



HAL
open science

Analyse des erreurs verbales dans un corpus d'apprenants sinophones de niveau B1-B2 du français langue étrangère

Jingrao Li

► **To cite this version:**

Jingrao Li. Analyse des erreurs verbales dans un corpus d'apprenants sinophones de niveau B1-B2 du français langue étrangère. Manon Boucharéchas, Iris Fabry, Marie Peuzin, Roxanne Comotti, Rim Abouwarda, Alexis Ladreyt. Sciences du langage : enjeux théoriques et pratiques méthodologiques. Actes du colloque CEDIL22, , 2024, 10.5281/zenodo.13623101 . hal-04701297

HAL Id: hal-04701297

<https://hal.science/hal-04701297v1>

Submitted on 18 Sep 2024

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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

ANALYSE DES ERREURS VERBALES DANS UN CORPUS D'APPRENANTS SINOPHONES DE NIVEAU B1-B2 DU FRANÇAIS LANGUE ETRANGERE

Jingrao Li

jingrao.li@univ-grenoble-alpes.fr

Laboratoire LiDiLEM, Univ. Grenoble Alpes, France

1. Introduction

Parce que les verbes sont invariables en chinois, la maîtrise des verbes en français avec un système complexe de temps/aspect/mode est un enjeu majeur pour les apprenants sinophones du français. Les données écrites ou orales produites par les apprenants constituent une ressource centrale pour étudier l'acquisition d'une langue seconde (Granger, Gilquin & Meunier, 2015), mais elles restent rares en français langue seconde, notamment pour les apprenants sinophones. Cette étude vise à montrer l'intérêt d'un corpus d'apprenants chinois de français afin de mieux identifier les régularités des erreurs verbales produites par les locuteurs des apprenants B1-B2 selon le Cadre européen commun de référence pour les langues (CECRL), à l'aide d'une annotation systématique. Elle tentera ensuite d'expliquer ces erreurs, ce qui permettra d'envisager la conception de supports pédagogiques adaptés au public sinophone. Dans cet article, nous commencerons par une brève présentation du corpus d'apprenants. Nous aborderons ensuite la question des erreurs lexicales dans les corpus d'apprenants. Dans un deuxième temps, nous présenterons la méthode de constitution et de traitement de notre corpus d'apprenants. Enfin, nous présenterons les résultats des analyses quantitative et qualitative.

2. Corpus d'apprenants et erreurs lexicales

Les corpus d'apprenants se sont beaucoup développés depuis les années 1980, notamment pour l'anglais (Granger et al., 2015), l'espagnol (Alonso Ramos et al., 2010), ainsi que pour le français (ex. : French Interlanguage Database [FRIDA, Dagneaux, Denness & Granger, 1998], Dire Autrement [Hamel & Milicevic, 2007]). Cependant, il existe peu de corpus concentrés sur les apprenants sinophones.

En raison de grandes différences entre le français et le chinois, la difficulté d'apprentissage en français chez les étudiants chinois fait l'objet de nombreuses études en français langue étrangère (Wang, 1993 ; Zhang & Wu, 2014 ; Zhang, 2016). L'une des méthodes les plus employées pour faire ressortir ces difficultés consiste à analyser les erreurs commises par les apprenants (Ying, 2013 ; Yan, 2017 ; Li, 2018). Cependant, peu d'études se concentrent sur les erreurs de verbe, bien que cette partie du discours soit le cœur de la phrase et qu'il existe des constructions verbales extrêmement complexes en français. Compte tenu de l'importance du verbe et de la grande différence entre le verbe en chinois et en français, les erreurs verbales méritent véritablement une attention particulière. La thèse de Yan (2017), qui s'est concentrée sur les erreurs verbales des apprenants sinophones dans les écrits académiques, fournit une très bonne base en combinant l'observation des erreurs verbales et le statut des apprenants sinophones, bien qu'elle n'ait pas développé une annotation détaillée dans le corpus. L'annotation des erreurs dans le corpus constitue une démarche originale dans le domaine de l'analyse des erreurs grâce au développement de la technologie. Nous parlerons de son intérêt plus tard.

En somme, étant donné le faible nombre de corpus d'apprenants sinophones du français langue étrangère et le peu d'études se concentrant sur l'analyse des erreurs verbales chez les apprenants sinophones, en adoptant une démarche d'annotation systématique dans le corpus, nous avons décidé de développer un corpus d'écrits académiques d'apprenants sinophones de niveau B1-B2, incluant une annotation détaillée des verbes et des

erreurs verbales. Les apprenants de niveau B1-B2 sont des « utilisateurs indépendants » selon les critères du CECRL. À ce niveau, ils acquièrent des constructions relativement complexes et ont tendance à les appliquer à l'écrit pour s'exprimer plus précisément. Ils ont donc davantage d'occasions de faire des erreurs que les débutants qui n'utilisent que des structures de base. C'est pourquoi nous avons choisi ce public comme sujet d'étude.

