impressions stéréotypées*, où les participes passés sont reconnus comme adjectif. L'attribution du statut verbal est également assurée par la double annotation des annotateurs.

Suite à l'annotation semi-automatique des verbes, nous avons annoté manuellement les erreurs verbales avec un guide d'annotation précis¹². Plusieurs annotateurs se sont engagés dans ce travail pour garantir la pertinence d'annotation. À la fin, les annotateurs ont discuté des désaccords sur l'annotation pour arriver à une décision finale.

Voici un tableau illustrant quelques exemples d'annotation :

Classes d'erreurs	Types d'erreurs	Exemples
Au niveau morphologique (MOPRHO)	Erreurs d'orthographe d'usage (ORT)	Ex : Demande de <w pos="v" lemma="continuer">continuer</w> à <w pos="v" lemma="organiser" type="MORPHO" subtype="ORT">orgniser</w> le concert gratuit à la fin.
Au niveau sémantique et des	Erreurs de sens	Ex : on <w pos=" v" lemma=" pouvoir">peut</w><w pos=" v" lemma=" dépasser" type=" SEMAN"

12 Voir annexe 5

cooccurrences lexicales (SEMAN)	inadapté (SIN)	<code>subtype=" SEN/SIN">dépasser</w></code> ce jour comme un fête.
Au niveau de la grammaire et de la syntaxe (GRASYN)	Redondance syntaxique (RS)	Ex : Je vous <code><w pos="v" lemma="prier">prie</w></code> d' <code><w pos="v" lemma="agr�er">agr�er</w></code> , le maire de la ville, je vous <code><w pos="v" lemma="pr�esente" type="GRAMSYN" subtype="SYN/RS">pr�esente</w></code> mes respectueuses salutations.
Au niveau du genre et du contexte (CONT)	Expression non conforme au registre de l'�crit au contexte scolaire (REG)	Ex : on <code><w pos="v" lemma="se souvenir" type="CONT" subtype="REG">se souvient</w></code> plus de soucis, de pression.

Tableau 3.2 : Grille d'illustration de l'annotation dans le corpus

En somme, nous voyons que les annotations valorisent le corpus pour la recherche, dans la mesure o  les informations ajout es permettent d'extraire les donn es du corpus qui ne sont pas accessibles par d'autres moyens. De plus, les  tiquettes homog enes facilitent l'analyse quantitative. Nous passons maintenant   la m thode du calcul des r sultats.

3.3.4 La m thode du calcul automatique des r sultats

  partir de notre annotation, nous avons calcul  automatiquement le nombre des verbes et de chaque type d'erreurs verbales. Puisque chaque verbe se voit attribuer un lemme, nous avons recherch  les occurrences de « pos= » dans le corpus. Il faut ajouter le signe « = » pour  liminer les cas o  les mots comportent les lettres « pos ». Pour compter le nombre des erreurs verbales, nous avons recherch  les occurrences de « subtype= ». Pour calculer le nombre d'erreurs de chaque cat gorie d'erreurs et de chaque type d'erreurs, nous avons recherch  par les codes des cat gories, comme « MORPHO », « SEMAN », « GRAMSYN », « CONT ». Nous avons aussi fait une v rification des r sultats de recherche pour exclure le

bruit. Les résultats ont été affichés sous forme de concordances, ce qui facilite notre vérification. Ci-dessus se présente une capture d'écran des résultats de recherche de « EP » :



Dans ce chapitre, nous avons présenté la méthodologie de collecte des données, d'annotation et de calcul des erreurs. Notre recherche est fondée sur un corpus d'apprenants sinophones de niveau B1 et B2, dans lequel nous avons classé les erreurs verbales selon les dimensions linguistiques, ce qui a débouché sur l'élaboration d'une grille d'erreurs descriptive. Nous avons aussi effectué un travail d'annotation systématique des verbes et des erreurs verbales dans les corpus en XML en collaboration avec plusieurs chercheurs. À l'aide de l'informatique, nous avons pu calculer le nombre des erreurs et révéler la régularité des erreurs commises à travers une analyse quantitative. Nous présenterons les résultats d'analyse dans le chapitre suivant.

Chapitre IV. L'analyse et l'interprétation des données

Suite à l'annotation des verbes et des erreurs verbales dans le corpus, nous procéderons à l'analyse et l'interprétation des résultats dans ce chapitre. Lors de la constitution du corpus, nous avons fait l'hypothèse selon laquelle les erreurs de temps/aspect/mode et de conjugaison et d'accord seront plus présentes chez les apprenants sinophones en raison du caractère invariable des verbes en chinois.

Dans un premier temps, nous observerons la répartition des erreurs selon les différents types dans le corpus à travers une analyse quantitative. Nous chercherons à identifier les difficultés principales des apprenants à travers une analyse qualitative. Dans un second temps, nous comparerons les textes rédigés sur table et les textes écrits à la maison afin d'observer l'influence de la condition de rédaction sur les erreurs commises. Enfin, nous essayerons de trouver les sources possibles des erreurs. L'identification des sources d'erreurs a pour but, d'une part, de sensibiliser les apprenants aux erreurs récurrentes, d'autre part, de mieux servir à la didactique et des verbes en français auprès des apprenants sinophones de niveaux B1 et B2.

4.1 L'analyse quantitative et qualitative des erreurs

Notre corpus comporte 5705 verbes au total. Rappelons que notre corpus comporte 50,014 mots, les verbes représentent donc une proportion de 11,4 % des mots, ce qui montre que les verbes occupent une place importante dans la rédaction. Nous avons relevé 875 occurrences d'erreurs verbales, c'est-à-dire que 15,3% de verbes sont mal employés. Cela rejoint les résultats de Ying (2013), qui montre que l'utilisation des verbes constitue une difficulté majeure pour les apprenants sinophones. Nous allons observer tout d'abord la répartition des erreurs selon les quatre grandes catégories, les types et les sous-types d'erreurs pour en avoir un aperçu général. Puis nous analyserons plus précisément les erreurs par catégorie.

4.1.1 La répartition des erreurs selon les catégories d'erreurs et les types d'erreurs

Après vérification, nous avons obtenu les résultats de la répartition des erreurs selon les quatre grandes catégories : 184 erreurs au niveau morphologique, 318 erreurs au niveau

sémantique et des cooccurrences lexicales, 340 erreurs au niveau de la grammaire et de la syntaxe, et 11 erreurs au niveau du genre et du contexte (cf. Figure 4.1).

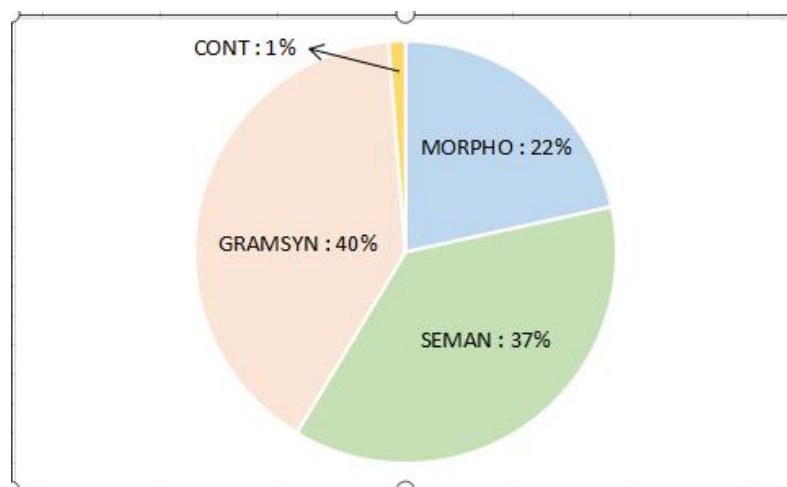


Figure 4.1 : La répartition des erreurs selon les quatre grandes catégories

Nous pouvons constater que les erreurs au niveau de la grammaire et de la syntaxe sont les plus fréquentes, puisqu'elles représentent un pourcentage de 40 % parmi toutes les erreurs. Elles sont suivies des erreurs au niveau sémantique et des cooccurrences lexicales avec un pourcentage presque équivalent (37 %). Ceci montre que les apprenants ont des difficultés à maîtriser la valeur sémantique, la combinatoire sémantique et syntaxique d'une unité verbale. Les erreurs morphologiques ne représentent que 22 %, soit environ une moitié des deux précédentes, ce qui montre que les erreurs morphologiques sont moins nombreuses que les erreurs syntaxiques et sémantiques pour les apprenants intermédiaires de niveau B1-B2. Mais il faut mentionner que les erreurs morphologiques occupent quand même un cinquième de toutes les erreurs, car les apprenants sont encore en phase d'acquisition de la compétence langagière. Conformément aux observations de Yan (2017), les erreurs au niveau du genre et du contexte sont les moins fréquentes avec un pourcentage de 1 %.

Il est intéressant de noter que dans l'étude de Yan (2017), qui a relevé 221 erreurs pour dix verbes scientifiques transdisciplinaires : *choisir, considérer, comparer, décrire, étudier, expliquer, montrer, noter, proposer, supposer*, les résultats sont similaires aux nôtres : les erreurs au niveau de la grammaire et de la syntaxe sont les plus nombreuses, suivies des erreurs au niveau sémantique et des cooccurrences lexicales. Les erreurs morphologiques arrivent en troisième place, et les erreurs du genre et du contexte sont peu présentes. Rappelons que le corpus de Yan est composé des mémoires rédigés par les apprenants

sinophones de niveau C1-C2, tandis que notre étude porte sur tous les verbes et concerne les apprenants sinophones de niveau B1-B2. Les résultats similaires montrent qu'avec une progression de la compétence langagière, les erreurs verbales des apprenants sinophones se concentrent toujours sur le plan syntaxique et sémantique. Le sens lexical, la combinatoire sémantique et syntaxique des unités verbales méritent donc toute notre attention.

Passons maintenant à l'analyse plus précise de la répartition des erreurs selon les types d'erreurs (cf. Figure 4.2) et les sous-types d'erreurs (cf. Figure 4.3).

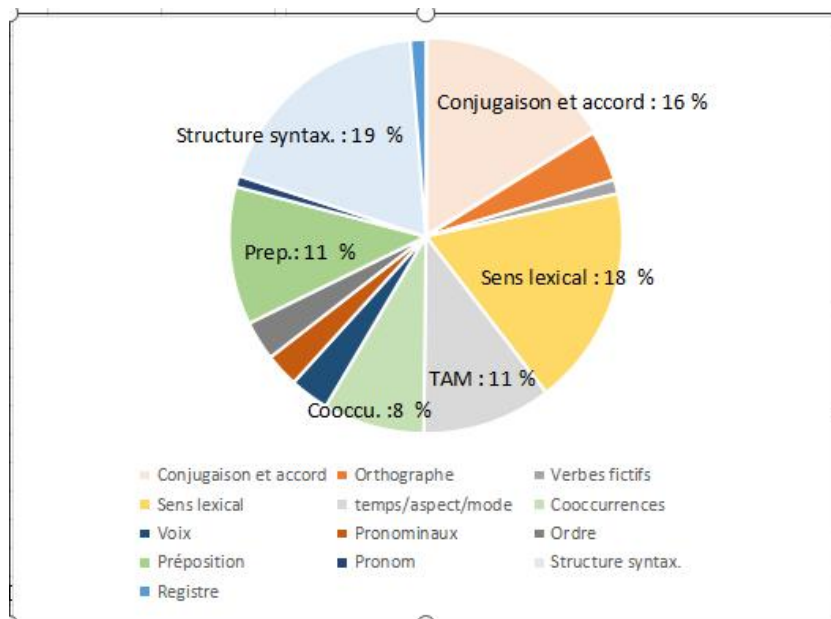


Figure 4.2 : La répartition des erreurs selon les types d'erreurs

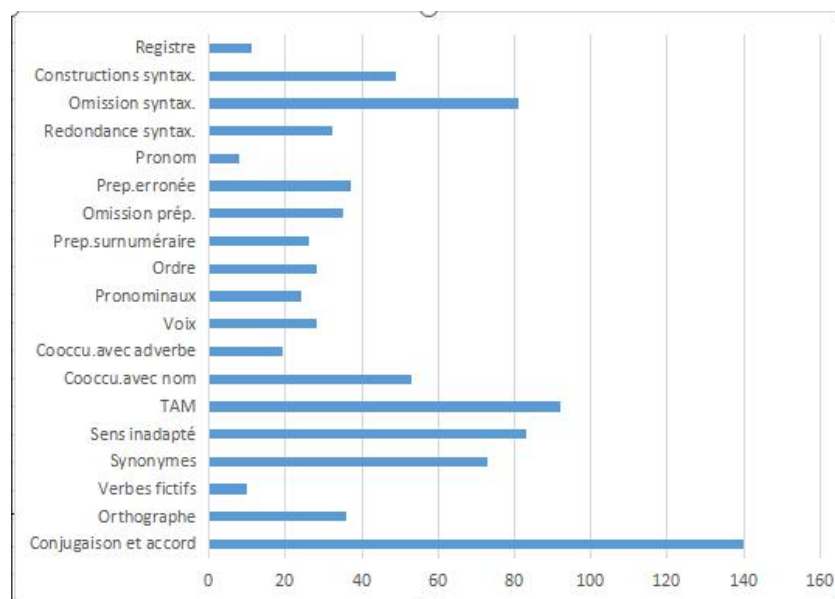


Figure 4.3 : La répartition des erreurs selon les sous-types d'erreurs

Dans la figure 4.2, d'une manière générale, les erreurs de structure syntaxique (SYN 19 %), comportant les erreurs de redondance syntaxique (RS, 4 %), les erreurs d'omission syntaxique (OS, 9 %) et les erreurs de constructions syntaxiques erronées (CS, 6 %) sont les plus importantes. Elles sont suivies des erreurs de sens lexical (SEN, 18 %) regroupant les erreurs de synonymes (SNM, 8 %) et les erreurs de sens inadapté (SIN, 10 %). Nous voyons que les erreurs de synonymes et sens inadapté occupent une proportion presque équivalente, ce qui justifie notre division des erreurs de sens lexical. Les erreurs de conjugaison et d'accord (CA, 16 %) arrivent en troisième place. Les erreurs de temps/aspect/mode (TAM) représentent un pourcentage de 11 %, identique avec les erreurs de préposition (PREP) regroupant la préposition surnuméraire (PS, 3 %), l'omission de préposition (OP, 4 %) et l'utilisation de préposition erronée (EP, 4%). Un autre type important est celui des erreurs de cooccurrences (COC, 8 %), y compris les erreurs du verbe en cooccurrence avec un nom (CONO, 6 %) et en cooccurrence avec un adverbe (COA, 2 %). Nous voyons que chaque type d'erreurs sous la catégorie d'erreurs sémantiques représente une proportion importante, tandis que sous la catégorie d'erreurs grammaticales et syntaxiques, ce sont les types d'erreurs de structures syntaxiques erronées et de prépositions qui méritent notre attention particulière.

Si nous regardons plus précisément les sous-types d'erreurs dans la figure 4.3, nous remarquons que les cinq sous-types d'erreurs les plus fréquentes sont les erreurs de conjugaison et d'accord (CA, 16 %), erreurs de temps/aspect/mode (TAM, 11%), les erreurs de sens inadapté (SIN, 10 %), l'omission syntaxique (OS, 9 %) et les erreurs de synonymes (SNM, 8 %). Leur nombre important nous incite à entrer en détail dans l'analyse qualitative de ces sous-types d'erreurs.

Il est intéressant de noter que si on calcule toutes les erreurs se rapportant aux conjugaisons des verbes, y compris les conjugaisons selon les personnes et selon le temps, l'aspect et le mode, on obtiendra un résultat de 27 % au total, c'est-à-dire que plus d'un quart des erreurs verbales sont relatives aux conjugaisons, ce qui est un chiffre non négligeable. Conformément à notre hypothèse, les erreurs de temps/aspect/mode et de conjugaison et d'accord sont très importantes chez les apprenants sinophones. Les verbes en chinois sont invariables et nous pouvons supposer que cela renforce la difficulté pour les apprenants sinophones à maîtriser les accords des verbes avec les personnes ou les TAM. Par ailleurs, nous avons remarqué que le lemme *pouvoir* est de 294 occurrences, qui représente 5,1 % de tous les verbes. Il est le deuxième verbe le plus utilisé, juste après le lemme *être* (957

occurrences, 16,7 %). À travers l'analyse de notre corpus avec Anatext¹³ (Kraif, 2019), nous avons constaté que la fréquence du lemme *pouvoir* est deux fois plus haute que celle dans le corpus de référence, ce qui montre un suremploi du verbe *pouvoir* dans la rédaction des apprenants. Ce recours excessif peut être considéré comme une stratégie des apprenants sinophones pour éviter de conjuguer les autres verbes.

En outre, le nombre important d'erreurs de prépositions confirme notre hypothèse émise pendant l'élaboration de la grille d'erreurs. Selon cette hypothèse, les prépositions régies par les verbes constituent une difficulté essentielle des apprenants sinophones. Ceci s'expliquerait par plusieurs raisons : d'une part, il n'existe pas de verbes transitifs indirects en chinois, d'autre part, les prépositions régies par les verbes en français sont généralement peu motivées et l'acquisition se réalise d'une manière mécanique, ce qui entraîne facilement des confusions.

Après avoir un aperçu général des erreurs, nous procédons maintenant à l'analyse quantitative et qualitative des erreurs par catégories d'erreurs.

4.1.2 Au niveau morphologique (MORPHO)

Les erreurs au niveau morphologique comportent les erreurs de conjugaison et d'accord, les erreurs d'orthographe d'usage, et les erreurs de verbes fictifs. Nous montrons leur proportion par la figure suivante :

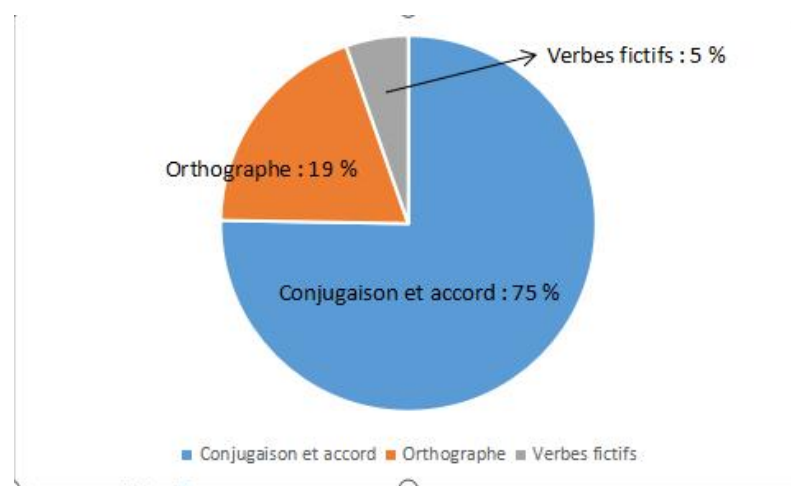


Figure 4.4 : La répartition des erreurs au niveau morphologique

¹³ <http://phraseotext.univ-grenoble-alpes.fr/anaText/index.php>

Nous voyons que les erreurs de conjugaison et d'accord représentent trois quarts des erreurs morphologiques. Les erreurs d'orthographe représentent un cinquième et les erreurs de verbes fictifs sont peu présentes.

4.1.2.1 Erreurs de conjugaison et d'accord (CA)

Parmi les erreurs de conjugaison et d'accord, nous avons relevé 16 erreurs de conjugaison et 124 erreurs d'accord. Pour les verbes du premier groupe dont les conjugaisons sont régulières, nous avons constatons plusieurs erreurs concernant l'omission de *e* à la fin de la première ou la troisième personne du singulier (60). Par exemple :

(60) *Mais dans le résumé, le locuteur **présent** (correction proposée : présente) aussi l'évaluation que on va dire plus tard pour montrer le but et l'intérêt du récit.*

Outre quelques erreurs de verbes du premier groupe causées peut-être par l'inattention, les verbes mal conjugués concernent en particulier ceux du deuxième et troisième groupe (61). Il existe également un transfert des règles de conjugaison des verbes du deuxième groupe à ceux du troisième groupe (62). Dans l'exemple 62, le verbe *suffire* se termine par *-re* et il relève du troisième groupe, mais il est évident que le scripteur l'a conjugué selon les règles de conjugaison du deuxième groupe.

(61) *Le narrateur **choit** (correction proposée : choisit) de faire une explication supplémentaire qui, néanmoins, ne se consacre pas au déroulement.*

(62) *Par conséquent, l'absence d'accident, de maladie professionnels et un bon salaire ne **suffissent** (correction proposée : suffisent) plus à le définir.*

Parmi les 16 erreurs de conjugaison, il y en a 8 qui ne nuisent pas à la prononciation (63), comme *recoure* à la place de *recourt*, *pris* à la place de *prie*, *résolut* à la place de *résolu*.

(63) *Je **croie** (correction proposée : crois) que ce concert gratuit peut attirer beaucoup de gens des autres villes venir ici (correction proposée : attirer beaucoup de gens.)*

Ce constat nous rappelle la complexité d'orthographe mentionnée dans l'étude de Ying (2013). L'écriture du français est caractérisée par la co-existence du système phonographique et idéographique ou logographique. C'est un système phonographique, car les unités de l'écrit transcrivent les éléments constitutifs de la chaîne sonore, mais c'est aussi un système idéographique ou logographique, car l'écriture du français comprend également les lettres muettes qui indiquent des informations grammaticales et sémantiques, comme le marqueur du pluriel « s » et celui de désinence verbale du troisième personne du pluriel « ent ». La double dimension du système d'écriture du français renforce la difficulté d'acquisition de l'orthographe. Par ailleurs, les phonèmes et les graphèmes n'entretiennent pas de relation

biunivoque : la plupart des phonèmes peuvent être écrits de plusieurs façons, par exemple : le code oral [e] peut être transcrit sous plusieurs formes graphiques : « é », « er », « es », ce qui complique la mémorisation de l'orthographe par l'accès phonologique (Ying, 2013, p.129). L'existence des lettres muettes ajoute sans doute les difficultés de l'orthographe. Une moitié d'erreurs de conjugaison (63) et d'erreurs d'orthographe (64) relevées ne nuisent pas à la prononciation le montre.

(64) *Nous vous persuadons de ne pas supprimer (correction proposée : supprimer) cet évènement et de ne pas nous arrêter de nous plaire.*

Pour les erreurs d'accord, nous pouvons observer que ce sont l'accord du verbe avec le sujet et l'accord du participe passé qui posent le plus de problèmes aux apprenants (cf. Tableau 4.5).

L'accord du verbe avec sujet	L'accord de l'auxiliaire	L'accord du participe passé	L'accord du pronom dans les pronominaux
80	6	35	3

Tableau 4.5 : La répartition des erreurs d'accord

Pour l'accord sujet-verbe, nous avons constaté deux cas où les erreurs d'accord sont plus susceptibles d'apparaître : l'un dans les phrases avec deux noms abstraits comme sujet (65), et l'autre est dans les phrases où le sujet est postposé au verbe (66 et 67).

(65) *La rapidité des réponses et l'efficacité est (correction proposée : sont) aussi un autre facteur attractif (correction proposée : d'autres facteurs attractifs) de téléconseil pour les malades en cas d'urgence.*

(66) *C'est (correction proposée : ce sont) les corrélateurs qui sont consacrés à relier deux événements réellement arrivés.*

(67) *Dans la colonne verticale de droite se trouve (correction proposée : se trouvent) les vidéos concernées.*

Pour le cas de deux noms abstraits, le scripteur suppose peut-être que les noms abstraits sont non comptables, il faut donc utiliser le singulier. Cependant, en français, la superposition des noms abstraits conduit au nombre pluriel.

