



HAL
open science

Approche méthodologique pour l'étude des constructions elliptiques dans les rédactions étudiantes

Laura Noreskal, Iris Eshkol-Taravella, Marianne Desmets

► To cite this version:

Laura Noreskal, Iris Eshkol-Taravella, Marianne Desmets. Approche méthodologique pour l'étude des constructions elliptiques dans les rédactions étudiantes. SHS Web of Conferences, 2020, 78, pp.06007. 10.1051/shsconf/20207806007 . hal-03565857

HAL Id: hal-03565857

<https://hal.parisnanterre.fr/hal-03565857v1>

Submitted on 24 Apr 2023

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 4.0 International License

Approche méthodologique pour l'étude des constructions elliptiques dans les rédactions étudiantes

Laura Noreskal¹, Iris Eshkol-Taravella¹, et Marianne Desmets²

¹MoDyCo, UMR 7114, 200 Avenue de la République, 92000 Nanterre, France

²LLF, UMR 7110, 8 Rue Albert Einstein 75013 Paris, France

Résumé. Située dans le cadre du projet *écri+* qui porte sur la maîtrise du français écrit à l'université, cette recherche présente une étude des constructions elliptiques, lesquelles comptent parmi les difficultés les plus révélatrices du niveau de compétence rédactionnelle. Un corpus composé de constructions erronées a été constitué à partir de copies d'étudiants. Différents choix méthodologiques ont été effectués pour repérer le type d'ellipse le plus souvent erroné et les types d'erreurs récurrents. Nous présentons une analyse quantitative des erreurs en fonction des types et des natures des rédactions ainsi que leur typologie.

Abstract. Methodological approach for the study of elliptical constructions in student writing. In the frame of the *écri+* research project that aims to increase written skills in French at the university, this research presents a study about elliptical constructions which are among the most revealing difficulties of the level of writing. A corpus of erroneous constructions has been built from student writings. Different methodological choices have been made in order to identify the most frequent faulty ellipsis and the most recurring type of errors. We present a quantitative analysis of these errors according to the type and nature of the writings and their typology.

1 Introduction

À l'heure actuelle, nombre de francophones natifs rencontrent des difficultés dans la maîtrise du français écrit à l'entrée de l'université. Or, la maîtrise des compétences rédactionnelles est essentielle pour qu'un étudiant soit pleinement acteur de sa formation. Le projet *écri+* (ANR17NCUN0015) a donc été lancé pour permettre aux étudiants francophones d'améliorer leurs compétences langagières, en proposant des outils d'évaluation, de formation et de certification pour l'expression et la compréhension écrites du français. Parmi les difficultés observées, on citera les constructions syntaxiques complexes et les séquences phrastiques longues, avec des coordinations ou des juxtapositions. C'est sur ces dernières que porte notre recherche, et plus spécifiquement sur les constructions elliptiques erronées. Relativement à l'ensemble des erreurs, tous types confondus, les occurrences d'erreurs pour ces constructions sont assez faibles, mais elles

sont qualitativement importantes car elles sont particulièrement révélatrices du niveau de maîtrise des compétences syntaxiques de l'étudiant.

Cet article fait état d'une première étape méthodologique. Nous cherchons à savoir si les conditions de production d'un texte ont une influence sur la fréquence et l'acceptabilité des ellipses produites par les étudiants. Après une présentation du phénomène linguistique (section 2), nous exposons la méthodologie de constitution du corpus d'ellipses erronées adoptée pour ce travail (section 3). Dans cette partie, le recueil des rédactions étudiantes ainsi que l'extraction et l'annotation des constructions erronées sont expliqués. Enfin, (section 4) nous présentons une analyse quantitative des données recueillies selon différents critères : nature de la rédaction, type de la rédaction, ellipse réalisée, et nous proposons une typologie des erreurs répertoriées.

2 À propos des constructions elliptiques

L'ellipse est un phénomène qui se produit dans presque toutes les langues. Ce phénomène a été très étudié en français (Zribi-Hertz 1986, Mouret 2007, Desmets 2008, Bilbiie 2011, Shiraishi 2019), mais aussi en anglais (Hardt 1997, Ginzburg et Cooper 2001, Merchant 2008).

L'ellipse peut être définie comme l'omission d'un ou de plusieurs éléments dans une construction syntaxique sans que le sens ne soit altéré. Abeillé & Mouret (2010) expliquent qu'« on parle d'ellipse, de manière générale, lorsque l'interprétation d'une forme syntaxique requiert plus que ce qui est fourni par les éléments qui la composent et que le matériel nécessaire pour obtenir cette interprétation est récupérable dans le contexte immédiat. ». Ainsi, au lieu d'être répétés, les éléments ayant un équivalent dans une séquence dite « source » sont omis dans la séquence dite « cible ». La phrase elliptique est régie par des conditions sémantiques, pragmatiques et syntaxiques (Desmets 2008).