La notion d'erreur est complexe. Il s'agit d'un concept relatif qui implique une déviation par rapport à une norme. Comme nous nous situons dans un contexte académique, nous avons adopté une norme prescriptive pour identifier les erreurs, selon laquelle une erreur est un écart par rapport à une forme normative et privilégiée. Nous nous appuyons ainsi sur des ouvrages de référence tels que le dictionnaire Petit Robert.

Les erreurs constituent des indicateurs significatifs du processus d'apprentissage, surtout quand elles sont catégorisées afin de révéler des régularités. Afin d'atteindre cet objectif, des grilles d'erreurs ont été développées, tant dans les études sur les langues premières que sur les langues secondes (Dagneaux *et al.*, 1998 ; Anctil, 2010 ; Ying, 2013 ; Yan, 2017). Notre étude portant particulièrement sur les erreurs verbales, nous avons largement adopté la grille d'erreurs de Yan (2017) en reprenant ses quatre grandes catégories d'erreurs : au niveau morphologique, au niveau sémantique et des cooccurrences lexicales, au niveau grammatical et syntaxique, et au niveau du registre de l'écrit au contexte scolaire. Nous avons modifié certains éléments, par exemple, les erreurs de temps ont ainsi été intégrées au niveau sémantique et non au niveau morphologique. Plusieurs typologies en français et en anglais prêtant une attention particulière à la combinatoire lexicale (Granger, 2007 ; Anctil, 2010) nous ont énormément inspirée. Dans la partie suivante, nous présentons la typologie détaillée des erreurs.

Cependant, les grilles d'erreurs seules ont des limites en tant qu'outil d'analyse. Quand les erreurs sont extraites de leur contexte, il est parfois difficile d'identifier les circonstances où elles se produisent et d'analyser les contextes sans erreurs. Pour pallier ces inconvénients, nous avons décidé de procéder à une annotation systématique des verbes et des erreurs verbales dans le corpus, comme dans d'autres études (Dagneaux *et al.*, 1998 ; Granger, 2007).

L'annotation des corpus d'apprenants, qui consiste à ajouter des informations interprétatives aux corpus sans modifier le contenu original, présente de nombreux avantages (Granger *et al.*, 2015). Tout d'abord, elle offre la possibilité de situer les verbes et les erreurs verbales dans le contexte. De plus, comme nous le verrons, les annotations nous aident à analyser des informations détaillées, par exemple pour déterminer la nature grammaticale d'un mot ambigu, comme *analyse* qui peut être un nom ou un verbe en français. Par ailleurs, l'annotation XML nous permet d'intégrer des métadonnées qui comprennent des informations sur les rédacteurs : âge, nombre d'années d'étude du français, autres langues parlées et autres informations supplémentaires, basées sur les fiches sociobiographiques des apprenants. En outre, la classification des erreurs verbales est abstraite et nécessite une analyse concrète : l'étape d'annotation joue un rôle important dans la validation empirique de la grille d'erreurs élaborée. Enfin, les textes annotés en XML facilitent une analyse quantitative pratique, et la vérification à l'aide d'un concordancier peut être effectuée rapidement.

Pour le français, à notre connaissance, peu d'annotations systématiques de corpus d'apprenants ont été réalisées. Parmi les travaux les plus significatifs, nous pouvons citer le corpus FRIDA de Granger (2007), dont l'annotation des erreurs utilise le langage XML. À la lumière de ce travail, étant donné que notre étude porte sur les verbes, nous indiquerons le lemme du verbe, la catégorie d'erreurs et les sous-catégories d'erreurs dans les étiquettes. En revanche, contrairement à Granger (2007), nous avons choisi de ne pas proposer de corrections dans notre étude, car il y a potentiellement plusieurs corrections possibles, notamment pour les corrections sémantiques.

Nous allons maintenant présenter la composition du corpus et le schéma d'annotation.

3. Méthodologie : constitution du corpus et principes d'annotation

Afin d'élaborer un corpus d'apprenants indiquant des erreurs verbales, suite à l'élaboration d'une grille d'erreurs, nous avons codé les textes au format XML et intégré les métadonnées des rédacteurs. Nous avons ensuite annoté de manière semi-automatique les verbes et les erreurs verbales dans le corpus en indiquant le lemme du verbe, le type d'erreur et le sous-type d'erreur.

3.1. *La constitution du corpus*

Afin d'élaborer un corpus d'apprenants, nous avons collecté des essais auprès des apprenants sinophones à l'Université Grenoble Alpes, avec l'aide des enseignants¹. Les fiches sociobiographiques comprenaient des informations sur les rédacteurs, telles que l'âge, le niveau d'études, le nombre d'années d'étude du français, les langues acquises et les moyens d'aide à la rédaction. Les apprenants sont de niveaux B1 et B2 selon le CECRL. Le corpus est composé de 95 textes et contient un total de 50 014 mots. Plusieurs types de textes sont inclus dans le corpus : textes argumentatifs, textes descriptifs, résumés, lettres et commentaires. Certains textes sont écrits à la main lors d'examens écrits et d'autres sont saisis à l'ordinateur comme devoirs. Les textes manuscrits, qui ont été rédigés en un temps limité sans recours à des dictionnaires ou autres ouvrages de référence, ont été transcrits en respectant la mise en page, les sauts de ligne et les ratures des textes.