Quant à l'exemple 66, parmi les 6 occurrences de la construction impersonnelle *c'est/ce sont...*, 3 sont erronés à cause de l'accord du verbe *être*. Généralement, le verbe *être* s'accorde avec le sujet apparent *ce*, Cependant, dans le cas où le sujet réel, à savoir le substantif situé après le verbe *être*, est un nom ou un groupe nominal de la troisième personne

du pluriel, le verbe *être* s'accorde avec son sujet réel à la troisième personne du pluriel. Nous pouvons supposer que le scripteur est trop habitué à la construction *c'est...* qu'il oublie d'accorder le verbe *être* au groupe de mots situé après. Il est aussi possible que le scripteur apprenne cette construction d'une façon mécanique et qu'il ne connaisse pas la variation possible. Il faut souligner ce point dans le cours du FLE. L'inversion du sujet-verbe illustrée par l'exemple 67 est un cas similaire auquel les apprenants doivent faire attention.

Pour l'accord du participe passé, nous avons relevé 15 erreurs se rapportant à l'accord du participe passé avec les compléments d'objets directs (COD) placés avant le verbe, qui représentent presque une moitié de toutes les erreurs d'accord du participe passé.

(68) *La manière que nous avons adopté (correction proposée : *adoptée*) ici pour calculer cette distance est « distance euclidienne ».*

(69) *Elle nous a d'abord laissé (correction proposée : *laissés*) marcher sur la scène et elle avec nous.*

Il faut faire attention aux situations où l'accord du participe passé avec les COD aura lieu, comme dans les propositions relatives introduites par « que » (68) ou dans l'utilisation des pronoms personnels compléments d'objet directs (69).

4.1.2.2 Erreurs de verbes fictifs (VFIC)

Pour les verbes fictifs, nous avons remarqué que l'apprenant crée par néologisme des signifiants inexistants, souvent à partir de racines nominales. Par exemple :

(70) *Également, ils sont capables d'exhibitionner (correction proposée : *montrer*) leur talent musical sur scène.*

Dans l'exemple 70, le verbe *exhibitionner* n'existe pas en français. Nous arrivons à inférer le sens cible du scripteur parce que le verbe *exhibitionner* ressemble au nom anglais *exhibition*, qui désigne 'une exposition' en français. Le verbe correspondant en anglais est *exhibit*, dont l'équivalent en français est *exposer*. Donc nous inférons que le sens cible est de 'montrer le talent au public'. Vu que les verbes en français se terminent souvent avec le suffixe *er*, nous pouvons supposer que cette erreur vient d'une influence de l'anglais et une surgénéralisation de la règle concernant la formation des verbes.

4.1.2.3 Conclusion provisoire

En somme, les erreurs de conjugaison et d'accord sont les plus importantes dans les erreurs morphologiques. Il faut sensibiliser les apprenants aux conjugaisons des verbes du deuxième et troisième groupe, aux accords du verbe aux noms abstraits en tant que sujets, au

sujet situé derrière le verbe, et au complément d'objet direct placé avant le verbe. La complexité du système d'écriture du français expliquerait certaines d'erreurs d'orthographe et de conjugaison.

4.1.3 Au niveau sémantique et des cooccurrences lexicales (SEMAN)

Les erreurs au niveau sémantique se composent d'erreurs de sens, d'erreurs temps/aspect/mode (TAM) et d'erreurs de cooccurrences (cf. Figure 4.6).

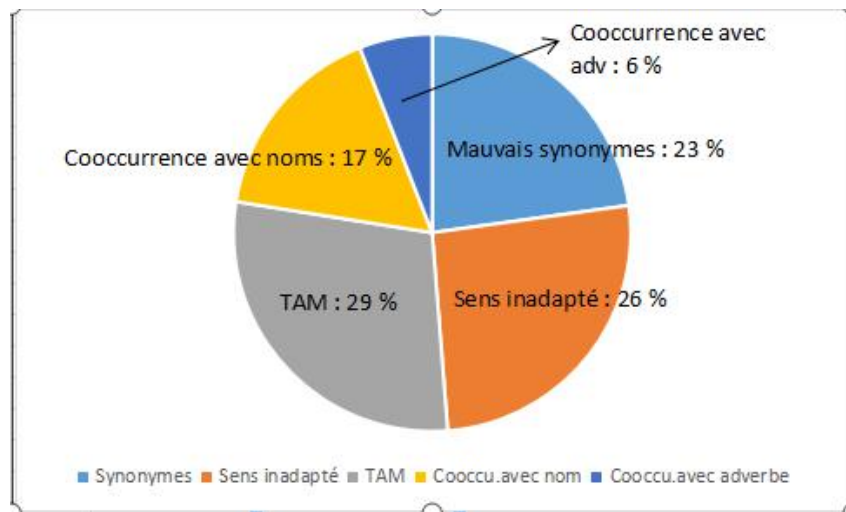


Figure 4.6 : La répartition des erreurs au niveau sémantique et des cooccurrences lexicales

Nous voyons que les erreurs de sens lexical représentent la moitié des erreurs sémantiques, tandis que les erreurs de TAM et les erreurs de cooccurrences en représentent respectivement environ un quart. Cela montre que le sens lexical, le temps/aspect/mode et la combinatoire lexicale sont des composantes indispensables pour l'appropriation du sens d'un verbe au contexte.

4.1.3.1 Erreurs de sens lexical (SEN)

Les erreurs de synonymes inadaptés sont liées aux verbes polysémiques, qui causeraient des difficultés dans l'apprentissage. Par exemple :

(71) Cette manière **présente** (correction proposée : *montre*) que pour un ensemble de mots des textes A et B, on peut calculer la distance entre deux textes.

(72) De plus, à l'aide du LPC, un support visuel avec la main **montrant** (correction proposée : *indiquant*) chaque son et enlève la confusion, les enfants sourds peuvent entendre par la vue.

Pour l'exemple 71, Les verbes *présenter* et *montrer* sont des verbes très utilisés et possèdent tous plusieurs acceptions. Ils sont synonymes dans le sens de ‘ faire voir quelque chose à quelqu’un en mettant sous les yeux’ et ‘faire connaître quelque chose au public en l’offrant aux regards ou par une démonstration, une manifestation exceptionnelles’. Mais seulement le verbe *montrer* possède le sens de ‘faire constater quelque chose’ exigé par le contexte. Pour l'exemple 72, le verbe *montrer* et *indiquer* sont synonymes dans le sens de ‘désigner quelque chose par un geste, un signe précis’. Cela renvoie à une action concrète, tandis que dans la phrase *le son* est un objet abstrait. Pour exprimer le sens de ‘dénoter, en être signe’ exigé, seulement le verbe *désigner* qui peut s’employer dans un sens abstrait convient au contexte.

Quant aux erreurs de sens inadapté, nous avons remarqué une confusion entre les verbes montrant une ressemblance morphologique, surtout pour les paires verbales qui diffèrent seulement par les préfixes. Par exemple :

(73) *Je me permet de vous bien vouloir examiner la proposition consistant à faire **endurer** (durer) le festival musical à la fin de l’été.*

(74) *On peut **dépasser** (passer) ce jour comme un fête, même si on aime plus la ville grâce au concert.*

Il nous semble que la distinction entre certains verbes et leur forme avec préfixe constitue une difficulté pour les apprenants sinophones. Nous avons relevé plusieurs confusions entre les verbes *jouir* et *réjouir*, *présenter* et *représenter*, *s’unir* et *se réunir*. Il faut faire prendre en conscience la distinction entre les paires verbales aux apprenants.

4.1.3.2 Erreurs de temps/aspect/mode (TAM)

À travers la répartition des erreurs de temps/aspect/mode (cf. Tableau 4.7), nous voyons que les erreurs de mode sont plus importantes, en particulier celles concernant le subjonctif, l’infinitif à la place des verbes conjugués ou inversement, et le participe présent et passé.

Temps	Aspects	Mode			
		Subjonctif	Infinitif	Participe	Les autres
13	2	18	28	15	16

Tableau 4.7 : La répartition des erreurs de temps/aspect/mode

Nous constatons que les apprenants ne commettent pas beaucoup d'erreurs de temps, ce qui montre qu'ils maîtrisent bien les concepts du « passé », « présent » et « futur » en français. Cependant, il est étonnant de remarquer un suremploi du temps du passé dans la rédaction des apprenants. Parmi les 13 erreurs de temps, il en a 8 qui sont liées avec le temps du passé à la place du temps du présent ou du futur. Par exemple :

(75) *Généralement, le récit a suivi (correction proposée : *suit*) une structure précise.*

(76) *L'expression utilisée « perdre la relation » favorise en quelque sorte une dramatisation de situation comme si dans (correction proposée : à) l'avenir, il n'y avait (correction proposée : *aurait*) plus de lecteurs lisant les articles issus des médias traditionnellement informatives comme *Le Monde*.*

Le suremploi du passé composé pourrait s'expliquer par une auto-correction du scripteur, car le passé composé est très fréquent en français (75). Nous supposons que la confusion entre l'imparfait et le conditionnel est attribuable à une ressemblance de leur conjugaison (76).

Nous avons remarqué plusieurs erreurs d'omission de conjugaison des verbes, qui apparaissent fréquemment dans l'inversion du sujet et du verbe (77). Rappelons que c'est aussi un cas soumis souvent aux erreurs d'accord. L'inversion du sujet et du verbe mérite donc d'être soulignée dans le cours du FLE.

(77) *En outre, en ce qui concerne les outils culturels et technologiques, les auteurs n'expliquent pas en détail comment fonctionner (correction proposée : *fonctionnent*) ces outils.*

Par ailleurs, les verbes conjugués à la place des verbes à l'infinitif concernent souvent l'utilisation des prépositions comme *pour*, *sans* (78), et dans les constructions verbales comme *pouvoir faire quelque chose* (79), *permettre à quelqu'un de faire quelque chose* (80). Il faut utiliser l'infinitif après les prépositions. C'est une règle facile à maîtriser, mais nous avons quand même relevé 15 erreurs de ce type. Nous pouvons voir que, dans l'exemple 79, l'apprenant connaît sans doute la règle qu'il faut utiliser le verbe à l'infinitif après *pouvoir*, car il l'utilise pour le premier verbe *tirer*. Nous supposons que l'erreur du deuxième verbe juxtaposé *partager* est plutôt causée par l'inattention.

(78) *Pour ajouter la force à cette simplicité sans la compliquer (correction proposée : *sans la compliquer*), on a 4 outils principaux distingués par leurs syntaxes divers.*

(79) *En dehors de ces trois domaines exigeants, je crois que la vie est un travail en progrès donc on peut tirer des leçons de ses erreurs et les partagent (correction proposée : *partager*) avec les autres pour transformer le droit aux erreurs en l'amélioration collective.*

(80) *L'évaluation par le fait est réalisée par la description de ce que les personnages ont fait, qui permet l'auditeur de se situe (correction proposée : *se situer*) dans une même situation et de produire les sentiments similaires.*

Nous avons remarqué un sous-emploi du subjonctif, car parmi les 18 erreurs du subjonctif, seulement 2 concernent le subjonctif à la place des autres modes, tout le reste est le cas inverse. Par exemple :

(81) *Bien que le développement de LSC peut (correction proposée : *puisse*) prendre les expériences des pays étrangers pour (correction proposée : *comme*) référence après 1840, la LSC a commencé à se développer seulement dans les années 1950, ce qui fait d'elle une langue récente.*

(82) *Et en pratique, il est ordinaire qu'une ou deux propositions qui résument toute l'histoire sont (correction proposée : *soient*) utilisées comme un commencement.*

L'usage du subjonctif nécessite à la fois une analyse syntaxique et une analyse sémantique. D'une part, dans les propositions introduites par les conjonctions comme *bien que, jusqu'à ce que, avant que*, il faut utiliser le subjonctif (81). D'autre part, le subjonctif est employé pour exprimer la subjectivité, un souhait, une incertitude, ou une éventualité. Comme dans l'exemple 82, *il est ordinaire que* exprime un avis du scripteur et exprime une subjectivité. Il faut donc utiliser le subjonctif. Nous avons constaté que les subjonctifs bien utilisés dans le corpus sont plutôt introduits par les conjonctions mentionnées ci-dessus, tandis que les erreurs du subjonctif apparaissent souvent dans les cas où il nécessite une analyse sémantique. Il y a lieu de croire que les apprenants retiennent l'usage du subjonctif d'une manière mécanique, mais qu'ils ne maîtrisent pas très bien la valeur sémantique exprimée par le subjonctif.

Les erreurs du participe se rapportent à la confusion entre l'indicatif, le participe présent et le participe passé. Comme nous le savons, l'indicatif fonctionne comme un verbe, et le participe pourrait s'employer comme des adjectifs. Par exemple :

(83) *Sa structure est composée de 8 éléments : conjonction de coordination et adverbe de temps (l'adverbe de phrase), sujet simple (syntagme nominale sujet), et des éléments correspondent (correction proposée : *correspondant*) au syntagme verbale.*

(84) *On trouvera une diversité d'articles lié au domaine médiatique classant (correction proposée : *classés*) selon la date de publication.*

Dans l'exemple 83, après les deux points, ce sont des substantifs juxtaposés, donc *correspondre* doit être un complément du nom *éléments*. Étant donné que le nom *éléments* est l'agent du verbe *correspondre*, il faut utiliser le participe présent *correspondant*. Dans l'exemple 84, le participe présent *classant* fonctionne en tant qu'un complément du nom

articles, mais *articles* est l'objet direct du verbe *classer*, il faut donc utiliser le participe passé *classés*, en s'accordant avec le nombre du nom *activités*, pour exprimer leur relation passive.

4.1.3.3 Erreurs de cooccurrences (COC)

Parmi les 53 erreurs de cooccurrences du type Verbe-Nom, seulement 2 sont les erreurs de noms, et toutes les autres concernent le verbe. Ce n'est pas surprenant, car pour les cooccurrences Verbe-Nom, surtout pour les collocations, le nom est souvent la base qui garde son sens normal, et le verbe est le collocatif qui est choisi en fonction du nom, d'où la possibilité de mal choisir le verbe. Anctil (2010) a aussi discuté de l'utilisation d'un mauvais collocatif dans les collocations. Il a proposé que le collocatif de sens vide est plus possiblement soumis aux erreurs. Par exemple, dans la phrase *Les publicitaires portent une très mauvaise influence sur les jeunes dans la société*, le verbe *porter* ne peut pas se combiner avec le nom *influence*. Anctil a supposé que l'apprenant natif a commis cette erreur afin d'éviter les verbes *avoir*, *être*, *faire* dans son texte. Pour les apprenants non natifs, nous avons également remarqué que les verbes erronés sont souvent ceux de sens général ou vide, par exemple :

(85) *Le lien entre impératif et futur **est** (correction proposée : porte) sur l'événements non réalisés, sous-entendus, mis en balance avec le contenu de l'énoncé.*

(86) *En concurrence, cette pratique risque d'**avoir** (correction proposée : présenter) un vide en matière de la responsabilité et peut glisser vers le dérive.*

(87) *Dans le paragraphe 6 et 7, l'auteur nous **fait** (correction proposée : donner) un exemple: on écrit un texte sur ordinateur, mais brutalement l'ordinateur ne marche plus.*

Dans les trois exemples, le verbe *être*, *avoir* et *faire* sont très fréquents et ne contribuent pas beaucoup à la construction du sens. C'est pourquoi leur utilisation erronée ne nous empêche pas trop d'accéder au sens cible, mais trahit le statut d'apprenants non natifs.

Dans les 19 erreurs de verbe en cooccurrence avec l'adverbe ou le complément circonstanciel, il y en a 6 qui se rapportent à l'utilisation de l'adjectif comme l'adverbe, en particulier pour l'adjectif *profond* (88). Nous pouvons supposer que les apprenants ne connaissent pas, ou ne sont pas très familiers avec sa forme adverbiale *profondément*.

(88) *Basée sur (correction proposée : En se basant sur) l'étude antérieure, on considère plus **profond** (correction proposée : réfléchit plus profondément) sur les éléments d'évaluation avec le degré d'élaboration syntaxique, afin d'analyser le style et la force au sein de la culture vernaculaire noire-américaine.*

Les erreurs de cooccurrences représentent 23 % des erreurs sémantiques. Cette présence non négligeable semble confirmer le constat de nombreuses études (Bolly 2011, Yan 2017), selon lesquelles ce type d'erreurs constitue des déficits « résistants » propres aux apprenants non-natifs.

4.1.3.4 Conclusion provisoire

Dans cette partie, nous voyons que chaque type d'erreurs sous la catégorie sémantique représente une proportion non négligeable, surtout les erreurs de sens lexical. Les erreurs de synonymes concernent souvent les verbes polysémiques, ce qui nous montre qu'il faut prêter attention aux nuances entre les synonymes et aux circonstances dans lesquelles les synonymes sont remplaçables. Les erreurs de sens inadapté attestent une influence de la ressemblance morphologique, surtout pour les verbes qui diffèrent seulement par des préfixes. Il faut donc sensibiliser les apprenants à la distinction entre les formes proches. Les erreurs de TAM reflètent une maîtrise insuffisante du subjonctif et du participe, surtout en ce qui concerne la valeur sémantique qu'ils véhiculent. De plus, les phrases où le sujet se trouve derrière le verbe sont souvent soumises aux erreurs d'omission de conjugaison des verbes. Nous avons aussi remarqué que les apprenants ont tendance à employer les verbes généraux dans les cooccurrences Verbe-Nom, car ils sont encore en phase d'acquisition des combinaisons lexicales. Il faut également sensibiliser les apprenants aux éléments phraséologiques, comme les collocations.

4.1.4 Au niveau de la grammaire et de la syntaxe (GRAMSYN)

Nous avons distingué six types d'erreurs grammaticales et syntaxiques : les erreurs de voix, les erreurs de pronominaux, les erreurs d'ordre, les erreurs de prépositions, les erreurs de pronoms, et les erreurs de structures syntaxiques (cf. Figure 4.8).

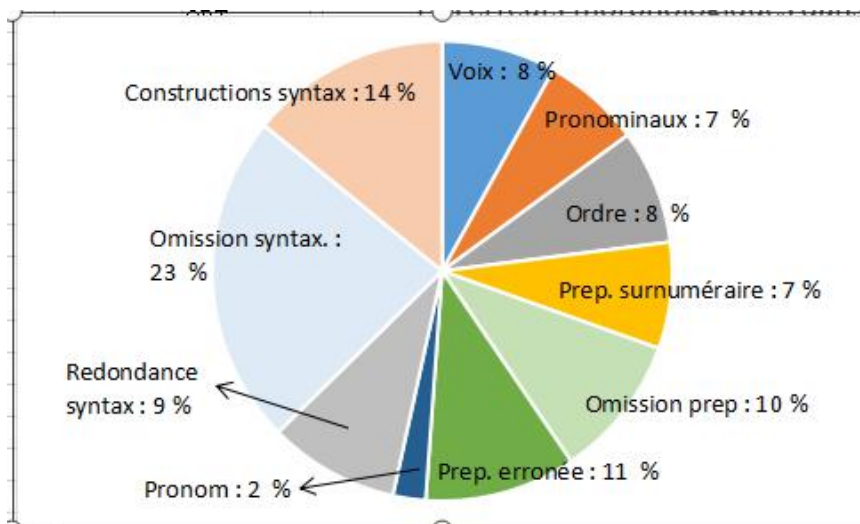


Figure 4.8 : La répartition des erreurs au niveau de la grammaire et de la syntaxe

Comme nous pouvons voir, les erreurs de structures syntaxiques représentent presque une moitié des erreurs grammaticales et syntaxiques, en particulier les erreurs d'omission syntaxique avec une proportion de 23 %. Les erreurs de prépositions représentent un quart des erreurs grammaticales et syntaxiques, qui méritent également notre attention.

4.1.4.1 Erreurs de voix et de pronominaux (VV et EPN)

Pour les 28 erreurs de voix, seulement 8 se rapportent à la voix active à la place du passif ou inversement, les autres concernent les pronominaux de sens passif. Si nous calculons les erreurs de pronominaux de sens passif avec les autres erreurs de pronominaux, à savoir les pronominaux de sens réfléchi, réciproque et absolu, nous obtiendrons une proportion de 13 % des erreurs grammaticales et syntaxiques. C'est une proportion importante qui montre la difficulté des apprenants face aux verbes pronominaux. D'une part, les verbes pronominaux sont très compliqués avec les quatre types d'emploi. D'autre part, il n'existe pas de constructions équivalentes en chinois ou en anglais. C'est un point grammatical tout neuf pour les apprenants sinophones qui constitue donc une spécificité dans l'apprentissage du français. Nous avons constaté que les erreurs de pronominaux concernent davantage les verbes qui possèdent à la fois la forme non pronominale et la forme pronominale. Pour les erreurs sur l'accord du pronom avec le sujet des verbes pronominaux, nous rejoignons l'idée de Ying (2013) en pensant que les apprenants sont si habitués à la mémorisation des pronominaux par *se + verbe à l'infinitif* qu'ils négligent parfois l'accord du pronom *se*. En revanche, les erreurs de la redondance ou l'absence du pronom dans les pronominaux reflètent une confusion sur la fonction du pronom. Pour les verbes qui ne possèdent pas la forme

pronominale ou pour ceux qui n'existent que sous forme pronominale, les apprenants pourraient les retenir par cœur, mais pour les verbes qui possèdent à la fois la forme non pronominale et la forme pronominale, il faut vraiment faire attention à analyser la fonction du pronom pendant l'apprentissage. Par exemple :

(89) *Dans ce type d'évaluation, le narrateur **s'arrête** (correction proposée : *arrête*) l'action d'événement pour attirer l'attention de l'auditeur.*

(90) *Les intensificateurs sont les événements qui veut mettre l'accent sur les propositions, y compris les gestes qui **ajoutent** (correction proposée : *s'ajoutent*) à la conversation.*

Pour la redondance du pronom comme l'exemple 89, on pourrait supposer que les apprenants retiennent le pronominal *s'arrêter* comme un ensemble sans l'analyser. Le pronominal *s'arrêter* peut être de sens réfléchi ou de sens passif si l'objet du verbe *arrêter* est le même que le sujet. Mais pour l'exemple 89, l'objet du verbe *arrêter* est *l'action d'événement* au lieu du sujet *le narrateur*, donc le pronom *se* ne fait pas de sens ici. Dans l'exemple 90, le verbe *ajouter* est un verbe transitif direct, mais son complément d'objet direct n'est pas présent dans la phrase. Nous pouvons inférer que le sens cible est 'on ajoute les gestes à la conversation', donc 'les gestes sont ajoutés à la conversation'. Le pronominal de sens passif *s'ajouter* est donc adapté au contexte, car son sujet est aussi son complément d'objet direct.

4.1.4.2 Erreurs d'ordre des mots (ORD)

Parmi les 28 erreurs d'ordre des mots, 10 se rapportent à l'ordre du verbe-adverbe, en particulier quand le verbe est au temps composé ou à l'infinitif. Par exemple :

(91) *L'anglais des signes (Signed English) est une langue des signes basée sur la grammaire anglaise et utilisant le vocabulaire gestuel. **Elle principalement est utilisée** (correction proposée : *elle est principalement utilisée*) dans l'enseignement de l'anglais pour aider les élèves sourds.*

(92) *L'auteur le cite pour **concrètement expliquer** (correction proposée : *expliquer concrètement*) à la fois des aspects négatifs et l'effet du succès.*

L'adverbe se place après le verbe conjugué au temps simple ou le verbe à l'infinitif, et entre l'auxiliaire et le participe passé au temps composé. Il faut sensibiliser les apprenants à cette différence d'usage. Les erreurs où l'adverbe est placé avant le verbe pourraient également s'expliquer par l'influence du chinois. Nous reviendrons sur ce point dans la section 4.3.3.2.

4.1.4.3 Erreurs de prépositions (PREP)

Pour les erreurs de prépositions, nous voyons que l'emploi erroné de prépositions est significatif. Nous avons relevé 6 erreurs concernant la confusion entre *par* et *de* à la voix passive parmi les 37 erreurs de prépositions erronées. Par exemple :

(93) *L'attache complexe est composée **par** (correction proposée : de) deux propositions enchâssées au même point dans la phrase matrice.*

Comme nous le savons, pour le passif, l'introduction de l'agent s'effectue à l'aide de la préposition *par* ou *de*. La plupart de verbes au passif sont suivis de *par* pour introduire l'agent, mais il existe une minorité de verbes qui sont suivis de *de*, souvent les verbes traduisant une émotion ou les verbes d'état. Pour l'exemple 93, le verbe *composer* est l'un des cas minoritaires qui s'emploie dans les constructions *A compose B*, *B est composé de A*.