Les recherches ont pu identifier plusieurs types d'ellipses (voir Dagnac 2019 pour un recensement des ellipses en français). Observons les exemples :

- (1) Jean boit du thé et Marie du café.
- (2) a. Marie viendra et Pierre aussi.
b. Marie ne viendra pas et Pierre non plus.
- (3) Paul fait et fera tout pour réussir.
- (4) Quelqu'un a frappé à la porte mais je ne sais pas qui.
- (5) a. Il est gentil comme son père.
b. Marie est aussi grande que lui.
- (6) Il a fait tout ce qu'il a pu.

L'exemple (1) illustre le gapping ou la phrase trouée (Abeillé & al. 2008, Abeillé & Mouret 2010, Dagnac 2016) qui est une ellipse du verbe principal dans la séquence cible. Dans les exemples (2) et (2), on observe le stripping ou l'ellipse polaire (Busquets 2004) qui a lieu quand le verbe principal et ses compléments sont ellipsés dans la séquence cible et qu'un adverbe est réalisé pour caractériser positivement (*aussi, également*) ou négativement (*non plus, pas*) la séquence cible. L'exemple (3) présente le right-node-raising ou phrase à factorisation à droite (Abeillé & al., 2008, Shiraishi 2019) qui est défini comme l'ellipse d'au moins un complément du verbe dans un constituant gauche. Le sluicing (Dagnac 2019 et Abeillé & al. 2008) qui est marqué par la réduction d'une construction à un syntagme interrogatif est présenté dans l'exemple (4). En (5) et (5), on peut observer des cas de gapping avec un seul constituant résiduel (Desmets 2008) qui se produisent dans une comparative en *que* ou en *comme*. L'exemple (6) illustre l'ellipse

modale (Busquet et Denis 2001, Dagnac 2008) qui survient lorsqu'un auxiliaire modal est réalisé sans argument.

Ces recherches ont également montré que tous les types d'ellipse répondent à des contraintes linguistiques qui assurent la stabilité de l'énoncé. Le non-respect de ces contraintes entraîne une instabilité et donc une incompréhension de l'énoncé (Desmets 2008, Bilbie, 2011). En l'occurrence, deux contraintes sont particulièrement importantes : la contrainte de parallélisme et le contraste sémantique.

La contrainte de parallélisme a des répercussions syntaxiques et sémantiques : elle impose la présence dans la séquence source d'un élément partageant les mêmes caractéristiques linguistiques que l'élément ellipsé. Ainsi, par exemple, dans le cas d'une coordination de deux séquences, chaque séquence doit pouvoir apparaître seule à la place de la coordination (Abeillé et Godard 2002, Mouret 2008) :

- (7) *Il pleut et Ø chante.
- (8) Il chante et Ø danse.
- (9) *Ces contraintes dépendent du contexte de la phrase et Ø la sémantique.
- (10) Ces contraintes dépendent du contexte de la phrase et Ø de la sémantique.

Dans l'exemple (7), l'élément ellipsé (pronom personnel) et l'élément réalisé (pronom impersonnel) ne partagent pas les mêmes caractéristiques linguistiques, ce qui provoque une instabilité de l'énoncé. Dans l'exemple (9), la contrainte non respectée est syntaxique. En effet, le verbe *dépendre* est bivalent, il appelle un syntagme nominal sujet et un complément prépositionnel en *de*. Or, l'argument réalisé dans la séquence ellipsée est un syntagme nominal (*la sémantique*), la sous-catégorisation du verbe mis en facteur commun n'est pas respectée. Ainsi, les exemples (7) et (9) ne respectent pas la contrainte de parallélisme, ce qui empêche la compréhension de l'énoncé ; en revanche, les exemples (8) et (10) montrent les constructions correctes attendues.

Enfin, la contrainte de parallélisme a également des répercussions sémantiques : les situations décrites doivent être parallèles.

- (11) #Jean a réparé la voiture et Pierre le volant.
- (12) Jean a réparé la porte de la voiture et Pierre le volant.

Hors contexte, les liens méronymiques et métonymiques entre les éléments réalisés (la voiture / le volant) peuvent empêcher la compréhension de l'énoncé parce qu'ils brouillent le parallélisme attendu. Ce qui n'est pas le cas en (12).

Il est également important de prendre en compte le contraste informationnel/sémantique qui doit s'appliquer entre la séquence source et la séquence cible. En effet, il est nécessaire que les éléments réalisés dans la séquence source et ceux réalisés dans la séquence cible soient semblables tout en étant distincts.

- (13) #Paul prépare des pâtes et Jean des pâtes.
- (14) #Jean fait la cuisine et Paul le repas.

Ici les éléments doivent être suffisamment discriminés, contrastés, pour que l'énoncé soit satisfaisant du point de vue informationnel.

Si les recherches citées ci-dessus font état des propriétés syntaxiques, pragmatiques et sémantiques de l'ellipse et établissent plusieurs contraintes mobilisées lors de la production, très peu de travaux cependant se sont intéressés aux discordances (Shiraishi 2019) ou aux ellipses erronées. Or, c'est un phénomène linguistique complexe qui demande un niveau de maîtrise important. Notre objectif est d'étudier les erreurs d'emploi d'ellipse produites dans

le contexte particulier de l'université. Pour ce faire, la première étape décisive est celle de la constitution du corpus de données produites par les étudiants.