3.2. *Grille d'erreurs et schéma d'annotation*

Notre grille d'erreurs s'inspire des travaux de Yan (2017), mais précise certains éléments. Par exemple, bien que les erreurs de pronoms soient également traitées dans l'étude de Yan (2017), elles ne concernent que l'ordre pronom-verbe-adverbe. Nous pensons que les pronoms clitiques, qui sont assez particuliers aux langues romanes, méritent un type d'erreur spécifique. Par conséquent, nous avons traité toutes les erreurs de pronoms régis par le verbe.

Nous résumons notre grille d'erreurs dans le Tableau 1, en illustrant chaque type d'erreur verbale par des exemples extraits des écrits des apprenants.

¹ Nous tenons à remercier les enseignants qui nous ont aidés dans la collection des données, en particulier Rui Yan et Anne-Cécile Perre.

Classes d'erreurs	Types d'erreurs	Exemples	
Au niveau morphologique (MORPHO)	Erreurs de conjugaison et d'accord (CA)	Le narrateur choit (choisit) de faire une explication supplémentaire qui, néanmoins, ne se consacre pas au déroulement.	
	Erreurs d'orthographe d'usage (ORT)	Nous vous persuadons de ne pas supprimer (supprimer) cet évènement et de ne pas nous arrêter de nous plaire.	
	Erreurs de verbes fictifs (VFIC : néologisme)	Également, ils sont capables d' exhibitionner (montrer) leur talent musical sur scène.	
	Erreurs de sens lexical (SEN)	Utilisation inappropriée de synonymes (SNM)	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.
Sens inadapté (SIN)		Je me permet de vous bien vouloir examiner la proposition consistant à faire endurer (durer) le festival musical à la fin de l'été.	
Au niveau sémantique et des cooccurrences lexicales (SEMAN)	Erreurs de temps/aspect/mode (TAM)	Supposons qu'il y a (ait) un concert gratuits, des artistes pourraient être invités pour participer au concert.	
	Erreurs de cooccurrences ² (COC)	Erreurs de cooccurrences avec un nom (CONO)	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.
		Erreurs de cooccurrence avec l'adverbe ou le complément circonstanciel (COA)	Ce changement de la langue du signe conduit exactement (inévitablement) à l'incompréhension des interprétations.
Au niveau de la grammaire et de la syntaxe (GRASYN)	Erreurs de voix (VV)	Les gens qui passionnent à (qui sont passionnés par/qui se passionnent pour) la musique viennent ici, ils ont besoin des endroits pour dormir.	
	Erreurs de pronominaux (EPN)	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.	
	Erreurs dans l'ordre des mots (ORD)	Il est significatif puisque nous pouvons ainsi développer mieux (mieux développer) l'industrie.	
		Prépositions surnuméraires (PS)	J'attends de votre réponse (j'attends votre réponse), M. le maire, et je vous remercie par avance.
Omission de prépositions (OP)		La grande majorité de notre citoyens ont envie de participer (à) cet événement musical.	

² La cooccurrence lexicale restreinte fait référence aux propriétés combinatoires des unités lexicales qui ont tendance à apparaître ensemble.

	Erreurs de prépositions (PREP)	Utilisation erronée de prépositions (EP)	On pourrait encourager notre peuple de (à) chanter.
	Erreurs de pronoms (PRO)	Le concert est important, ne seulement pas les commerçant, mais aussi les artistes, peuvent le bénéficier (en bénéficier).	
	Structures syntaxiques erronées (SYN)	Redondance syntaxique (RS)	Je vous prie d'agrèer , le maire de la ville, je vous présente mes respectueuses salutations. (Je vous prie d'agrèer mes respectueuses salutations)
Omission syntaxique (OS)		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.	
Erreurs de constructions syntaxiques (CS)		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/pour nous permettre d'aller plus loin)	
Au niveau du registre de l'écrit au contexte scolaire (CONT)	Expressions non conformes au registre de l'écrit au contexte scolaire (REG)	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.	

Tableau 1 – Grille d'erreurs verbales

Nous avons intégré la grille d'erreurs dans un schéma d'annotation, suivant les recommandations de la Text Encoding Initiative, qui vise à fournir des recommandations pour l'encodage des données numériques textuelles (Burnard, 2014). Deux annotateurs ont été impliqués dans le travail de vérification de l'annotation.

