



Année 2021-2022

**Évaluation de la place des constructions typiques
de l'immédiat dans les écrits scolaires**
Le cas des constructions clivées

Maria Lidén

M1 – Linguistique Informatique et Technologies du Langage
UFR Langues, Littératures et Civilisations étrangères
Département des Sciences du Langage

Sous la direction de Josette Rebeyrolle et Silvia Federzoni

Résumé

Dans cette étude, un corpus d'écrits scolaires nommé Résolco est utilisé dans le but de décrire et de comparer l'usage des constructions associés à l'*oral*, ou plutôt à l'*immédiat communicatif*. Après avoir consacré l'état de l'art aussi bien aux constructions disloquées que clivées, nous nous limitons aux constructions clivées par la suite. Nous émettons deux hypothèses pour les constructions clivées, fondées sur des études antérieures : 1) Les sujets et les circonstants constitueront les fonctions syntaxiques les plus fréquentes de l'élément clivé et 2) La fréquence des constructions clivées diminuera graduellement en fonction de l'augmentation du niveau de scolarité.

L'étude a également pour but de proposer une méthode permettant de détecter ces constructions automatiquement – un aspect important en TAL et en linguistique outillée qui a été peu traité jusqu'à maintenant.

L'extraction des constructions clivées se fait en plusieurs étapes. Dans un premier temps, nous constituons un gold-standard à l'aide des patrons génériques formulés dans un outil d'analyse textuelle (TXM). Ensuite, nous élaborons un script Python qui extrait les clivées à l'aide des dépendances syntaxiques que nous évaluons à partir de notre gold-standard.

Les résultats confirment l'hypothèse que les sujets et les circonstants sont les fonctions syntaxiques faisant le plus souvent l'objet du clivage. Cependant, les résultats ne sont pas conformes à l'hypothèse que la fréquence d'usage des clivées diminue graduellement selon l'augmentation du niveau de scolarité. Dans la discussion, nous évoquons entre autres le rôle que joue le type d'écrits scolaires analysé, à savoir les textes narratifs.

Concernant les méthodes d'extraction des clivées, notre analyse des sorties de l'outil d'analyse textuelle ainsi que le script Python permet de dégager des points d'amélioration.

Mots clés : écrits scolaires, Résolco, constructions clivées, immédiat communicatif, oral/écrit

Remerciements

En préambule de ce travail, je souhaite exprimer ma gratitude à un certain nombre de personnes qui m'ont guidée, soutenue et inspirée tout au long de cette première année de master.

Mes remerciements vont en premier lieu à mes directrices de mémoire, Josette Rebeyrolle et Silvia Federzoni. Nos échanges ont toujours été enrichissants ainsi qu'encourageants. Une attention particulière à Josette Rebeyrolle pour son expertise dans les aspects théoriques, et à Silvia Federzoni pour son expertise dans les aspects techniques.

Je tiens également à remercier l'ensemble des enseignants qui ont assuré les cours que j'ai suivis durant l'année. Grâce à mes nouvelles compétences en statistiques et en TAL, j'ai pu découvrir un nouvel angle de la recherche en linguistique. Ma reconnaissance va surtout vers Ludovic Tanguy d'avoir consacré du temps à répondre à mes nombreuses questions envoyées par courriel.

Par ailleurs, je souhaite remercier mon ancienne directrice de mémoire, Maria Svensson à l'université d'Uppsala, qui m'a rapidement fourni une source dont j'avais besoin et grâce à qui j'ai eu le courage de continuer mes études.

Pour finir, je tiens bien évidemment à remercier mes camarades de classe et un certain Geoffrey sans qui je n'aurais pas eu le courage d'aboutir ce travail.

Table des matières

Introduction	0
PARTIE 1 – Deux constructions typiques de l’immédiat	2
1.2 Les constructions disloquées	2
1.2.1 L’élément disloqué	3
1.2.2 Les dislocations non-prototypiques	4
1.2.3 Analyse syntaxique	5
1.2.4 Analyse pragmatique et informationnelle	6
1.2.5 Le statut des dislocations à l’oral et à l’écrit	7
1.3 Les constructions clivées	8
1.3.1 Les clivées non-prototypiques	8
1.3.2 L’élément clivé	9
1.3.3 Les constructions homonymes	10
1.3.4 Analyse syntaxique	13
1.3.5 Analyse pragmatique et informationnelle	14
1.3.6 Le statut des clivées à l’oral et à l’écrit	15
PARTIE 2 – Objectifs et données	17
2.1 Objectifs et hypothèses	17
2.2 Données : le corpus RésolCo	17
2.2.1 Constitution	18
2.2.2 Normalisation	19
PARTIE 3 – Extraction des constructions clivées	21
3.1 Première étape : constitution du gold-standard avec TXM	21
3.1.1 Extraction des constructions avec TXM	22
3.1.2 Analyse des faux positifs	23
3.1.3 Amélioration des patrons	27
3.1.4 Annotation des vrais positifs	28
3.2 Deuxième étape : modèle d’extraction des clivées avec Python	30
3.2.1 Post-traitement des données avec Stanza	30
3.2.2 Extraction avec un script Python	32
3.2.3 Performance des scripts Python	32
3.2.4 Analyse du bruit et du silence dans le script Python	33
PARTIE 4 – Analyse descriptive des clivées	36
4.1 Description globale	36
4.2 Résultats de l’annotation	38
4.2.1 L’élément clivé : fonction syntaxique	38
4.2.2 L’élément clivé : constituant syntaxique	40
4.2.3 L’élément clivé : nombre de mots	42
4.2.4 Comparaison avec l’étude de Roubaud et Sabio (2018)	43
PARTIE 5 – Analyse contrastive des clivées	47
5.1 Fréquence relative des clivées selon le niveau de scolarité	47
5.2 Type de clivée selon le niveau de scolarité	50
5.2.1 Clivées prototypiques	51

5.2.2 Il y a X qu-.....	52
5.2.3 Pseudo-clivées	52
<i>PARTIE 6 – Discussion</i>	53
6.1 La détection semi-automatique des constructions clivées : une tâche complexe	53
6.2 L’usage des clivées	54
6.3 La corrélation entre la fréquence des clivées et le niveau de scolarité.....	55
<i>Conclusion</i>	58
<i>Annexes</i>	62
Annexe 1 – Extrait de l’annotation manuelle des sorties de TXM.....	62
Annexe 2 – Extrait du tableau quantitatif	63

Introduction

Le langage oral et le langage écrit, bien que constituant des codes distincts, sont en continuité et en interdépendance constantes dans la société. Mais le langage oral est premier aussi bien au plan anthropologique que personnel : les enfants apprennent à parler bien avant de percevoir ce qu'est un écrit et a fortiori, bien avant d'apprendre à lire et écrire. (Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche 2016¹)

La citation ci-dessus, tirée d'une ressource pédagogique publiée par le Ministère de l'Éducation nationale, met le doigt sur la relation complexe entre l'oral et l'écrit et son acquisition. En effet, d'un point de vue occidental, l'écriture et ses codes prennent une grande importance aussi bien à l'école que dans la société, bien que le langage oral soit acquis en premier.

Le Ministère de l'Éducation nationale n'est pas le seul à appuyer sur la nécessité de percevoir les différences des deux codes : dans un manuel pédagogique, Béguelin et al. (1988:15) constatent que « les élèves ne sont pas assez conscients des connotations d'oralité qui accompagnent l'usage de certaines tournures ».

Dans notre étude, nous nous intéressons à cette problématique. En utilisant un corpus d'écrits scolaires, notre étude vise à observer deux constructions qui se produisent fréquemment en oral spontané, à savoir les dislocations et les clivées (Abeillé et Godard 2021, 117). Ces deux types de constructions sont illustrés par les exemples suivants :

Dislocation : **moi je** libérai ses enfants qui coururent jusqu'à chez eux. (CO-4e-2018-LSPJJRC-D1-R18-V1_N²)
Clivée : **C'est** à 19h45 **que** les enfants se retrouvèrent enfin dans l'obscurité. (CO-4e-2018-LSPJJRC-D1-R4-V1_N)

Comme le montrent les exemples ci-dessus, les constructions disloquées et clivées divergent de la structure des « phrases de base », à savoir celle où un groupe nominal sujet est suivi d'un groupe verbal (Béguelin 2000, 210). Dans le cas des dislocations, la divergence se manifeste par l'ajout d'un élément détaché lié à un argument du verbe. Dans le cas des clivées, une construction *simple* est devenue *binnaire* par la mise en relief d'un de ses constituants.

Plusieurs études ont été consacrées aux dislocations et aux clivées (De Cat 2007 ; Karssenberget Lahousse 2015 ; Lambrecht 2001a, 2001b), souvent avec l'objectif de donner une description pragmatique ou syntaxique du phénomène. Des travaux ont été menés aussi bien avec des données écrites qu'orales (De Cat 2007 ; Karssenberget Lahousse 2015 ; Lambrecht 2001a, 2001b). Lorsqu'il s'agit d'un corpus écrit, les textes appartiennent dans de nombreux cas aux genres journalistiques ou littéraires, mais également aux genres plus informels (Karssenberget Lahousse 2015 ; Roubaud et Sabio 2018). Néanmoins, à notre connaissance, les dislocations et les clivées n'ont pas été étudiées de façon systématique dans un corpus des écrits scolaires. Par ailleurs, il semble que ces phénomènes n'ont pas été suffisamment explorés à l'aide de la linguistique outillée ou dans le domaine de TAL.

Cette étude a pour objectif de décrire et de comparer l'usage des dislocations et des clivées dans des écrits scolaires. Nous souhaitons également contribuer aux études de linguistique outillée et de TAL en proposant une méthode pour détecter de façon semi-automatique ces deux constructions. Le corpus

¹ Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. (2016, avril). *Ressources maternelle. Mobiliser le langage dans toutes ses dimensions. Partie II - Lien oral-écrit - Texte de cadrage*. Disponible à :

https://cache.media.eduscol.education.fr/file/Langage/18/3/Ress_c1_langage_oralecrit_cadrage_529183.pdf

² Ce type d'identifiant signifie que l'exemple est extrait du corpus RésolCo. Dans la partie 2.2.1 Constitution, les identifiants seront explicités davantage. Dans la mesure du possible, nous prendrons des exemples issus de ce corpus pour illustrer les notions que nous utilisons dans l'étude.

RésolCo, composé d'un vaste ensemble de textes rédigés par des élèves de différents niveaux scolaires, de la maternelle à l'université, sert de base à la description ainsi qu'à l'élaboration de la méthode.