3 Méthodologie de constitution de corpus

Pour étudier les constructions elliptiques dans les productions des étudiants, il est nécessaire, en premier lieu, de répertorier les usages en contexte. Ainsi, nous avons décidé de collecter des données qui répondent à deux problématiques principales. Premièrement, nous cherchons à savoir quels sont les différents types de constructions elliptiques erronées qui apparaissent dans les productions écrites. Deuxièmement, nous voulons obtenir une typologie des erreurs réalisées. Nous avons donc construit un premier corpus de données.

3.1 Collecte de données

Afin d'avoir une vision réaliste des difficultés rédactionnelles des étudiants, nous avons fait le choix de recueillir des productions réalisées dans le cadre d'une évaluation. Nous avons cherché à obtenir un corpus suffisamment conséquent et diversifié. Le recueil de données s'est donc fait sur plusieurs rédactions issues de différents niveaux d'études. Le fait de collecter des productions de la Licence 1 au niveau Master 2 a également permis d'observer l'évolution des compétences rédactionnelles au fil des années d'études. Le corpus est constitué à partir de sources typographiques et de sources manuscrites. Les sources manuscrites ont été scannées pour permettre l'exploitation numérique de l'ensemble des rédactions. En ce qui concerne les données, l'ensemble des constructions elliptiques ont été extraites, qu'elles soient erronées ou non.

3.2 Typologie des rédactions recueillies

Dans le cadre de la formation universitaire, les rédactions produites varient. Les étudiants produisent des écrits personnels (la prise de notes par exemple) mais également des écrits évaluatifs. Pour mener cette recherche, ce sont les rédactions évaluatives qui ont été privilégiées parce qu'elles correspondent à un contexte de production écrite très spécifique. De ce fait, nous n'avons pas pris en considération les notes obtenues par l'étudiant¹. Nous avons considéré les rédactions évaluatives en tant qu'elles permettent d'observer les compétences langagières des étudiants lorsque l'écrit est non seulement un moyen de communiquer mais également un moyen de montrer sa capacité à s'exprimer et à répondre aux attentes de l'enseignant.

La typologie des rédactions recueillies est présentée dans la section suivante.

3.2.1. Les différentes « natures de rédactions »

En premier lieu, nous distinguons cinq sortes ou *natures de rédactions*² évaluatives :

- Les **examens** peuvent être définis comme des écrits réalisés dans le cadre de l'évaluation d'une partie (partiel) ou de l'ensemble (examen final) d'un enseignement universitaire dans une salle dédiée. Le temps de production de ces écrits est limité. Chaque étudiant est censé mener une réflexion personnelle et n'a généralement pas le droit de consulter ses cours ou une tierce personne. Les examens recueillis dans le cadre de cette étude ne dépassent généralement pas une dizaine de pages.

- Les **devoirs maison** sont des écrits produits en dehors du cours et servent à évaluer l'autonomie des étudiants. L'étudiant a accès à de multiples sources (documents, Internet, etc.), des contraintes de taille peuvent aussi être imposées par l'enseignant. Contrairement à l'examen, la durée du devoir est limitée à la date de dépôt et sa rédaction n'est pas spontanée.
- Les **exercices** sont produits dans le cadre de l'évaluation d'une partie précise d'un enseignement universitaire lors d'une séance. Ils ne sont pas toujours notés mais, la plupart du temps, des corrections sont faites pour que chacun vérifie qu'il a bien acquis les compétences demandées. Le temps consacré à la réflexion n'est généralement pas très long et les exercices peuvent être réalisés en groupe ou seul. Les écrits sont généralement courts car le temps limité ne permet pas d'être exhaustif.
- Les **rapports de stage** visent la validation des acquis d'une formation et l'obtention d'un diplôme (licence, master etc.). Les rapports de stage permettent aussi de vérifier plusieurs compétences rédactionnelles de l'étudiant : présenter le lieu du stage (entreprise, institution etc.), introduire son sujet, décrire le travail réalisé etc. Généralement, l'étudiant est encadré par son tuteur de stage ou son tuteur universitaire qui peuvent l'aider dans la rédaction. Le temps de rédaction et donc sa taille sont plus longs que pour les rédactions précédentes puisque le rapport peut être écrit pendant toute la période de stage.
- Les **mémoires de recherche** ressemblent aux rapports de stage car ils permettent également d'obtenir un diplôme, cependant ils doivent répondre à une problématique scientifique. Les mémoires de recherche sont issus de réflexions de longue haleine. L'étudiant doit réaliser une analyse approfondie sur un sujet en lien avec sa formation. La rédaction se fait tout au long du travail de recherche en collaboration avec le directeur de recherche. Le temps de rédaction s'étend sur plusieurs mois, voire sur une année entière. La taille du mémoire de recherche est donc assez importante.

La première classification établie dans cette section présente les différentes *natures de rédactions* qui sont examinées dans ce travail. Mais, il est également possible de classer les rédactions différemment en prenant en compte le temps imparti et les possibilités de relecture.