L'élément <teiHeader> a été utilisé pour caractériser les métadonnées du fichier, notamment les informations sur les rédacteurs. Nous avons utilisé l'élément <w> pour les mots, avec l'attribut « pos » (catégorie grammaticale) et « lemma » (lemme du verbe). Nous avons annoté tous les verbes afin de calculer la proportion de verbes erronés, tandis que l'annotation du lemme nous aide à mettre en évidence les verbes les plus fréquemment utilisés. Nous avons attribué un code spécifique à chaque type et sous-type de notre grille d'erreurs, ce code étant destiné à être utilisé lors de l'annotation (voir Tableau 1). Par exemple, nous avons employé le code « MORPHO » pour désigner les erreurs au niveau morphologique, et le code « CA » pour les « erreurs de conjugaison et d'accord ». La Figure 1 est une illustration de notre schéma d'annotation :

```
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 oeuvre</w> dans quel
contexte les langues des signes <w pos="v" lemma="évoluer">évoluent</w>.
```

Figure 1 – Un exemple de texte annoté

Pour les verbes correctement employés, la balise « pos » indique le statut verbal, et la balise « lemma » présente le lemme du verbe. Pour les verbes erronés, comme la locution à verbe support *mettre en oeuvre*, outre le statut verbal et le lemme, la valeur « SEMAN » de l'attribut « type » indique qu'il s'agit d'une erreur sémantique, et le sous-type d'erreurs « SEN/SIN » signifie que le sens est inadapté.

Le corpus élaboré est en phrase de préparation pour une diffusion publique.

4. Analyse des résultats

À partir des textes annotés, nous avons pu calculer automatiquement la proportion de verbes utilisés incorrectement et observer la distribution des erreurs verbales afin d'identifier les principales difficultés des apprenants. Nous avons également procédé à une analyse qualitative des erreurs afin de mettre en évidence les circonstances dans lesquelles les erreurs se produisent fréquemment. Dans un troisième temps, nous avons essayé d'identifier les sources possibles d'erreurs pour anticiper les erreurs potentielles.

4.1. Analyse quantitative

Au total, notre corpus contient 5601 verbes, parmi lesquels 962 peuvent être considérés comme incorrects (17,2 %). Ceci est en accord avec les résultats de Ying (2013), qui montre que l'utilisation des verbes représente une difficulté majeure pour les apprenants sinophones. Nous allons analyser la distribution des erreurs selon les quatre catégories principales, les types et sous-types d'erreurs de notre grille.

Parmi les quatre grandes catégories, les erreurs « au niveau sémantique et des cooccurrences lexicales » (SEMAN) sont les plus fréquentes (41 %). Elles sont suivies par les erreurs de grammaire et de syntaxe avec un pourcentage presque égal (37 %). Cela 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 21 %, soit environ la moitié des deux précédentes. Si elles sont moins nombreuses chez les apprenants intermédiaires de niveau B1-B2, elles représentent tout de même un cinquième des erreurs, car les compétences linguistiques des apprenants sont encore en voie d'acquisition. Conformément aux observations de Yan (2017), les erreurs de registre de l'écrit au contexte scolaire sont les moins fréquentes avec un pourcentage de 1 %.

Quant à la répartition des erreurs en fonction des types d'erreurs, les erreurs de sens lexical (SEN, 25 %) sont les plus fréquentes. Parmi celles-ci, les erreurs de synonymes (SNM, 11%) et les erreurs de sens inadapté (SIN, 14 %) sont presque aussi fréquentes. Elles sont suivies par les erreurs de structure syntaxique (SYN, 17%), parmi lesquelles les erreurs d'omissions syntaxiques sont les plus fréquentes (OS, 8 %). Les erreurs de conjugaison et d'accord du verbe (CA, 16 %) arrivent en troisième position. Les erreurs de temps/aspect/mode (TAM) représentent 10 % du nombre total, presque équivalentes aux erreurs de préposition (PREP, 9%). Un autre type fréquent est celui des erreurs de cooccurrence (COC, 7 %).

Il est intéressant de noter que si nous calculons toutes les erreurs relatives aux conjugaisons du verbe, selon la personne, le temps, l'aspect et le mode, nous obtenons un résultat de 26 % au total, ce qui est un chiffre assez important. Conformément à notre hypothèse, les erreurs de temps/aspect/mode et de conjugaison et d'accord restent fréquentes chez les apprenants sinophones.

Après cet aperçu général des erreurs, nous allons maintenant présenter une analyse qualitative de chaque type d'erreurs.

4.2. Analyse qualitative

Les analyses qualitatives suivantes sont présentées dans l'ordre de la grille d'erreurs que nous avons élaborée (Tableau 1).

4.2.1. Erreurs morphologiques

Les erreurs de conjugaison et d'accord représentent les trois quarts des erreurs morphologiques. Parmi les erreurs de conjugaison, beaucoup sont dues à une confusion entre le deuxième et le troisième groupe de

conjugaison³. Il faut rappeler qu'une des principales difficultés du français est que l'accord n'est souvent pas entendu, mais seulement visible à l'écrit. Les erreurs d'accord concernent principalement la relation sujet-verbe et le participe passé. Les apprenants ont souvent commis des erreurs d'accord sujet-verbe avec des sujets postposés, comme dans l'exemple (1).

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

Cependant, il faut souligner que ce type de phénomène pose également des difficultés aux apprenants natifs (Anctil, 2005, 2010 ; Li, 2018).

Pour l'accord du participe passé, nous avons relevé 15 erreurs relatives à des compléments d'objet direct placés avant le verbe (exemple 2), ce qui représente presque la moitié des erreurs d'accord du participe passé. Généralement, ce type d'accord n'est pas phonétiquement perceptible, ce qui peut accroître la difficulté dans la tâche d'écriture.