Nous constatons aussi une confusion sur les verbes qui peuvent se combiner avec plusieurs prépositions, mais avec des sens différents, comme dans l'exemple 94. Le verbe *servir* peut être suivi de *à* ou de *de*. *Servir à* signifie 'être d'un emploi ou d'un recours utiles à une fonction', et souvent s'emploie dans la construction *quelque chose A sert à faire quelque chose B*. *Servir de* signifie 'avoir la fonction de quelque chose pour quelqu'un', souvent s'emploie dans la construction *quelque chose A sert de quelque chose B pour quelqu'un*. Dans l'exemple, nous pouvons inférer que le sens cible est 'les outils du passé sont utilisés comme références pour étudier les outils du présent', il est plus proche de la construction *servir de*.

(94) *Les différents outils pris dans la recherche muséale dans le passé peuvent également servir **aux** (correction proposée : de) références pour nous aider à étudier les outils actuels.*

D'une manière générale, les erreurs se concentrent sur les prépositions *de*, *à* et *pour*. Comme nous le savons, les prépositions possèdent aussi une valeur sémantique. Par exemple, *pour* exprime souvent un but, *à* désigne souvent une destination. Cependant, les prépositions régies par les verbes ne sont pas toujours motivées. Par exemple :

(95) *Le narrateur raconte son histoire au niveau d'intermédiaires et présente au moment des événement, c'est-à-dire utiliser le présent **à** (correction proposée : pour) raconter son histoire.*

(96) *La distance statistique entre deux distributions peut nous aider **pour** (correction proposée : à) identifier la ressemblance intertextuelle.*

(97) *Par exemple les liens à droite qui nous **guident aux** (correction proposée : vers) autres vidéos concernées.*

Dans l'exemple 95, il faut employer la préposition *pour* afin d'exprimer un but. L'utilisation de la préposition *pour* dans la construction verbale *utiliser quelque chose pour faire quelque chose* s'avère donc logique dans une certaine mesure. En revanche, dans l'exemple 96, il y a lieu de croire que le scripteur utilise *pour* afin d'introduire un but, mais l'usage du verbe *aider* est *aider quelqu'un à faire quelque chose*. L'exemple 97 est un cas similaire, le scripteur aurait utilisé *à* pour exprimer une destination, mais le verbe *guider* contrôle la préposition *vers*. Nous voyons que le choix des prépositions est souvent peu prédictible, comme ce que montrent les exemples 96 et 97. Cela rend son acquisition difficile. Plusieurs chercheurs ont proposé de réaliser l'apprentissage des prépositions régies par les verbes à travers la mémorisation des constructions verbales (Yan 2017, Ying 2013). Par exemple, au lieu de retenir le verbe isolé *aider*, il vaut mieux mémoriser *aider quelqu'un à faire quelque chose*. On pourrait recourir au sens des prépositions afin de faciliter la mémorisation, mais il faut accorder plus d'attention au cas où le sens de la préposition n'explique pas son intégration dans la construction verbale, comme ce que montrent les exemples 96 et 97.

4.1.4.4 Erreurs de structures syntaxiques (SYN)

Voici la répartition des erreurs de redondance syntaxique (cf. Tableau 4.9).

Redondance du pronom relatif « qui » ou « que »	Redondance du complément	Répétition Paraphrase Pléonasme	Formes pronominales inexistantes
7	5	16	4

Tableau 4.9 : La répartition des erreurs de redondance syntaxique

Il y a 3 erreurs de redondance du pronom relatif dans le corpus B2 et 4 dans le corpus B1 malgré sa petite taille. Par exemple :

(98) *La femme qui porte un turban jaune* (correction proposée : *La femme porte un turban jaune*).

(99) *La femme d'une quarantaine d'années qui est plutôt bien enrobée* (correction proposée : *La femme d'une quarantaine d'années est un peu grosse*).

(100) *L'évaluation par le fait est plus dramatique que celles de précédentes et qui s'intéresse* (correction proposée : *et s'intéresse*) *plutôt à ce que le narrateur à fait que(qu'à) ce qu'il a dit dans l'histoire.*

Les exemples 98 et 99 sont extraits du corpus B1. Nous pouvons constater que les apprenants de niveau B1 ne distinguent pas très bien le verbe principal et le verbe subordonné. Dans l'exemple 100, le scripteur ne s'est pas rendu compte qu'avec la conjonction de coordination *et*, ce sont deux propositions juxtaposées, et le *qui* ne fonctionne pas ici.

Nous voyons que les erreurs de répétition, de paraphrases et de pléonasme représentent une moitié des erreurs de redondance syntaxique. Par exemple :

(101) *Enfin, il y a sans doute qu'il existe (correction proposée : *il y a sans doute/il existe sans doute*) des avantages pour des artiste, ce concert est une bonne chance de faire l'échange.*

(102) *Au 18ème siècle, le système initial de LSF commence à se former, tandis que LSC ne fait du progrès (correction proposée : *progresses*) qu'au 19ème siècle sous l'influence occidentale.*

Dans l'exemple 101, les constructions *il y a* et *il existe* possèdent le même sens. Donc on ne doit garder que l'une des deux. Dans l'exemple 102, la construction *faire des progrès* s'emploie souvent avec un adjectif, comme *faire des progrès notables*, *faire des progrès fulgurants*. C'est le verbe *progresser* qui s'adapte mieux au contexte dans le sens 'avancer, se mouvoir, gagner du terrain'.

Comme nous le voyons, les erreurs d'omission syntaxique sont les plus importantes dans les erreurs de structures syntaxiques. Nous présenterons ci-dessous la répartition des erreurs d'omission syntaxique (cf. Tableau 4.10).

Omission du pronom relatif « qui » ou « que »	Omission du complément (attribut, COD, COI)	Omission du verbe	Omission de la reprise du sujet dans l'inversion sujet-verbe
42	21	12	6

Tableau 4.10 : La répartition des erreurs d'omission syntaxique

L'omission des pronoms relatifs « qui » ou « que » dans les propositions relatives représente presque une moitié d'erreurs d'omission syntaxique. Par exemple :

(103) *Le phonétisme expressif, dans lequel le plus courant est l'allongement des voyelles est superposé (correction proposée : *l'allongement des voyelles qui est superposé*) aux mots de la proposition.*

(104) *Dans une période courte, **il semble un concert gratuit** (correction proposée : *il semble qu'un concert gratuit*) gagne très peu d'argent.*

L'omission des pronoms relatifs pourrait s'expliquer par l'influence du chinois. Nous l'analyserons plus profondément dans la section 4.3.3.3.

Par ailleurs, nous avons aussi remarqué 6 cas d'omission de la reprise du sujet dans l'inversion sujet-verbe parmi les 81 erreurs d'omission syntaxique. C'est un point grammatical à souligner dans le cours du FLE. Par exemple :

(105) *Au niveau de l'échange socio-économique, comment les réseaux sociaux se **présentent** (correction proposée : *se présentent-ils*) ?*

(106) *Pourquoi les jeunes adultes **prennent** (correction proposée : *prennent-ils*) une alimentation néfaste à la santé en sachant clairement l'importance de l'alimentation ?*

Ce type d'erreurs n'apparaît que dans le corpus B2 et nous pourrions supposer que les apprenants de niveau B1 ont le souci d'éviter ces structures compliquées. Il est aussi possible que les phrases interrogatives ou les phrases introduites par les adverbes *peut-être* ou *sans doute* où la reprise du sujet est obligatoire sont peu présentes dans le corpus B1 en raison de son caractère descriptif.

Les erreurs de constructions syntaxiques occupent aussi une place importante dans les erreurs grammaticales et syntaxiques. La plupart des erreurs de constructions syntaxiques renvoient aux constructions verbales inexistantes, dont la correction porte sur la phrase au lieu d'un seul verbe. Par exemple :

(107) *La question **consistant à demander durant la conversation repose sur** le « danger de mort » (correction proposée : *La question centrale dans la conversation porte sur le « danger de mort »*).*

(108) *Avant toute chose, le moment, le lieu, les personnes concernées, leur activité ou leur situation, on les appelle « Les indications », **elles ont besoin de nous pour préciser dans un récit** (correction proposée : *elles sont nécessaires à la poursuite du récit*).*

Nous reviendrons sur les sources possibles de ce type d'erreurs dans la section 4.3.2.2.

4.1.4.5 Conclusion provisoire

Pour résumer cette section, nous avons vu que les erreurs de structures syntaxiques et de prépositions sont les plus importantes dans les erreurs grammaticales et syntaxiques. Les pronominaux constituent une grande difficulté pour les apprenants sinophones, surtout en ce qui concerne les verbes qui possèdent à la fois la forme pronominale et la forme non pronominale. Il faut sensibiliser les apprenants à la fonction du pronom dans les pronominaux.

Pour l'ordre des mots, influencés par leur langue maternelle, les apprenants sinophones ont tendance à placer les adverbes avant les verbes. Les prépositions sont aussi difficiles à manipuler car elles sont peu prédictibles. Les apprenants doivent faire attention à la distinction entre *de* et *par* pour introduire l'agent à la voix passive, et aux verbes qui sont capables de se combiner avec plusieurs prépositions mais avec des sens différents. Nous proposons de prêter plus d'attention aux prépositions, surtout quand on les observe dans les constructions verbales. Les erreurs de redondance syntaxique se concentrent sur les paraphrases, tandis que celles qui concernent l'omission syntaxique se rapportent davantage à l'omission des pronoms relatifs « qui » ou « que ». Nous avons aussi relevé plusieurs points grammaticaux à souligner dans le cours du FLE, comme la reprise du sujet dans l'inversion sujet-verbe, l'adverbe se place derrière le verbe au temps simple ou à l'infinitif, mais entre l'auxiliaire et le participe au temps composé.

4.1.5 Récapitulation

Dans cette section, nous avons observé la répartition des erreurs selon les catégories et les types d'erreurs. À travers l'analyse quantitative, nous voyons que les erreurs au niveau sémantique et syntaxique représentent plus trois quarts d'erreurs au total. Pour les types d'erreurs, les erreurs de constructions syntaxiques (SYN), les erreurs de sens lexical (SEN), les erreurs de conjugaison et d'accord (CA), les erreurs de temps/aspect/mode (TAM) et les erreurs de prépositions (PREP) sont les plus importantes. Nous avons analysé précisément chaque sous-type d'erreurs, surtout les cinq premiers sous-types où les erreurs sont les plus importantes, à savoir les erreurs de conjugaison et d'accord (CA, 16 %), erreurs de temps/aspect/mode (TAM, 11%), les erreurs de sens inadapté (SIN, 10 %), l'omission syntaxique (OS, 9 %) et les erreurs de synonymes (SNM, 8 %). Nous avons résumé les circonstances dans lesquelles les erreurs apparaissent fréquemment à travers une analyse qualitative. Les erreurs de conjugaison concernent surtout les verbes du deuxième et troisième groupe, et les erreurs d'accord apparaissent souvent dans les cas où le sujet se situe après le verbe, où le complément d'objet direct se place avant le participe passé, et où plusieurs noms abstraits fonctionnent en tant que sujet. Les erreurs de temps/aspect/mode concernent plutôt le sous-emploi du subjonctif et l'utilisation erronée du participe. Le sous-emploi du subjonctif apparaît souvent dans les cas où le choix du subjonctif se justifie par sa valeur sémantique, à savoir pour exprimer la subjectivité, l'incertitude ou un souhait. Les erreurs de sens inadapté se rapportent à la proximité morphologique, en particulier pour les verbes qui diffèrent

seulement par les préfixes. L'omission des pronoms relatifs « qui » ou « que » est la plus importante dans l'omission syntaxique, tandis que la redondance syntaxique se concentre sur la répétition et la paraphrase. Les erreurs de synonymes inadaptés sont liées aux verbes polysémiques. Il faut donc faire attention aux circonstances où les synonymes sont remplaçables. Par ailleurs, nous avons aussi constaté une difficulté des apprenants sur les verbes qui possèdent à la fois une forme pronominale et une forme non pronominale. Les points grammaticaux qui posent des problèmes aux apprenants que nous avons relevés pourraient orienter la conception du programme didactique du FLE auprès les apprenants sinophones, comme les apprenants ont l'habitude de placer l'adverbe avant le verbe, et de négliger les pronoms relatifs « qui » ou « que » dans les propositions relatives.

Après avoir analysé les erreurs dans le corpus entier, nous passons maintenant à la comparaison entre les textes écrits sur table et à la maison pour observer l'influence de la condition de rédaction sur les erreurs.

4.2 La comparaison entre les textes écrits sur table et à la maison

Notre corpus se constitue de textes rédigés sur table et à la maison et les deux types ont un nombre de mots presque équivalent (textes sur table 22,584 mots dont 2759 sont les verbes et textes à la maison 27, 430 mots dont 2946 verbes). Il est donc possible de comparer les deux types afin d'observer l'influence de la condition de rédaction sur les erreurs commises. Pour les dossiers rédigés à la maison, parmi les 20 scripteurs, 19 ont recouru au dictionnaire bilingue comme moyen d'aide à la rédaction. De plus, tous les scripteurs ont indiqué dans leurs fiches sociobiographiques que leurs textes avaient été relus par des locuteurs natifs du français. C'est pourquoi nous formulons l'hypothèse qu'il y aura moins d'erreurs dans les textes écrits à la maison, surtout pour les erreurs morphologiques qui sont facilement corrigés à l'aide du dictionnaire et des correcteurs automatiques.

Dans les textes écrits sur table, nous avons relevé 517 erreurs, qui représentent 18,74 % de tous les verbes, tandis que dans ceux écrits à la maison, 340 erreurs représentant 11,54 % de tous les verbes. Avec moins de verbes au total, les textes rédigés sur table présentent plus d'erreurs que ceux à la maison, ce qui confirme notre hypothèse selon laquelle il y a moins d'erreurs dans les textes écrits à la maison. Les conditions de rédaction, comme le recours aux moyens d'aide à la rédaction, la relecture par les francophones, exercent aussi une forte influence sur la qualité de la rédaction.

Nous commencerons par la comparaison aux niveaux morphologique, sémantique et syntaxique, puis nous passerons à la comparaison des types d'erreurs.

4.2.1 La comparaison selon les quatre grandes catégories

Voici la répartition des erreurs selon les quatre catégories dans les textes écrits sur table (cf. Figure 4.11) et dans les devoirs à la maison (cf. Figure 4.12) :

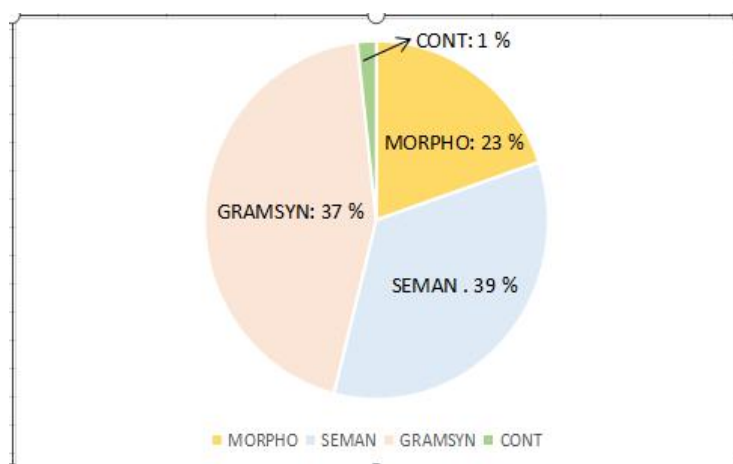


Figure 4.11 : La répartition des erreurs selon les quatre catégories : textes sur table

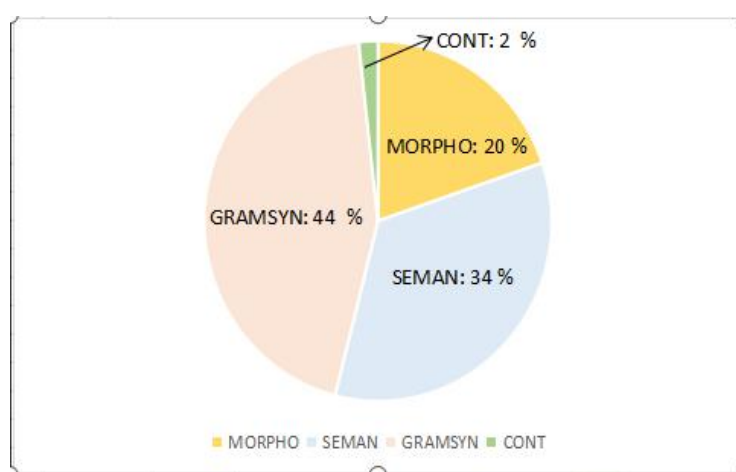


Figure 4.12 : La répartition des erreurs selon les quatre catégories : textes à la maison

Nous constatons que les erreurs morphologiques, conformément à notre hypothèse, sont moins présentes dans les textes écrits à la maison grâce à l'utilisation du dictionnaire. Les erreurs sémantiques occupent aussi une proportion moins importante dans les textes écrits à la maison, ce qui pourrait s'expliquer par la relecture des locuteurs natifs.

En revanche, il n'y a pas de grandes différences pour les erreurs morphologiques et les erreurs sémantiques dans les deux types de textes. Les erreurs de grammaire et de syntaxe varient davantage. Nous pouvons supposer que pendant les examens sur tables, le temps est limité et les apprenants ont tendance à employer des structures syntaxiques plus simples, tandis que pour les devoirs à la maison, il n'y a pas de limite de temps, les apprenants ont plus de volonté d'utiliser les structures plus complexes pour obtenir de meilleurs résultats, ce qui engendrerait davantage d'erreurs au niveau grammatical et syntaxique. Pour vérifier cette hypothèse, nous avons analysé le corpus sur table et le corpus à la maison à l'aide de l'outil Anatext¹⁴. Nous avons remarqué que le nombre moyen de formes par phrase est de 20.6 dans les corpus écrits sur table, et de 24.4 dans ceux rédigés à la maison. Quand les phrases sont plus longues, elles seront plus complexes au niveau syntaxique, et il y aura plus de possibilités de commettre les erreurs de grammaire et de syntaxe. Notre hypothèse est donc validée.

4.2.2 La comparaison au niveau du type d'erreurs

Pour les erreurs morphologiques, si nous regardons plus précisément les trois types d'erreurs, nous constatons une grande différence dans les erreurs d'orthographe d'usage. Les erreurs d'orthographe d'usage représentent une proportion beaucoup plus grande dans les textes écrits sur tables (26 %) que dans les textes écrits à la maison (7 %).

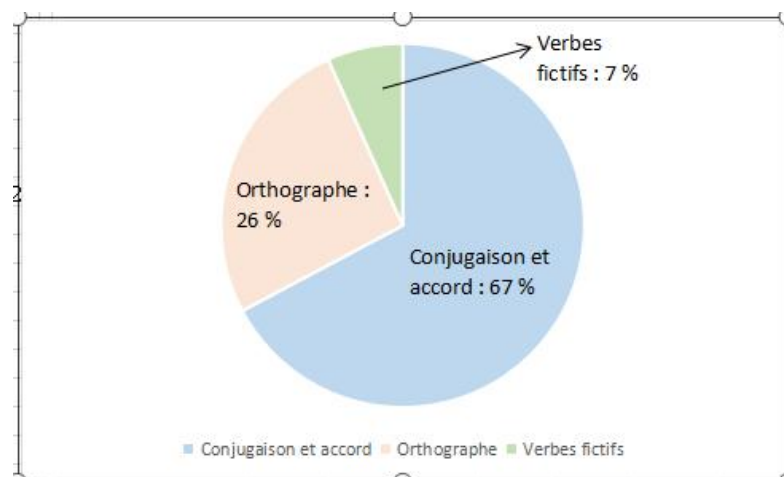


Figure 4.13 : Les erreurs morphologiques dans les textes écrits sur table

¹⁴ Olivier Kraif, laboratoire LIDILEM, université Grenoble-Alpes. Site : <http://phrasotext.univ-grenoble-alpes.fr/anaText/>

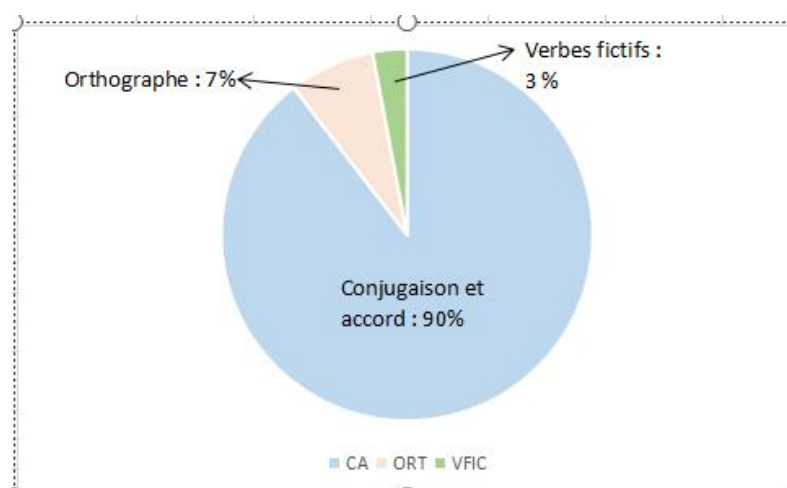


Figure 4.14 : Les erreurs morphologiques dans les textes écrits à la maison

Les devoirs à la maison sont tapés sur l'ordinateur, et comme nous le savons, les outils de rédaction comme Word ou Libre Office intègrent des correcteurs orthographiques, ce qui pourrait expliquer la réduction des erreurs d'orthographe. Il faut mentionner que cinq moyens figurent dans la rubrique « moyens d'aide à la rédaction » des fiches sociobiographiques : dictionnaire bilingue, dictionnaire monolingue, grammaire, traducteur automatique et autres, mais peu d'apprenants ont indiqué les autres moyens potentiels. Nous pouvons regretter que le choix « les correcteurs automatiques » n'ait pas été proposé.

Si nous comparons plus précisément la proportion de chaque type d'erreurs dans les deux types de textes, la différence principale existe dans les erreurs de temps/aspect/mode, qui sont beaucoup plus présentes dans les textes à la maison. Nous supposons que cela pourrait s'expliquer par la nature des textes. Les textes écrits sur table, à savoir les lettres aux maires, les commentaires des chiffres donnés, les descriptions d'une photo, impliquent davantage le temps au présent, tandis que les textes écrits à la maison, y compris les résumés des récits d'expériences personnelles, la présentation des langues des signes et de leur histoire, et la présentation d'une approche pour identifier l'auteur d'un texte mystérieux, font appel à des temps, aspects et modes beaucoup plus variés, ce qui entraînerait plus de possibilités de commettre les erreurs de TAM. Nous n'avons pas constaté d'autres différences importantes au niveau du type d'erreurs entre les deux types de textes.

4.2.3 Conclusion provisoire

En somme, à travers la comparaison entre les textes écrits sur table et à la maison, nous avons révélé qu'il y a moins d'erreurs morphologiques et sémantiques dans les devoirs à la maison, en particulier les erreurs d'orthographe, ce qui pourrait s'expliquer par l'utilisation du dictionnaire, du correcteur orthographique et la relecture des francophones natifs. En revanche, les erreurs grammaticales et syntaxiques, ainsi que les erreurs de temps/aspect/mode se voient augmenter dans les textes à la maison. Nous pouvons supposer que cela est attribuable à la complexité et la diversité des phrases écrites à la maison, car le nombre moyen de formes par phrase est plus important dans les textes à la maison que ceux sur table.

Après avoir regardé la répartition des erreurs, les circonstances dans lesquelles les erreurs apparaissent fréquemment et l'influence de moyens d'aide à la rédaction sur les erreurs, nous procéderons maintenant à l'interprétation des origines possibles d'erreurs.