3.2.2 Les différents « types » de rédactions

L'analyse des données collectées a permis de séparer les rédactions en deux *types* : les rédactions spontanées et les rédactions préparées.

Les rédactions spontanées sont considérées comme des productions effectuées pendant de courtes périodes. Les étudiants doivent composer dans un temps très limité, ce qui ne leur permet pas de mener une longue réflexion. Le temps étant limité, les informations peuvent parfois être présentées d'une façon totalement déstructurée. La spontanéité se manifeste dans ce cas au travers des contraintes temporelles mais aussi au travers des contraintes liées au lieu de rédaction (université), à la présence d'enseignants et à certains interdits comme l'impossibilité de consulter ses cours, de parler avec des camarades, de sortir, etc. La relecture et la révision ne sont pas toujours possibles. En s'appuyant sur la classification *supra*, les examens et les exercices peuvent être considérés comme des rédactions spontanées.

Les rédactions préparées sont des productions qui sont effectuées sur des périodes beaucoup plus longues puisqu'il s'agit de travaux préparés pendant une période déterminée

qui dépasse largement le temps des rédactions spontanées. Elles sont censées être réalisées chez soi ou dans un endroit comparable. Les étudiants ont un temps de révision, de relecture qu'ils n'ont pas forcément lors des rédactions spontanées. Elles peuvent être relues ou revues avec d'autres. Nous considérons les devoirs maison, les rapports de stage et les mémoires comme des rédactions préparées.

Les conditions de réalisation de ces deux *types* de rédaction sont donc différentes et peuvent, selon nous, influencer la qualité de rédaction. Notre hypothèse est que la *nature* d'une rédaction ainsi que son *type* peuvent rendre favorable l'apparition de constructions elliptiques erronées. Par exemple, les rédactions spontanées peuvent être plus sujettes aux erreurs que les rédactions préparées en raison de leur nature spontanée. Ainsi, le temps imparti, la taille de la rédaction et le lieu de production peuvent être des facteurs d'apparition d'erreurs. Pour vérifier cette hypothèse, nous avons procédé à l'analyse quantitative des erreurs d'ellipses en fonction de la *nature* et du *type de rédaction*. Cette analyse est présentée section 4.1.2.

Le corpus constitué est composé de 309 rédactions comprenant 7 rapports de stage, 5 mémoires, 90 exercices, 122 examens et 85 devoirs maison. Il y a donc 97 rédactions préparées et 212 rédactions spontanées.

3.3 Annotation de constructions elliptiques

Pour étudier les constructions elliptiques, nous utilisons la démarche fondée sur l'annotation. Ce processus consiste à marquer par un signe distinctif le phénomène étudié afin de le repérer plus facilement.

L'annotation est effectuée manuellement dans un format XML. Ainsi, dans la figure 1, la phrase dans laquelle l'erreur est produite (*C'est-à-dire qu'un signifiant est constitué par ses différences qui séparent son phonème de tous les autres tout comme la valeur d'une lettre dépend de sa différence graphique avec les autres et non sa forme*) est contenue dans l'élément <doc> qui a plusieurs attributs : la nature de la rédaction (Rédaction-Examen), le niveau d'étude (L1), la présence d'une ellipse (ellipse="true" ou ellipse="false"), le type de l'ellipse (stripping) s'il y en a une et la présence d'une erreur (erreur="true" ou erreur="false"). Dans cet exemple, la construction elliptique attendue est « [...] et non de sa forme ». L'erreur produite est due à l'absence de la préposition *de*. La balise <abs> précise donc la place de l'erreur, le mot manquant (word="de") et sa catégorie morphosyntaxique (type="prep").

```
<doc nature="Rédaction-Examen" niveau="L1" ellipse="true" ellipse-type="stripping"
erreur="true">C'est-à-dire qu'un signifiant est constitué par ses différences qui séparent
son phonème de tous les autres tout comme la valeur d'une lettre dépend de sa différence
graphique avec les autres et non <abs type="prep" word="de"/> sa forme.</doc>
```

Fig. 1. Exemple d'annotation manuelle des constructions elliptiques erronées.

Les conventions d'annotation ainsi définies portent sur la phrase comprenant une construction elliptique en indiquant s'il y a une ellipse ou non, le type d'ellipse réalisé (gapping, stripping, right-node-raising, sluicing, comparative, ellipse modale), et si l'ellipse est erronée ou non.

À partir de ces données, de premières analyses peuvent d'ores et déjà être présentées sur les critères d'apparition d'ellipse.

4 Analyse quantitative

4.1 Rédactions évaluatives

Pour rappel, le corpus constitué est composé de 309 rédactions comprenant 7 rapports de stage, 5 mémoires, 90 exercices, 122 examens et 85 devoirs maison. Il y a donc 97 rédactions préparées et 212 rédactions spontanées. La répartition des données collectées est présentée en figure 2. Nous pouvons voir que le corpus n'est pas homogène puisque les examens sont plus présents que les autres avec 39%. Les exercices et les devoirs maison avec respectivement 29% et 28% sont également très présents. *A contrario*, les rapports de stage et les mémoires sont très peu représentés avec 2%.