Ce travail est organisé de la manière suivante : dans la partie 1, l'intérêt est tout d'abord de nuancer l'opposition entre l'oral et l'écrit. Dans cette partie, nous donnerons une définition du terme de *l'immédiat* qui figure dans le titre de ce travail. Nous continuerons la partie 1 en établissant le cadre théorique des constructions disloquées et clivées. Dans la partie 2, les objectifs et les hypothèses de l'étude seront présentés. Cette partie contient également une description du corpus de l'étude. La partie 3 est consacrée aux méthodes employées afin d'extraire les constructions. L'analyse descriptive de ces constructions sera présentée dans la partie 4. Ensuite, la partie 5 sera consacrée à l'analyse inférentielle de la fréquence des constructions selon le niveau de scolarité. Les résultats de la partie 4 et 5 seront discutés dans la partie 6. Étant donné que les aspects méthodologiques de ce travail font partie de nos objectifs, la discussion traitera également la partie 3. Le travail s'achèvera par une conclusion.

PARTIE 1 – Deux constructions typiques de l’immédiat

Cette étude part de la présomption que les constructions disloquées et clivées font partie des traits caractéristiques de l’oral. Cependant, en employant la notion d’oral, il est facile de tomber dans la dichotomie souvent faite entre l’oral et l’écrit. En effet, il faut faire attention à faire une opposition nette entre ces deux modalités de communication.

Il est un fait que la langue orale et la langue écrite ne se réalisent pas de la même manière, la langue orale étant une chaîne de phonèmes produite par les organes de la parole, et la langue écrite étant une chaîne de graphèmes. Cependant, il ne suffit pas de distinguer l’oral de l’écrit au niveau du *moyen* de communication car une production orale peut bel et bien être réalisée à partir d’un texte, au même titre qu’une chaîne énoncée à l’oral peut être transcrite à l’écrit. Ainsi, Koch et Oesterreicher (2001) insistent sur le *mode* de communication pour cerner des différences entre l’oral et l’écrit. Selon ces mêmes auteurs (2001, 586), chaque production, écrite ou orale d’un locuteur, se situe dans un schéma des paramètres liés à l’*immédiat* et la *distance* où les paramètres concernent le contexte communicatif. L’ensemble des propriétés de l’immédiat correspond au pôle extrême du langage oral, tandis que l’ensemble des propriétés de la distance correspond au pôle extrême du langage écrit. En revanche, une production langagière est souvent un mélange des paramètres dont certains appartiennent à l’immédiat, et d’autres à la distance, d’où la pertinence de la notion du *continuum* (Koch et Oesterreicher 2001). Les paramètres de l’immédiat et de la distance sont repris ci-dessous :

	Immédiat communicatif	Distance communicative	
1.	Communication privée	Communication publique	1.
2.	Interlocuteur intime	Interlocuteur inconnu	2.
3.	Émotionnalité forte	Émotionnalité faible	3.
4.	Ancrage actionnel et situationnel	Détachement actionnel et situationnel	4.
5.	Ancrage référentiel dans la situation	Détachement référentiel de la situation	5.
6.	Coprésence spatio-temporelle	Séparation spatio-temporelle	6.
7.	Coopération communicative intense	Coopération communicative minimale	7.
8.	Dialogue	Monologue	8.
9.	Communication spontanée	Communication préparée	9.
10.	Liberté thématique	Fixation thématique	10.

Tableau 1 Paramètres communicatifs, tels qu’ils sont présentés dans Koch et Oesterreicher (2001, 586)

Les propriétés présentées sous chaque pôle ci-dessus permettent de nuancer et d’approfondir l’analyse des productions langagières. À l’aide du schéma, il est possible de démontrer que le langage des SMS n’est pas à considérer comme de l’oral dans l’écrit, mais plutôt comme de l’écrit qui porte des traits de l’immédiat communicatif, entre autres les traits de la communication spontanée et dialoguée.

Plusieurs auteurs, dont Koch et Oesterreicher (2001) et Rouquier (2020), soulignent qu’il n’y a pas de dichotomie absolue entre l’oral et l’écrit. Riegel et al. (2009, 60) pensent même qu’il est « artificiel d’opposer la syntaxe de l’oral et la syntaxe de l’écrit ». Pourtant, on ne peut pas nier qu’il y ait des différences par rapport aux constructions syntaxiques fréquentes à l’oral et à l’écrit, comme le montre Blanche-Benveniste et al. (1997) en analysant la syntaxe des énoncés d’un corpus parlé. Parmi les constructions souvent mentionnées comme faisant partie des traits caractéristiques de l’oral spontané, on trouve les constructions disloquées et clivées (Abeillé et Godard 2021 ; Blanche-Benveniste et al. 1997). Ces dernières constituent l’intérêt de cette étude. Dans les parties qui suivent, ces deux constructions seront examinées plus précisément.

1.2 Les constructions disloquées

Selon la définition donnée dans *La Grande Grammaire Historique du Français*, GGHF (Marchello-Nizia et al. 2020, 1315), le phénomène de dislocation est caractérisé « par le détachement [...] d’un

élément qui n'est pas impliqué directement dans la relation prédicative, mais qui est lié à un élément de reprise qui, lui, l'est ». Voici quelques exemples de constructions disloquées, suivies par leur construction canonique :

Constructions disloquées :

- 1) **La glace**, je l'ai mangée.
- 2) Je l'ai mangée, **la glace**.

Construction canonique :

- 3) J'ai mangé la glace.

Comme le montrent les exemples, le constituant disloqué peut être placé à gauche 1) (DG) ou à droite 2) (DD). Selon les études citées par Berrendonner (2021), la DG est considérablement plus fréquente que la DD à l'écrit comme à l'oral. En employant la formule proposée par Berrendonner (2021) la DG et la DD peuvent être illustrées par les schémas suivants :

- 4) A#[...a...]_Z
- 5) [...a...]_Z#A

Le A représente l'argument détaché, régi par le verbe du noyau, et le Z représente ce noyau. Le a est le pronom clitique qui occupe une relation coréférentielle avec A. En effet, a occupe le même rôle d'argument que A, ce qui a pour conséquence que l'un des deux peut être omis sans que la grammaticalité de la phrase soit affectée. L'élément détaché est séparé du reste des constituants par un marquage intonatif, illustré par # (Berrendonner 2021). Dans ce qui suit, le constituant A sera appelé *élément disloqué* ou *élément détaché* et a sera appelé *élément doublé*.

Il faut souligner que la nomenclature employée pour décrire ce phénomène n'est pas homogène. Parmi les notions utilisées, on trouve entre autres le *redoublement* (Gadet 1997), le *détachement* (Riegel et al. 2009), ou encore la *phrase segmentée* (Roulet 2001).

Dans ce qui suit, l'élément disloqué sera décrit plus en profondeur. Nous enchaînerons avec une sous-partie consacrée aux dislocations non-prototypiques pour ensuite traiter des aspects syntaxiques et pragmatiques du phénomène. La partie se terminera par un aperçu des études menées sur leur statut à l'écrit et à l'oral.

1.2.1 L'élément disloqué

Dans les exemples ci-dessus, l'élément disloqué est un syntagme nominal (SN), or cela n'est pas toujours le cas. En effet, tous les types de syntagmes aptes à porter le rôle d'argument d'un verbe peuvent être disloqués. Par conséquent, l'élément disloqué peut être un syntagme prépositionnel 6), une proposition à l'infinitif 7) ou une proposition subordonnée relative 8). L'élément disloqué peut également être un syntagme adjectival comme dans 9) (Berrendonner 2021):

- 6) **Aux aristocrates**, on leur coupera le cou. (Berrendonner 2021)
- 7) **Faire le tri**, ça m'amuse. (Notre exemple)
- 8) Je le sais enfin, **que François vous aime**. (Riegel et al. 2009, 430)
- 9) **Tranquilles**, Gabriel et Mercier ne l'étaient guère. (Berrendonner 2021)

À cette liste s'ajoutent les éléments disloqués du type [de N], qui ne sont considérés ni comme un SP ni un SN. Lambrecht (2001a, 1063) fait la distinction entre les SN et les constituants de type [de N] en raison du statut « non-maximal » de l'élément doublé. Ainsi, en comparant la phrase canonique dans 10) aux versions disloquées dans 11), la dislocation n'est grammaticale que lorsque l'élément doublé (*laquelle/quelle*) peut remplir à lui seul le rôle de l'objet. Comme l'écrit Lambrecht (2001a), la séquence [de N] est une construction syntaxique très particulière qui existe exclusivement pour la dislocation des noms.

- 10) Je mets quelle robe ? / * je mets quelle ?
11) Je mets laquelle, de robe ? / *Je mets quelle, de robe ?

En ce qui concerne la fonction grammaticale de l'élément disloqué, il a été montré qu'il est le plus souvent un SN sujet, ce qui correspond à des constructions de type *Pierre, il aime le cinéma* (Lambrecht 2001a, 1054).

1.2.2 Les dislocations non-prototypiques

Les linguistes semblent être d'accord sur le fait de classer les exemples 1) et 2) ci-dessus comme des cas de dislocations. Cependant, le consensus fait défaut en ce qui concerne les cas moins prototypiques. Dans ce qui suit, quelques-uns des cas non-prototypiques seront abordés. De plus, une précision sera faite sur les types de dislocations inclus dans notre étude.

Le premier cas qui pose problème est celui où l'élément détaché semble n'occuper aucune fonction syntaxique, bien qu'il y ait un lien sémantique *implicite* :

- 12) **La résurrection de Jésus**, on saura sans doute jamais la vérité. (Berrendonner 2021)

Dans l'exemple ci-dessus, le SN détaché *la résurrection de Jésus* est complètement coupé au niveau syntaxique du noyau *saura sans doute jamais la vérité*. Parfois, l'élément détaché ne porte pas la marque casuelle normalement requise pour y remplir une fonction d'argument, bien qu'il coréfère avec un clitique marqué en cas. Ceci est le cas dans 13) où l'argument du noyau « dire bonjour » nécessite la préposition *à*. Notons que le clitique qui reprend l'élément détaché est bien marqué en cas (*leur*). Si l'élément détaché était *à mes collègues*, la phrase ne ferait pas l'objet d'une ambiguïté.