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

En bref, la plupart des erreurs d'accord sont également des erreurs commises par des apprenants natifs, ce qui montre que les règles morphologiques de base semblent à peu près maîtrisées à ce niveau.

4.2.2. Erreurs sémantiques

Les « erreurs de sens lexical » qui concernent l'utilisation inappropriée des synonymes et le choix d'un verbe dont le sens est inadapté au contexte représentent plus de la moitié des erreurs sémantiques, tandis que les erreurs de temps/aspect/mode représentent environ un quart, et les erreurs de cooccurrence près d'un cinquième. En ce qui concerne les verbes dont le sens ne convient pas au contexte, nous avons remarqué des confusions entre des verbes présentant une similarité morphologique, notamment pour les paires verbales qui ne diffèrent que par le préfixe comme *endurer/durer* (3).

(3) *Je me permet de vous bien vouloir examiner la proposition consistant à faire endurer* (correction proposée : *durer*) *le festival musical à la fin de l'été.*

Il nous semble que la distinction entre certains verbes et leur forme avec préfixe constitue une difficulté majeure pour les apprenants sinophones. Nous avons observé plusieurs confusions entre les verbes *jouir* et *réjouir*, *présenter* et *représenter*, *s'unir* et *se réunir*. Les apprenants doivent être sensibilisés à la distinction entre les paires verbales.

Les erreurs relatives au mode verbal sont très fréquentes, notamment celles concernant le subjonctif et la confusion entre le participe présent et le participe passé. Sans surprises, le subjonctif pose de nombreuses difficultés. Ce mode est fortement sous-utilisé : dans beaucoup de cas où l'utilisation du subjonctif est nécessaire, les étudiants optent pour l'indicatif, ce qui représente 88,9 % des erreurs de subjonctif, comme on peut le voir dans l'exemple (4).

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

³ En français, les verbes sont classés en trois groupes principaux selon la terminaison de leur infinitif. Le premier groupe comporte tous les verbes en *-er*, à l'exception du verbe *aller*. Le deuxième groupe englobe les verbes en *-ir* qui se conjuguent sur le modèle de verbes tels que *finir* ou *choisir*. Les verbes appartenant aux deux premiers groupes suivent des règles de conjugaisons spécifiques. Tous les autres verbes sans règles de conjugaison particulières sont regroupés dans le troisième groupe (*être*, *aller*, *sortir*...). Les conjugaisons des verbes appartenant aux deux derniers groupes sont souvent sujettes à des erreurs.

L'emploi du subjonctif nécessite une analyse à la fois syntaxique et sémantique. Nous avons observé que les emplois adéquats du subjonctif sont plutôt introduits par des conjonctions très caractéristiques comme *bien que* ou *avant que* (où le subjonctif est obligatoire), alors que les erreurs liées au subjonctif apparaissent souvent dans les cas où il nécessite une analyse sémantique de la part des apprenants. Il y a lieu de penser que les apprenants retiennent l'emploi du subjonctif de manière mécanique, mais qu'ils ne maîtrisent pas encore très bien la valeur sémantique qu'il exprime.

Pour les collocations verbe-nom, le nom garde souvent son sens habituel et constitue généralement la base, tandis que le verbe est choisi en fonction du nom et est sujet à des erreurs. Anctil (2010) a observé que le collocatif de sens « vide » est plus susceptible d'être soumis à des erreurs. Par exemple, dans (5), le verbe *faire* ne peut pas être combiné avec le nom *exemple*.

(5) *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.*

Les verbes comme *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. Les erreurs de collocation de type Verbe+Nom et Verbe+Adjectif représentent 17 % des erreurs sémantiques. Cette présence significative semble confirmer le constat de nombreuses études (Bolly, 2008 ; Yan, 2017), selon lequel ce type d'erreur constitue des déficits résistants propres aux apprenants non natifs.

4.2.3. Erreurs grammaticales et syntaxiques

Parmi les 22 erreurs de voix, seules huit concernent l'utilisation de la voix active au lieu de la passive ou vice-versa, et les autres concernent les verbes pronominaux à valeur de passif (par exemple, *ce livre s'achète*). De manière générale, les erreurs de verbes pronominaux représentent 13 % des erreurs totales. C'est une proportion importante qui montre la difficulté des apprenants face à ce type de verbes, du fait de leur complexité avec quatre sens : sens réfléchi (*se lever, se coiffer*), sens réciproque (*se téléphoner, se rencontrer*), sens passif (*se prononcer, se vendre*) et sens absolu (*se passer, se trouver*). Les verbes pronominaux représentent un point grammatical tout nouveau pour les apprenants sinophones et constituent donc une spécificité dans leur apprentissage du français. Pour les erreurs sur l'accord du pronom réfléchi avec le sujet des verbes pronominaux, nous pensons comme Ying (2013) que les apprenants sont tellement habitués à mémoriser les pronoms *se* + verbe infinitif et qu'ils négligent parfois l'accord du pronom. Il s'agit dans ce cas d'erreurs de conjugaison et d'accord. Les autres erreurs liées aux verbes pronominaux, comme la redondance ou l'absence du pronom réfléchi, relèvent des erreurs grammaticales et syntaxiques, qui traduisent une confusion sur la fonction du pronom réfléchi. Cela montre qu'il y a un réel besoin de sensibiliser les apprenants à la fonction du pronom réfléchi *se* lors de l'apprentissage des verbes pronominaux.