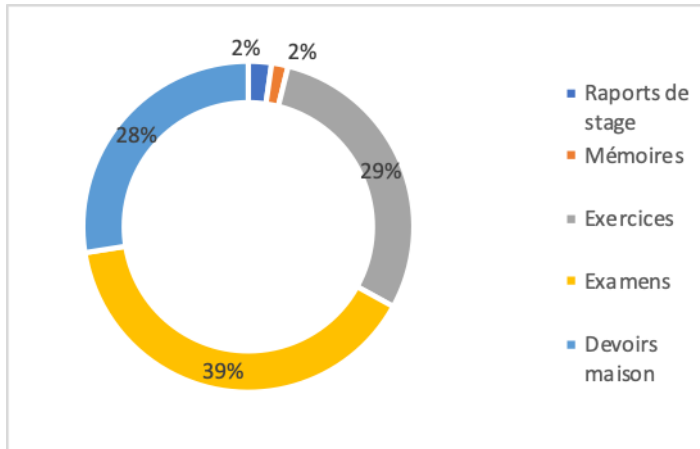


Fig. 2. Répartition des rédactions collectées

Comme il a été mentionné, les rédactions évaluatives diffèrent par leur taille. Pour comparer les tailles, nous avons décidé de prendre en compte la moyenne de phrases par rédaction et la moyenne de mots par phrase. Cette comparaison est présentée dans le tableau 1.

Tableau 1. Moyenne de phrases et longueur moyenne d'une phrase par rédaction³

Rédactions	Moyenne de phrases par rédaction	Longueur moyenne d'une phrase (en mots)	Longueur minimum (en mots)	Longueur maximum (en mots)
Exercices	6	23	3	159
Devoirs maison	52	21	4	108
Mémoires	486	24	3	110
Rapports de stage	424	21	6	98

Nous pouvons voir que les mémoires et les rapports de stage sont les rédactions les plus longues avec respectivement 486 et 424 phrases et une longueur de 24 et 21 mots par phrase. Les devoirs maison contiennent en moyenne 52 phrases contenant environ 21 mots. Les exercices sont les rédactions les plus courtes avec une moyenne de 6 phrases et une

longueur moyenne de 23 mots. Contrairement aux nombres de phrases dans les rédactions, les longueurs moyennes des phrases pour chaque rédaction ne diffèrent pas énormément puisqu'elles restent entre 21 et 24 mots.

Les longueurs minimum et maximum des phrases du corpus pour chacune des rédactions ont également été récupérées. Les phrases les plus courtes ont été trouvées dans les exercices et les mémoires (3 mots). Dans les devoirs maison, la phrase la plus courte avait une longueur de 4 mots et pour les rapports de stage une longueur de 6 mots. La plus longue phrase du corpus a été repérée dans les exercices avec une longueur de 159 mots. Les devoirs maisons et les mémoires ont eu des longueurs maximums de 108 et 110 mots. La plus longue phrase des rapports de stage est constituée de 98 mots. En observant les phrases, nous avons remarqué que les longues phrases étaient principalement dues à l'utilisation de points-virgules.

En collectant ces informations, nous voulions savoir si elles pouvaient permettre d'émettre des hypothèses sur les erreurs dans les constructions elliptiques pour chaque rédaction. Nous remarquons ainsi que les exercices se démarquent des autres rédactions puisqu'ils contiennent peu de phrases. En revanche, la longueur moyenne des phrases dans les exercices est plus élevée que pour les devoirs maison et les rapports de stage. Nous pouvons donc nous demander si les exercices ne seront pas plus sujets aux erreurs que les autres rédactions.

4.2 Critères d'apparition d'ellipse

Deux critères d'apparition d'ellipse ont été proposés selon la *nature de la rédaction* (examen, exercice, mémoire, rapport de stage ou devoir maison) et selon le *type de la rédaction* (spontanée ou préparée).

4.2.1 Selon la « nature de la rédaction »

231 constructions elliptiques erronées ont été extraites des rédactions. 7 erreurs ont été extraites des 7 rapports de stage, 40 des 90 exercices, 29 des 5 mémoires, 94 des 85 devoirs maison et 61 des 122 examens (voir figure 4).

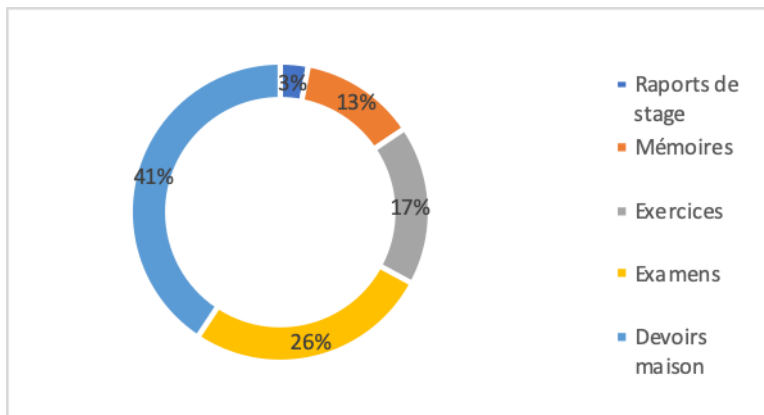


Fig. 3. Fréquence des erreurs dans l'ensemble des données

Parmi les rédactions annotées, les devoirs maison sont les plus représentés dans le corpus d'erreurs puisqu'ils représentent 41% des erreurs. Les examens représentent 26%

des erreurs, les exercices 17%, les mémoires 13% et les rapports de stage 3%. Nous pouvons ainsi dire que la présence de constructions elliptiques erronées n'est pas la même selon la *nature de la rédaction*. Les caractéristiques liées à chaque *nature de rédactions* influencent donc la production ou non d'ellipses erronées.