- 13) **Mes collègues**, je leur disais bonjour. (Berrendonner 2021)

Les constructions exemplifiées ci-dessus sont connues sous le nom de *Nominativus Pendens* (Berrendonner 2021), ou bien *topique suspendu* (Riou et Hemforth 2015). Riou et Hemforth (2015) fournissent une liste des critères pour distinguer ce phénomène de la dislocation :

- Le constituant disloqué ne peut être qu'un syntagme nominal
- La nature de l'élément doublé n'est pas limitée à celle de pronom clitique
- L'absence de cas entre le constituant disloqué et l'élément doublé
- La relation entre le constituant disloqué et l'élément doublé n'est pas sensible aux contraintes d'ilots
- La périphérie gauche doit contenir plus d'un syntagme³
- Le détachement du constituant est le plus souvent limité à la périphérie gauche d'une phrase racine

Bien que les critères de la liste de Riou et Hemforth ne soient pas accompagnés de descriptions plus détaillées, ils peuvent être interprétés à l'aide de l'article encyclopédique de Berrendonner (2021). En ce qui concerne b), l'élément doublé peut être un SN ou même absent (voir 12), contrairement aux dislocations où il s'agit toujours d'un pronom clitique. Cela permet de distinguer la construction illustrée en 14) à celle que l'on trouve en 15) :

- 14) Topique suspendu avec élément doublé SN : **Paul**, j'ai vu **cet idiot** au café hier. (Berrendonner, 2021)
15) Version DG: **Paul**, je l'ai vu au café hier. (Notre exemple)

Le critère c) explique la construction de 13) tandis que le critère d) renvoie au fait que la relation entre l'élément disloqué et l'élément doublé peut passer par-dessus une frontière de proposition relative,

³ Dans l'article de Riou et Hemforth (2015), ce critère était formulé de la manière suivante : « Il ne peut pas n'y avoir qu'un seul syntagme dans la périphérie gauche ». Cependant, suite à un échange avec un des auteurs, une reformulation du critère a été faite par nous pour ne pas confondre le lecteur.

comme c'est le cas dans 16). Contrairement aux topiques suspendus, les DG sont sensibles à cette contrainte, raison pour laquelle 17) est jugée agrammaticale.

16) Georges, j'ai connu la fille qui lui a écrit hier. (Berrendonner, 2021)

17) *À Georges, j'ai connu la fille qui lui a écrit hier. (Berrendonner, 2021)

Si les topiques suspendus sont exclus de l'étude de Riou et Hemforth (2015), il y a d'autres chercheurs qui les incluent dans l'analyse des dislocations, notamment ceux qui s'appuient sur une approche macro-syntaxique (Blanche-Benveniste et al. 1997 ; Blasco-Delbecco 1999). Nous y reviendrons dans la partie 1.2.3 Analyse syntaxique.

Il n'y a pas que les topiques suspendus qui peuvent poser problème dans l'analyse. Un deuxième cas atypique est lorsque l'élément disloqué ne peut pas être déplacé dans l'autre périphérie de la phrase. Regardons l'exemple ci-dessous :

Personne il fiche rien

* Il fiche rien **personne**

Contrairement aux phrases 1) et 2), l'élément disloqué ci-dessus ne peut pas être placé à la fin de la phrase, mais uniquement en tête. Il est à noter que ce type de construction n'est pas non plus détaché par l'intonation et il n'est donc pas considéré comme une dislocation. À la place, on parle d'une variante de la proposition canonique, souvent signalée comme fautif ou populaire (Berrendonner 2021).

Ajoutons maintenant encore quelques cas particuliers. Dans un article de Nølke (1998), plusieurs structures qui sont proches de la dislocation sont mentionnées, à savoir :

18) **Quant à Pierre**, il est toujours prêt à nous aider (Nølke 1998, 388)

19) **Pour ce qui est de Pierre**, il est toujours prêt à nous aider (Nølke 1998, 388)

20) **Pour moi**, je suis né dans votre pays (Nølke 1998, 388)

21) Elle est charmante **comme fille** (Nølke 1998, 392)

22) C'est une fleur **que la rose** (Nølke 1998, 392)

Dans 18)-20), la présence d'un « introducteur » (*quant à, pour ce qui est de, pour moi, etc.*) sert à mettre cet élément en relief à l'instar d'une dislocation. Nølke (1998) soulève le fait que les constructions dans 18)-20) ne semblent pas pouvoir être disloquées à droite, ce qui pourrait indiquer quelque chose sur leur fonction. En ce qui concerne les structures dans 21) et 22), Nølke (1998) prétend qu'elles peuvent être rapprochées à la dislocation à droite. Dans 21), *comme*, qui est vidé du sens, peut être interprété comme une sorte d'article dans le disloqué. Lorsque l'élément disloqué est doublé par le démonstratif *ce*, ce premier peut être introduit par *que* comme dans 22).

1.2.3 Analyse syntaxique

Il n'y a pas d'unanimité quant à la description syntaxique des dislocations. Selon certaines approches, la dislocation fait exclusivement partie des propositions principales, ce qui veut dire que le noyau Z doit être un verbe de rang maximal (Berrendonner 2021). Selon Lambrecht (2001a, 1050) qui s'oppose à ces approches, les éléments disloqués ont toujours un statut *extra-propositionnel* : ils sont placés en dehors de la proposition. En suivant cette logique, un exemple avec la séquence [de N] comme dans 23) ne peut pas être traité comme une dislocation puisque le groupe détaché se trouve à l'intérieur d'une proposition (« le vôtre a toujours été bien personnel »). Notons que la phrase serait peu acceptable si l'élément disloqué était placé à l'extérieur de la proposition.

23) Le vôtre, **de style**, a toujours été bien personnel. (Berrendonner 2021)

24) ? **De style**, le vôtre, a toujours été bien personnel.

Comme l'écrit Lambrecht (2001a, 1066) : « le statut extra-propositionnel des syntagmes disloqués est manifesté aussi dans leur position relative au focus de la phrase » [notre traduction]. Cela veut dire que l'élément disloqué doit obligatoirement précéder ou suivre le focus. S'il suit le focus, il s'agit d'une DD, et s'il le précède il s'agit d'une DG. Dans 25), *mon argent* constitue le focus de la phrase. Par conséquent, l'élément disloqué ne peut pas être inséré immédiatement avant celui-ci.

25) Elle ne m'a pas rendu mon argent, **Nicole** / * Elle ne m'a pas rendu, **Nicole**, mon argent.

Si Lambrecht insiste sur le statut extra-propositionnel de l'élément disloqué, le Groupe Aixois de Recherche en Syntaxe (GARS) de son côté propose une approche *macro-syntaxique* dans l'analyse des dislocations. D'après Blanche-Benveniste, la syntaxe, ou bien la micro-syntaxe, composée en unités syntaxiques, est « fondée sur l'organisation des catégories grammaticales et leur rection », tandis que la macro-syntaxe sert à cerner les « relations qu'on ne peut pas décrire à partir des rections de catégories grammaticales » (Blanche-Benveniste et al. 1997 : 113). En d'autres termes, la micro-syntaxe permet d'analyser les relations entre le verbe et ses arguments, tandis que la macro-syntaxe permet d'analyser les dépendances au niveau énonciatif et pragmatique entre un noyau et des éléments prosodiquement détachés, à savoir des éléments disloqués. Avec cette approche, les DG et les topiques suspendus peuvent être rapprochés.

Dans l'approche macro-syntaxique, le *noyau* constitue l'unité minimale. Le noyau peut donc apparaître tout seul, mais il se voit souvent accompagné par des éléments avant ou après, appelés des préfixes ou des suffixes. Les DG font partie des préfixes, et les DD des suffixes (Blanche-Benveniste et al. 1997).

1.2.4 Analyse pragmatique et informationnelle

La littérature sur les dislocations semble être d'accord sur leur fonction de thématization (Berrendonner 2020 ; Lambrecht 2001a). Néanmoins, comme le remarque De Cat (2007, 2), l'analyse pragmatique des dislocations doit aller au-delà d'une description simpliste selon laquelle l'élément détaché serait nécessairement le thème, le thème étant « ce dont on parle ». La même autrice appuie sur le fait que cette logique ne s'applique pas à tous les cas de dislocation dans la mesure où Lambrecht (2001a, 1072) définit le thème en citant Gundel (1988) :

« An entity, *E*, is the topic of a sentence, *S*, if in using *S* the speaker intends to increase the addressee's knowledge about, request information about, or otherwise get the addressee to act with respect to *E*. »

Comme le montre la citation, le thème, qui porte sur l'intention du locuteur, est fortement pertinent pour la phrase dans laquelle il se trouve. Lambrecht (2001a, 1073) ajoute qu'une entité peut uniquement être interprétée comme le thème de la phrase lorsqu'elle contient un référent *identifiable*. En français, l'identifiabilité d'un référent dépend de la définitude de ce dernier. Par conséquent, si l'élément disloqué est un SN, il faut qu'il soit défini 26). Pourtant, une exception se trouve dans les cas où le référent peut porter un sens *générique* 27) :

26) Le garçon, il attend devant la porte/ *un garçon, il attend devant la porte (Lambrecht, 2001a)

27) Un garçon, ça n'attend pas devant la porte, ça entre. (Lambrecht 2001a)

Dans 27), le sens générique est motivé par l'usage du pronom *ça* à la place du pronom *il*.

Le thème ne doit pas seulement être identifiable, il doit aussi être un référent saillant, c'est-à-dire assez *actif* dans le contexte du discours (De Cat 2007, 67). Néanmoins, le fait d'être actif ne doit pas forcément dire qu'il a été mentionné plus tôt dans le discours. Il peut également s'agir d'un référent qui fait partie du savoir commun partagé par les locuteurs (De Cat 2007, 68).

Jusqu'ici, la pragmatique des dislocations a été traitée au niveau général, or une distinction entre les DG et les DD peut être ajoutée. Selon Nølke (1998), la fonction de la DG est de thématiser l'élément disloqué, tandis que la fonction de la DD est le rappel du thème ou l'introduction d'une valeur émotionnelle. La fonction de rappel du thème s'explique par la relation cataphorique inévitable entre le disloqué et le sujet clitique (Nølke 1998, 391). La valeur émotionnelle potentielle des DD s'explique par leurs traits prosodiques et par le fait de répéter un élément de la phrase (Nølke 1998, 391). Dans un énoncé tel que « Je ne suis pas bête, moi ! », l'élément disloqué est non-accentué par défaut du fait qu'il est placé après l'élément focalisé (*bête*). De plus, l'émotion du locuteur se manifeste par la répétition (*je - moi*).

Une autre définition est donnée par Lambrecht (2001a, 1074). Selon ce dernier, la fonction de la DG est d'annoncer le thème, tandis que la fonction de la DD est de maintenir le thème. Il exemplifie ces fonctions avec l'échange suivant :

Une femme et son mari sont assis à table pour dîner. Le mari regarde le poulet dans son assiette :
M : **ça** n'a pas de goût, **ce poulet**.
F : **Le veau**, c'est pire.

Dans l'échange ci-dessus, la DD dans l'énoncé du mari est justifiée par sa saillance pragmatique dans le contexte discursif vu que le poulet se trouve dans son assiette. La réponse de la femme contient une DG, et cela puisqu'elle annonce un thème générique qui n'a pas été établi avant dans le discours.

1.2.5 Le statut des dislocations à l'oral et à l'écrit

Il suffit de consulter une des grammaires courantes du français telle que la *Grande grammaire du français*, GGF (Abeillé et Godard 2021, 117) ou la *Grammaire méthodique du français* (Riegel et al. 2009, 63-66), pour constater que les dislocations y figurent souvent avec la notion d'oral spontané. La notion d'oral est reprise dans un ouvrage entièrement dédié aux dislocations (De Cat 2007, 1), où l'auteur affirme que « l'usage omniprésent⁴ » des dislocations constitue un des traits caractéristiques du français oral.