Outre les verbes pronominaux, nous avons observé des confusions dans l'ordre des mots, assez souvent avec la suite verbe-adverbe, surtout avec le verbe composé et l'infinitif (exemples 6 et 7).

(6) *L'anglais des signes principalement est utilisée (correction proposée : est principalement utilisée) dans l'enseignement de l'anglais pour aider les élèves sourds.*

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

Les erreurs d'omission syntaxique sont fréquentes, comme l'omission des verbes, des prépositions régies par le verbe, ou du complément des verbes. Nous avons également remarqué de nombreuses confusions entre *de* et *par* à la voix passive, un point grammatical complexe pour les apprenants non natifs

5. Interprétation des sources d'erreurs

Comme nous l'avons mentionné, l'identification des sources d'erreurs nous aide à anticiper les erreurs potentielles commises par les apprenants sinophones. Il y a lieu de penser que certains types d'erreurs sont simplement dus à un manque d'attention, comme les fautes d'orthographe d'usage. Hormis cela, nous avons identifié trois principales sources d'erreurs : une connaissance incomplète des règles grammaticales de la langue française, l'influence de l'anglais et l'influence du chinois.

5.1. Connaissance incomplète des structures de la langue française

Certaines erreurs sont dues à une maîtrise insuffisante de certaines structures du système linguistique français. Par exemple, le pronom *qui* peut être difficile pour les apprenants sinophones, en raison de son ambiguïté : toujours singulier quand il s'agit d'un pronom interrogatif (comme en 8) et parfois pluriel dans certains contextes comme pronom relatif (comme en 9). Les apprenants ne comprennent pas que les pronoms relatifs transfèrent les caractéristiques morphologiques de l'antécédent.

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

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

Outre les erreurs probablement dues à une maîtrise insuffisante des structures de la langue française, nous pouvons également observer l'influence de la langue anglaise, première langue étrangère de nos apprenants sinophones.

5.2. L'influence de l'anglais

Comme mentionné dans l'introduction, la plupart des Chinois commencent dès l'école primaire à apprendre l'anglais, dont l'apprentissage est obligatoire jusqu'à l'enseignement supérieur. Par rapport au chinois, l'anglais présente davantage de similitudes avec le français sur le plan morphologique et syntaxique puisqu'il s'agit de deux langues indo-européennes. Il est donc inévitable que l'anglais influence l'apprentissage du français. Dans notre corpus, nous avons observé des traces d'anglais, notamment dans les erreurs de verbes fictifs et de constructions syntaxiques. Parmi 10 erreurs de verbes fictifs relevées dans notre corpus, trois reflètent une influence de l'anglais avec une adaptation à la morphologie du français (10) :

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

En français, le verbe *exhibitionner* n'existe pas, tandis que le substantif *exhibition* peut désigner l'action de montrer au public, ce qui correspond au contexte de la phrase. Une des hypothèses suggère que l'apprenant ajoute *-er*, le suffixe verbal le plus fréquent en français, directement après le nom *exhibition* afin de le transformer en un verbe. L'autre hypothèse peut être l'influence de l'anglais, car il existe des mots morphologiquement similaires qui ont un sens correspondant : le substantif *exhibition* et le verbe *exhibit*. Par conséquent, il est également possible que l'apprenant crée un verbe français en ajoutant le suffixe *-er* après le substantif anglais *exhibition*. Dans ce cas, il s'agit d'une combinaison du sens en anglais et des règles morphologiques du français pour inventer le verbe *exhibitionner*.

À côté des erreurs de verbes fictifs, nous avons également remarqué une plus forte influence de l'anglais dans les erreurs sur les constructions syntaxiques (11).

(11) *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.

Parmi les sept occurrences de la construction *rendre+Nom+Adj*, quatre sont mal employées, avec une surutilisation de cette construction, qui semble calquée de la construction anglaise *make+Nom+Adj*, par exemple, *The news made us very happy*. Cependant, l'emploi du verbe *make* en anglais est plus varié que

celui du verbe *rendre* en français, puisque l'adjectif peut être remplacé par un participe passé : *The strange noise made us frightened*, ce qui n'est pas permis en français avec *rendre*. À part l'influence de l'anglais, nous pouvons également constater l'influence du chinois, la langue maternelle des apprenants sinophones, surtout au niveau sémantique.

5.3. L'influence du chinois

Comme mentionné ci-dessus, le chinois et le français étant deux langues morphologiquement éloignées, l'influence du chinois est plus perceptible au niveau sémantique. Dans ce cas, nous identifions deux explications principales possibles : la confusion sur la connotation des synonymes et la polysémie en chinois transférée de manière inappropriée en français.