En prenant chaque *nature de rédactions* individuellement, les résultats changent considérablement. En effet, les erreurs sont très fréquentes dans les mémoires car il y a environ 6 erreurs par rédaction. Les rapports de stage ont une fréquence relative de 1, ce qui signifie qu'il y a une erreur par rapport. Les devoirs maison ont une fréquence presque similaire à celle des rapports de stage (1.11). Les examens qui sont les données les plus représentées dans notre corpus de rédactions ont une fréquence de 0.5, ce qui signifie qu'une erreur est présente une fois sur deux dans les examens. Pour finir, les exercices ont la fréquence la plus basse (0.44), ainsi une erreur survient moins d'une fois sur deux dans les exercices.

Jusqu'alors la taille des rédactions n'a cependant pas été prise en compte lors des calculs. Malgré la forte probabilité d'avoir des erreurs dans les mémoires ou dans les rapports, dans la perspective de la constitution du corpus, ils représentent une tâche de lecture plus coûteuse que ne le sont les exercices ou les examens. Le rapport entre le temps de traitement et le nombre de données extraites a dû être pris en compte. Par conséquent, une analyse a été faite en prenant en compte la taille des rédactions (voir tableau 2).

Tableau 2. Répartition des erreurs selon la taille des rédactions

Rédactions	Fréquence relative nombre d'erreurs/taille	1 erreur toutes les x phrases environ
Exercices	7%	14
Devoirs maison	2%	47
Mémoires	1%	84
Rapports de stage	0.2%	424

Par taille de rédactions, nous entendons le nombre de phrases que contiennent les rédactions. Ce choix est motivé par le fait que l'ellipse est une construction qui se produit dans une phrase. Ainsi, il est plus pertinent de comparer le nombre de phrases contenues dans une rédaction que de comparer le nombre de mots.

Ainsi, dans les exercices, une erreur de ce type apparaît toutes les 14 phrases, toutes les 47 phrases dans les devoirs maison, toutes les 84 phrases dans les mémoires et toutes les 424 phrases dans les rapports de stage. Les données montrent que les erreurs ont un taux d'apparition plus élevé dans les exercices que dans les autres rédactions (7%). Les devoirs maison et les mémoires ont respectivement un pourcentage d'apparition de 2% et 1%. Les rapports de stage ont la fréquence la plus basse (0.2%).

Les examens n'étant pas pris en compte dans ce calcul, les exercices se révèlent être les rédactions pour lesquelles le relevé d'erreurs est le plus rentable, et cela malgré la basse fréquence obtenue lors du calcul des erreurs par rapport à l'effectif des rédactions (0.44). En effet, les exercices étant plus courts que les autres rédactions, il est plus aisé de lire un nombre plus conséquent de rédactions. Au contraire, les rapports stage contiennent certes

une erreur par rédaction mais nécessitent de lire 424 phrases en moyenne, ce qui les rend moins sujets aux erreurs.

Toutes ces analyses ont permis d'observer la composition du corpus sous différents angles. Les résultats de l'analyse faite selon la taille des rédactions sont cependant ceux que nous privilégions car ils permettent d'avoir une vision plus précise du phénomène. La comparaison par taille est donc celle qui répond le mieux à notre recherche.

4.1.2 Selon le « type de rédaction »

Nous avons auparavant défini deux *types de rédactions* : les rédactions spontanées comprenant les examens et les exercices et les rédactions préparées comprenant les devoirs maison, les rapports de stage et les mémoires. Nous avons collecté 97 rédactions préparées et 212 rédactions spontanées.

La figure 4 montre que malgré le grand nombre de rédactions spontanées (212), il y a davantage de constructions elliptiques erronées dans les rédactions préparées (130 contre 101). De plus, la fréquence relative des rédactions préparées (1.34 erreurs par rédaction) est beaucoup plus importante que celle des rédactions spontanées (0.48 erreur par rédaction).