Néanmoins, toutes les études menées sur les dislocations ne sont pas d'accord sur la fréquence soi-disant élevée à l'oral. Selon Blasco-Dulbecco (1999, 83), les dislocations ne semblent pas fortement présentes à l'oral. Au contraire, elles restent relativement rares aussi bien à l'écrit qu'à l'oral. Dans son étude, l'autrice a trouvé seulement 2% de ces constructions à l'écrit et 10% à l'oral. Ceci mentionné, il faut toutefois soulever le fait que ces observations ont été faites sur un corpus écrit et un corpus oral faisant partie d'une collection de données d'un mémoire de maîtrise fait en 1996. En d'autres termes, il s'agit d'un travail relativement ancien, fait au deuxième cycle universitaire. On peut également s'interroger sur la pertinence des données puisque l'étude ne fournit aucun élément de description de la manière dont le corpus a été constitué.

Quoi qu'il en soit, la quantification du phénomène des dislocations constitue une tâche complexe et cela puisque la définition même du phénomène n'est pas univoque. De plus, comme le constate Berrendonner (2021) la multitude d'approches des dislocations rend impossible la comparaison des résultats des études.

Au lieu d'insister sur la fréquence des constructions disloquées dans de différents corpus, une alternative est de conduire des études expérimentales basées sur des jugements d'acceptabilité, comme le font Riou et Hemforth dans leur étude de 2015. L'objectif de leur étude est d'évaluer à quel point le poids du canal écrit joue sur le jugement d'acceptabilité d'une construction syntaxique. Les auteurs estiment que la dislocation du clitique à gauche « n'est jamais aussi bien acceptée à l'écrit par rapport à des

⁴ Notre traduction. Dans le texte original: « pervasive use » (De Cat 2007,1).

constructions perçues comme appartenant au français standard ou familier » (Riou et Hemforth 2015, 34). Les résultats montrent que les phrases avec dislocation à gauche sont fortement défavorisées par rapport aux constructions qui appartiennent au français standard et familier. Même dans un contexte favorable, la dislocation du clitique à gauche est toujours défavorisée. En effet, les dislocations sont jugées au même niveau de grammaticalité que des constructions agrammaticales. Ainsi, les expériences menées par Riou et Hemforth confirment leur hypothèse.

1.3 Les constructions clivées

Les constructions clivées, ou les *clivées* tout court, sont des propositions coupées en deux. Ce type de construction est employé pour mettre en relief un des éléments de la proposition. Les cas prototypiques prennent la forme « c'est X qu- + V », dont l'élément clivé est le constituant X. Ci-dessous se trouve un exemple d'une clivée prototypique, suivi de sa réécriture canonique.

Construction clivée : **c'était** le maire **qui** tenait un fusil dans ses mains (CO-3e-2016-VTAC305-D1-R17-V1_N)

Construction canonique : Le maire tenait un fusil dans ses mains

Avec plusieurs attestations en latin et quelques-unes en ancien français, il est prouvé que les clivées ne constituent pas un nouveau phénomène (Rouquier 2020, 1309). Il faut également souligner que ce phénomène existe dans plusieurs langues. Parmi les auteurs cités dans cette partie, se trouvent des linguistes qui ont travaillé principalement sur les constructions clivées en anglais, les *cleft constructions* (Lambrecht 2004 ; Carter-Thomas 2009).

Dans ce qui suit, nous allons d'abord élargir la définition des constructions clivées en introduisant des constructions non-prototypiques. Nous allons continuer en précisant la nature du constituant X des clivées pour ensuite distinguer les clivées des constructions homonymes. Ensuite, nous allons résumer les diverses conceptions syntaxiques et pragmatiques des clivées avant de survoler ce que la littérature dit sur leur statut à l'écrit et à l'oral.

1.3.1 Les clivées non-prototypiques

Aux clivées prototypiques s'ajoutent des clivées de type non-prototypique, c'est-à-dire des clivées qui ne commencent pas par un pronom clitique et la copule *être* :

- 1) **Il y a** Sarah **qui** a eu un bébé. (Karssenber 2017)
- 2) **J'ai** mon amie **qui** habite à Paris. (Conti 2010)

À l'instar des clivées prototypiques, les deux exemples ci-dessus ont la même valeur de vérité que leurs équivalents canoniques. Dans 1), la partie clivée est introduite par la locution verbale impersonnelle *il y a*. Dans 2), la partie clivée est introduite par un pronom clitique suivi du verbe *avoir*. À cause de leur statut non-prototypique, les structures dans 1) et 2) ne sont pas catégorisées comme des clivées par tous les linguistes. Parmi les linguistes qui analysent 1) comme une clivée, nous trouvons Karssenber et Lahousse (2015) et Lambrecht (200b). L'analyse de l'exemple 2) comme une clivée est surtout menée par Conti (2010), mais on la trouve également chez Lambrecht (2001b). En effet, dans l'article de Lambrecht (2001b, 508), les clivées dans 1)-2) sont regroupées ensemble comme des clivées construites avec *avoir*. Néanmoins, dans GGF (Abeillé et Godard 2021, 1964), ces deux constructions sont catégorisées comme des constructions présentatives (voir 1.3.3 Les constructions homonymes). Dans ce travail, nous partageons l'analyse de Karssenber, Lahousse, Lambrecht et Conti plutôt que ces derniers.

Les clivées non-prototypiques peuvent également inclure la construction dite *pseudo-clivée* de type A, *c'est B* (Rouquier 2018) qui est illustré en 3) :

3) **Ce que j'aime, c'est** la glace au chocolat (Apothéloz et Roubaud 2018)

Ce dernier type est décrit comme une combinaison entre une phrase clivée et une dislocation dans GGF (Abeillé et Godard 2021, 1940). Il est à noter que certains chercheurs ont une définition plus restreinte des pseudo-clivées que d'autres. Dans leur article de 2006, Legallois et Grea font la distinction entre les énoncés spécificationnels et les pseudo-clivées. D'après eux, la première partie d'une pseudo-clivée (la partie A), contient nécessairement un verbe tandis qu'un énoncé spécificationnel représente une forme condensée où la première partie est un nom. Ainsi, 4) est à considérer comme un énoncé spécificationnel, et 5) comme une pseudo-clivée.

- 4) **Le piquant de l'affaire**, c'est que Martine Anzani vient justement d'être nommée à la chambre criminelle de la Cour de cassation... (Legallois et Gréa 2006)
- 5) **Ce qui est piquant dans l'affaire**, c'est que Martine Anzani vient justement d'être nommée à la chambre criminelle de la Cour de cassation... (Legallois et Gréa 2006)

Dans le cadre de cette étude, nous nous intéressons uniquement aux pseudo-clivées dont la première partie contient le relatif *que* ou *qui* (*ce que* ou *ce qui*), comme dans l'exemple 5).

1.3.2 L'élément clivé

L'élément clivé, c'est-à-dire le constituant X, peut porter la fonction du sujet de même que la fonction de complément d'objet direct ou indirect (Rouquier 2020, 1309). Dans une étude de Carter-Thomas (2009, 8-9), l'élément clivé est le plus fréquemment un sujet ou un circonstant. Ces résultats sont conformes aux résultats de Katz (1997, cité dans Carter-Thomas 2009) qui montrent que les sujets et les circonstants sont les rôles syntaxiques les plus souvent clivés en français parlé⁵.

L'élément X peut appartenir aux différentes catégories grammaticales. Rouquier (2018) en liste 8 pour les constructions clivées prototypiques : les syntagmes nominaux 6), les pronoms accentués 7), les noms propres 8), les adverbes 9), les syntagmes prépositionnels 10), les subordonnées complétives ou autre 11)-12), les syntagmes adjectivaux 13) ou les infinitifs 14)⁶.

- 6) C'était **la voisine** qui habitait la maison au fond de la rue (CO-3e-2016-VTAC305-D1-R25-V1_N)
- 7) Il demanda si c'était **lui** qui l'avait tuée (CO-6e-2016-PJPR1-D1-R21-V1_N)
- 8) En fait, c'était **Levy** qui avait cassé son verre (CO-3e-2018-FSBJC6-D1-R9-V2_N)
- 9) C'était **là** que, avant leur décès, ses parents l'avaient élevé (CO-3e-2018-VTAC305-D1-R1-V1_N)
- 10) C'est **à 19h45** que les enfants se retrouvèrent enfin dans l'obscurité. (CO-4e-2018-LSPJJRC-D1-R4-V1_N)
- 11) C'est **qu'il puisse faire naître en nous-même l'infinie diversité** qui donne du prix à nos plus précieux rouleaux de soie. (Rouquier 2018)
- 12) Je sais que c'est **juste pour que je t'obéisse** que tu as inventé cette histoire. (UN-M2-2020-UCL-D1-R80-V1_N)
- 13) C'est **bien mûres** que je les préfère (Rouquier 2018)
- 14) C'est **partir très tôt le matin** que je redoutais. (Blanche-Benveniste 2002, 93)

Notons que les pronoms accentués et les noms propres font partie des syntagmes nominaux. Il faut également souligner, comme le font Abeillé et Godard (2021, 1958), que certains adverbes sont plus acceptables dans l'élément clivé que d'autres. Les auteurs citent les adverbes de manière, de temps et de fréquence parmi les plus acceptables, et les adverbes de degré, tel que *beaucoup*, parmi les moins acceptables.

⁵ Notons que ces résultats ne concernent uniquement les clivées prototypiques.

⁶ Blanche-Benveniste (2002, 93) soulève le fait que certains verbes à l'infinitif sont rejetés en position X de la clivée. Elle exemplifie avec la phrase *« c'est mentir qui est honteux ».

Si les catégories de syntagmes peuvent varier dans l'élément clivé des clivées prototypiques, les clivées non-prototypiques sont plus restreintes. En effet, dans le cas des clivées introduites par *il y a*, l'élément X ne peut être qu'un constituant nominal, tel qu'un syntagme nominal 15), sauf dans les emplois avec une négation où l'on trouve également des syntagmes prépositionnels 16) (Karssenberg 2017). Cependant, Karssenberg suggère que l'acceptabilité des clivées *il y a X qui* avec un pronom accentué dépend de sa structure informationnelle. En effet, il semble que la clivée doit introduire un contraste entre deux éléments dans l'emploi des pronoms accentués. Regardons l'exemple 17) : la phrase sans mise en contraste serait peu acceptable, tandis que la phrase avec contraste constitue une phrase attestée.