4.3 L'interprétation des sources d'erreurs

L'identification des sources d'erreurs comporte deux étapes. Tout d'abord, nous avons analysé chaque type d'erreurs et proposé leurs explications possibles. Dans un second temps, nous avons regroupé les explications pour en trouver la régularité. Il y a lieu de croire que certains types d'erreurs sont dus à l'inattention, comme les erreurs d'orthographe d'usage et les erreurs concernant les accords simples, comme l'accord du pronom avec le sujet pour les verbes pronominaux. Mais outre l'inattention, les sources d'erreurs sont variées. Dans la partie suivante, nous essayerons d'expliquer ces sources possibles en les illustrant avec des exemples du corpus.

4.3.1 La déficience dans la connaissance langagière du français

Certains types d'erreurs reflètent une déficience dans la connaissance langagière des apprenants en phase d'acquisition de la langue, comme la maîtrise inachevée des règles grammaticales du français, la surgénéralisation des règles grammaticales, et le manque du vocabulaire approprié.

4.3.1.1 La maîtrise inachevée des règles grammaticales du français

Nous avons remarqué quelques types d'erreurs attribuables à une maîtrise inachevée des règles en français, en particulier celles de conjugaison et d'accord, de temps/aspect/mode, et d'ordre des mots.

4.3.1.1.1 Erreurs de conjugaison et d'accord (CA)

Nous supposons que pour la transgression des règles simples, comme l'accord du pronom avec le sujet pour les verbes pronominaux, la cause pourrait être l'inattention des apprenants, mais pour les règles plus complexes, il se peut que les apprenants ne les maîtrisent pas. Prenons un exemple : nous avons constaté que les erreurs d'accord du verbe avec le sujet apparaissent souvent dans la proposition relative introduite par « qui », « ce qui » ou « que ». Parmi les 140 erreurs de conjugaison et accord, nous avons remarqué 23 erreurs liées à la proposition relative. Par exemple :

(109) *Ce degré d'enchâssement, selon l'auteur, est souvent maîtrisé par les narrateurs avertis, âgés et qui appartient (correction proposée : *appartiennent*) à la working classe traditionnelle.*

(110) *Qui parlent (correction proposée : *parle*) ? Qui ne parlent (correction proposée : *parle*) pas ?*

(111) *Aux yeux d'Aristote, ceux qui ne parle (correction proposée : *parlent*) pas, ne peut (correction proposée : *peuvent*) pas penser.*

(112) *La Télévision centrale de Chine, réseau principal de télévision publique, ainsi que la plupart des chaînes de télévision provinciale ont commencé à lancer des émissions d'information avec des interprétations en LSC, ce qui permettent (correction proposée : *permet*) aux sourds et malentendants de mieux connaître le monde et la société à travers les actualités.*

Le pronom relatif « qui » remplace son antécédent qui s'emploie comme sujet d'une proposition relative. Il prend le genre et le nombre de cet antécédent et le verbe de la proposition relative s'accorde avec l'antécédent. Dans l'exemple 109, « qui » représente *les narrateurs*, donc le verbe doit s'accorder à la troisième personne du pluriel. Cependant, dans une phrase interrogative avec le pronom interrogatif « qui » comme sujet, le verbe reste toujours au singulier. L'exemple 110 correspond à cette règle. Pour les pronoms relatifs avec un pronom démonstratif antécédent, comme « ce qui », « ceux qui », « celles qui », le verbe s'accorde avec le pronom démonstratif. C'est pourquoi dans l'exemple 111, le verbe *parler* doit s'accorder au pluriel tandis que dans l'exemple 112, le verbe *permettre* doit s'accorder au singulier. Ici c'est une règle d'accord complexe, ce qui peut expliquer la difficulté des apprenants.

4.2.3.1.2 Erreurs de temps/aspect/mode (TAM)

Les erreurs relatives au temps, à l'aspect et au mode trahissent une confusion sur la valeur sémantique exprimée par des différents TAM du verbe. La flexion verbale est une caractéristique importante des langues indo-européennes comme le français, mais le chinois est une langue isolante qui n'a pas de flexion verbale. En chinois, le temps, l'aspect et le mode d'une action sont marqués par l'ajout des adverbes, des morphèmes aspectuels ou des ponctuants¹⁵, tandis qu'en français, les verbes assurent la fonction de dénoter le temps, l'aspect et le mode de l'action à travers la conjugaison. C'est la différence principale entre les verbes en français et en chinois. L'absence de conjugaisons de verbes en chinois pourrait expliquer pourquoi la maîtrise des valeurs sémantiques de différents TAM paraît complexe pour les apprenants sinophones (Ying 2013, Yan 2017).

(113) *Puisque la situation financière de notre ville **soit** (n'est) pas si satisfaisant, nous pouvons toujours le (le concert) représente par une réduction du budget.*

(114) *Supposons qu'il y **a** (ait) un concert gratuits, des artistes pourraient être invités pour participer au concert.*

(115) *Tout le monde, y compris les personnes qui sont chanteurs, étudiants, professeurs **étudiés** (étudiant) la musiques, les amateurs, vont au concert.*

Nous proposons deux exemples pour illustrer la distinction entre le subjonctif et l'indicatif. Le subjonctif est un mode utilisé pour exprimer une hypothèse, un doute ou un souhait, tandis que l'indicatif exprime un fait, une certitude, une réalité. Dans l'exemple 113, la conjonction *puisque* introduit une réalité, donc sa proposition est censée être à l'indicatif. Dans l'exemple 114, le verbe *supposer* signifie 'admettre quelque chose comme hypothèse, comme prémisse d'un raisonnement', donc il introduit souvent une proposition au subjonctif. Comme ce que nous avons mentionné dans la section 4.1.3.2, on pourrait penser que les étudiants chinois apprennent le subjonctif sans avoir bien compris leur usage, c'est-à-dire

¹⁵ Les verbes sont invariables en chinois, par exemple :

(1) Phrase en chinois : 我/读/这本/书; transcription phonétique du chinois : wǒ/dú/zhè běn/shū; traduction littérale en français : Je/lire/ce/ livre; phrase en français : *Je lis ce livre.*

(2) Phrase en chinois : 我/读/过/这本/书/了; transcription phonétique en chinois : wǒ/dú/guò/zhè běn/shū/le; traduction littérale en français : Je/lire/morphème du passé/ce/livre/morphème d'accompli; phrase en français : *J'ai lu ce livre.*

(3) Phrase en chinois : 我/将/读/这本书 ; transcription phonétique en chinois : wǒ/jiāng/dú/zhè běn/shū; traduction littérale en français : Je/morphème du futur/lire/ce/livre; phrase en français : *Je vais lire ce livre.*

Nous pouvons voir que le verbe *dú* reste invariable au présent, au passé et au futur. Le temps et l'aspect sont indiqués par des morphèmes, comme *guò* marque le passé, *le* marque une action accomplie, et *jiāng* marque le futur.

qu'ils retiennent par cœur les expressions ou les conjonctions qui introduisent normalement une proposition au subjonctif, comme *bien que*, *malgré que*, *il se peut que*, sans réfléchir plus profondément sur leur motivation sémantique. C'est pourquoi les apprenants se trompent facilement d'expressions et suremployent le subjonctif avec *puisque*.

L'exemple 115 illustre une confusion entre le participe présent et le participe passé. Le participe présent dénote une relation active, et le participe passé est le plus souvent passif. Dans cet exemple, la relation entre *professeurs* et le verbe *étudier* est active, car ce sont les professeurs qui étudient la musique, donc il faut utiliser le participe présent au lieu du participe passé.

4.3.1.1.3 Erreurs d'ordre des mots (ORD)

Certaines erreurs d'ordre des mots reflètent aussi une mauvaise maîtrise des règles concernant l'ordre des mots en français, comme les pronoms se placent avant le verbe ou l'auxiliaire, les adverbes se situent entre l'auxiliaire et le participe passé dans les temps composés, les deux éléments la négation se placent avant le verbe à l'infinitif.

Il est important de noter que les erreurs d'ordre concernant l'utilisation de la construction *faire faire* sont récurrentes dans le corpus. Parmi les 6 constructions *faire faire* que nous avons récupérées, 4 ont subi les erreurs d'ordre des mots. Par exemple :

(116) *Par ailleurs, dans le monde scientifique, on est tenu d'être presque parfait pour ne pas **faire les autres souffrir*** (correction proposée : *faire souffrir les autres*).

(117) *L'évaluation est de (correction proposée : *consiste à*) découvrir au préalable ce qui rend les éléments mémorables et est une façon de **faire l'auditeur comprendre*** (correction proposée : *faire comprendre à l'auditeur*) *pourquoi on le raconte*.

Pour la construction *faire faire*, selon les règles grammaticales, si le verbe à l'infinitif est un verbe intransitif ou transitif sans complément, l'ordre est *faire+verbe à l'infinitif+sujet du verbe à l'infinitif*, si le verbe à l'infinitif est un verbe transitif suivi de complément, l'ordre devient *faire+verbe à l'infinitif+à+sujet du verbe à l'infinitif+complément du verbe à l'infinitif*. C'est une règle relativement facile à manipuler, mais nous avons constaté qu'elle fait quand même l'objet de plusieurs erreurs. Cette difficulté pourrait s'expliquer par le fait que la construction *faire faire* n'a pas d'équivalents en chinois. Nous pourrions aussi supposer que les apprenants sont trop habitués à la logique où le verbe est directement suivi de son complément, donc cette construction qui implique la superposition de deux verbes s'avère

irrégulière pour eux. Il faut accorder plus d'attention à cette construction dans le cours du FLE.

4.3.1.2 La surgénéralisation des règles en français

Comme nous l'avons mentionné dans la théorie d'interlangue (voir la section 2.2), les apprenants formulent sans cesse de nouvelles hypothèses sur les règles de la langue cible en les examinant dans la pratique. Cela entraîne facilement une surgénéralisation des règles acquises. Par exemple, nous avons noté des néologismes inventés à partir des formes verbales existantes :

(118) *La chute **globe** (englobe) toutes les formes, des plus simples aux plus complexes.*

(119) *Pour **profondir** (approfondir) sa recherche, elle se charge aussi d'une collection sur les musées dans monde entier qui s'appelle « Musées-Mondes ».*

Dans les exemples 118 et 119, nous pourrions restituer intuitivement les verbes cibles à partir des formes erronées en nous basant sur leur ressemblance morphologique. Nous savons qu'en français il existe des paires ayant même racine mais qui diffèrent par leur préfixe ou suffixe, comme *présenter-représenter, joindre-rejoindre*. Cependant, dans les deux exemples au-dessus ce n'est pas le cas, la forme *globe, profondir* n'existent pas. Il s'agit peut-être d'une surgénéralisation des règles morphologiques en français.

4.3.1.3 Le manque de vocabulaire approprié en français

Le répertoire lexical restreint des apprenants sinophones se manifeste par un vocabulaire moins riche et des redondances syntaxiques, en particulier la paraphrase.

Nous avons remarqué que les apprenants non-natifs ont tendance à paraphraser. Il en résulte que leur rédaction semble maladroite et lourde. Par exemple :

(120) *Lorsque les professeurs enseignent des mots ou des phrases anglais, ils présentent d'abord en langue écrite en anglais, puis joignent (correction proposée : ajoutent) une (la) langue écrite en chinois, ensuite utilisent le chinois des signes **pour aider les élèves à comprendre afin de favoriser la compréhension par les élèves de la signification anglaise** (correction proposé : *pour aider les élèves à comprendre la signification anglaise*) .*

Les constructions *aider les élèves à comprendre* et *favoriser la compréhension des élèves* expriment le même sens, nous pouvons supposer que la répétition est une manifestation de l'insécurité linguistique des apprenants non-natifs, car ils ont peur de ne pas pouvoir bien exprimer le sens cible.

De plus, nous avons aussi constaté des paraphrases attribuables à un manque du vocabulaire. Par exemple :

(121) *Les paroles de Juana nous ont donné une profonde influence (correction proposé : *profondément influencés*).*

Dans l'exemple 121, la construction *donner une influence* peut être paraphrasée par un seul verbe *influencer*. Il existe plusieurs cas similaires, comme *arriver à la conclusion* à la place de *conclure*, *faire du progrès* à la place de *progresser*, *être correspondant* à la place de *correspondre*. Les apprenants ont souvent recours à un verbe support de sens vide, pour combler un manque dans leur répertoire lexical.

En somme, nous voyons que la plupart des erreurs sont attribuables aux connaissances insuffisantes du système langagier du français. Comme nos sujets d'étude sont les apprenants sinophones ayant l'anglais comme la première langue étrangère, nous avons également aperçu une influence de l'anglais.

4.3.2 L'influence de l'anglais

Comme nous l'avons mentionné dans l'introduction, la plupart de Chinois commencent à apprendre l'anglais depuis l'école primaire, et son apprentissage est obligatoire pendant 15 ans jusqu'à l'éducation supérieure. De plus, par rapport au chinois, l'anglais ressemble davantage au français au niveau morphologique et syntaxique en tant que langues indo-européennes, donc il est inévitable que l'anglais influence l'apprentissage du français. Dans notre corpus, nous avons aperçu les traces de l'anglais surtout dans les erreurs de verbes fictifs et celles de constructions syntaxiques.

4.3.2.1 Erreurs de verbes fictifs (VFIC)

Parmi les 10 erreurs de verbes fictifs relevées dans notre corpus, il y en a trois qui reflètent autant une influence de l'anglais qu'une surgénéralisation des règles grammaticales en français. Comme nous l'avons expliqué à la section 2.2, l'imperfection de l'interlangue des apprenants consiste en un transfert négatif des langues acquises et une surgénéralisation des règles de la langue cible. Par exemple :

(122) *Les artistes gagnent la réputation à cause du grand nombre de spectateurs et beaucoup d'opéra hors de la province les invitent aussi à performer (représenter leurs musiques).*

Dans l'exemple 122, le verbe *performer* n'existe pas en français, mais cette forme existe en anglais en tant que nom, qui désigne la personne qui représente le spectacle. En anglais, les mots qui se terminent avec *-er* renvoient souvent à la personne qui effectue l'action. Tandis qu'en français, le *-er* est souvent un suffixe du verbe. Sous l'influence du français, l'apprenant a pris le mot *performer* comme un verbe. Le verbe qui correspond à ce nom en anglais est *perform*, qui signifie l'action de représenter le spectacle devant le public. C'est le sens qui convient au contexte. Ainsi, nous pouvons penser que le verbe *performer* est créé en combinant la forme en anglais, le sens en anglais, mais la nature en tant que verbe sous l'influence du français.

Ce constat nous rappelle la théorie de l'activation des lemmes plurilingues (Wei, 2009). Selon cette théorie, les concepts à exprimer peuvent activer parallèlement des lemmes de différentes langues, et à cause du manque d'informations du lemme de la langue cible, l'apprenant transfère celles du lemme équivalent dans les langues acquises. Dans notre cas, le concept de 'représenter le spectacle devant le public' active à la fois le lemme *biǎo yǎn* en chinois et *perform* en anglais, du fait que le caractère chinois est très éloigné de l'écriture en français au niveau morphologique, c'est le lemme *perform* qui est choisi pour combler le manque du signifiant en français. Puisque beaucoup de verbes sont similaires en français et en anglais, il y a lieu de croire que l'apprenant tente alors sa chance en ajoutant le suffixe *-er* après le verbe *perform*.

4.3.2.2 Erreurs de constructions syntaxiques erronées (CS)

Outre les erreurs de verbes fictifs, c'est dans les erreurs de constructions syntaxiques erronées que l'on aperçoit le plus la forte influence de l'anglais sur le français. Par exemple :

(123) *Nous reproduisons trois récits de bagarre et montrons leurs styles et forces pour **rendre lecteurs habitués*** (correction proposée : *pour habituer les lecteurs*) *aux caractères généraux des récits vernaculaires noirs.*

(124) *L'auteur continue à nous présenter les différents types d'évaluation, qui **rendent les évènements différenciés*** (correction proposée : *qui différencie les évènements particuliers des évènements ordinaires*) *des autres plutôt ordinaires.*

Parmi les 7 occurrences de la construction *rendre quelque chose+adj*, 4 sont des usages erronés. Nous avons constaté un suremploi de cette construction. Elle est une construction équivalente de celle en anglais *make+nom+adj*, par exemple, *The news that our team had won made us very happy*. Toutefois, l'usage du verbe *make* en anglais est plus varié

que celui du verbe *rendre* en français. Il existe une construction *make+nom+participe passé*, par exemple, *The strange noise made us frightened*, mais cette construction ne fonctionne pas pour le verbe *rendre*. Si nous traduisons les exemples 123 et 124 en anglais, les parties erronées correspondent à *to make readers accustomed to the general characters of black vernacular tales* et *which make events differentiated from others rather ordinary*, donc nous pourrions penser que c'est un calque de la construction *make+nom+participe passé* en anglais.

(125) *Plus profondément, cela va ouvrir une plus grande porte **pour nous d'aller plus loin*** (correction proposée : *pour aller plus loin/pour que nous puissions aller plus loin*).

Si nous traduisons l'exemple 125 en anglais, nous obtiendrons *this will open a bigger door for us to go further*, qui correspond mot à mot à *pour nous d'aller plus loin* sur le plan syntaxique. Cependant, la syntaxe de l'anglais n'est pas adaptée au français. En français, nous disons soit *pour aller plus loin*, soit *pour que nous puissions aller plus loin*. Nous avons aussi constaté que, si l'influence de l'anglais se manifeste au niveau syntaxique, l'influence du chinois est plus évidente au niveau sémantique, surtout pour les métaphores. Par exemple, l'expression *ouvrir une plus grande porte* est une traduction littérale d'une métaphore en chinois qui signifie 'donner plus d'opportunités à quelqu'un'.

4.3.3 L'influence du chinois

Comme nous l'avons mentionné précédemment, du fait que le chinois et le français sont deux langues éloignées au niveau morphologique, l'influence du chinois se manifeste plutôt au niveau sémantique, surtout sur les erreurs de synonymes inadaptés. De plus, les apprenants sinophones connaissent le monde et établissent un système conceptuel au cours de l'acquisition de leur langue maternelle, soit le chinois, donc il est ordinaire que pendant l'apprentissage du français, les apprenants ont tendance à penser dans leur langue maternelle et sans se débarrasser du système sémantique chinois (Liu & Pu, 2019).

4.3.3.1 Erreurs de synonymes inadaptés (SNM)

Pour les erreurs de synonymes inadaptés, nous pouvons identifier trois explications possibles : la confusion sur la connotation des synonymes, sur le typage des actants ou sur l'usage concret ou abstrait. Étant donné que presque tous les apprenants ont recours aux dictionnaires bilingues comme des outils d'aide à la rédaction, nous pensons que certaines erreurs pourraient s'expliquer par l'acquisition du sens lexical par la traduction littérale en chinois :

(126) *Les danseurs interprètes sont chargés non seulement d'exécuter des mouvements, mais aussi d'incarner des personnages, de jouer des situations dramatiques, voire de **proférer** (correction proposée : *réciter*) du texte.*

Dans l'exemple 126, le verbe *proférer* est traduit en chinois par le verbe *dà shēng shu* 大声说, qui signifie 'parler à haute voix'. Si nous cherchons dans le dictionnaire monolingue, la définition du *proférer* est 'prononcer à haute voix des paroles, en particulier quelque chose d'hostile'. La connotation négative du verbe *proférer* n'a pas été traduite en chinois. Si le scripteur a appris le verbe *proférer* par l'accès sémantique en chinois, il se peut qu'il ne comprenne même pas la différence entre *prononcer* et *proférer*. Il y a lieu de croire que c'est la raison pour laquelle le scripteur a employé le verbe *proférer* pour exprimer le sens de 'réaliser oralement une parole ou un discours à voix haute et les articuler' dans sa rédaction.

Par ailleurs, l'utilisation du verbe *refléter* constitue une erreur récurrente pour les apprenants, car parmi les 4 occurrences du verbe *refléter* dans notre corpus, 3 correspondent des usages erronés. Le verbe *refléter* possède trois acceptions en français. La première est 'renvoyer, réfléchir la lumière'. La deuxième est 'renvoyer l'image des êtres et des choses plus ou moins nettement par réverbération lumineuse'. La troisième est 'présenter une image, une expression plus ou moins fidèle de quelque chose, révéler quelque réalité intérieur'. La troisième acception est la plus fréquemment employée par les apprenants, alors que les deux premiers de sens concret sont moins utilisés. La forme pronominale *se refléter* signifie donc 'être illustré indirectement de manière concrète ou caractérisée'. Dans un dictionnaire bilingue, la troisième acception est traduite en chinois par le verbe *fǎn yìng*. Toutefois, le verbe *fǎn yìng* est polysémique en chinois. Il a deux acceptions, l'une est le synonyme de *refléter* dans sa troisième acception, l'autre signifie 'apparaître, révéler son existence', qui est synonyme de *se manifester*. Par exemple :

(127) *Les médiats doivent prendre la conscience (correction proposée : *prendre conscience*) du service public et celle de responsabilité sociale, celles qui devraient être **reflétées** (correction proposée : *intégrées*) dans la conception et la mise en œuvre du programme.*

(128) *La suspension de l'action **se reflète** (correction proposée : *se manifeste*) dans presque tous les procédés d'évaluation.*

Dans l'exemple 127, le sens visé est 'la prise en conscience de responsabilité civile doit se concrétiser dans la conception et la mise en œuvre du programme'. Il s'agit de la deuxième acception du verbe *fǎn yìng* qui est synonyme de *se manifester*. Nous proposons donc la correction *celles qui devraient être intégrées dans la conception et la mise en œuvre*

du programme. Dans l'exemple 128, le sens cible est 'la suspension de l'action est présente dans presque tous les procédés d'évaluation'. Il s'agit aussi de la deuxième acception du verbe *fǎn yìng* qui est synonyme de *se manifester*.

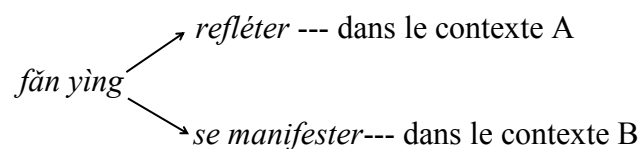
De ces deux exemples ci-dessus, il y a lieu de croire que l'utilisation erronée du verbe *réfléter* est imputable à la confusion entre les deux acceptions de sa traduction en chinois *fǎn yìng*. Ce constat nous rappelle l'hypothèse d'équivalence proposée par Swan (1997), selon laquelle l'apprenant tend à calquer la structure sémantique des mots de la langue étrangère et celle des mots de la langue maternelle. Il s'agit en effet d'une stratégie cognitive consistant à favoriser ou simplifier l'acquisition du lexique de L2 en utilisant les connaissances de L1. Il faut noter que les synonymes sont équivalents pour certaines de leurs acceptions, mais pas toutes. La traduction en chinois d'un verbe français est souvent un verbe polysémique dont les autres acceptions correspondent aux autres verbes en français. L'acquisition du sens lexical par la traduction en chinois entraîne facilement une surgénéralisation de la relation d'équivalence.

Ying a aussi mentionné la traduction littérale attribuable à la polysémie de la traduction chinoise dans son étude (2013). Elle propose que dans la production des apprenants, à cause du manque de la capacité à penser en langue cible, les apprenants ont tendance à formuler tout d'abord le sens à exprimer ou les phrases en langue maternelle. Et puis vient une traduction mot à mot fondée sur des correspondances déjà établies entre les verbes chinois et les verbes français dans l'apprentissage antérieur. Dans ce processus, les apprenants négligent souvent la polysémie des verbes en chinois, et trouvent leurs soi-disant équivalents en français, d'où vient une erreur de synonymes. Nous présentons le schéma suivant pour illustrer la confusion concernant le verbe *réfléter* :

Condition 1 : équivalence établie dans l'apprentissage antérieur

réfléter = *fǎn yìng*

Condition 2 : négligence de la polysémie du verbe *fǎn yìng*



Étape 1 : l'apprenant formule la phrase en chinois en utilisant le verbe *fǎn yìng* dans le contexte B.