- Confusion sur la connotation

(12) *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 12, le verbe *proférer* est traduit en chinois par 大声说 (*dà shēng shuō*), ce qui signifie « parler à haute voix ». Selon le dictionnaire Le Robert, la définition de *proférer* est « prononcer des paroles à haute voix, surtout quelque chose d'hostile ». La connotation négative du verbe *proférer*, à savoir « surtout pour quelque chose d'hostile », n'a pas été repérée dans sa traduction en chinois 大声说 (*dà shēng shuō*). Si l'étudiant a appris le verbe *proférer* par le biais de la traduction chinoise, il se peut qu'il ne connaisse pas sa connotation négative. Il y a lieu de penser que c'est la raison pour laquelle l'auteur a utilisé le verbe *proférer* pour exprimer le sens de « faire un discours à haute voix et l'articuler ».

- Polysémie du chinois transférée de manière inappropriée au français

Nous avons également remarqué que la polysémie du chinois est souvent transférée de manière inappropriée en français. Par exemple, l'utilisation du verbe *refléter* est une erreur récurrente chez les apprenants, car parmi les quatre occurrences du verbe *refléter* dans notre corpus, trois correspondent à des utilisations erronées. Le verbe *refléter* a le sens de « présenter une image, une expression plus ou moins fidèle de quelque chose, révéler une réalité intérieure » en français. La forme pronominale *se refléter* signifie donc « être indirectement illustrée de manière concrète ou caractérisée ». Ce sens est traduit en chinois par le verbe 反映 (*fǎn yìng*). Cependant, le verbe 反映 (*fǎn yìng*) est polysémique en chinois, et signifie aussi : « apparaître, révéler son existence », qui peut être traduit par *se manifester*. Dans l'exemple (13), on peut supposer que l'emploi erroné du verbe est dû à la confusion entre les deux sens de sa traduction chinoise 反映 (*fǎn yìng*).

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

Cette observation est cohérente avec l'hypothèse d'équivalence proposée par Swan (1997), selon laquelle l'apprenant a tendance à établir une équivalence stricte entre un mot dans la langue étrangère et son équivalent dans la langue maternelle, en considérant ces deux mots comme identiques. Il s'agit en fait d'une stratégie cognitive consistant à simplifier l'acquisition du lexique de la L2 en utilisant les connaissances de la L1. L'étude de Ying (2013) confirme également cette observation. Elle suggère que dans la production, à cause de compétences langagières insuffisantes, les apprenants n'arrivent pas à penser directement en français. Ils ont tendance à formuler d'abord les phrases ou le sens à exprimer en chinois, puis à les traduire en français. Vient ensuite une traduction mot à mot basée sur les correspondances déjà établies entre les verbes chinois et français dans les apprentissages précédents. Dans ce processus, les apprenants négligent souvent la polysémie des verbes en chinois, et trouvent leurs supposés équivalents en français, ce qui entraîne un transfert inapproprié.

L'examen des exemples ci-dessus permet de présumer que certaines pratiques d'apprentissage ont une forte influence sur l'acquisition lexicale, comme l'utilisation de dictionnaires bilingues. Plusieurs études consacrées à l'analyse des erreurs commises par les apprenants sinophones ont abordé le problème de l'utilisation excessive des dictionnaires bilingues, alors que les dictionnaires monolingues sont peu utilisés en Chine (Luste-Chaâ, 2010 ; Bohec, 2019). Yan (2017) émet l'hypothèse que les apprenants chinois qui ont l'habitude de se servir des dictionnaires sino-français ne maîtrisent pas pleinement les subtilités sémantiques des mots, car ces dictionnaires présentent souvent des traductions littérales sans contexte. Yan (2017) propose donc de sensibiliser les apprenants à la nature phraséologique de la langue et suggère de faciliter l'accès aux informations phraséologiques, par l'utilisation de corpus dans les apprentissages. En résumé, les traces de la langue chinoise que nous avons observées se concentrent sur des erreurs sémantiques, notamment en ce qui concerne la connotation différente entre le verbe en français et son équivalent en chinois, ainsi que la polysémie en chinois transférée de manière inappropriée en français.