15) il y a **le téléphone** qui sonne. (Rouquier 2018)

16) il n'y a qu'**avec cette voiture** que je conduis. (Rouquier 2018)

17) ? Y'a toi qui fêtes la chandeleur normalement. / Y'a toi qui fêtes la chandeleur normalement, puis y'a eux... (Karssenberg 2017)

Quant aux clivées de type *pronom avoir X qui*, l'élément clivé ne peut être qu'un syntagme nominal (Conti 2010) :

18) J'ai **tous mes amis** qui sont ici (EC-CM2-2016-SGLEB-D1-R12-V1_N)

Enfin, en ce qui concerne l'élément clivé des pseudo-clivées, il faut souligner qu'il s'agit non pas de l'élément à gauche mais du segment à droite. En effet, nous ne retrouvons pas la notion de *l'élément clivé* dans les études concernant les pseudo-clivées. À la place, les auteurs semblent préférer la notion du *segment à droite* ou *segment droit* (Müller Blaser 2007 ; Apothéloz et Roubaud 2018). Selon Müller Blaser (2007), l'élément clivé des pseudo-clivées est typiquement un syntagme nominal 19), une proposition infinitive 20) ou une subordonnée 21) :

19) Ce qui l'effrayait davantage c'était **le déménagement**. (UN-M2-2020-UCL-D1-R123-V1_N)

20) [...] ce qu'il aimait le plus c'est **engloutir les enfants lors de grandes pluies afin de les piéger** à [...] (UN-M2-2021-UCL-D1-R63-V1_N)

21) ce qu'ils ignoraient c'est **qu'il y avait une vieille sorcière qui y habitait**. (EC-CM2-2016-SGLEA-D1-R22-V1_N)

1.3.3 Les constructions homonymes

Un enjeu dans l'analyse des clivées concerne le fait que sa forme est identique à d'autres types de constructions. Afin de pouvoir distinguer les constructions clivées aux constructions homonymes, il faut prendre en compte le contexte. En plus, nous notons que la fonction des cas homonymes diffèrent selon la construction qui introduit la clivée, à savoir *c'est*, *il y a* ou *j'ai*. Dans cette partie, nous commençons par les cas introduits par *c'est*, pour ensuite traiter les ambiguïtés liées aux constructions avec *avoir*.

En ce qui concerne les constructions de type *c'est X qu-*, Rouquier (2018, 7) cite les quatre constructions homonymes données par Rouget et Salze (1986) en utilisant l'exemple suivant:

22) C'est le prunier qui a des fleurs blanches

Sans contexte autour, il est impossible de déterminer s'il s'agit d'une clivée ou non. La construction ne peut être considérée comme clivée que lorsque la proposition est interprétée de la manière suivante: « quelque chose a des fleurs blanches et ce quelque chose est le prunier ». Cette interprétation est possible dans la mise en contraste, comme dans l'exemple suivant:

23) C'est le prunier qui a des fleurs blanches, et non le cerisier.

Cependant, 22) ne peut pas être interprété comme une clivée dans les exemples 24)-26).

24) [Mon arbre préféré/ c'] est [le prunier /celui qui a des fleurs blanches]

- 25) [Mon arbre préféré /, c'] est [le prunier / celui-ci / lui], [qui / lequel] a des fleurs blanches
 26) Pour moi, depuis toujours, l'arrivée du printemps, c'est le prunier qui a des fleurs blanches.

Dans 24), il s'agit d'une *construction copulative canonique*, également traitée avec le *terme proposition relative* (ce + être + SN) (Scappini 2007). Dans une telle construction, le sujet est le pronom *ce*, et l'attribut est un syntagme nominal qui contient une proposition relative déterminative. Cette dernière permet d'identifier l'antécédent qui dans ce cas est *le prunier*. Contrairement aux clivées, le pronom *ce* a un sens anaphorique dans la construction copulative canonique. Il renvoie à un élément déjà mentionné, par exemple à *mon arbre préféré*. 25) ressemble au 24), or dans 25) la proposition relative est non-déterminative, ce qui veut dire que l'information donnée sur l'antécédent n'est pas indispensable. Finalement, il y a des constructions comme dans 26) où le SN attribut peut être interprété comme le nom d'un état-de-choses. En d'autres termes, le locuteur dans 26) parle « du fait que le prunier ait des fleurs blanches » (Rouquier 2018). Rouget et Salze (1986, 123) exemplifient ce type de construction avec la phrase suivante : « la corrida c'est le matador qui s'amuse avec le taureau ». Cette phrase peut être réécrite en « la corrida, c'est *quand* le matador s'amuse avec le taureau ». De la même manière, la phrase « l'arrivée du printemps, c'est le prunier qui a des fleurs blanches » peut être réécrite en « l'arrivée du printemps, c'est *quand* le prunier a des fleurs blanches. ». Par ailleurs, dans l'exemple 26), le démonstratif peut aussi être analysé comme anaphorique, au même titre qu'en 24) et 25).

Quand on consulte le corpus sur lequel repose notre étude, à savoir le corpus Résolco, des constructions homonymes aux clivées sont rapidement détectées. Il s'agit souvent des cas similaires à 24), 25) et à 26) où le pronom est porteur d'un sens anaphorique.

Il se retourna en entendant ce grand bruit. C'était un éboulement de pierres qui allait droit sur eux. (EC-CE2-2016-VRX-D1-R1-V1)

Dans l'exemple ci-dessus, le pronom de la copule n'est pas vide du sens, mais il réfère à *ce grand bruit* de la phrase précédente. Dans la GGF, les auteurs distinguent les clivées des *constructions présentatives*. Nous considérons que la définition correspond à celle des constructions copulatives canoniques. Les auteurs insistent sur le fait que la présentative sert à introduire quelqu'un ou quelque chose dans le discours, sans pour autant mettre en élément en relief comme dans une clivée. En effet, aucun élément de la phrase présentative constitue un focus ou est particulièrement saillant (Abeillé et Godard 2021, 1968). Les arbres syntaxiques dans la figure 1 tirés de GGF permettent de mieux concevoir la distinction entre la construction clivée et la construction présentative (à savoir une proposition relative), décrite ci-dessus.

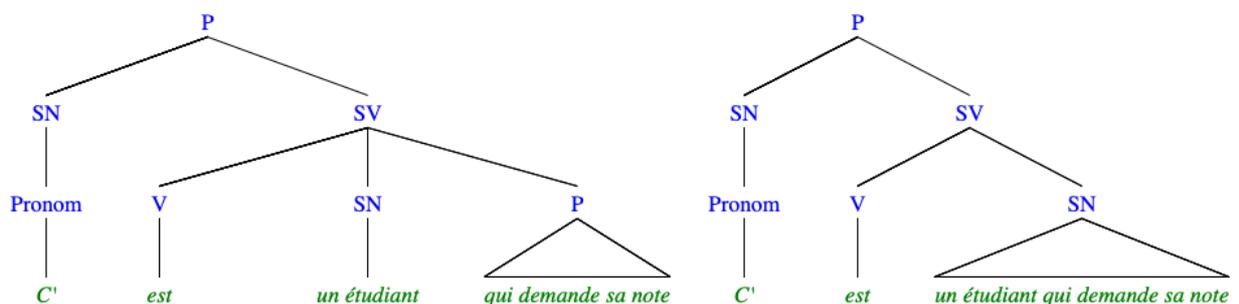


Figure 1 Arbre syntaxique d'une construction clivée et d'une proposition relative (Abeillé et Godard 2021, 1960)

Revenons maintenant à un exemple traité dans la partie 1.2.2 Les dislocations non-prototypiques, notamment *c'est une belle fleur que la rose*. Cette phrase n'appartient pas aux constructions homonymes traitées ci-dessus, bien qu'elle apparaisse dans des requêtes du type *c'est X qu-*. En effet, ce type de phrase est facilement distinguée des clivées puisque la subordonnée relative n'est pas suivie d'un noyau verbal. La subordonnée relative dans cette phrase est uniquement employée comme une sorte d'introducteur à un élément disloqué à droite.

Pour finir cette partie sur les constructions homonymes introduites par *c'est*, il nous semble important de soulever une cinquième construction homonyme, notamment les constructions de type *il + être + adjectif + que* où le pronom impersonnel *il* a été remplacé par le pronom démonstratif *ce*. Ces constructions sont attestées dans le corpus Résolco comme le montre l'exemple suivant :

27) C'est normal qu'ils sortent plus. (EC-CM2-2016-SGLEB-D1-R6-V1_N)

Notre étude portant également sur les clivées construites avec *il y a*, il nous semble pertinent de les distinguer de leurs constructions homonymes. Karssenber (2018, 3) distingue trois types de constructions avec *il y a* : la clivée 28), la phrase existentielle 29) et la phrase locative 30).

28) **Y'a** Jean **qui** s'est cassé la jambe. (Karssenber 2018, 3)

29) Le vendeur demanda s'**il y avait** une chance **qu'**elle revive ou pas. (CO-6e-2016-VTAC603-D1-R2-V1_N)

30) Et à côté de la maison, **il y avait** un garçon **qui** s'appelait Jean. (EC-CE2-2016-SGLEB-D1-R1-V1_N)

La clivée est la seule à pouvoir être transformée en ordre SV sans perdre son sens (*Jean s'est cassé la jambe*). Notons également que seule la relative de la clivée a une fonction prédicative. Dans les phrases existentielles et locatives la relative a une fonction appositive ou déterminative. La clivée dans 28) est décrite comme une phrase *événementielle*, or lorsque l'élément clivé présente une valeur pour une variable elle est *spécificionnelle*. Dans 31), l'élément clivé *celui d'la Roquette* est une valeur pour la variable *espaces verts* :

31) A : et alors vous manquez un peu **d'espaces verts** tu dis vers le square

B : on a un petit square

C : euh oui oui y a celui-là y a **celui d'la Roquette** qu'est pas loin (...) (Karssenber 2018, 4)

Comme nous l'avons évoqué sous 1.3.1 Les clivées non-prototypiques, les clivées peuvent également être construites avec le verbe *avoir*. Cependant, la construction *J'AI X qui*, ne peut pas toujours être analysée comme une clivée. Selon Conti (2010) qui a beaucoup travaillé sur cette construction, il ne s'agit uniquement d'une clivée lorsque la partie déclenchée par *qui* a une fonction prédicative et pas une fonction déterminative. Cela veut dire que l'exemple ci-dessous n'est pas une clivée, bien qu'il corresponde à la construction *J'AI X qui*.

32) **Elle avait** deux enfants **qui** s'appelaient Valentin et Alexandre. (CO-3e-2016-VTAC305-D1-R23-V1_N)⁷

Afin de cerner les cas où la construction *J'AI X qui* est une véritable clivée, nous aurons recours à une liste de contraintes qui pèsent sur la construction en question, citée dans Conti (2010) :

- Le sujet du verbe AVOIR désigne toujours un être ANIMÉ.
- L'antécédent doit effectuer une référence au sujet. « J'ai ma femme (ma fille, mon amie) qui m'attend », mais non pas * « J'ai votre cousine qui m'attend. »
- Le pronom relatif doit être *qui*.
- Le verbe de la relative peut dénoter une activité ou un état.
- Il peut y avoir une négation dans la relative ; il ne peut en revanche y en avoir une qui porte sur le verbe *avoir*.
- Les temps verbaux possibles sont les suivants : présent ou imparfait dans les deux propositions (situation la plus fréquente) ; éventuellement passé composé dans la relative.