Étape 2 : l'apprenant cherche la traduction du verbe *fǎn yìng* en français, mais à cause de la négligence de sa polysémie, l'apprenant trouve l'équivalent établi dans l'apprentissage antérieur : *réfléter*, qui fonctionne cependant dans le contexte A.

Étape 3 : une erreur de surgénéralisation de relation d'équivalence apparaît.

Par ailleurs, l'emploi concret ou abstrait des synonymes constitue aussi une difficulté pour les apprenants. Par exemple :

(129) *Les chutes « disjonctives » renvoient auditeur au présent et **ferment** (correction proposée : conclure) les événements racontés au passé.*

(130) *Pendant la République de Chine, en 1930, les chercheurs réussissent à créer les gestes destinés à **manifeste** (correction proposée : représenter) les signes avec intonation en chinois.*

Dans l'exemple 129, le verbe *fermer* signifie ici 'mettre un terme dans le temps à quelque chose, mettre définitivement fin à l'activité'. Dans cette acception, il est synonyme des verbes *conclure*, *mettre un terme à*. Cependant, le verbe *fermer* s'emploie seulement dans un sens concret, par exemple *fermer une souscription*, *fermer un compte*, tandis que le verbe *conclure* peut s'employer avec un objet concret ou abstrait. Les verbes *fermer* et *conclure* correspondent à *shǐ jié shù* en chinois, qui pourrait aussi être suivi d'un objet concret ou abstrait.

Dans l'exemple 130, d'après *le Petit Larousse*, le verbe *manifeste* signifie 'rendre perceptible ou faire connaître un sentiment, une opinion, un désir, une manière d'être, un état par la parole ou un signe extérieur'. Son usage est *manifeste quelque chose d'abstrait par quelque chose de concret*. Le sens de 'faire connaître' convient au contexte, mais son objet dans la phrase *les signes avec intonation* est concret, ce qui ne convient pas à l'usage du verbe *manifeste*. Nous proposons de corriger la phrase par *les gestes destinées à représenter les signes avec intonation en chinois*. Le verbe *représente* s'emploie dans son sens de 'correspondre à quelque chose, apparaître comme son équivalent' pour signifier que 'les gestes sont équivalents des signes avec intonation'. Les verbes *représente* et *manifeste* correspondent à un même verbe *biǎo xiàn* en chinois dans un dictionnaire bilingue. Le verbe *biǎo xiàn* s'emploie à la fois avec un objet concret ou abstrait.

En somme, pour *fermer* et *conclure*, *représente* et *manifeste*, si le scripteur les apprend à partir de la traduction en chinois, il est possible qu'il ne se rend pas compte de leur usage différent.

À partir des exemples précédents, il y a lieu de croire que certains moyens d'apprentissage ont une forte influence sur la pratique d'acquisition lexicale, comme l'utilisation des dictionnaires bilingues. Plusieurs études consacrées à l'analyse des erreurs commises par les apprenants sinophones ont abordé le problème d'utilisation excessive des dictionnaires bilingues (Yan 2017, Luste-Chaâ 2010). Yan (2017) a supposé que les étudiants chinois qui ont l'habitude de recourir à des dictionnaires sino-français ne maîtrisent pas totalement les subtilités sémantiques des mots, car les dictionnaires sino-français présentent souvent des traductions littérales sans contexte.

Par l'observation des erreurs de synonymes, nous supposons également que, d'une part, la traduction en chinois ne peut pas révéler toutes les composantes sémantiques ou la connotation d'un mot, de sorte que les apprenants n'arrivent pas à bien différencier les synonymes. D'autre part, la traduction en chinois est souvent polysémique et cela entraîne facilement une surgénéralisation de la relation d'équivalence entre le verbe en français et sa traduction en chinois.

Yan (2017) propose donc de sensibiliser les apprenants au phénomène de la phraséologie, comme la combinatoire lexicale et syntaxique, et de suggérer de meilleurs moyens de consultation qui facilitent l'accès aux informations phraséologiques, comme l'utilisation du corpus.

4.3.3.2 Erreurs d'ordre des mots (ORD)

Nous constatons parfois une influence du chinois sur l'apprentissage du français concernant l'ordre des mots. À l'inverse du français, les adverbes en chinois se placent en général devant les verbes. À l'instar de Yan (2017), nous pensons que cette différence pourrait expliquer certaines erreurs où les apprenants sinophones mettent les adverbes devant les verbes. Par exemple :

(131) *Je voudrais vous expliquer les avantages que ce concert représente et vous peut-être changez (correction proposée : *changez peut-être*) votre avis,*

L'adverbe *peut-être* est placé avant le verbe *changez*. Cela correspond à l'ordre syntaxique en chinois, mais en français, il faut dire *vous changez peut-être votre avis*.

4.3.3.3 Erreurs d'omission syntaxique (OS)

L'omission syntaxique est très fréquente chez les apprenants étrangers. Nous avons constaté que l'omission des pronoms relatifs « qui » ou « que » dans les propositions relatives représente presque une moitié d'erreurs d'omission syntaxique. Par exemple :

(132) *Pour les commerçants, il peuvent vendre plus de choses puisqu'il y a beaucoup de gens viennent à la ville* (correction proposée : *beaucoup de gens qui viennent à la ville*), *dans ce concert*.

(133) *A vrai dire, je n'ai pas bien compris cette théorie mais j'ai compris quand même dans cette partie* (correction proposée : *j'ai quand même compris que dans cette partie*), *les auteurs évoquent une temporalité dans l'échange*.

Nous supposons que certains apprenants ne peuvent pas distinguer le verbe principal et le verbe subordonné, et qu'ils prêtent peu d'attention à la règle qu'il existe un seul verbe principal dans une proposition. L'absence des pronoms relatifs « qui » ou « que » pourrait aussi s'expliquer par le fait qu'il n'existe pas de pronoms relatifs en chinois. Nous pouvons voir la traduction littérale de la proposition relative dans l'exemple 132 en chinois (cf. Tableau 4.15) :

Traduction en chinois	有	很多	人	来	这个	城市
Transcription phonétique en chinois	Yǒu	hěnduō	rén	lái	zhè gè	chéng shì
Traduction littérale	avoir	beaucoup de	gens	venir	ce	ville
Phrase en français	Il y a beaucoup de gens qui viennent à la ville					

Tableau 4.15 : Traduction de l'exemple 132 en chinois

Nous pouvons voir que l'ordre des mots en chinois est le même qu'en français, mais il n'existe pas de pronoms relatifs. Il se peut que l'omission de « qui » ou « que » soit attribuable à la syntaxe en chinois. C'est le même cas pour l'exemple 133 : l'ordre des mots est aussi le même en chinois outre l'absence de pronom relatif « que ».

4.3.3.4 Erreurs de prépositions et erreurs de pronoms (PREP et PRO)

Les erreurs de prépositions et de pronoms découlent en général d'une confusion sur la transitivité du verbe ou sur la préposition qu'il régit.

Il existe seulement des verbes transitifs directs ou des verbes intransitifs en chinois, mais pas de verbes transitifs indirects. Les équivalents des verbes transitifs indirects sont

souvent utilisés comme les verbes transitifs directs en chinois, ce qui pourrait expliquer plusieurs erreurs d'omission de prépositions. Par exemple :

(134) *Ainsi, je vous prie de reprendre et de bien **réfléchir** (correction proposée : réfléchir à) votre décision de supprimer cette fête.*

Le verbe *réfléchir* est un verbe transitif indirect dans son sens de 'penser plus profondément sur quelque chose' qui régit la préposition *à*, tandis qu'en chinois, le verbe correspondant est *kǎo lù*, qui est un verbe transitif direct suivi directement d'un complément d'objet direct sans préposition.

Pour résumer, les traces du chinois que nous avons observées se concentrent sur les erreurs de sens lexical, surtout sur les mauvais synonymes, et parfois sur les erreurs d'ordre adverbe et verbe, et sur l'omission de pronoms relatifs et de prépositions régies par les verbes transitifs indirects.

4.3.4 La proximité morphologique

Comme ce qui est indiqué dans la section 4.1.3.1, nous avons constaté beaucoup d'erreurs de sens inadapté qui sont attribuables à la proximité morphologique. Par exemple :

(135) *Je me permets de vous écrire, en tant que citoyenne grenobloise, pour vous demander de **réserver** (conserver) l'organisation annuelle du concert estival dans notre belle ville.*

(136) *En l'utilisant, la locutrice, qui emploie d'ordinaire plutôt l'évaluation externe, **se relève** (correction proposée : se révèle) capable d'enchâsser plus profond (correction proposée : profondément) ses commentaires.*

Parmi les 83 erreurs de sens inadapté relevées dans notre corpus, 20 erreurs peuvent être considérées comme attribuables à une proximité morphologique existant entre le verbe employé et le verbe visé. Comme dans les exemples 135 et 136, les verbes erronés et les verbes appropriés *réserver-conserver*, *relever-révéler* montre une ressemblance morphologique.

4.3.5 Conclusion provisoire

Pour résumer, dans cette section, nous avons identifié cinq sources d'erreurs possibles : l'inattention, la déficience des connaissances langagières du français, l'influence de l'anglais, l'influence du chinois, et la proximité morphologique entre le verbe visé et le verbe employé. La plupart d'erreurs sont attribuables aux connaissances insuffisantes du français, y compris

la maîtrise inachevée des règles grammaticales, la surgénéralisation des connaissances en français ainsi que le manque du vocabulaire en français. Il est intéressant de noter que les traces de l'anglais se présentent plutôt au niveau syntaxique, tandis que l'influence du chinois se manifeste davantage au niveau sémantique. Nous pouvons penser que les méthodes d'apprentissage exercent une forte influence sur la pratique d'acquisition lexicale, comme l'utilisation excessive des dictionnaires sino-français pour accéder au sens lexical.

4.4 Conclusion

Dans ce chapitre, nous avons tout d'abord abordé la répartition des erreurs selon les catégories et les types d'erreurs à travers une analyse quantitative. Les erreurs au niveaux sémantique et syntaxique sont les plus importantes, en particulier les erreurs de structures syntaxiques (SYN), les erreurs de sens lexical (SEN), les erreurs de conjugaison et d'accord (CA), les erreurs de temps/aspect/mode (TAM) et les erreurs de prépositions (PREP). Notre hypothèse, selon laquelle la conjugaison, l'accord et le TAM du verbe constituent une grande difficulté pour les apprenants sinophones, se voit valider. Nous voyons que le sens lexical, la combinatoire sémantique et syntaxique s'avèrent problématiques pour les apprenants sinophones.

Dans un second temps, nous avons effectué une analyse qualitative des sous-types d'erreurs. À travers la comparaison avec les usages verbaux corrects, nous avons relevé, d'une part, les circonstances où les erreurs apparaissent fréquemment, d'autre part, les points grammaticaux à souligner dans l'enseignement du FLE auprès des apprenants sinophones.

Par exemple, le sous-emploi du subjonctif est plus présent dans les cas où le choix du subjonctif se justifie par sa valeur sémantique. Nous avons constaté dans le corpus que les subjonctifs bien utilisés sont souvent combinés avec les conjonctions comme *bien que*, *jusqu'à ce que*. Nous supposons que les apprenants retiennent les circonstances où le subjonctif s'emploie d'une manière mécanique. La valeur sémantique véhiculée par le subjonctif est donc un point à souligner pendant l'apprentissage. La majorité des erreurs de pronominaux concerne les verbes qui possèdent à la fois une forme pronominale et une forme non pronominale. Il existe peu d'erreurs sur les verbes qui ne possèdent pas de forme pronominale ou qui, à l'inverse, n'existent que sous forme pronominale. Il se peut que les apprenants mémorisent les verbes pronominaux comme un ensemble sans analyser la fonction du pronom. Il faut donc sensibiliser les apprenants au rôle du pronom dans les pronominaux.

Les erreurs de prépositions apparaissent souvent dans les cas où les prépositions régies par les verbes sont peu prédictibles, c'est-à-dire que l'intégration des prépositions dans les constructions verbales ne peut pas être expliquée par le sens des prépositions. Nous proposons donc d'accorder plus d'attention aux prépositions dans les constructions verbales.

Par ailleurs, nous avons relevé plusieurs points grammaticaux auxquels il faut prêter attention dans le cours du FLE auprès des apprenants sinophones de niveau B1-B2, comme l'accord du verbe avec le sujet postposé au verbe, l'accord du participe passé avec le complément d'objet direct déplacé avant le verbe, l'accord avec deux noms abstraits comme sujet, la reprise du sujet dans l'inversion sujet-verbe. Il existe aussi des points à souligner tout particulièrement auprès des apprenants sinophones, car ils sont liés à l'influence du chinois. Par exemple, en français, l'adverbe doit se placer après le verbe au temps simple et à l'infinitif, et entre l'auxiliaire et le participe passé au temps composé, mais les apprenants sinophones ont tendance à mettre l'adverbe avant le verbe sous l'influence du chinois. Il faut aussi mettre en avant la fonction des pronoms relatifs « qui » ou « que » en français, car les apprenants sinophones les oublient souvent sous l'influence de la syntaxe en chinois.

À travers la comparaison entre les textes écrits sur table et à la maison, nous confirmons notre hypothèse selon laquelle les conditions de rédaction exercent une influence sur les erreurs commises. Les erreurs morphologiques sont moins présentes dans les textes écrits à la maison grâce à l'utilisation des dictionnaires, surtout les erreurs d'orthographe d'usage. Cependant, les erreurs grammaticales et syntaxiques augmentent, car les phrases dans les textes à la maison s'avèrent plus compliquées sur le plan syntaxique. Pour les types d'erreurs, nous avons constaté que les erreurs de temps/aspects/mode sont plus importantes dans les textes écrits à la maison. Cela s'expliquerait par la nature des textes écrits à la maison qui exige des TAM plus variés. Il n'y a pas de grandes différences entre les deux types de textes pour les autres types d'erreurs. Nous concluons donc que l'utilisation des dictionnaires réduirait surtout les erreurs morphologiques, mais ne permet pas bien de corriger les erreurs sur le plan syntaxique, comme les erreurs de voix, les erreurs d'ordre des mots.

Pour l'origine des erreurs, nous avons identifié que la plupart d'erreurs sont attribuables aux connaissances insuffisantes du français, en raison de la complexité des règles, le manque d'informations des combinatoires lexicales. Des influences du chinois et d'anglais sont aussi non négligeables. Notre hypothèse, selon laquelle les traces de l'anglais se

présentent plutôt au niveau morphologique et syntaxique, tandis que l'influence du chinois se manifeste davantage au niveau sémantique, a aussi été confirmée.

Conclusion

Notre étude consiste à analyser les erreurs verbales commises par les apprenants sinophones de niveau B1-B2. Pour ce faire, nous avons élaboré une grille d'erreurs et constitué un corpus d'apprenants sinophones indiquant les erreurs verbales.

Les corpus d'apprenants jouent un rôle important dans l'étude de l'acquisition de la langue seconde et dans la didactique des langues. Ils se sont bien développés ces dernières années : les langues s'étendent de l'anglais à une grande variété de langues étrangères comme l'espagnol et le français ; il y a des corpus comportant plusieurs nationalités d'apprenants, et aussi des corpus destinés aux apprenants partageant une même langue maternelle pour répondre à leurs besoins spécifiques. Malgré leur développement remarquable, les corpus d'apprenants sinophones du français langue seconde restent peu développés. Nous sommes convaincue que le corpus d'apprenants sinophones que nous avons élaboré dans le cadre de notre mémoire servira de base empirique pour la recherche sur l'acquisition du français des apprenants sinophones, et constituera une ressource applicable dans la didactique du français auprès des apprenants sinophones de niveau B1-B2.

Notre étude se situant dans un cadre universitaire, nous avons décidé de faire référence à la norme définie par les grammaires et dictionnaires de référence, comme le dictionnaire *Robert*, en tant que critère d'évaluation des erreurs. Notre grille d'erreurs est fondée sur celle de Yan (2017) et en adopte les quatre grandes catégories : les erreurs au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la grammaire et de la syntaxe, et au niveau du genre et du contexte. Nous y avons ajouté les types et sous-types en fonction des erreurs relevées dans notre corpus afin de mieux cibler notre public d'étude. Les verbes et les erreurs verbales dans le corpus sont annotés en indiquant le lemme, le type et le sous-type d'erreurs. L'annotation a été effectuée semi-automatiquement avec Nooj, accompagnée d'une vérification manuelle pour trouver les silences et éliminer les bruits. Plusieurs annotateurs se sont engagés dans le travail d'annotation avec un guide précis pour garantir la pertinence d'annotation.

La différence principale entre les verbes du français et ceux du chinois porte sur la conjugaison et l'accord du verbe avec la personne, le temps, l'aspect et le mode. Nous avons donc formulé l'hypothèse selon laquelle la conjugaison et l'accord, le TAM des verbes constitueraient une difficulté majeure pour les apprenants sinophones. Elle a été validée dans

nos résultats, car les erreurs verbales se rapportant à la conjugaison, l'accord et le temps/aspect/mode représentent 27 % de toutes les erreurs verbales. Par ailleurs, étant donné que notre public est constitué d'apprenants de niveau intermédiaire, nous avons supposé que les erreurs morphologiques seraient moins présentes que les erreurs sémantiques et syntaxiques. Cette hypothèse a aussi été confirmée. Les erreurs syntaxiques et sémantiques sont les plus importantes avec une proportion respective de 40 % et 37 %, tandis que les erreurs morphologiques ne représentent que 22 %, soit environ une moitié des deux précédentes. L'hypothèse qu'il aurait moins d'erreurs morphologiques dans les textes écrits à la maison en raison d'utilisation des dictionnaires a également été validée. Nous avons conclu que l'utilisation des dictionnaires diminue davantage les erreurs morphologiques, surtout les erreurs d'orthographe, mais n'influence pas trop les erreurs syntaxiques.

Pour répondre à nos problématiques d'étude, les six sous-types d'erreurs les plus fréquentes sont les erreurs de structures syntaxiques, les erreurs de sens lexical, les erreurs de conjugaison et d'accord, les erreurs de temps/aspect/mode, les erreurs de prépositions et les erreurs de cooccurrences. Le fait que les apprenants ne maîtrisent pas bien le sens lexical, la combinatoire sémantique et syntaxique d'une unité verbale devrait inciter à proposer des moyens d'apprentissage plus efficaces. Yan (2017) a ainsi proposé dans son étude l'introduction du corpus dans le cours du FLE, où les apprenants pourraient mieux observer la combinatoire des unités verbales.

À travers l'analyse qualitative de chaque type d'erreurs, nous avons relevé les circonstances où les erreurs apparaissent fréquemment, comme le sous-emploi du subjonctif plus présent dans les cas où ce choix se justifie par sa valeur sémantique. Nous avons aussi mis en avant des points grammaticaux à souligner dans la didactique du FLE auprès des apprenants sinophones, comme la fonction du pronom dans les pronominaux, l'accord avec deux noms abstraits comme sujet. La mise en lumière des circonstances où les erreurs apparaissent a pour but, d'une part, de sensibiliser les apprenants à ces situations, et d'autre part, de pouvoir mettre en avant les points grammaticaux qui poseraient des problèmes aux apprenants sinophones dans l'enseignement du FLE.

À la lumière de la théorie d'interlangue, nous avons identifié cinq sources d'erreurs possibles : l'inattention, la déficience des connaissances langagières du français, l'influence de l'anglais, l'influence du chinois, et la proximité morphologique entre le verbe visé et le verbe employé. La plupart des erreurs sont attribuables aux connaissances insuffisantes du

français, comme la maîtrise inachevée des règles grammaticales, la surgénéralisation des connaissances en français ou le manque du vocabulaire du français. Si nous réfléchissons plus profondément à l'influence du chinois et de l'anglais sur le français, nous constatons que le transfert du chinois sur le français se manifeste plutôt au niveau sémantique, alors que le transfert de l'anglais sur le français apparaît tant au niveau morphologique qu'au niveau syntaxique. L'anglais entraîne aussi des confusions sémantiques sur les verbes proches morphologiquement des mots en anglais, mais avec un sens différent.

Au terme de ce travail, nous avons pu observer que les erreurs commises par les apprenants dévoilent bien leurs difficultés d'apprentissage, dont l'étude pourra être utile pour la didactique du FLE auprès des apprenants sinophones de niveau B1-B2. Notre étude n'a pas proposé de pistes pédagogiques systématiques, car nous n'avons pas pu effectuer une enquête sur terrain. C'est un point qui méritera d'être approfondi dans les recherches à l'avenir.

Bibliographie

Ågren, M. (2008), À la recherche de la morphologie silencieuse. Sur le développement du pluriel en français L2 écrit, *Études Romanes de Lund 84*, Thèse de doctorat, Université de Lund.

Alonso Ramos, M. L. , Wanner, O. , Vincze, G. , Casamayor, N. , Vázquez, E. & Mosqueira, S. Prieto. (2010). Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora”, *7th International Conference on Language Resources and Evaluation (LREC)*, La Valetta, Malta, pp. 3209-3214.

Anctil, D. (2005). *Maîtrise du lexique chez les étudiants universitaires : typologie des problèmes lexicaux et analyse des stratégies de résolution de problèmes lexicaux* (Mémoire de Master 1 : Science de l'éducation). Montréal : Université de Montréal.

Anctil, D. (2010). *L'erreur lexicale au secondaire : analyse d'erreurs lexicales d'élèves de 3e secondaire et description du rapport à l'erreur lexicale d'enseignants de français* (Thèse de doctorat : Science de l'éducation). Montréal : Université de Montréal.

Anctil, D. (2012). « Portrait des erreurs lexicales d'élèves de 3e secondaire en production écrite et proposition de pistes didactiques », *Pratiques* [En ligne], 155-156 | 2012, mis en ligne le 18 décembre 2017, consulté le 14 novembre 2019. URL : <http://journals.openedition.org/pratiques/3447> ; DOI : 10.4000/pratiques.3447

Biskri, I. & Jebali. A. (2010). *Multilinguisme et traitement des langues naturelles*, Presse de l'université de Québec.

Bohec, J. (2019). « Les erreurs lexicales des étudiants chinois spécialistes de français en 2e année de licence en expression écrite », *Recherches en didactique des langues et des cultures [En ligne]*, 16-2 |2019, mis en ligne le 05 septembre 2019, consulté le 20 novembre 2019. URL : <http://journals.openedition.org/rdlc/6758> ; DOI : 10.4000/rdlc.6758

Bolly, C. (2008) Les unités phraséologiques : un phénomène linguistique complexe? Séquences (semi-)figées construites avec les verbes "prendre" et "donner" en français écrit L1 et L2 : Approche descriptive et acquisitionnelle. Prom. : Klein, Jean René <http://hdl.handle.net/2078.1/19625>

Burnard, L. (2014). *What is the texte encoding initiative ? How to add intelligent markup to digital resources*. OpenEdition Press. Marseille. p114. <https://books.openedition.org/oep/426>

Dagneaux, E., Denness, S. & Granger, S. (1998). Computer-aided error analysis. *System*, 26, pp. 163–174.

De Cock, S. & Tyne, H. (2014). « Corpus d'apprenants et acquisition des langues », *Recherches en didactique des langues et des cultures* [En ligne], 11-1 | 2014, mis en ligne le 07 janvier 2014, consulté le 19 avril 2020. URL : <http://journals.openedition.org/rdlc/1716> ; DOI : 10.4000/rdlc.1716

Granger, S., Vandeventer, A. & Hamel, M-J (2001). Analyse des corpus d'apprenants pour l'ELAO basé sur le TAL. *Traitement automatique des langues* 42(2), Paris: Hermès, 609-621.

Granger, S. (2007). Corpus d'apprenants, annotations d'erreurs et ALAO: une synergie prometteuse, *Cahiers de lexicologie*, 91/2, pp. 465–480.