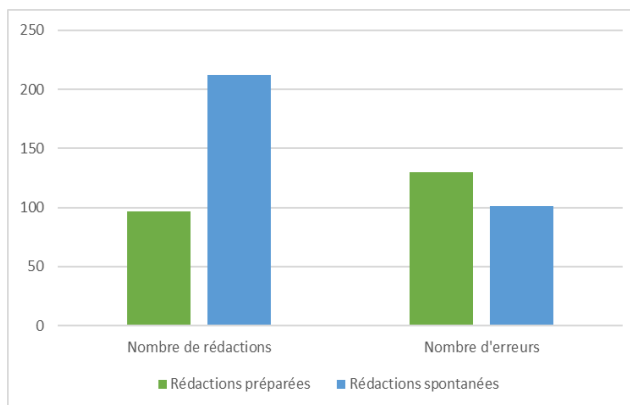


Fig. 4. Comparaison spontané/préparé

Cette observation invalide notre première hypothèse selon laquelle il y aurait plus d'erreurs dans les rédactions spontanées. Cette hypothèse se fondait sur l'étape de la relecture qui est en effet différente selon le *type de rédaction*. Toutefois, les analyses précédentes ont montré que les devoirs maison -qui ont été catégorisés comme des rédactions préparées- sont les rédactions dans lesquelles il y a le plus d'erreurs. En confrontant ces deux observations, il apparaît évident que le principe de relecture n'est pas le seul critère qui doit être pris en compte dans la catégorisation des rédactions. En effet, les devoirs maison, les rapports de stage et les mémoires ne sont pas soumis à la même relecture. Les rapports de stage et les mémoires sont des travaux qui demandent beaucoup plus d'implication que les devoirs maison puisqu'ils permettent de valider un diplôme. La qualité du travail est donc plus minutieuse que pour les devoirs maison. L'enjeu est un critère qui pourrait être pris en compte pour de prochaines analyses.

4.2 Le type d'erreurs

L'observation des données annotées a permis de constater que le principal problème dans les constructions elliptiques provenait du non-respect de la valence des verbes. Ainsi, les verbes bivalents exigeant un syntagme prépositionnel (par exemple *dépendre de*, *consister à*, *permettre de* etc.) sont les plus touchés par les erreurs lors de l'ellipse. Par ailleurs, la non-maîtrise de la valence des verbes touche également les autres types de verbes.

Trois types d'erreurs récurrentes dans les rédactions ont été observées : l'absence d'une préposition ou d'une conjonction attendue (1), le remplacement d'une préposition par une autre ou par un mot d'une autre nature (2) et l'emploi d'une préposition non attendue (3).

(1) * Il est chargé d'un programme éditorial mais également Ø la bonne entente de l'équipe. (DM-L1)

(2) * Il ne me parle pas d'un divorce en particulier, mais que l'action de divorcer était interdite. (Exercice-L1)

(3) * Ses missions témoignent du métier de l'avocat qui doit savoir accompagner le client en lui apportant des solutions juridiques pour les problèmes soumis, à travers des conseils, mais aussi de prendre en charge sa défense en cas de litige. (Rapport-M2)

Les principales prépositions touchées par ces instabilités sont les prépositions *à* et *de*. Dans l'exemple 1, la préposition *de* n'est pas répétée après l'ellipse alors qu'elle était attendue. Dans l'exemple 2, *de* est remplacé par *que*. Ainsi, pour avoir la construction attendue, il ne suffit pas juste de modifier *que* par *de* (**Il ne me parle pas d'un divorce en particulier, mais de l'action de divorcer était interdite*), il est nécessaire de modifier tout le syntagme (*Il ne me parle pas d'un divorce en particulier, mais de l'action de divorcer qui est interdite*). Dans l'exemple 3, la préposition *de* apparaît alors qu'elle n'était pas attendue. La suppression de la préposition suffit à avoir une construction stable.

Pour des analyses plus approfondies, l'influence des éléments précédents (telle que la présence de syntagmes prépositionnels juste avant, par exemple) pourrait être prise en compte lors des ajouts de préposition. De même, pour les deux autres types d'erreurs (l'absence de préposition et le remplacement d'une préposition), il pourrait être intéressant d'analyser les structures précédentes également mais aussi les synonymes des verbes impliqués (*parler de/dire que* pour l'exemple 2), et les paronymes qui sont des sources fréquentes de confusion.

5 Conclusion

L'objectif de notre recherche est d'étudier la production de constructions elliptiques erronées dans les rédactions des étudiants. Pour la constitution du corpus de données, nous avons collecté les productions écrites selon deux critères : (1) *nature de rédaction* (5 mémoires, 7 rapports de stage, 122 examens, 85 devoirs maison et 90 exercices faits en classe) et (2) *type de rédaction* (spontanée vs non spontanée). Nous avons cherché à savoir quel était le type d'ellipse le plus sujet aux erreurs dans les rédactions, quels types d'erreurs pouvaient être produits et quel type de rédaction pouvait contenir le plus de constructions erronées. Nous avons vu que sur 231 constructions elliptiques erronées extraites du corpus 17% proviennent des exercices, 13% des mémoires, 3% des rapports de stage, 26% des examens et 41% des devoirs maison. L'analyse quantitative des constructions elliptiques erronées a montré que les erreurs observées étaient principalement issues des exercices. La comparaison entre les rédactions préparées et les rédactions spontanées a invalidé notre hypothèse selon laquelle il y aurait plus d'erreurs dans les rédactions spontanées. De plus, la catégorisation des devoirs maison comme rédactions préparées a permis de mettre en évidence les variétés d'implication entre les différentes *natures de rédactions*. En observant les rédactions, il est apparu que les devoirs maison étaient plus sujets aux erreurs que les autres rédactions préparées. Nous avons alors réalisé que les rapports de stage et les mémoires demandaient plus d'implication que les devoirs maison. La classification que nous avons établie s'est donc révélée insatisfaisante pour nos recherches. Concernant les erreurs, trois erreurs récurrentes ont été relevées : l'omission d'une préposition attendue, l'ajout d'une préposition non attendue et le remplacement d'une préposition attendue par une autre ou par un mot d'une autre nature.