⁷ Conti (2010) donne son propre exemple tiré de ses données. Cependant, nous estimons que son exemple n'illustre pas un exemple dont la forme surface est identique à une clivée de type *pronom + avoir X qui* et cela parce que le pronom relatif est *que* pas *qui* : « eux ils ont le logiciel qu'on a développé » (Conti 2010)

1.3.4 Analyse syntaxique

La construction clivée fait l'objet de plusieurs interprétations syntaxiques. Dans cette partie, nous présentons deux manières de concevoir les propriétés syntaxiques des clivées, dont la principale différence concerne la fonction de la copule. Commençons par un rappel de la structure schématisée de la clivée :

$$\frac{C'est X \quad Qu- + Verbe}{S1 \quad S2}$$

La première partie est celle qui inclut le clivé, l'élément mis en relief. La deuxième partie est déclenchée par le pronom relatif *qui* ou *que*.

Selon Lambrecht (2004), la clivée fait preuve d'une construction syntaxique « bi-phrastique » dont X est un argument de la copule, faisant de la copule un verbe constructeur. En revanche, même si le verbe de la relative ne semble pas occuper un rôle *syntactique*, il donne un rôle *sémantique* au constituant X. En revanche, selon Lambrecht (2004), la copule du clivé est vide au niveau sémantique.

Dans d'autres approches, notamment dans celle proposée par Blanche-Benveniste et l'équipe de GARS (1997), le constituant X n'est pas un argument de la copule, mais du verbe qui suit le pronom relatif. Au lieu de parler des constructions « bi-phrastiques » qui s'opposent aux constructions « mono-phrastiques », Blanche-Benveniste et al. (1997, 55) partent de l'idée qu'un verbe recteur et ses éléments régis peuvent être arrangés de différentes manières, nommées des « dispositifs ». Ainsi, dans la terminologie introduite par Blanche-Benveniste, les deux constructions ci-dessous font preuve de deux dispositifs différents:

- 33) Le maire tenait le fusil
- 34) C'était le maire qui tenait le fusil

Lorsque la relation de rection ne repose que sur le verbe recteur lui-même, comme dans 33), il s'agit d'un dispositif « direct ». Cependant, lorsque la rection est divisée en deux parties, comme dans 34), il est question d'un dispositif d'extraction. Dans un dispositif d'extraction, la copule est à considérer comme un « auxiliaire de dispositifs » (Blanche-Benveniste et al. 1997, 55), non comme un verbe constructeur.

Lorsque la copule n'est pas un verbe constructeur mais un auxiliaire de dispositif, certains traits s'ensuivent, notamment un « faible degré de verbalité » de la copule (Blanche-Benveniste 2002). Blanche-Benveniste (2002) constate que la copule est souvent au présent, même si elle a la possibilité d'être conjuguée en différents temps verbaux. Dans les cas où le verbe de la relative est dans un autre temps, la copule peut donc rester au présent, comme dans 35) :

- 35) C'est Lulli qui **inventa** ces symphonies.

De plus, l'auteur remarque que la copule est conjuguée au singulier dans la plupart des cas. Quant au pronom clitique *ce*, Blanche-Benveniste et al. (1997, 59) prétendent qu'il n'est pas un véritable pronom sujet puisque sa forme est constante⁸, peu importe de ce qui suit. Cela veut dire que le pronom n'est jamais remplacé par *cela* ou *ceci*.

⁸ À l'exception de la forme *ça* dans «ça sera...» (Blanche-Benveniste 1997, 59)

1.3.5 Analyse pragmatique et informationnelle

Du point de vue informationnel, la littérature distingue généralement deux, ou parfois trois, types de clivées (Karssenberget Lamiroy 2017 ; Prince 1978 ; Declerck 1988). Dans ce qui suit, un bref aperçu des trois structures informationnelles sera donné.

Dans le cas prototypique, le locuteur recourt à une clivée quand il veut focaliser un élément plutôt qu'un autre (Karssenberget Lamiroy 2018). Il s'agit en d'autre terme d'une fonction contrastive (Karssenberget Lamiroy 2017). Pour exemplifier, regardons les deux réponses possibles à la question suivante :

- 36) – Qui a mangé le gâteau ?
a) Pierre l'a mangé.
b) **C'est Pierre qui** l'a mangé.

À la différence de la phrase canonique 29a), la phrase clivée 29b) permet de mettre *Pierre* en relief et laisser le verbe de la relative constituer l'arrière-fond. Néanmoins, selon Lambrecht (2004) la véritable « newsworthiness » ou focus, ne se trouve pas dans le constituant clivé à lui seul, mais dans la relation entre le constituant clivé et la proposition subordonnée relative. Prince (1978) emploie le terme « stress-focus » pour décrire ce même phénomène. Dans la description de Prince (1978), la deuxième partie de la clivée est porteuse de l'information ancienne ou de l'information dite présupposée.

Cependant, il serait faux de dire que la deuxième partie des clivées contient toujours de l'information ancienne. En effet, les clivées peuvent être employée avec un deuxième type de structure informationnelle qui sert à enchaîner le discours. Dans ces clivées, parfois appelées des *clivées cohésives*⁹ (Karssenberget Lamiroy 2017), la première partie est liée au contexte antérieur. Par conséquent, la deuxième partie de la construction est porteur de l'information nouvelle. Les clivées cohésives sont souvent adverbiales (Rouquier 2018) :

- 37) En quarante-deux, les bombardements! Et **c'est là que** beaucoup de gens sont partis de Toulon.
(Karssenberget Lamiroy 2017)

Il existe également un troisième type de clivée dont *aucun* élément contient de l'information déjà donnée. Voici un exemple :

- 38) – Qu'est-ce qui se passe ?
– **Il y a** le chien **qui** aboie.

Dans la réponse à la question ci-dessus, l'interlocuteur apprend deux nouvelles informations : il s'agit d'un *chien* et ce chien *aboie*. Selon Karssenberget Lamiroy (2018), cette phrase fait preuve d'un « tout-focus » puisque toute l'information est nouvelle. Prince (1978) appelle ce type de clivée des clivées « à présupposition informative ». Cependant, comme le constate Rouquier (2018), cette notion peut sembler contradictoire étant donné que l'information de ces clivées ne sont pas présupposée, mais nouvelle. Dans sa définition des clivées « à présupposition informative », Prince met l'accent sur le degré de nouveauté dans la deuxième partie de la clivée:

« With these sentences, not only is the hearer not expected to be thinking about the information in the that-clause, but s/he is not expected even to KNOW it. In fact, the whole point of these sentences is to INFORM the hearer of that very information » (Prince 1978, 898)

⁹ Cette structure informationnelle est traitée sous plusieurs noms. Karssenberget Lamiroy (2017) en citent plusieurs : *clivées anaphoriques*, *clivées à enchaînement* ou *clivées à effet de récapitulation*. Durant un séminaire le 24 mai 2022, Lahousse a employé le terme de *topique – focus* pour désigner le même phénomène (Lahousse, 2022).

Outre la nature de l'information donnée dans la deuxième partie, la clivée « à présupposition informative » diffère de la clivée « stress-focus » dans sa possibilité d'être placée à l'initiale d'un texte (Rouquier 2018). L'exemple donné par Prince (1978) est tiré d'un journal, ou la phrase constitue le début d'un article de presse :

39) It was just about 50 years ago **that** Henry Ford gave us the weekend. (Prince 1978, 483)

Plusieurs auteurs, dont Declerck (1988), et Karssenberget Lahousse (2015) sont d'accord pour distinguer les clivées « stress-focus » des clivées « à préposition informative », or les termes employés varient. Karssenberget Lahousse (2015) utilisent le terme « all focus » pour désigner les clivées dont aucun élément n'est présupposé.

Karssenberget Lahousse (2015) ont observé les structures informationnelles qui sont favorisées par les clivées avec *c'est* et les clivées avec *il y a*. Les autrices ont constaté que les clivées introduites par *il y a* ont typiquement une structure tout-focus tandis que les clivées prototypiques ont tendance à avoir une structure focus-arrière-fond (contrastive).

Néanmoins, certains auteurs préfèrent ne pas opposer les clivées stress-focus aux clivées tout-focus, notamment Lambrecht (2001b) qui soulève de nombreuses difficultés liées à la catégorisation des clivées. Une des difficultés concerne l'établissement d'un critère formel et sémantique qui permettrait de différencier les différents types de clivées, sans se reposer sur l'intuition du locuteur. Pour Lambrecht (2001b, 484), il suffit de constater que toutes les clivées font partie d'une catégorie discursive qui sert à mettre un élément en relief. En revanche, il distingue les clivées prototypiques des pseudo-clivées (2004) en employant les termes *clivées à relative finale* pour désigner cette première et *clivée à relative initiale* pour désigner cette dernière. La terminologie employée par Lambrecht (2004) renvoie au fait que le pronom relatif se trouve dans la deuxième partie d'une clivée prototypique, mais dans la première partie d'une pseudo-clivée. Voici un exemple tiré de l'article de Lambrecht (2004, 23) :

40) Clivée à relative finale : C'est le vin **que** j'aime

41) Clivée à relative initiale : Ce **que** j'aime c'est le vin

Comme déjà soulevé dans 1.3.1 Les clivées non-prototypiques, il est à noter que toutes les pseudo-clivées ne suivent pas le schéma *ce qu- Y c'est X* comme les notions de Lambrecht le laissent sous-entendre.

1.3.6 Le statut des clivées à l'oral et à l'écrit

En ce qui concerne le statut des clivées à l'oral et à l'écrit, les travaux de Robaud et Sabio (2015, 2018) et de Karssenberget Lahousse (2015) sont centraux. Dans ce qui suit nous allons présenter les résultats pertinents de ces études.

Dans l'étude de Robaud et Sabio publiée en 2015, la fréquence des constructions clivées de type *c'est là où V* et *c'est là que V* est analysée et comparée dans un corpus oral et écrit. Les résultats montrent que ces types de clivées sont trois fois plus nombreux à l'oral qu'à l'écrit, la construction avec *que* étant plus utilisée que celle avec *où*.

Trois ans plus tard, en 2018, les mêmes auteurs observent un corpus oral et un corpus écrit de 2 millions de mots chacun, afin de comparer la répartition à l'oral et à l'écrit de cinq éléments clivés courants, à savoir *c'est pour ça que*, *c'est là que*, *c'est ainsi que*, *c'est alors que* et *c'est comme ça que*. Ils constatent que certaines de ces constructions sont majoritairement employées à l'oral tandis que d'autres sont plus fréquentes à l'écrit. Les constructions *c'est pour ça que*, *c'est là que* et *c'est comme ça que* figurent dans plus de 70% des cas à l'oral, tandis que les constructions *c'est ainsi que* et *c'est alors que* apparaissent

dans plus de 90% des cas à l'écrit. Les études de Robaud et Sabio (2015, 2018) montrent l'importance d'une analyse plus détaillée des clivées où l'élément clivé doit être pris en compte.