- Granger, S., Dagneaux, E., Meunier, F. & Paquot, M. (2009). *International Corpus of Learner English. Version 2. Handbook and CD-ROM*. Presses universitaires de Louvain
- Granger, S., Gilquin, G., & Meunier, F. (Éd.). (2015). *The Cambridge Handbook of Learner Corpus Research*. Cambridge : Cambridge University Press.
- Hamel, M.-J. et Milicevic, J. (2007). Analyse d'erreurs lexicales d'apprenants du FLS: démarche empirique pour l'élaboration d'un dictionnaire d'apprentissage. *Canadian Journal of Applied Linguistics*, 10/1, pp. 25-45.
- Hausmann, F. J. (1989). Le dictionnaire de collocations. In F.J Hausmann *et al.* (Eds.) *Wörterbücher - Dictionaries - Dictionnaires*, vol. 1. Berlin: de Gruyter, pp.1010-1019
- Huangpu, Q. & Di, Y. & Dong, M. & al. (2000). *Dictionnaire concis, français-chinois chinois-français* (2000), édition corrigée. Paris, Pékin : éditions La Presse Commerciale & Larousse.
- 何雯雯. (2010). *基于错误分析和中介语理论的法语专业写作研究[D]*. (mémoire de master). L'Université de Hebei.
- 金艳妮. (2009). *探究法语习得过程中汉语和英语的角色[D]*. (mémoire de master). L'Université de Dalian.
- 贾秋月. (2017). *法语专业初学者写作中的英语负迁移现象探究[D]*. (mémoire de master). L'Université de Sichuan.
- Labov, W., 1976, *Sociolinguistique*, Trad. Fr.d'A.Kihm, Paris, Editions de Minui, P. 176.
- Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press: Ann Arbor.
- Le Petit Robert, Paris, dictionnaires, (2009). Paris : Le Robert.
- Le Petit Larousse illustré et en couleurs, (2007). Paris : éditions Larousse.
- Li, Y-L. (2018). L'apprentissage de l'orthographe chez les apprenants sinophones du français L2 : une analyse portant sur un corpus écrit. *Linguistique*. Université Sorbonne Paris Cité, Français. ffnNT :2018USPCA071ff. fftel-02147421f
- Liu, X & Pu, Z-H (2019). Extension sémantique de lexie chez des apprenants sinophones plurilingues. *Francisola: Revue Indonésienne de la langue et la littérature françaises*, 4(1), 22-29. doi: 10.17509/francisola.v4i1.20338.
- Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and Second Language Acquisition: The design and collection of CEDEL2. In A. Díaz-Negrillo, N. Ballier & P. Thompson (Eds.), *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins, pp. 65-100.
- Luste-Chaâ, O. (2010). « L'erreur lexicale dans l'analyse des productions écrites en FLE », *Pratiques* [En ligne], 145-146 | 2010, mis en ligne le 15 juin 2010, consulté le 30 avril 2019. URL : <httpjournals.openedition.org/pratiques/1556> ; DOI : 10.4000/pratiques.1556
- Mel'čuk, I., A. Clas et A. Polguère (1995). *Introduction à la lexicologie explicative et combinatoire*, Coll. Champs linguistiques/Universités francophones, Louvain-la-Neuve/Paris, Éditions Duculot/AUPELF-UREF.
- Mel'čuk, I (1998), *Collocations and Lexical Functions*. In A.P. Cowie (Ed.) *Phraseology. Theory, Analysis, and Applications*. pp. 23-53. Oxford: Clarendon Press.

- Mendikoetxea, A & O'Donnell, M-J. (2009). WriCLE: A learner corpus for Second Language Acquisition Research. Conference: Corpus Linguistics 2009, At University of Liverpool, UK.
- Na Phuket, P-R. (2015). Understanding EFL Students' Errors in Writing. *Journal of Education and Practice*. Vol.6, No.32, 2015
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24/2. Oxford University Press. p.223-242
- Omidipour, M. (2014). An analysis of errors in writing among adult persian learners of English. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, Volume 5 (3). pp. 176-187
- Park, T-J. & Zheng, C. (2013). An Analysis of Errors in English Writing Made by Chinese and Korean University Students. *Theory and Practice in Language Studies*. Vol. 3, No. 8, pp. 1342-1351
- Riegel, M., Pellat, J.-C., & Rioul, G. (1994). *Grammaire méthodique du français*. Presses Universitaires de France.
- Ringbom, H. (1999). 'High-frequency verbs in the ICLE corpus'. In Antoinette Renouf (ed.), *Explorations in Corpus Linguistics*. Amsterdam: Rodopi: 191-200.
- Siouffi, G., & Raemdonck, D.-V. (2017). 100 fiches pour comprendre les notions de grammaire. Clamecy : Presses de la Nouvelle Imprimerie Laballery.
- Silberztein, M. (2015). *La formalisation des langues: l'approche de NooJ*. Londres : ISTE Group.
- Stormbom, C. (2018). 'Epicene pronouns in intermediate to advanced EFL writing'. *International Journal of Learner Corpus Research* 4(1): 1-22.
- Swan, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. In Schmitt, N. et McCarthy, M. (eds.), *Vocabulary, description, acquisition and pedagogy*, 156-180. Cambridge : Cambridge University Press.
- Virtanen, T. (1996). 'Exploiting the International Corpus of Learner English (ICLE)'. In Terttu Harakka & Merja Koskela (eds.) *Kieli ja tietokone: AFinLAN vuosikirja*. Jyväskylä: Jyväskylän yliopistopaino: 157-166
- 王晓慧.(2018). 探讨法语学习中英语负迁移现象及策略[J]. *文学教育(下) Literature Education*, 04, pp.138-139.
- Yan, R. (2017). *Étude des constructions verbales scientifiques dans une perspective didactique : utilisation des corpus dans le diagnostic des besoins langagiers du FLE à l'aide des techniques de TAL* (Thèse de Doctorat en Sciences du langage publiée). Université Grenoble Alpes, Grenoble.
- Ying, X. (2013). *Approche cognitive en didactique des langues : analyse et interprétation d'erreurs écrites prototypiques en français langue étrangère par des apprenants sinophones et remédiation* (Thèse de doctorat : Linguistique française). Université de Bordeaux 3, Université de Wuhan.
- 张淑琴.(1993). 法语动词常见错误浅析[J]. *山东外语教学 (Shandong Foreign Language Teaching Journal)*, 01, pp.40-43

Sitographie

1. Bibliothèques Universitaires Grenoble Alpes. (2009-2018). Le Petit Robert de la langue française. Repéré à l'adresse : <https://pr-bvdep-com.sid2nomade-2.grenet.fr/robert.asp>. Consulté le 10 mai 2020.
2. Hamel, M-J. & Milićević, J. (2010). DIRE autrement. Repéré à l'adresse : <http://web5.uottawa.ca/direautrement/index.html>. Consulté le 15 mai 2020.
3. Agren, M. (2009). Corpus Écrit de Français Langue Étrangère. Repéré à l'adresse : <https://projekt.ht.lu.se/cefle/information>. Consulté le 15 mai 2020.
4. Université catholique de Louvain. (2019). Learner Corpora around the World. Repéré à l'adresse : <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>. Consulté le 13 mai 2020.

Glossaire

- Cooccurrence lexicale : la cooccurrence lexicale est le simple fait pour deux ou plusieurs lexies d'apparaître ensemble dans un même énoncé (Anctil, 2005, p. 22). La cooccurrence lexicale restreinte est cette propriété qu'ont les unités lexicales d'apparaître ensemble de façon contrainte dans un même énoncé (Anctil, 2005, p. 38)
- Collocation : l'expression AB (ou BA), formée des lexies A et B, est une collocation si, pour produire cette expression, le locuteur sélectionne A librement d'après son sens 'A', alors qu'il sélectionne B pour exprimer un sens 'C' en fonction de A. Nous dirons que la lexie choisie librement est la base de la collocation, alors que les expressions contraintes seront appelées collocatifs (Polguère, cité par Anctil, 2005, p. 25)

Sigles et abréviations utilisés

LEC : lexicologie Explicative et Combinatoire

CECRL : le cadre européen commun de référence pour les langues.

FLE : français langue étrangère

CUEF : centre universitaire d'études françaises

MORPHO : morphologique

SEMAN : sémantique

GRAMSYN : grammatical et syntaxique

CONT : contexte

CA : conjugaison et accord

ORT : orthographe d'usage

VFIC : verbe fictif

SEN : sens lexical

SNM : synonymes

SIN : sens inadapté

TAM : temps, aspect, mode

COC : cooccurrence

CONO : cooccurrence avec le nom

COA : cooccurrence avec l'adverbe ou le complément circonstanciel

VV : voix des verbes

EPN : erreurs de pronominaux

ORD : ordre des mots

PREP : prépositions

PS : préposition surnuméraire

OP : omission de préposition

EP : préposition erronée

PRO : pronom

SYN : syntaxique

RS : redondance syntaxique

OS : omission syntaxique

CS : construction syntaxique

REG : registre

RP : le Petit Robert

PL : le Petit Larousse

Table des illustrations

Tableau 3.1 : Grille d'erreurs élaborée pour l'analyse des erreurs verbales.....	48
Tableau 3.2 : Grille d'illustration de l'annotation dans le corpus.....	60
Figure 4.1 : La répartition des erreurs selon les quatre grandes catégories.....	64
Figure 4.2 : La répartition des erreurs selon les types d'erreurs.....	65
Figure 4.3 : La répartition des erreurs selon les sous-types d'erreurs.....	65
Figure 4.4 : La répartition des erreurs au niveau morphologique.....	67
Tableau 4.5 : La répartition des erreurs d'accord.....	69
Figure 4.6 : La répartition des erreurs au niveau sémantique et des cooccurrences lexicales	71
Tableau 4.7 : La répartition des erreurs de temps/aspect/mode.....	72
Figure 4.8 : La répartition des erreurs au niveau de la grammaire et de la syntaxe.....	77
Tableau 4.9 : La répartition des erreurs de redondance syntaxique.....	80
Tableau 4.10 : La répartition des erreurs d'omission syntaxique.....	81
Figure 4.11 : La répartition des erreurs selon les quatre catégories : textes sur table.....	85
Figure 4.12 : La répartition des erreurs selon les quatre catégories : textes à la maison.....	85
Figure 4.13 : Les erreurs morphologiques dans les textes écrits sur table.....	86
Figure 4.14 : Les erreurs morphologiques dans les textes écrits à la maison.....	87
Tableau 4.15 : Traduction de l'exemple 132 en chinois	100

Table des annexes

Annexe 1 Exemple de texte écrit sur table.....	117
Annexe 2 Exemple de texte écrit à la maison (extrait).....	119
Annexe 3 Exemple de corpus annoté (texte sur table).....	120
Annexe 4 Exemple de corpus annoté (texte à la maison).....	124
Annexe 5: Guide d'annotation.....	129

Annexe 1 : Exemple de texte écrit sur table

Grenoble, le 30 août 2018

██████████
Monsieur ██████████
le maire de la ville

Objet : la maintenance du concert gratuit de l'été

Bonjour monsieur le maire,

C'est au nom de ma famille et mes amis que je vous écris pour deman exprimer mes opinions sur l'annulation du concert gratuit de l'été. Cet concert, ayant eu lieu pendant des années, nous a apporté pleins de joies. Donc, je voudrais vous proposer quelques raisons pour le maintenir si vous pouvez réfléchir avant de faire la décision.

En effet, cet concert a déjà devenu un événement important à la fin de l'été pour nous tous. Grâce aux expériences des années passées, le concert est maintenant le plus réputé de la région. Chaque année, il y a de plus en plus de touristes qui viennent d'ailleurs pour le voir.

En tant qu'un concert gratuit, il nous permet de se retrouver et écouter la musique ensemble sous une atmosphère agréable. C'est absolument salubre pour améliorer l'image de la ville. De plus, les artistes, même si moins connus que les stars, ont l'opportunités de se présenter devant le public. Personne ne peut nier que les artistes de la région pr font aussi des très bonnes musiques qui nous plaisent beaucoup. C'est non seulement une chance des artistes pour communiquer avec ses fans potentiels mais aussi une chance des commerçants pour les activités accessoires comme les petits restos autour du lieu de concert.

Pour toutes ces raisons, nous sommes nombreux pour demander la maintenance de cet concert. Les problèms financières peuvent être éventuellement résolus en rendant compte de l'essor du tourisme de cette période dans l'avenir. Après tout, l'idée initiale du concert est de proposer aux habitants des musiques et une détente avant d'entrer dans la saison d'automne.

Dans l'attente de réponse de votre part, je vous prie d'agrèer,

monsieur le maire, l'expression de mes salutations distinguées.



Annexe 2 : Exemple de texte écrit à la maison (extrait)

Dans nos études sur la langue vernaculaire, pour enregistrer un plus grand nombre de discours familiers, nous avons utilisé des techniques, parmi lesquelles, les récits d'expériences personnelles, entre autres la question sur le « danger de mort » et sur les « bagarres » s'avèrent le plus efficace. Nos interlocuteurs viennent de plusieurs tranches d'âge, plusieurs classes sociales et plusieurs groupes ethniques. Puisque leurs expériences et émotions jouent un rôle important dans leurs biographies, il faut mettre l'interview dans la même situation qu'ils ont connue. Nous nous intéressons surtout à la partie évaluative dans nos précédentes études et maintenant nous comptons regarder plus près des éléments d'évaluation, comme l'utilisation des procédés syntaxiques qui peuvent déboucher sur une conclusion inattendue.

Avant de commencer l'analyse, nous reproduisons trois récits de bagarre et montrons leurs styles et forces pour rendre lecteurs habitués aux caractères généraux des récits vernaculaires noirs. Le premier est écrit par Boot, le second Larry H et le troisième John.L.

Définition du récit

Nous le définirons comme un moyen de récapitulation de l'expérience passée consistant à faire correspondre à une suite d'événements (supposés) réels une suite identique de propositions verbales. Le premier et le deuxième exemple sont à l'ordre des événements rapportés alors que l'on inverse l'ordre des faits dans le troisième. Les propositions narratives dans le récit sont ordonnées temporellement et nous nommons deux propositions unies (et séparées) par une jonction temporelle « récit minimal ». Par contre, on nomme une proposition non fixée par une jonction temporelle « proposition libre ». Prenant un exemple du récit de John L, nous prouvons que les propositions libres et les propositions subordonnées ne peuvent pas changer l'interprétation par inversion et que seules les propositions indépendantes ont la fonction narrative.

Structure générale du récit

Les 6 parties du récit sont résumé, indication, développement, évaluation, résultat ou conclusion et chute. Les narrateurs commencent souvent leurs récits par un résumé, qui donne un aperçu de l'histoire. Ensuite, c'est l'indication qui précise le moment, le lieu, les personnes, leur activité ou leur situation détaillée en recourant aux propositions libre avant le commencement de l'histoire ou tout au long de celle-ci. Et puis, nous présentons la chute qui signifie qu'un récit est achevé et qui clôt le développement. Les chutes « disjonctives » renvoient auditeur au présent et ferment les événements racontés au passé. Prenons une interview avec des pré-adolescents, nous trouvons que les récits non évalués sont obscurs et ambiguë alors que chaque élément dans les récits évalués aide à mettre en lumière les points principaux et mettent en valeur un aspect particulier des personnages. Puisque l'argumentation, l'interprétation de l'histoire se trouvent tout au long du récit, l'évaluation ne doit pas être limitée seulement dans sa partie, elle est présentée aussi sous formes variables qui pénètrent dans le récit par la structure interne des propositions narratives et par leur enchaînement. Donc, un récit complet est composé de quelques indications au début, un développement, un foyer d'évaluation qui l'arrête, un résultat qui le résume et enfin une chute au présent. D'un autre point de vue, on peut considérer le récit comme des réponses à des questions. Le développement est l'identification cruciale d'un récit alors que les autres parties touchent au plan référentiel. A part la reprise des indications et l'annonce du développement, l'évaluation est aussi incluse dans le résumé. La chute exclut toutes les questions.

Annexe 3 : Exemple de corpus annoté (texte sur table)

```
<?xml version="1.0" encoding="UTF-8"?>
<!--odt2tei: 2019-04-08T10:03:53Z-->
<?xml-stylesheet type="text/css" href="../../annotation_verbes.css"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title/>
      </titleStmt>
      <editor>
        <name>LIDILEM</name>
        <address>
          <addrLine>Université Grenoble Alpes</addrLine>
          <addrLine>Bâtiment Stendhal</addrLine>
          <addrLine>CS40700</addrLine>
          <addrLine>38058 Grenoble cedex 9</addrLine>
        </address>
      </editor>
    </titleStmt>
    <editionStmt>
      <edition>Edition électronique <date>2018-2019</date>
    </edition>
    <respStmt>
      <resp>Responsable du projet</resp>
      <name>
        <persName>[REDACTED]</persName>
        <email>[REDACTED]</email>
        <affiliation>
          <name>LIDILEM</name>
          <orgName>Université Grenoble ALpes</orgName>
          <address>
            <addrLine>Université Grenoble Alpes</addrLine>
            <addrLine>Bâtiment Stendhal</addrLine>
            <addrLine>CS40700</addrLine>
            <addrLine>38058 Grenoble cedex 9</addrLine>
          </address>
        </affiliation>
      </name>
    </respStmt>
    <respStmt>
      <resp>Transcription ODT</resp>
      <name>
        <persName>Jingrao LI</persName>
        <email>[REDACTED]</email>
        <affiliation>
          <orgName>Université Grenoble Alpes</orgName>
        </affiliation>
      </name>
    </respStmt>
    <respStmt>
      <resp>Encodage XML TEI</resp>
    </respStmt>
  </teiHeader>

```

```

<name>
  <persName>Jingrao LI</persName>
  <email>[REDACTED]</email>
  <affiliation>
    <orgName>Université Grenoble Alpes</orgName>
  </affiliation>
</name>
</respStmt>
</editionStmt>
<publicationStmt>
  <distributor>
    <name type="projet">Projet corpus d'apprenants </name>
    <address>
      <addrLine>LIDILEM</addrLine>
      <addrLine>Université Grenoble Alpes</addrLine>
      <addrLine>Bâtiment Stendhal</addrLine>
      <addrLine>CS40700</addrLine>
      <addrLine>38058 Grenoble cedex 9</addrLine>
    </address>
  </distributor>
  <availability>
    <p>Disponible gratuitement sous licence Creative Commons :
      http://creativecommons.org/licenses/by/2.0/ </p>
    <p>Cette licence implique la libre reproduction et diffusion de l'œuvre au public,
      notamment par téléchargement, mais est soumise aux conditions suivantes : </p>
    <p>Respect du droit à la Paternité : l'Utilisateur est tenu de citer le nom de l'auteur
      original de la manière indiquée par l'Auteur de l'oeuvre (mais pas d'une manière qui
      suggérerait que l'Auteur soutient ou approuve l'utilisation de l'œuvre par
      l'Utilisateur) ; </p>
    <p>Pas d'Utilisation Commerciale : l'Utilisateur ne peut exploiter ou diffuser l'œuvre à
      des fins commerciales ; </p>
    <p>Pas de Modification : l'Utilisateur ne peut modifier l'œuvre originale sans l'accord
      express de l'Auteur, quelle que soit la nature et l'importance de cette modification ; </p>
    <p>Partage des Conditions à l'Identique : L'Utilisateur ne peut diffuser l'œuvre que sous
      une licence identique ; et aux mêmes conditions. </p>
  </availability>
</publicationStmt>
<sourceDesc>
  <p/>
</sourceDesc>
</fileDesc>
<profileDesc>
  <textClass>
    <keywords>
      <term type="dateProduction">2018</term>
      <term type="nomEtablissement">Centre universitaire d'études françaises - UGA</term>
      <term type="typeEtablissement">centre universitaire français de formation en français langue
étrangère</term>
      <term type="pays">France</term>
      <term type="ville">Grenoble</term>
      <term type="discipline">sciences humaines et sociales</term>
      <term type="sousDiscipline">Français langue étrangère</term>
      <term type="niveauDiscipline"></term>
      <term type="nomDiplôme"> DU Lettres, Langue Française, Communication </term>
      <term type="genre">texte argumentatif</term>

```

<term type="evaluation">oui</term>
<term type="modaliteRedaction">examen sur table</term>
<term type="tempsDeProduction"></term>
<term type="consigne">Entraînement à l'épreuve de production écrite du DELF B2 -Prise de position argumentée : vous habitez dans une ville qui organise chaque année un grand concert gratuit pour marquer la fin de l'été. Pour des raisons financières, votre ville vous annonce qu'elle veut supprimer cet événement musical. Vous écrivez au maire de la ville pour le persuader, à l'aide d'arguments et d'exemples précis, des avantages culturels et touristiques que ce concert représente. Vous insistez également sur l'intérêt économique de cette manifestation pour les commerçants et les artistes de la région. (250 mots minimum) </term>

<term type="modeRedaction">rédaction manuscrite</term>

<term type="sexe">F</term>
<term type="âge">20</term>
<term type="nationalité">chinoise</term>
<term type="langueMaternelle">chinois</term>
<term type="paysDeRésidence">France</term>
<term type="langueParléeAvecParents|frères|sœurs">cantonais</term>
<term type="langueMaternelleParents|frères|sœurs">cantonais</term>
<term type="langueParléeAvecConjointEt|OuEnfants"></term>
<term type="langueMaternelleConjointEt|OuEnfants"></term>

<term type="étudesEnCours">DULLFC</term>
<term type="niveauEtudesDansLePaysD'origine">Licence</term>
<term type="spécialitéDansLePaysD'origine">français</term>
<term type="nombreAnnéesD'étudesDuFrançais">3</term>
<term type="duréeDeSéjourEnFrance">1 an</term>
<term type="autresSéjoursEnFrance/paysFrancophone">Non</term>
<term type="niveauCECRLestimé">B2</term>
<term type="autresLanguesConnues">anglais</term>
<term type="séjoursEtranger"></term>
<term type="langueEtudesTravail">français</term>
<term type="langueEchangesAmis">chinois</term>
<term type="langueLecture">chinois</term>
<term type="langueTélévisionCinéma"></term>
<term type="nombreLivresEnFrançaisParAn">2</term>
<term type="moyenAideRédaction"></term>
<term type="relecturesPersonnesExtérieures"></term>

</keywords>

</textClass>

<langUsage>

<language ident="fr">français</language>

</langUsage>

</profileDesc>

</teiHeader>

<text>

<body>

<p>██████████ Grenoble, le 24 Avril, 2018</p>

<p>à ██████████ le maire de Grenoble</p>

<p>Objectif: Demande de <w pos="v" lemma="continuer">continuer</w> à <w pos="v" lemma="organiser" type="MORPHO" subtype="ORT">organiser</w> le concert gratuit à la fin de l'été.</p>

<p>Monsieur le maire</p>

<p>En tant qu'une citadine dans cette ville, nous <w pos="v" lemma="organiser">organisons</w> le concert gratuit chaque année, c'<w pos="v" lemma="être">est</w> la marque qui <w pos="v" lemma="être">est</w>

lemma="signifier">signifie</w> que l'été <w pos="v" lemma="terminer" type="MORPHO" subtype="CA">est terminée</w> Chacun <w pos="v" lemma="être">est</w> fier de cet évènement musical. En revanche, vous <w pos="v" lemma="vouloir" type="SEMAN" subtype="TAM">voudriez</w><w pos="v" lemma="annuler">annuler</w> cet évènement en raison des problèmes financiers. En fait, ce concert musical <w pos="v" lemma="pouvoir">pourrait</w> nous <w pos="v" lemma="apporter">apporter</w> beaucoup d'avantages sous plusieurs d'aspects.</p>

<p>Nous <w pos="v" lemma="savoir">savons</w><w pos="v" lemma="organiser" type="MORPHO" subtype="ORT">que, organiser</w> un concert gratuit <w pos="v" lemma="exiger">exige</w> pas mal d'argent. Cela <w pos="v" lemma="faire" type="SEMAN" subtype="COC/CONO">fait</w> une partie du budget de notre ville. Par exemple, il <w pos="v" lemma="falloir">faut</w> <w pos="v" lemma="recruter">recruter</w> beaucoup de serveurs, <w pos="v" lemma="préparer">préparer</w> des desserts et <w pos="v" lemma="inviter">inviter</w> les chanteurs.</p>