Le travail en perspective doit à présent porter sur la prise en compte du contexte (les verbes et les prépositions impliqués, la place de l'erreur dans les rédactions) et sur l'analyse de critères supplémentaires (enjeu du travail, implication dans la rédaction...) pour faire de nouvelles hypothèses sur les mécanismes en jeu au moment de la production d'une construction elliptique et avoir une meilleure compréhension de la nature des difficultés rencontrées par les étudiants. Il est également nécessaire d'augmenter la taille du corpus d'erreurs et de constituer un corpus d'ellipses bien formées pour créer un outil de détection automatique du phénomène. Un tel outil doit permettre aux étudiants d'identifier leurs erreurs et peut être intégré à différents dispositifs d'apprentissage du français écrit.

Références bibliographiques

- Abeillé, A. (2021). Les phrases elliptiques, in A Abeillé et D Godard (eds) la grande grammaire du français, Arles, Actes sud
- Abeillé, A., Bilbâie, G., Mouret, F. (2008). Gapping in Romance : a fragment analysis, *International Conference on Elliptical constructions*, LLF & Université Paris Diderot, Chicago Center, Paris.
- Abeillé, A., Godard, D. (2002). The syntactic structure of French auxiliaries. *Language*, 78(3), 404-452.
- Abeillé, A., Mouret, F. (2010). Quelques contraintes sur les coordinations elliptiques en français. *Revue de sémantique et de pragmatique*, 24(177-206), 89.
- Bilbâie, G. (2011). Grammaire des constructions elliptiques. Une étude comparative des phrases sans verbe en roumain et en français. Thèse de Doctorat. Université Paris 7.
- Busquet, J., Denis, P. (2001). L'ellipse modale en français. *Cahiers de Grammaire*, 26, 55-74.
- Busquets, J. (2004). Position du Focus et la distinction Stripping/Ellipse du Groupe Verbal en Catalan. *Cahiers de grammaires*, 29, 41-57.
- Dagnac, A (2008). L'Ellipse modale en français : argument pour une ellipse du TP. *Actes du Congrès Mondial de Linguistique Française CMLF'08, Jul 2008, Paris, France. pp.2453-2465, 10.1051/cmlf08211. hal-00881837*
- Dagnac, A. (2016). Gapping also needs vP-coordination: An argument from French NPI licensing. *The Linguistic Review*, 33(4), 503-530.
- Dagnac, A. (2019). Typological case studies: French. Jeroen van Craenenbroeck; Tanja Timmermann. *The Oxford Handbook of Ellipsis*, Oxford University Press, pp.785-814, 2019, 9780198712398.
- Desmets, M (2008). Ellipses dans les constructions comparatives en comme. *LINX*, 58, 47-74.
- Ginzburg, J., Cooper, R. (2001). Resolving ellipsis in clarification. Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, 236-243.
- Hardt, D (1997). An empirical approach to VP Ellipsis. *Computational Linguistics*, 23(4), 525-541.
- Merchant, J. (2008). An asymmetry in voice mismatches in VP-ellipsis and pseudogapping. *Linguistic Inquiry*, 39(1), 169-179.
- Mouret, F. (2007). Grammaire des constructions coordonnées en français. Thèse de Doctorat d'état. Université Paris 7.
- Mouret, F. (2008). Les coordinations de termes dissemblables sont-elles elliptiques ? *Actes du Congrès Mondial de Linguistique Française CMLF'08, Jul 2008, Paris, France. EDP Sciences, 221.*
- Shiraishi, A. (2019). Discordances dans l'ellipse périphérique en français. Thèse de Doctorat. Université Paris 7.
- Zribi-Hertz, A. (1986). Relations anaphoriques en français. Thèse de Doctorat. Université Paris 8.

¹ Le terme évaluation est défini par le CNRTL comme l'« Action d'évaluer, d'apprécier la valeur (d'une chose) » et parfois comme la « Valeur résultant de l'estimation ». Le terme évaluer est défini à son tour comme le fait d' « Estimer, juger pour déterminer la valeur. » Dans le cadre de l'enseignement supérieur, les évaluations vont donc servir à estimer les connaissances des étudiants en donnant une valeur à leurs écrits.

² Nous choisissons de parler ici de nature et non de genre de rédactions car nous estimons que les rédactions des étudiants sont un genre à part entière.

³ Le nombre de phrases étant calculé automatiquement, les examens ont été exclus car ils n'ont pas été transcrits.