Par ailleurs, dans une étude présentée par Karssenbergh et Lahousse (2015) à la conférence *Incontro di Grammatica Generativa 41*, la répartition des clivées de type *il y a X qu-* est analysée dans un corpus écrit et oral. Les résultats montrent une fréquence élevée de ce type de construction dans le corpus écrit informel et dans le corpus oral par rapport au corpus écrit formel. En d'autres termes, leurs résultats sont conformes à ce que les grammairiens (Abeillé et Godard 2021, 117) constatent sur le statut à l'écrit et à l'oral de ce type de construction.

PARTIE 2 – Objectifs et données

Dans cette section, les objectifs et les hypothèses seront présentés, avant de décrire les données de l'étude.

2.1 Objectifs et hypothèses

L'objectif de notre étude peut être divisée en trois parties. Le premier objectif est de donner une description des usages des constructions orales dans les écrits scolaires, et plus précisément des constructions clivées et disloquées. Nous nous intéresserons également à la fréquence de ces usages. Le deuxième objectif est de comparer, en fonction du niveau scolaire, les usages et les fréquences observés des constructions clivées et disloquées. Le troisième et le dernier objectif est de proposer une méthode pour détecter de façon automatique les constructions décrites et observées dans l'étude. Nos questions de recherche sont les suivantes :

- Quels sont les usages des dislocations et des clivées dans des écrits scolaires ?
- Parmi ces usages, lesquels sont favorisés ?
- Il y a-t-il un lien entre la fréquence d'emploi des dislocations et des clivées et le développement de la littérature ? Autrement dit, les dislocations et les clivées disparaissent-elles des écrits des apprenants au fur et à mesure que leur maîtrise de l'écrit se développe ?
- Comment peut-on détecter de façon automatique les constructions clivées et disloquées dans un corpus écrit ?

En considérant que des études antérieures (Carter-Thomas 2009 ; Katz 1997) constatent que les sujets et les circonstants sont les fonctions syntaxiques les plus souvent clivées en français parlé, notre hypothèse est que les éléments clivés étudiés dans notre étude seront le plus souvent des sujets ou des circonstants. Pour ce qui est des dislocations, notre hypothèse est que la dislocation à gauche sera plus fréquente que la dislocation à droite, en accord avec les études antérieures (Blasco-Dulbecco 1999). Vu que l'élément disloqué porte le plus souvent le rôle syntaxique d'un SN sujet (Berrendonner 2021), une fréquence élevée de SN sujet en position disloquée est anticipée. Par ailleurs, nous nous attendons à ce que les dislocations soient moins fréquentes que les clivées.

Quant à la troisième question de recherche, notre prédiction se fonde sur les résultats d'une étude menée par Katharine Perera en 1986 sur l'acquisition de la différenciation grammaticale entre l'oral et l'écrit chez les enfants anglais. Dans cette étude, Perera montre que l'écriture de l'enfant devient de plus en plus éloignée de leur parole au fur et à mesure que l'enfant grandit. En d'autres termes, dans notre étude nous nous attendons à ce que la fréquence d'emploi des constructions disloquées et clivées diminue dans les écrits d'élèves tout au long de la scolarité.

En considérant que la dernière question de recherche est plutôt exploratoire, aucune hypothèse n'a été formulée. Les hypothèses sont résumées ci-dessous :

- Les sujets et les circonstants constitueront les fonctions syntaxiques les plus fréquentes de l'élément clivé.
- La dislocation à gauche sera plus fréquente que la dislocation à droite, et l'élément disloqué portera le plus souvent le rôle d'un SN sujet.
- Les clivées seront plus fréquemment employées que les dislocations.
- La fréquence des constructions disloquées et clivées diminuera graduellement en fonction de l'augmentation du niveau de scolarité.

2.2 Données : le corpus RésolCo

Cette partie contient deux sous-parties. La première sous-partie sert à introduire le corpus utilisé ainsi que de décrire sa constitution. Dans la deuxième sous-partie, l'étape de la normalisation du corpus sera précisée.

2.2.1 Constitution

Dans cette étude, la ressource RésolCo servira de base. RésolCo est un corpus qui a été conçu dans le cadre du projet É:CALM, à savoir *Écritures scolaires : Corpus, Analyses Linguistiques, Modélisations didactiques*. Le projet É:CALM¹⁰ regroupe plusieurs corpus dédiés à l'analyse des écrits scolaires. Dans la présentation du projet, il est constaté que les corpus d'apprenant en français langue maternelle sont rares, bien qu'ils deviennent de plus en plus nombreux.

Le corpus RésolCo, disponible sous licence Creative Commons¹¹, est constitué des productions écrites d'élèves et d'étudiants de différents niveaux de scolarité. Tous les textes recueillis dans ce corpus ont été rédigés en suivant la même consigne. Dans la consigne, on demandait aux participants d'écrire une histoire en y insérant, séparément et dans l'ordre donné, trois phrases spécifiques. La tâche donnée aux participants a été conçue pour favoriser l'étude de la cohérence discursive et de la continuité référentielle¹². Un aperçu de la composition du corpus est présenté dans le tableau ci-dessous :

<i>Niveaux scolaires</i>	<i>Années de collecte</i>	<i>Nb Textes</i>	<i>Nb Mots-formes</i>
CE2	2015-2018	55	6495
CM1	2015-2018	53	6467
CM2	2016	59	11275
6 ^e	2016-2018	114	21868
5 ^e	2018	4	1182
4 ^e	2018	47	13576
3 ^e	2016, 2018	54	17808
M2	2018, 2020-2021	231	102577
TOTAL	2015-2021	617	181248

Tableau 2 Composition du corpus RésolCo

Il faut noter que ce tableau n'est pas tiré directement du site du corpus. Cela s'explique pour plusieurs raisons. La première raison étant que ce dernier regroupe uniquement les textes qui ont été annotés et analysés dans les études sur la continuité référentielle. Vu que notre étude ne s'intéresse pas à la continuité référentielle, nous pouvons également analyser les textes qui ne contiennent pas ce type d'annotations. La deuxième raison est que le tableau disponible sur le site ne prend pas en compte les fichiers où les phrases de la consigne n'avaient pas été traitées dans l'ordre. Cet aspect ne nous semble pas critique dans notre étude. Finalement, nous notons que la taille du corpus du M2 a augmenté depuis la dernière publication sur le site. Afin de prendre en compte tous les textes normalisés du corpus, nous avons mis à jour le tableau.

Si l'on considère le tableau ci-dessus, on peut noter que le corpus est déséquilibré quant au nombre de textes et de mots-formes dans chaque niveau de scolarité. Le sous-corpus du M2 comporte le plus grand nombre de textes et de mots-formes, tandis que le sous-corpus du 5^e en comporte le moins. Nous estimons que les textes du 5^e ne sont pas suffisamment nombreux pour être inclus dans l'analyse. Le tableau 2 montre également les différences par rapport aux années de collecte de différents niveaux scolaires. Les textes du CE2 ont été collectés pendant 4 ans, de 2015 au 2018, tandis que tous les textes du CM2 ont été collectés pendant la même année, en 2016. Il faut également évoquer le fait que le corpus RésolCo ne contient aucun texte produit au lycée, ce qui veut dire que nous ne pouvons pas suivre l'évolution de l'usage des clivées par tous les niveaux scolaires.

Le corpus à télécharger contient les textes originaux en version scannée, des fichiers en version non-normalisée de même que normalisée en TXT, et finalement des fichiers XML balisés et annotés. Les

¹⁰ É:CALM, disponible à : <http://e-calm.huma-num.fr/le-projet/>

¹¹ Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage à l'Identique 3.0 France. Plus d'informations disponibles à : <https://creativecommons.org/licenses/by-nc-sa/3.0/fr/>

¹² Le corpus RésolCo est disponible à : <http://redac.univ-tlse2.fr/corpus/resolco/exploration.html>

fichiers XML ne sont pas normalisés. Dans notre étude, l'analyse portera sur les textes normalisés afin de diminuer le risque d'une annotation morphosyntaxique erronée due aux ambiguïtés des mots mal orthographiés. Les critères de normalisation seront décrits davantage sous 2.2.2.

Afin de pouvoir repérer plus facilement l'information métalinguistique de chaque texte du corpus, telle que le niveau scolaire du texte et l'élève qui l'a écrit, il faut connaître le système d'encodage du nom de chaque fichier. Ce nom contient 7 identifiants, présentés dans la Figure 2 Explication des identifiants des fichiers du corpus RésolCo:

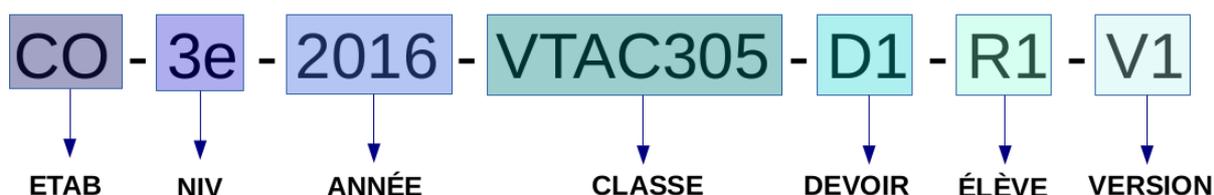


Figure 2 Explication des identifiants des fichiers du corpus RésolCo

Comme l'indique la Figure 2, certains textes apparaissent en plusieurs versions. Cela veut dire que ces textes sont passés par une étape de réécriture. Parmi les fichiers TXT normalisés, 32 textes ont subi une étape de réécriture. En d'autres termes, il y a 32 textes avec un identifiant qui finit par -V2. Il ne s'agit que des textes des élèves du 3^e et du CM1. En considérant que cette étude ne s'intéresse pas à l'impact des interventions pédagogiques sur l'écriture, les réécritures ne seront pas incluses dans l'étude. Par conséquent, elles ne figurent pas dans Tableau 2.

2.2.2 Normalisation

Chaque texte du corpus a subi une étape de normalisation où les erreurs au niveau de l'orthographe, des accords et de la segmentation en mots ont été corrigées. Ce traitement permet une meilleure annotation des textes par les outils tels que TreeTagger et facilite notre analyse des données.

Voici quelques exemples tirés du manuel de normalisation, disponible sur le site du corpus¹³:

- 1) Il etait une foie
[Normalisation] il était une fois
- 2) Ses maitres etait très gentille
[Normalisation] Ses maitres étaient très gentils
- 3) Il leprend
[Normalisation] Il le prend

Si un mot a été omis par l'élève, mais qu'il est identifiable de manière non ambiguë, il est ajouté dans l'étape de la normalisation. Tel est le cas pour l'article *la* dans l'exemple suivant :

- 4) la maman le pran par bouche
[Normalisation] la maman le prend par **la** bouche

Si un mot a été rajouté de manière non stylistique et non ambiguë, il est supprimé. Pourtant, s'il y a une redondance caractéristique de l'oral, telle qu'une dislocation à gauche ou à droite, elle est conservée.