<p>Mais, en même temps, cet évènement musical <w pos="v" lemma="permettre">permet</w> à notre ville d'<w pos="v" lemma="obtenir" type="SEMAN" subtype="COC/CONO">obtenir</w> plus de profits. Tout d'abord, des citoyens <w pos="v" lemma="pouvoir">pourraient</w><w pos="v" lemma="profiter">profiter</w> du concert gratuit pour <w pos="v" lemma="connaître">connaître</w> et <w pos="v" lemma="apprendre">apprendre</w> profondément des cultures musicales. Par exemple, tout le monde, y compris les personnes qui <w pos="v" lemma="être">sont</w> chanteurs, étudiants, professeurs <w pos="v" lemma="étudier" type="SEMAN" subtype="TAM">étudiés</w> la musiques, les amateurs, <w pos="v" lemma="aller" type="MORPHO" subtype="CA">vont</w> au concert. Les personnes professionnelles <w pos="v" lemma="pouvoir">peuvent</w><w pos="v" lemma="diffuser">diffuser</w> des connaissances de la culture musicale aux citoyens. Ensuite, cet évènement musical <w pos="v" lemma="pouvoir">pourrait</w><w pos="v" lemma="être">être</w> le symbole de notre ville. C'<w pos="v" lemma="être">est</w> la démarche qui <w pos="v" lemma="être">est</w> atout du tourisme. Grâce au concert gratuit, cela <w pos="v" lemma="pouvoir">pourrait</w><w pos="v" lemma="attirer">attirer</w> beaucoup de gens. En outre, <w pos="v" lemma="organiser">organiser</w> le concert gratuit <w pos="v" lemma="promouvoir">promeut</w> le développement économique. Autrement dit, des marchands et des artistes <w pos="v" lemma="pouvoir">pourraient</w><w pos="v" lemma="obtenir" type="SEMAN" subtype="COC/CONO">obtenir</w> plus d'argent. Supposons qu'il y <w pos="v" lemma="être" type="SEMAN" subtype="TAM">a</w> un concert gratuits, des artistes <w pos="v" lemma="pouvoir">pourraient</w><w pos="v" lemma="inviter">être invités</w> pour <w pos="v" lemma="participer">participer</w> au concert. En même temps, des marchands <w pos="v" lemma="pouvoir">peuvent</w><w pos="v" lemma="aller">aller</w> au concert pour <w pos="v" lemma="vendre">vendre</w> des pizzas, des spécialités etc.</p>

<p>En somme, chacun <w pos="v" lemma="pouvoir">pourrait</w><w pos="v" lemma="profiter">profiter</w> du concert gratuit pour <w pos="v" lemma="obtenir" type="SEMAN" subtype="COC/CONO">obtenir</w> des profits. Et cet évènement <w pos="v" lemma="être">est</w> important pour les citadins dans notre ville. J'<w pos="v" lemma="espérer">espère</w> que le concert gratuit <w pos="v">continuera</w> à <w pos="v" lemma="organiser" type="GRAMSYN" subtype="VV">organiser</w>.</p>

<p>Dans l'attente de votre réponse, je vous <w pos="v" lemma="prier">prie</w> d'<w pos="v" lemma="accepter">accepter</w>, Monsieur le maire, l'expression de mes sentiments respectueux.</p>

<p rend="right">[REDACTED]</p>

<p rend="right">(289 mots)</p>

</body>

</text>

</TEI>

Annexe 4 : Exemple de corpus annoté (texte à la maison)

```
<?xml version="1.0" encoding="UTF-8"?>
<!--odt2tei: 2019-04-08T10:03:53Z-->
<?xml-stylesheet type="text/css" href="../../annotation_verbes.css"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title/>
      </titleStmt>
      <editor>
        <name>LIDILEM</name>
        <address>
          <addrLine>Université Grenoble Alpes</addrLine>
          <addrLine>Bâtiment Stendhal</addrLine>
          <addrLine>CS40700</addrLine>
          <addrLine>38058 Grenoble cedex 9</addrLine>
        </address>
      </editor>
    </titleStmt>
    <editionStmt>
      <edition>Edition électronique <date>2018-2019</date>
    </edition>
    <respStmt>
      <resp>Responsable du projet</resp>
      <name>
        <persName>[REDACTED]</persName>
        <email>[REDACTED]</email>
        <affiliation>
          <name>LIDILEM</name>
          <orgName>Université Grenoble ALpes</orgName>
          <address>
            <addrLine>Université Grenoble Alpes</addrLine>
            <addrLine>Bâtiment Stendhal</addrLine>
            <addrLine>CS40700</addrLine>
            <addrLine>38058 Grenoble cedex 9</addrLine>
          </address>
        </affiliation>
      </name>
    </respStmt>
    <respStmt>
      <resp>Encodage XML TEI</resp>
      <name>
        <persName>Jingrao LI</persName>
        <email>[REDACTED]</email>
        <affiliation>
          <orgName>Université Grenoble Alpes</orgName>
        </affiliation>
      </name>
    </respStmt>
  </editionStmt>
  <publicationStmt>
```

```

<distributor>
  <name type="projet">Projet corpus d'apprenants </name>
  <address>
    <addrLine>LIDILEM</addrLine>
    <addrLine>Université Grenoble Alpes</addrLine>
    <addrLine>Bâtiment Stendhal</addrLine>
    <addrLine>CS40700</addrLine>
    <addrLine>38058 Grenoble cedex 9</addrLine>
  </address>
</distributor>
<availability>
  <p>Disponible gratuitement sous licence Creative Commons :
  http://creativecommons.org/licenses/by/2.0/ </p>
  <p>Cette licence implique la libre reproduction et diffusion de l'œuvre au public,
  notamment par téléchargement, mais est soumise aux conditions suivantes : </p>
  <p>Respect du droit à la Paternité : l'Utilisateur est tenu de citer le nom de l'auteur
  original de la manière indiquée par l'Auteur de l'oeuvre (mais pas d'une manière qui
  suggérerait que l'Auteur soutient ou approuve l'utilisation de l'œuvre par
  l'Utilisateur) ; </p>
  <p>Pas d'Utilisation Commerciale : l'Utilisateur ne peut exploiter ou diffuser l'œuvre à
  des fins commerciales ; </p>
  <p>Pas de Modification : l'Utilisateur ne peut modifier l'œuvre originale sans l'accord
  express de l'Auteur, quelle que soit la nature et l'importance de cette modification ; </p>
  <p>Partage des Conditions à l'Identique : L'Utilisateur ne peut diffuser l'œuvre que sous
  une licence identique ; et aux mêmes conditions. </p>
</availability>
</publicationStmt>
<sourceDesc>
  <p/>
</sourceDesc>
</fileDesc>
<profileDesc>
  <textClass>
  <keywords>
    <term type="dateProduction">2018</term>
    <term type="nomEtablissement">Centre universitaire d'études françaises - UGA</term>
    <term type="typeEtablissement">centre universitaire français de formation en français langue
étrangère</term>
    <term type="pays">France</term>
    <term type="ville">Grenoble</term>
    <term type="discipline">sciences humaines et sociales</term>
    <term type="sousDiscipline">Français langue étrangère</term>
    <term type="niveauDiscipline"></term>
    <term type="nomDiplôme"> DU Lettres, Langue Française, Communication </term>
    <term type="genre">texte argumentatif</term>
    <term type="evaluation">oui</term>
    <term type="modaliteRedaction">dossiers à la maison</term>
    <term type="tempsDeProduction"></term>
    <term type="consigne"></term>
    <term type="modeRedaction">rédaction à l'ordinateur</term>

    <term type="sexe">F</term>
    <term type="âge">21</term>
    <term type="nationalité">chinoise</term>
    <term type="langueMaternelle">chinois</term>
    <term type="paysDeRésidence">France</term>

```

```

<term type="langueParléeAvecParents|frères|sœurs">chinois</term>
<term type="langueMaternelleParents|frères|sœurs">chinois</term>
<term type="langueParléeAvecConjointEt|OuEnfants">chinois</term>
<term type="langueMaternelleConjointEt|OuEnfants">chinois</term>

<term type="étudesEnCours">droit et français et sciences du langage</term>
<term type="niveauEtudesDansLePaysD'origine">Licence</term>
<term type="spécialitéDansLePaysD'origine">français</term>
<term type="nombreAnnéesD'étudesDuFrançais">3</term>
<term type="duréeDeSéjourEnFrance">1 an</term>
<term type="autresSéjoursEnFrance/paysFrancophone">Non</term>
<term type="niveauCECRLestimé">B2</term>
<term type="autresLanguesConnues">anglais/italien</term>
<term type="séjoursEtranger"></term>
<term type="langueEtudesTravail">français</term>
<term type="langueEchangesAmis">français/chinois</term>
<term type="langueLecture">français/chinois/anglais</term>
<term type="langueTélévisionCinéma"></term>
<term type="nombreLivresEnFrançaisParAn"></term>
<term type="moyenAideRédaction">dictionnaire bilingue/dictionnaire monolingue</term>
<term type="relecturesPersonnesExtérieures">locuteurs natif du français/locuteurs non natifs du
français</term>
</keywords>
</textClass>
<langUsage>
<language ident="fr">français</language>
</langUsage>
</profileDesc>
</teiHeader>
<text>
<body>

```

<p><w pos="v" lemma="remonter">Remontons</w> le fil de l'histoire, la génération de nos grands-
 parents <w pos="v" lemma="revendiquer">revendique</w> rarement le bonheur au travail car on <w
 pos="v" lemma="bossier" type="CONT" subtype="REG">bossait</w> juste pour <w pos="v"
 lemma="subvenir">subvenir</w> à ses besoins. Mais le critère du bonheur <w pos="v"
 lemma="évoluer">évolue</w> tout le temps : de nos jours, le travail n'<w pos="v"
 lemma="considérer">est plus considéré</w> comme une activité avilissante mais bien au contraire, comme
 un vecteur d'émancipation. Par conséquent, l'absence d'accident, de maladie professionnels et un bon
 salaire ne <w pos="v" lemma="suffir" type="MORPHO" subtype="CA">suffissent</w> plus à le <w
 pos="v" lemma="définir">définir</w>. Par contre, le bien-être spirituel et physique <w pos="v"
 lemma="jouer">joue</w> un rôle de plus en plus important. On <w pos="v"
 lemma="constater">constate</w> que le monde du travail <w pos="v" lemma="se diviser">se divise</w>
 en deux catégories : ceux qui <w pos="v" lemma="aimer">aiment</w> leurs boulots et ceux qui <w
 pos="v" lemma="se contenter">se contentent</w> de <w pos="v" lemma="gagner">gagner</w> sa croûte.
 Ce phénomène nous <w pos="v" lemma="amener">amène</w> à <w pos="v" lemma="s'interroger"
 type="MORPHO" subtype="CA">s'interroger</w> : est-ce que <w pos="v" lemma="être">être</w>
 heureux au boulot <w pos="v" lemma="être">est</w> une utopie que <w pos="v"
 lemma="faire">fait</w> <w pos="v" lemma="rêver">rêver</w> tout les salariés ?</p>

<p>Avant de <w pos="v" lemma="commencer">commencer</w>, <w pos="v"
 lemma="permettre">permettez</w>-moi de vous <w pos="v" lemma="annoncer">annoncer</w> mon
 plan. D'abord, je <w pos="v" lemma="aller">vais</w> <w pos="v" lemma="parler">parler</w> de deux
 paradoxes dans le monde du travail en <w pos="v" lemma="donner">donnant</w> quelques exemples du
 malheur au travail. Après, je <w pos="v" lemma="aller">vais</w> <w pos="v"
 lemma="proposer">proposer</w> comment <w pos="v" lemma="argumenter">augmenter</w> le bonheur.
 Enfin, cet exposé <w pos="v" lemma="déboucher">débouche</w> sur une conclusion qui <w pos="v"
 lemma="résumer">résume</w> mes idées principales.</p>

<p>Le fait que la plupart d'entre nous <w pos="v" lemma="avoir">a</w> la hantise de <w pos="v" lemma="reprendre">reprendre</w> le travail chaque dimanche soir <w pos="v" lemma="témoigner">témoigne</w> que notre travail nous <w pos="v" lemma="apporter">apporte</w> plus d'aliénation que la gratification.</p>

<p>Le paradoxe c'<w pos="v" lemma="être">est</w> que le code de la déontologie <w pos="v" lemma="préconiser">préconise</w> le bonheur au travail et l'épanouissement des salariés alors qu'au sein de l'entreprise, on <w pos="v" lemma="mettre">met</w> en avant les intérêts économiques et la productivité qui <w pos="v" lemma="obliger" type="GRAMSYN" subtype="PREP/EP">obligent</w> aux employés de <w pos="v" lemma="sacrifier">sacrifier</w> leur bien-être.</p>

<p>D'une enquête de France 2, j'<w pos="v" lemma="apprendre">apprends</w> que dans les supermarchés drive, les employés <w pos="v" lemma="travailler">travaillent</w> à des cadences effrénées et sous une pression forte pour <w pos="v" lemma="trouver">trouver</w> les courses que les clients <w pos="v" lemma="commander">ont commandées</w> en ligne. Cette condition abominable du travail <w pos="v" lemma="transformer">transforme</w> les salariés en robots. De plus, la contrainte et le stress au travail <w pos="v" lemma="nuire" type="GRAMSYN" subtype="ORD">nuisent</w> non seulement à la performance, mais aussi <w pos="v" lemma="forcer">forcent</w> certains salariés à <w pos="v" lemma="se renfermer">se renfermer</w>, qui en <w pos="v" lemma="résulter" type="GRAMSYN" subtype="SYN/CS">résulte</w> une baisse de capacité et la détérioration de la santé. C'<w pos="v" lemma="être">est</w> la raison pour laquelle, certains entre eux <w pos="v" lemma="porter">portent</w> plainte contre leurs sociétés pour <w pos="v" lemma="dénoncer">dénoncer</w> ce calvaire. Malheureusement, ce n'<w pos="v" lemma="être">est</w> pas un cas isolé. Je <w pos="v" lemma="souvenir">me souviens</w> qu'en Chine, quelques travailleurs de la chaîne de montage <w pos="v" lemma="succomber">ont succombé</w> à leurs tâches monotones et ensuite <w pos="v" lemma="se suicider" type="MORPHO" subtype="CA">s'est suicidé</w>.</p>

<p>Le deuxième paradoxe c'<w pos="v" lemma="être">est</w> que la quête incessante au bonheur, ce dogme qui <w pos="v" lemma="imposer">impose</w> l'euphorie permanente dans tous les métiers <w pos="v" lemma="être">est</w> quelques fois un fardeau au détriment du dynamisme des salariés. Il <w pos="v" lemma="être">est</w> incontestable que, dans le secteur tertiaire où la communication et relation sociale <w pos="v" lemma="être">sont</w> au coeur de la réussite, les employés <w pos="v" lemma="devoir">doivent</w> <w pos="v" lemma="arborer">arborer</w> un sourire et <w pos="v" lemma="transmettre">transmettre</w> la bonne humeur au client. En revanche, on <w pos="v" lemma="pouvoir">peut</w> <w pos="v" lemma="garder">garder</w> notre humeur réelle quand on <w pos="v" lemma="être">est</w> devant un ordinateur froid où d'autres objets inhumains. Si dans ce cas-là, on <w pos="v" lemma="prôner">prône</w> encore des formules comme « il <w pos="v" lemma="falloir">faut</w> <w pos="v" lemma="sourire">sourire</w> tous les jours », ce <w pos="v" lemma="guinder">sera guindé</w> et ne <w pos="v" lemma="servir" type="SEMAN" subtype="TAM">sert</w> qu'à <w pos="v" lemma="repousser" type="SEMAN" subtype="SEN/SIN">repousser</w> les employés. En outre, notre expression d'une vraie émotion <w pos="v" lemma="servir">sert</w> d'un feedback qui <w pos="v" lemma="permettre">permet</w> aux managers de <w pos="v" lemma="modifier">modifier</w> la condition du travail et qui <w pos="v" lemma="contribuer" type="MORPHO" subtype="CA">contribuent</w> à la maintenance de la relation entre les collègues. Par ailleurs, Je <w pos="v" lemma="convaincre">suis convaincue</w> que le bonheur au travail ne <w pos="v" lemma="provenir">provient</w> pas de l'injonction, car il <w pos="v" lemma="être">est</w> une sensation spontanée qui <w pos="v" lemma="émaner">émane</w> de l'autonomie, de l'intégration à l'équipe et à la vie communautaire.</p>

<p>A l'heure numérique, je <w pos="v" lemma="croire">crois</w> que le télétravail <w pos="v" lemma="pouvoir">peut</w> <w pos="v" lemma="apporter">apporter</w> le bonheur aux salariés car il leur <w pos="v" lemma="permettre">permet</w> une grande flexibilité et une conciliation entre leurs vies professionnelle et familiales. De plus, toutes les tâches <w pos="v" lemma="dénouer">dénuées</w> de réflexion, fastidieuses et ingrates <w pos="v" lemma="pouvoir" type="MORPHO" subtype="CA">peut</w> <w pos="v" lemma="déléguer" type="GRAMSYN" subtype="PREP/EP">être déléguées</w> par des machines. Donc les sociétés <w pos="v" lemma="devoir">doivent</w> <w pos="v" lemma="mettre">mettre</w> en place la robotisation pour que les hommes <w pos="v" lemma="pouvoir">puissent</w> <w pos="v" lemma="accomplir">accomplir</w> le travail compliqué et <w pos="v" lemma="sublimier">sublime</w> qui <w pos="v" lemma="pouvoir">peut</w> leur <w

```
pos="v" lemma="procurer">procurer</w> le bonheur.</p>
  <p>Pour <w pos="v" lemma="conclure">conclure</w>, le bonheur au travail <w pos="v"
lemma="être">est</w> primordial, néanmoins, il ne <w pos="v" lemma="pouvoir">peut</w> pas <w
pos="v" lemma="être">être</w> un objectif mais un résultat. Les sociétés <w pos="v"
lemma="devoir">doivent</w> <w pos="v" lemma="mettre">mettre</w> au point de nouvelles
technologies et <w pos="v" lemma="mettre">mettre</w> en place de nouvelles façons de travail qui <w
pos="v" lemma="permettre">permettent</w> aux salariés de <w pos="v"
lemma="s'épanouir">s'épanouir</w> au boulot.</p>
  </body>
</text>
</TEI>
```

Annexe 5: Guide d'annotation

Classes d'erreurs	Types d'erreurs	Exemples
<p>Au niveau morphologique (MORPHO)</p>	<p>Erreurs de conjugaison et accord (CA)</p> <ul style="list-style-type: none"> - mal conjugué - l'accord de l'auxiliaire - l'accord du participe passé - l'accord du pronom dans les verbes pronomiaux 	<p>Le narrateur choit (choisit) de faire une explication supplémentaire qui, néanmoins, ne se consacre pas au déroulement.</p> <p>ça pourra promouvoir l'économie car le tourisme deviennera (deviendra) plus florissant.</p> <p>Les artistes de la région obtient (obtiennent) la chance de se présenter</p> <p>Comme vous le savez, ce concert a (est) devenu une partie importante dans notre vie.</p> <p>C'est (ce sont) les corrélateurs qui sont consacrés à relier deux événements réellement arrivés.</p> <p>La manière que nous avons adopté (adoptée) ici pour calculer cette distance est « distance euclidienne ».</p> <p>En tant qu'un concert gratuit, il nous permet de se retrouver (nous retrouver) et écouter la musique ensemble sous une atmosphère agréable.</p>

		C'est la marque qui signifie que l'été est terminée . (terminé)
	Erreurs d'orthographe d'usage (ORT)	<p>Nous vous persuadons de ne pas supprimer (supprimer) cet évènement et de ne pas nous arrêter de nous plaire.</p> <p>Nous ne trovons (trouvons) pas supprimer le concert est une bien décision.</p> <p>Il vaut mieux de le maintenir pour ouvrire (ouvrir) progressivement le marché du tourisme.</p> <p>Je vous pris d'agéer (d'agréer), monsieur le maire, l'expression de mon profond respect.</p> <p>Il y a beaucoup d'informations que nous ne pouvons pas utiliser directement, mais cela peut nous aider à mieux réflichir (réfléchir) à notre sujet.</p> <p>C'est très utile de obtenir (d'obtenir) une bonne réputation de notre ville.</p>
	<p>- les erreurs d'écriture qui ne nuisent pas énormément à la prononciation</p> <p>- l'élision</p>	
	Erreurs de verbes fictifs, néologismes (VFIC)	<p>Également, ils sont capables d'exhibitionner (montrer) leur talent musical sur scène.</p> <p>Ce concert est une bonne chance de faire l'échange et se progressiver (progresser).</p> <p>Mais je sais que la ville est en train de facier (est face) aux problèmes financiers, je pense qu'on peut gagner de l'argent par le concert quand même.</p> <p>La chute, l'un des procédés grâce auxquels le narrateur signale que son récit</p>
	- la création des formes non existantes	

			<p>est achevé, en pratique, elle globe (englobe) toutes les formes, des plus simples aux plus complexes.</p> <p>Pour profondir (approfondir) sa recherche, elle se charge aussi d'une collection sur les musées dans monde entier qui s'appelle « Musées-Mondes ».</p>
<p>Au niveau sémantique et des cooccurrences lexicales (SEMAN)</p>	<p>Erreurs de sens lexical (SEN)</p>	<p>L'utilisation inappropriée des synonymes (SNM)</p>	<p>Il y a beaucoup de techniques pour enrichir la syntaxe narrative mais nous confinons (limitons) nos études aux éléments évaluatifs qui concernent directement à la structure interne.</p> <p>Ce graphique démontre (montre) le nombre des temps utilisés par le texte mystérieux.</p> <p>Les danseurs interprètes sont chargés non seulement d'exécuter des mouvements, mais aussi d'incarner des personnages, de jouer des situations dramatiques, voire de proférer (réciter) du texte.</p>
		<p>- ceux qui partagent des mêmes composantes du sens et substituables dans certains contextes</p>	<p>En 1951, L'éducation des sourds est intégrée dans le système national d'éducation, les associations de la protection des sourds se propagent (se développent).</p> <p>Cette manière présente (montre) que pour un ensemble de mots des textes A et B, on peut calculer la distance entre deux textes</p> <p>Un lien faible ne nous conduit pas à maintenir une relation dente mais il peut amener (apporter) toutes sortes d'informations à valeur importante.</p>
		<p>Le sens inadapté (SIN)</p>	<p>Nous allons nous pencher sur la surdité en matière identitaire et culturelle, et</p>

	<p>- le sens inadapté</p>	<p>- mauvais attribution d'une acception à une forme existante</p>	<p>mettre en œuvre (mettre en évidence) dans quel contexte les langues des signes évoluent et comment l'identité sourde se construit.</p> <p>Et d'après les colonnes dans le graphique, les phrases utilisées dans le texte mystère sont plutôt courtes, ce qui nous constate (montre) que la façon d'écriture du texte mystère est plus proche de celle de Zola.</p> <p>Grâce à ce site web, la chaîne TV5 MONDE arrive à se transformer, en conquérant (dépassant) les limites de la télévision traditionnelle.</p> <p>Puisque vous êtes en face des raisons financières, j'ai trouvé quelque solutions pour rédigier (résoudre) ce problème.</p> <p>Je me permet de vous bien vouloir examiner la proposition consistant à faire endurer (durer) le festival musical à la fin de l'été.</p> <p>On peut dépasser (passer) ce jour comme un fête, même si on aime plus la ville grâce au concert.</p> <p>Je me permets de vous écrire, en tant que citoyenne grenobloise, pour vous demander de réserver (conserver) l'organisation annuelle du concert estival dans notre belle ville.</p>
	<p>Erreurs de temps/d'aspect/de mode des verbes (TAM)</p>	<p>- Temps : le passé, le présent, le futur</p>	<p>Comme vous le saviez (savez), depuis l'existence de cette fête, un nouveau moyen de se reposer s'ajoute dans la vie des citoyens.</p> <p>Supposons qu'il y a (ait) un concert gratuits, des artistes pourraient être invités pour participer au concert.</p>