6. Conclusion

Les corpus d'apprenants jouent un rôle important dans l'étude de l'acquisition d'une langue seconde et dans l'enseignement des langues. Malgré l'accroissement du nombre des corpus d'apprenants, les ressources concernant les apprenants sinophones de français langue seconde restent peu fréquentes. Nous sommes convaincue que ce corpus pourra servir de base empirique à la recherche sur l'acquisition du français et constituera une ressource utile pour la recherche sur l'enseignement du français aux apprenants sinophones au niveau B1-B2. L'annotation dans le corpus constitue également une approche efficace pour étudier le lexique en contexte de manière détaillée, au-delà des généralités. Grâce à l'analyse quantitative et qualitative de chaque type, nous avons identifié les circonstances dans lesquelles les erreurs se produisent fréquemment, et mis en évidence les points grammaticaux qui posent des problèmes aux apprenants sinophones. Nos résultats montrent que les erreurs concernant la morphologie flexionnelle sont relativement peu nombreuses, en dehors des cas complexes liés aux sujets inversés ou à l'accord avec le complément d'objet direct, que les locuteurs natifs eux-mêmes ne maîtrisent pas toujours. Les problèmes se situent plutôt au niveau syntaxique et sémantique. Les structures syntaxiques complexes restent des questions importantes, comme les verbes pronominaux ou les constructions causatives avec *faire*, structurellement éloignées des structures chinoises. Plusieurs erreurs syntaxiques révèlent aussi souvent l'influence de l'anglais. En ce qui concerne les difficultés sémantiques, nous avons observé une mauvaise maîtrise de la connotation des verbes français ou des transferts inappropriés de sens polysémiques, où l'influence du chinois apparaît déterminante. Du point de vue discursif, en raison d'une hétérogénéité textuelle de notre corpus, nous avons simplifié le niveau de genre textuel en nous concentrant uniquement sur la conformité au registre de l'écrit dans un contexte scolaire. À l'avenir, nous envisageons d'harmoniser les genres textuels du corpus afin de mieux prendre en compte les erreurs liées au registre discursif.

Les résultats de cette étude suggèrent des pistes pour mieux orienter l'enseignement du français auprès des étudiants sinophones. Il semble important de mettre l'accent sur certaines structures syntaxiques et points grammaticaux qui posent particulièrement problème à ces apprenants. En ce qui concerne les erreurs sémantiques, qui sont les plus nombreuses, la question du sens des mots devra être abordée en lien étroit avec les collocations lexicales. Des exercices issus du corpus annoté pourront s'avérer utiles pour travailler ces questions en détail. Enfin, de manière générale, nos observations plaident en faveur d'une pédagogie plus axée sur l'usage, utilisant de manière plus systématique les données de corpus authentiques, ce qui est encore peu répandu dans l'enseignement du français langue étrangère.

Références

- ALONSO RAMOS, M., WANNER, L., VINCZE, O., CASAMAYOR, G., VAZQUEZ VEIGA, N., MOSQUEIRA SUAREZ, E. & PRIETO GONZALEZ, S. (2010). Towards a Motivated Annotation Schema of Collocation Errors in Learner Corpora. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC)*, La Valetta, Malta, 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, Université de Montréal]. <https://olst.ling.umontreal.ca/static/pdf/AncilMA2005.pdf>.
- 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, Université de Montréal]. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/5077>.
- 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*. [Thèse de doctorat, Université catholique de Louvain]. <https://dial.uclouvain.be/pr/boreal/object/boreal:19625>.
- 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. Les cahiers de l'Acedle*, 16(2) : 373-420. <https://doi.org/10.4000/rdlc.6758>.
- BURNARD, L. (2014). *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*. OpenEdition Press. <https://doi.org/10.4000/books.oep.679>.
- DAGNEAUX, E., DENNESS, S., & GRANGER, S. (1998). Computer-aided error analysis. *System*, 26(2). 163-174.
- GRANGER, S. (2007). Corpus d'apprenants, annotations d'erreurs et ALAO : une synergie prometteuse. *Cahiers de lexicologie*, 91. 465-480.
- GRANGER, S., GILQUIN, G., & MEUNIER, F. (2015). *Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.
- HAMEL, M. J., & 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. 25-45.
- LE PETIT ROBERT. (2020). *Dictionnaire alphabétique et analogique de la langue française*. Le Robert.
- LI, Y. (2018). Une erreur peut en cacher une autre : les apprenants chinois du français L2 face aux problèmes de morphologie verbale. *SHS Web of Conferences*, 46. <https://doi.org/10.1051/shsconf/20184607014>.
- LUSTE-CHAA, O. (2010). L'erreur lexicale dans l'analyse des productions écrites en FLE. *Pratiques*, 145-146. 197-210. <https://doi.org/10.4000/pratiques.1556>.
- SWAN, M. (1997). The influence of the mother tongue on second language vocabulary acquisition and use. In SCHMITT, N. & MCCARTHY, M. (Eds.) *Vocabulary: description, acquisition and pedagogy* (p.156-180). Cambridge University Press.
- WANG, P. (1993). L'apprentissage du français chez les immigrants d'origine chinoise : les difficultés d'apprentissage d'ordre linguistique. *Québec français*, 89. 52–55.
- 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, Université Grenoble Alpes]. HAL. <https://inria.hal.science/tel-01691923v1>.
- YING, X-H. (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, Université de Bordeaux 3]. <https://theses.fr/2013BOR30010>.
- ZHANG, L. (2016). *Analyse des difficultés rencontrées par les étudiants chinois au cours de leur apprentissage du français et réflexions didactiques (à partir de la voix passive)*. [Thèse de doctorat, Université Sorbonne Paris Cité & Wuhan : Université de Wuhan]. <https://theses.fr/2016USPCF018>.
- ZHANG, L., & WU, H-M. (2014). Difficultés d'apprentissage du français chez les étudiants chinois. *Synergies China*, 9. 165-173.