	<p>- Aspect : l'action accomplie, l'action inaccomplie</p> <p>- Mode : l'indicatif, le subjonctif, l'impératif, l'infinitif, le conditionnel et le participe.</p>	<p>Puisque la situation financière de notre ville soit (n'est) pas si satisfaisant, nous pouvons toujours le(le concert) représente par une réduction du budget.</p> <p>À part la passion que les habitants avoir (ont), de nombreux raisons sont quand même raisonnable.</p> <p>Tout le monde, y compris les personnes qui sont chanteurs, étudiants, professeurs étudiés (étudiant) la musiques, les amateurs, vont au concert.</p> <p>On trouvera une diversité d'articles lié au domaine médiatique classant (classés) selon la date de publication.</p>
	<p>Erreurs de cooccurrence avec les noms (CONO)</p> <p>- erreurs du verbe dans les cooccurrences V+N</p> <p>- erreurs du nom dans les cooccurrences V+N</p>	<p>La deuxième manière est les comparateurs qui désigne (mesure) l'écart des événements narratifs pour envisager des possibilités non réalisées.</p> <p>Les adultes one déjà appris à bien s'exprimer et à donner (raconter) des vécus sous une structure plus complexe.</p> <p>Après, tout le monde dit dans un monologue (prononce un monologue).</p> <p>Les amateurs de la musique, notamment ceux qui ne sont pas très habituels à assister à un concert, sont (ont) accès à divers genres de musiques grâce à la gratuité de cet événement.</p> <p>Ce n'est pas un moyen pertinent de faire toujours des compris (compromis) avec les concurrents tels que Facebook.</p>
	<p>Erreurs de</p>	<p>En l'utilisant, la locutrice, qui emploie d'ordinaire plutôt l'évaluation</p>

	Erreurs de cooccurrence (COC)	cooccurrence avec les adverbes et les compléments circonstanciels (COA)	<p>externe, se relève (se révèle) capable d'enchâsser plus profond (profondément) ses commentaires.</p> <p>Ce type peut se constituer d'ordinaire (facilement) mais manifeste profondément ce que le locuteur a évalué.</p> <p>Ce changement de la langue du signe conduit exactement (inévitablement) à l'incompréhension des interprétations.</p> <p>Ces parties se lient avec des façons (d'une façon) les plus simples (plus simple).</p>
<p>Au niveau de la grammaire et de la syntaxe</p> <p>(GRAMSYN)</p>	<p>Erreurs de voix des verbes (VV)</p> <p>(erreur de construction passive à la place active ou inversement, erreurs de pronominaux de sens passif)</p> <ul style="list-style-type: none"> - la voix passive - la voix active - la voix moyenne (pronominal de sens passif) 	<p>Nous pouvons comparer les manières d'étude qui restent là ou qui ont été évoluées (qui ont évolué) avec les outils utilisés actuellement dans les musées de Grenoble.</p> <p>Sous le couvert de l'interrogation, les actes peuvent interpréter (peuvent être interprétés) comme demande d'information, demande d'action, défi, passage à l'acte.</p> <p>Ces événements présentent (se présentent) également au présent.</p>	
	<p>Erreurs de pronominaux (EPN)</p> <ul style="list-style-type: none"> - pronominal de sens réfléchi 	<p>Elle vêt (se vêt) d'un t-shirt à manche court rose moulant.</p> <p>Par exemple, on peut se rencontrer (rencontrer) les gens qui viennent des</p>	

	<ul style="list-style-type: none"> - pronominal de sens réciproque - pronominal de sens absolu 	<p>régions différentes et renforcer des amitiés.</p> <p>Il faut rendre compte (se rendre compte) que tous les verdicts ne sont pas sans appel car la cour d'appel et la cour de cassation sont là pour contester des décisions litigieuses et inopportunes.</p>
	<p>Erreurs d'ordre des mots (ORD)</p> <ul style="list-style-type: none"> - l'adverbe - la négation - les pronoms 	<p>Il est significatif puis-que nous pouvons ainsi développer mieux (mieux développer) l'industrie.</p> <p>Je voudrais vous expliquer les avantages que ce concert représente et vous peut-être changez (changez peut-être) votre avis.</p> <p>Elle principalement est utilisée (elle est principalement utilisée) dans l'enseignement de l'anglais pour aider les élèves sourds.</p> <p>Demande de ne supprimer pas (ne pas supprimer) le concert</p> <p>L'année dernière, plus de 1000 tourists sont y allés (y sont allés) et des artistes les plus célèbres se sont concentrés dans notre ville.</p> <p>Par ailleurs, dans le monde scientifique, on est tenu d'être presque parfait pour ne pas faire les autres souffrir (faire souffrir les autres).</p>
	<p>Prépositions surnuméraire</p> <p>(PS)</p>	<p>J'attends de votre réponse (J'attends votre réponse), M. le maire, et je vous remercie par avance.</p> <p>Il vaut mieux de le (le concert) maintenir (Il vaut mieux le maintenir) pour ouvrir progressivement le marché du</p>

	Erreurs de préposition (PREP)		<p>tourisme.</p> <p>Chaque année, il y a beaucoup de français qui viennent de tenter (viennent tenter) leur chance.</p> <p>Comme on a dit dans la partie de l'analyse de texte, cet article visé non seulement aux travailleurs (visé non seulement les travailleurs) dans l'édition et les plateformes des réseaux sociaux, mais aussi au grand public (mais aussi le grand public) qui lit les lectures.</p>
		Omission des prépositions (OP)	<p>La grande majorité de notre citoyens ont envie de participer (à) cet événement musical.</p> <p>Cette partie sert à s'arrêter le récit pour faire l'auditeur mieux savoir (faire mieux savoir à l'auditeur) où est l'intérêt de ce récit.</p> <p>La répétition est un genre où on répète les événements qu'on veut insister (sur lesquels on veut insister).</p> <p>Il est possible que le narrateur parle plus de l'évaluation que le récit (que du récit).</p> <p>À la fin, les trois sortent (la scène ensemble.</p>
		Utilisation erronée des prépositions	<p>On pourrait encourager notre peuple de (à) chanter.</p> <p>Des artistes pourraient être invités pour (à) participer au</p>

		(EP)	<p>concert.</p> <p>L'attache complexe est composée par (de) deux propositions enchâssées au même point dans la phrase matrice.</p> <p>Les différents outils pris dans la recherche muséale dans le passé peuvent également servir aux (de) références pour nous aider à étudier les outils actuels.</p>
	<p>L'utilisation erronée de pronom (PRO)</p> <p>- pronom erroné découlant d'une confusion sur les propriétés combinatoires syntaxiques du verbe</p>		<p>Le concert est important, ne seulement pas les commerçant, mais aussi les artistes, peuvent le bénéficier (en bénéficiant).</p> <p>Il est plutôt comme un déclencheur du texte, ce qui permet de donner envie aux lecteurs de savoir ce qui se passe en réalité et leur (les) pousse à continuer de (à) lire le contenu suivant.</p> <p>Le narrateur s'arrête l'action d'événement pour attirer l'attention de l'auditeur et les (lui) faire mieux comprendre que ce qui s'est passé.</p>
	<p>Structure syntaxique erronée (SYN)</p>	<p>Redondance syntaxique (RS)</p> <p>- deux verbes principaux dans une phrase mais un seul est nécessaire.</p>	<p>Je vous prie d'agréer, le maire de la ville, je vous présente mes respectueuses salutations. (Je vous prie d'agréer mes respectueuses salutations)</p> <p>Enfin, il y a sans doute qu'il existe des avantages pour des artiste, ce concert est une bonne chance de faire l'échange. (Il y a sans doute des avantages/Il existe des avantages)</p> <p>Notre ville plein de patrimoines culturels nécessite un moyen efficace en vue</p>

		<p>- paraphrase</p> <p>- pléonasme</p>	<p>de garder et conserver (conserver) tout cela.</p> <p>Les paroles de Juana nous ont donné une profonde influence (profondément influencés).</p> <p>Chaque mot du modèle de phrase est correspondant à (correspond à) un signe afin que l'élève puisse s'exercer à utiliser le modèle de phrase après avoir appris le nouveau mot.</p> <p>Je croie que ce concert gratuit peut attirer beaucoup de gens des autres villes venir ici. (peut attirer beaucoup de gens des autres villes.)</p>
		<p>Omission syntaxique (OS)</p> <p>- l'omission de la conjonction</p> <p>- l'omission du verbe</p> <p>- l'omission du complément</p> <p>- l'omission de la reprise du sujet dans l'inversion sujet-verbe</p>	<p>Nous ne trouvons pas (que) supprimer le concert est une bien décision.</p> <p>Pour les commerçants, il peuvent vendre plus de choses puisqu'il y a beaucoup de gens (qui) viennent à la ville.</p> <p>Je sais que cette décision (est prise) à cause des raisons financières, mais je voudrais vous expliquer les avantages que ce concert représente et vous peut-être changez votre avis.</p> <p>Je me permet de vous bien vouloir examiner (de vous demander de bien vouloir examiner) la proposition consistant à faire endurer le festival musical à la fin de l'été.</p> <p>Les musiciens auront la motivation à créer. (créer les musiques)</p> <p>Pourquoi les jeunes adultes prennent (prennent-ils) une alimentation néfaste à la santé en sachant clairement l'importance de l'alimentation?</p>

			<p>Peut-être ce documentaire veut (veut-il) nous montrer ce qu'il faut savoir avant de plonger dans la Silicon Valley pour les entrepreneurs français.</p>
		<p>Erreurs de construction syntaxique (CS)</p> <p>- fusion de plusieurs structures syntaxiques</p> <p>- construction verbale inexistant</p>	<p>Plus profondément, cela va ouvrir une plus grande porte pour nous d'aller plus loin (pour aller plus loin/ pour que nous puissions aller plus loin)</p> <p>L'auteur nous met en garde (met en garde en expliquant que) que l'évaluation est parsemé (présente) dans tout le récit, non pas dans un seul endroit, bien qu'elle soit centrée dans le "foyer".</p> <p>Je semble (il m'a semblé) vraiment m'intégrer dans cet ensemble.</p> <p>Un enchâssement plus profond s'agit de (permet de) faire intervenir une tierce personne qui évalue l'histoire à la place de narrateur.</p> <p>Enfin, il alerte les éditeurs de faire attention aux risques (alerte les éditeurs sur les risques/alerte les éditeurs pour qu'ils fassent attention aux risques) et les sensibilise à agir comme ce qu'a déjà fait le gros diffuseur Buzzfeed."</p> <p>Cela nous amène à réfléchir si on doit reconnaître le droit à l'erreur dans tous les cas ? (cela nous amène à réfléchir sur la reconnaissance du droit à l'erreur/Cela nous amène à nous demander si nous devons reconnaître le droit à l'erreur).</p>
Au niveau du genre et du contexte	Expression non conforme au registre de l'écrit au contexte scolaire (REG)		<p>On laisse tomber les travaux, les champs, les devoirs, juste pour écouter de la musique créée par nos musiciens locaux, on se souvient plus de soucis, de pression (on ne se souvient plus)</p>

(CONT)		<p>Remontons le fil de l’histoire, la génération de nos grands-parents revendique rarement le bonheur au travail car on bossait (travaillait) juste pour subvenir à ses besoins.</p> <p>Dans le monde scientifique, Edison a inventé l’ampoule après avoir tenté et fait capoter (raté) plein d’expériences.</p> <p>Par exemple, un livreur de repas a forte chance (a de fortes chances) de se faire virer (se faire licencier) s’il est en retard ou prend la mauvaise bouffe.</p>
--------	--	--

MORPHO :

CA : conjugaison du verbe avec le sujet, l’accord de l’auxiliaire et du participe passé, l’accord du pronom dans les verbes pronominaux.

ORT : les erreurs d’orthographe qui ne nuisent pas énormément à la prononciation ; l’elision

VFIC : la forme verbale inexistante, néologisme

SEMAN :

SEN : SNM, SFIC

SNM : l’utilisation inappropriée des synonymes (ceux qui partagent des mêmes composantes du sens et substituables dans certains contextes)

SFIC : sens lexical fictif, mauvais attribution d'une lexie à une forme existante.

TAM : le temps du verbe : le passé, le présent, le futur.

le mode du verbe : l'indicatif, le subjonctif, l'impératif, l'infinitif, le conditionnel et le participe.

l'aspect du verbe : l'action accomplie ou l'action inaccomplie

COC : CONO, COA

CONO : Erreurs de cooccurrence avec les noms : erreurs du verbe dans les cooccurrences V+N, erreurs du nom dans les cooccurrences V+N

COA : Erreurs de cooccurrence avec les adverbes et les compléments circonstanciels

GRAMSYN :

VV : la voix du verbe : la voix active, la voix passive, la voix moyenne (le pronominal de sens passif)

(Erreurs de construction passive à la place active ou inversement, erreurs des pronominaux de sens passif)

EPN : le pronominal de sens réfléchi, le pronominal de sens réciproque, le pronominal de sens absolu

(la fonction sémantique erronée des verbes pronominaux)

ORD : l'ordre des noms, pronoms , adverbes (y compris adverbe de négation) avec le verbe

PREP : RP, OP, EP

PS : les prépositions régies par les verbes en trop (utilisation transitive indirect des verbes transitif direct)

OP : l'omission des prépositions régies par les verbes (utilisation transitive direct des verbes transitif indirect)

EP : le choix erroné des prépositions régies par les verbes.

PRO : La pronominalisation erronée due à la confusion sur les propriétés combinatoires syntaxiques des verbes (maîtrise insuffisante sur la transitivité des verbes, ou les prépositions régies par les verbes).

SYN : RS, OS, CS

RS : la redondance du verbe principal dans une phrase, paraphrase, pléonasme

OS : l'omission des composantes régies par les verbes qui sont indispensables, comme l'omission du complément des verbes, l'omission de la conjonction *qui* ou *que* dans une proposition subordonnée suivie d'un verbe, l'omission des verbes eux-mêmes dans une phrase, l'omission de la reprise du sujet dans l'inversion sujet-verbe

CS : les constructions syntaxiques erronées liées aux verbes outre la redondance syntaxique et omission syntaxique, comme la fusion de plusieurs structures syntaxiques dans une phrase, construction verbale inexistant.

CONT :

REG : les expressions non conforme au registre formel, comme l'omission de *ne* dans la négation, l'utilisation des termes ou des structures familiers

Table des matières

Introduction.....	8
Chapitre I. Corpus d'apprenants.....	11
1.1 Le corpus d'apprenants et l'acquisition de la langue seconde.....	11
1.2 Le corpus d'apprenants et la didactique des langues.....	13
1.3 Les corpus d'apprenants dans les différentes langues.....	13
1.4 Les corpus d'apprenants pour le français.....	15
Chapitre II. Les erreurs dans les corpus d'apprenants.....	18
2.1 La notion d'« erreur ».....	18
2.1.1 L'« erreur » et la « norme ».....	18
2.1.2 La « maladresse » et la « faute ».....	20
2.2 L'intérêt d'étudier l'« erreur » et le système de l'interlangue.....	20
2.3 Les grilles d'erreurs des chercheurs.....	23
2.3.1 La typologie d'erreurs de Dagneaux pour l'anglais.....	23
2.3.2 Les typologies d'erreurs pour le français.....	24
2.3.2.1 Les caractéristiques sémantiques des grilles d'erreurs.....	24
2.3.2.2 Les caractéristiques syntaxiques des grilles d'erreurs.....	27
2.3.2.3 La typologie d'erreurs pour le corpus Frida.....	29
Chapitre III. Méthodologie : constitution du corpus et principes d'annotation.....	31
3.1 La méthodologie de traitement des données.....	31
3.2 La constitution d'une grille d'erreurs.....	34
3.2.1 Au niveau morphologique (MORPHO).....	34
3.2.1.1 Erreurs de conjugaison et d'accord (CA).....	34
3.2.1.2 Erreurs d'orthographe d'usage (ORT).....	35
3.2.1.3 Erreurs de verbes fictifs (VFIC).....	35
3.2.2. Au niveau sémantique et des cooccurrences lexicales (SEMAN).....	35
3.2.2.1 Erreurs de sens (SEN).....	35
3.2.2.1.1 Erreurs de synonymes (SNM).....	36
3.2.2.1.2 Erreurs de sens inadapté (SIN).....	37
3.2.2.2 Erreurs de temps/d'aspect/de mode (TAM).....	37
3.2.2.3 Erreurs de cooccurrences (COC).....	38
3.2.2.3.1 Erreurs de cooccurrence avec les noms (CONO).....	38
3.2.2.3.2 Erreurs de cooccurrence avec les adverbes ou les compléments circonstanciels (COA).....	40
3.2.3 Au niveau de la grammaire et de la syntaxe (GRAMSYN).....	41
3.2.3.1 Erreurs de voix (VV).....	41
3.2.3.2 Erreurs de pronominaux (EPN).....	41
3.2.3.3 Erreurs d'ordre des mots (ORD).....	42
3.2.3.4 Erreurs de prépositions (PREP).....	43

3.2.3.4.1	Prépositions surnuméraires (PS).....	43
3.2.3.4.2	Omission de prépositions (OP).....	43
3.2.3.4.3	Utilisation erronée de prépositions (EP).....	44
3.2.3.5	Erreurs de pronoms (PRO).....	44
3.2.3.6	Structures syntaxiques erronées (SYN).....	45
3.2.3.6.1	Redondance syntaxique (RS).....	45
3.2.3.6.2	Omission syntaxique (OS).....	47
3.2.3.6.3	Erreurs de constructions syntaxiques (CS).....	47
3.2.4	Au niveau du genre textuel et du contexte : expression non conforme au registre de l'écrit au contexte universitaire (REG).....	48
3.2.5	Récapitulation de la grille d'erreurs.....	49
3.3	L'annotation systématique dans le corpus en XML.....	54
3.3.1	Les propositions d'annotation des erreurs en anglais et en français.....	55
3.3.2	Le balisage d'annotation des erreurs verbales.....	57
3.3.3	L'annotation semi-automatique des verbes avec NooJ.....	59
3.3.4	La méthode du calcul automatique des résultats.....	62
Chapitre IV.	L'analyse et l'interprétation des données.....	64
4.1	L'analyse quantitative et qualitative des erreurs.....	64
4.1.1	La répartition des erreurs selon les catégories d'erreurs et les types d'erreurs.....	64
4.1.2	Au niveau morphologique (MORPHO).....	68
4.1.2.1	Erreurs de conjugaison et d'accord (CA).....	69
4.1.2.2	Erreurs de verbes fictifs (VFIC).....	71
4.1.2.3	Conclusion provisoire.....	71
4.1.3	Au niveau sémantique et des cooccurrences lexicales (SEMAN).....	72
4.1.3.1	Erreurs de sens lexical (SEN).....	72
4.1.3.2	Erreurs de temps/aspect/mode (TAM).....	73
4.1.3.3	Erreurs de cooccurrences (COC).....	76
4.1.3.4	Conclusion provisoire.....	77
4.1.4	Au niveau de la grammaire et de la syntaxe (GRAMSYN).....	77
4.1.4.1	Erreurs de voix et de pronominaux (VV et EPN).....	78
4.1.4.2	Erreurs d'ordre des mots (ORD).....	79
4.1.4.3	Erreurs de prépositions (PREP).....	80
4.1.4.4	Erreurs de structures syntaxiques (SYN).....	81
4.1.4.5	Conclusion provisoire.....	83
4.1.5	Récapitulation.....	84
4.2	La comparaison entre les textes écrits sur table et à la maison.....	85
4.2.1	La comparaison selon les quatre grandes catégories.....	86
4.2.2	La comparaison au niveau du type d'erreurs.....	87
4.2.3	Conclusion provisoire.....	89

4.3 L'interprétation des sources d'erreurs.....	89
4.3.1 La déficience dans la connaissance langagière du français.....	89
4.3.1.1 La maîtrise inachevée des règles grammaticales du français.....	90
4.3.1.1.1 Erreurs de conjugaison et d'accord (CA).....	90
4.3.1.1.2 Erreurs de temps/aspect/mode (TAM).....	91
4.3.1.1.3 Erreurs d'ordre des mots (ORD).....	92
4.3.1.2 La surgénéralisation des règles en français.....	93
4.3.1.3 Le manque de vocabulaire approprié en français.....	93
4.3.2 L'influence de l'anglais.....	94
4.3.2.1 Erreurs de verbes fictifs (VFIC).....	94
4.3.2.2 Erreurs de constructions syntaxiques erronées (CS).....	95
4.3.3 L'influence du chinois.....	96
4.3.3.1 Erreurs de synonymes inadaptés (SNM).....	96
4.3.3.2 Erreurs d'ordre des mots (ORD).....	100
4.3.3.3 Erreurs d'omission syntaxique (OS).....	101
4.3.3.4 Erreurs de prépositions et erreurs de pronoms (PREP et PRO).....	101
4.3.4 La proximité morphologique.....	102
4.3.5 Conclusion provisoire.....	102
4.4 Conclusion.....	103
Conclusion.....	106
Bibliographie.....	109
Sitographie.....	112
Glossaire.....	113
Sigles et abréviations utilisés.....	114
Table des illustrations.....	116
Table des annexes.....	117

MOTS-CLÉS : erreurs verbales, corpus d'apprenants, apprenants sinophones, écrits universitaires

RÉSUMÉ

La présente étude porte sur les erreurs verbales dans les productions écrites d'apprenants sinophones. Les erreurs verbales renvoient à un usage de l'unité verbale non conforme à ceux de la norme, comme ceux que l'on observe dans les dictionnaires et les grammaires de référence. Les erreurs verbales concernent plusieurs dimensions: au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau de la grammaire et de la syntaxe et au niveau du contexte et du genre. En nous basant sur des travaux d'analyse des erreurs dans les corpus d'apprenants, nous avons élaboré une grille d'erreurs en fonction des erreurs relevées. Cette grille d'erreurs nous a aidés à élaborer semi-automatiquement un corpus d'apprenants sinophones indiquant les verbes et les erreurs verbales. Notre objectif était de révéler les difficultés principales des apprenants sinophones en ce qui concerne les unités verbales, mettre en avant les circonstances où les erreurs apparaissent fréquemment, et en identifier les sources possibles. Les résultats ont montré que les erreurs verbales commises par les apprenants sinophones se concentrent sur la conjugaison et l'accord du verbe, le temps/aspect/mode du verbe, le sens lexical du verbe, les prépositions régies par les verbes et les structures syntaxiques erronées liées au verbe. Outre l'inattention, nous avons identifié d'autres sources d'erreurs possibles : le transfert négatif du chinois et de l'anglais, la déficience dans la connaissance langagière du français et la proximité morphologique entre le verbe visé et le verbe employé.

KEYWORDS : verbal errors, learner corpus, Chinese learners, academic writings

ABSTRACT

This study focuses on verbal errors in the written production of Chinese learners. Verbal errors refer to a use of the verbal unit that does not conform to those of the norm, such as those observed in reference dictionaries and grammars. Verbal errors concern several dimensions: at the morphological level, at the semantic level and lexical co-occurrences, at the level of grammar and syntax and at the level of context and gender. On the basis of error analysis research and learner's corpus research, we have developed a grid of errors based on the errors identified in the corpus. This grid of errors helped us to semi-automatically build a corpus of Chinese learners indicating verbs and verbal errors. Our objective was to reveal the main difficulties of Chinese learners with regard to the verbal units, to highlight the circumstances where errors appear frequently, and to identify possible sources of the difficulties. The results showed that verbal errors made by Chinese learners focus on the conjugation and the agreement of the verb, the time, the aspect and the mode of the verb, the lexical meaning of the verb, the prepositions governed by verbs, and the wrong syntactic structures related to the verb. Inattention aside, we have identified others possible sources of errors: the negative transference of Chinese and English, the deficiency in language knowledge of French and the morphological proximity between the target verb and the verb used.