¹³ Le manuel de normalisation sous la rubrique « Normalisation des textes d'élèves » est disponible à l'adresse suivante : <http://redac.univ-tlse2.fr/corpus/resolco/constitution.html>.

- 5) le chat pannicai le chat le chat dit « c'est pour quoi »
[Normalisation] le chat paniquait le chat dit « c'est pourquoi »
- 6) Le cha il marche
[Normalisation] le chat il marche

En ce qui concerne la ponctuation, la majuscule en début de production et le point en fin de production sont rétablis. Outre cela, la virgule est rajoutée uniquement dans les cas d'énumération ou dans les cas où un autre signe de ponctuation a été employée à la place de manière fautive. Ainsi, dans 6), l'absence de virgule entre l'élément disloqué et le pronom clitique n'est pas corrigé. Cela veut dire que la détection des dislocations ne peut pas se faire à l'aide des virgules entre un nom/pronom et un autre pronom.

Dans les cas où plusieurs propositions sont enchaînées sans segmentation ou conjonctions, la balise <segmentation/> a été rajoutée. Cette balise peut remplacer un signe de ponctuation ou un connecteur. Voici un exemple avec la balise <segmentation/> :

- 7) Oh c'est la fille de l'ancienne propriétaire elle habitait dans cette maison depuis longtemps et depuis son décès c'est Pauline sa fille qui vit ici (CO-3°-2018-VTAC305-D1-R2-V1)

[Normalisation] Oh c'est la fille de l'ancienne propriétaire <segmentation/> elle habitait dans cette maison depuis longtemps et depuis son décès c'est Pauline sa fille qui vit ici

Néanmoins, les balises <segmentation/> ne sont pas présentes dans les fichiers en format TXT utilisés dans ce travail, et nous avons choisi de ne pas les inclure. En effet, les balises n'indiquent que les endroits où une ponctuation pourrait être mise, mais nous ne savons pas comment l'auteur de texte aurait souhaité segmenter son texte. Le fait de ne pas inclure ces balises permet donc de rester plus fidèle aux textes de base. De plus, nous faisons en sorte que les patrons soient réutilisables pour d'autres corpus ne comportant pas nécessairement ce type de balise.

PARTIE 3 – Extraction des constructions clivées

Dans cette étape de l'étude, l'analyse des clivées a rapidement pris de l'ampleur. Dans le cadre de ce travail, le choix a donc été fait de mettre l'analyse des dislocations de côté. Ceci dit, la méthode employée pour détecter les clivées a le potentiel d'être modifiée pour détecter les dislocations dans une étude ultérieure.

Dans ce chapitre, nous allons commencer par une description de l'outil TXM, utilisé pour une première extraction des constructions clivées. En considérant que la priorité de nos premiers patrons était le rappel et pas la précision, cette partie contient une analyse des faux positifs ainsi que des patrons améliorés. De plus, nous décrirons notre annotation des vrais positifs.

Étant donné qu'un des objectifs de cette étude est d'explorer les méthodes possibles afin de détecter de façon semi-automatique les clivées des écrits scolaires, l'usage de TXM est complété par l'extraction automatique des clivées par un script Python qui repose sur le modèle d'annotation de Stanza. Après avoir décrit la nature des annotations faites par Stanza et le fonctionnement du script Python, nous allons évaluer la performance de cette méthode. L'évaluation sera suivie d'une observation plus en détail du bruit et du silence de ce script.

3.1 Première étape : constitution du gold-standard avec TXM

La figure 3 ci-dessous permet d'illustrer le déroulement de la constitution de notre gold-standard, ainsi première élaboration des patrons pour extraire les clivées. Ensuite, nous avons ajouté des patrons supplémentaires afin d'identifier des clivées plus longues. Les résultats de ces deux versions des patrons ont ensuite été triés manuellement en deux catégories : les vrais positifs (les clivées) et les faux positifs (d'autres constructions). Cette étape a donné lieu à l'établissement de notre gold-standard¹⁴. De plus, à l'aide de notre analyse des faux positifs, nous avons pu améliorer la précision de nos patrons TXM. Une fois les patrons améliorés, les vrais positifs ont été annotés plus en détail. Toutes les étapes de la figure 3 seront décrites plus précisément dans les sous-parties qui suivent (3.1.1 Extraction des constructions avec TXM-3.1.4 Annotation des vrais positifs).

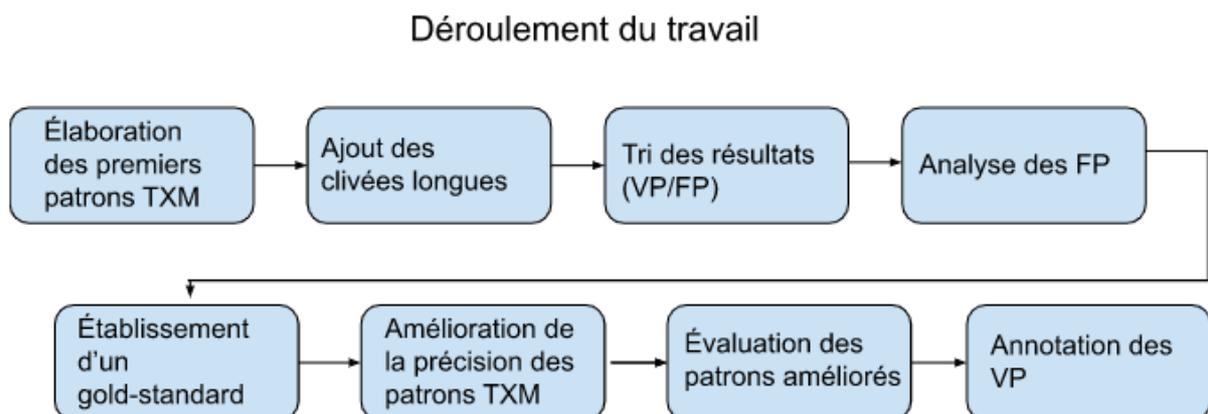


Figure 3 Schéma de la méthode d'extraction

¹⁴ Nous avons pris le choix de ne pas procéder à l'annotation manuelle d'un échantillon des copies pour la constitution du gold-standard, et cela parce que le phénomène qui nous intéresse n'est pas suffisamment fréquent. Il aurait fallu annoter un nombre considérable de copies avant de tomber sur des clivées.

3.1.1 Extraction des constructions avec TXM

TXM (Heiden et al. 2010) est une plateforme d'analyse textuelle créée dans le cadre d'un projet financé par ANR¹⁵ intitulé *Textométrie*¹⁶. TXM permet d'analyser des données textuelles de façon quantitative et qualitative. Cet outil a été choisi en raison de la diversité des fonctionnalités proposées.

Le corpus normalisé sous format txt a été importé dans TXM. Ensuite, une annotation morphosyntaxique du corpus a été effectuée sur le corpus avec TreeTagger. L'annotation morphosyntaxique nous a permis de faire des requêtes à partir des parties du discours.

L'étape suivante a été d'exploiter la possibilité de détecter des clivées à l'aide des patrons exprimés en langage CQL (*Corpus Query Language*). Ce dernier est proposé dans TXM et permet d'exprimer des motifs linguistiques, tels que des constructions syntaxiques.

Les résultats de nos patrons en CQL ont été exportés dans un document csv pour être annotées manuellement afin de distinguer les vrais positifs des faux positifs. Étant donné que le contexte est nécessaire pour déterminer s'il s'agit d'une clivée ou une construction homonyme (cf. 1.3.3 Les constructions homonymes) nous avons utilisé les concordances dans TXM qui permettent de voir le contexte à gauche et à droite de la construction.

En considérant que notre premier objectif est de décrire les constructions clivées identifiées dans le corpus, l'important en premier lieu était d'assurer que les requêtes identifient autant de clivées que possible. Cela veut dire que les premiers patrons n'ont pas forcément une bonne précision. Cette dernière a été améliorée de façon réitérée dans une étape ultérieure. À l'aide de nos premières annotations, présentées dans ce qui suit, nous avons pu établir un gold-standard à partir duquel nous avons pu évaluer les versions améliorées des patrons, ainsi que notre script Python.

Nous rassemblons dans le tableau 3 les premières versions des patrons que nous avons construits :

Nb	Construction	Patron	Total	VP ¹⁷	Précision
1	C'est X que	[frlemma="ce"][frlemma="être"][]{}{0,7}[frlemma="que"]	133	69	0.52
2	C'est X qui	[frlemma="ce"][frlemma="être"][]{}{0,7}[frlemma="qui"]	114	41	0.36
3	Il y avoir X que	[frlemma="y"][frlemma="avoir"][]{}{0,7}[frlemma="que"]	20	1	0.05
4	Il y avoir X qui	[frlemma="y"][frlemma="avoir"][]{}{0,7}[frlemma="qui"]	37	21	0.57
5	Pronom avoir X qui	[frpos="PRO:PER" &! frlemma="y"][frlemma="avoir"][frpos!="SENT" &! frpos="VER.*"]{}{1,7}[frlemma="qui"]	20	2	0.1
6	Pseudo-clivée	[frlemma="ce"][frlemma="que" "qui"][frpos!="SENT"]{}{1,20}[frlemma="ce"][frlemma="être" & frpos="VER.*"]	19	17	0.89
TOTAL			343	151	

Tableau 3 Résultats des premiers patrons TXM.

Bien que nous estimions que nos premiers patrons sont suffisamment génériques pour détecter la grande majorité des clivées dans le corpus, il est possible de constater que nos premiers patrons ne sont pas suffisamment performants au niveau de la précision. Notons que le bruit constitue plus que la moitié de nos résultats.

Dans nos premières requêtes, nous nous sommes limitées à 7 mots dans l'élément clivé, et cela pour ne pas avoir trop de bruit dans nos résultats. Cependant, il y a effectivement un nombre restreint de phrases

¹⁵ ANR= Agence Nationale de la Recherche.

¹⁶ <http://textometrie.ens-lyon.fr/>

¹⁷ VP = Vrais positifs

dans le corpus où l'élément clivé comporte plus de sept mots. Ces occurrences ont été identifiées dans des requêtes à part, à savoir :

Nb	Construction	Patron	Total	VP	Précision
7	C'est X qu- (> 7 mots)	[frlemma="ce"][frlemma="être"][frpos!="SENT"]{7,20}[frlemma="que" "qui"]	67	10	0.15
8	Il y a X qu- (> 7 mots)	[frlemma="y"][frlemma="avoir"][frpos!="SENT"]{7,20}[frlemma="que" "qui"]	32	0	0
		TOTAL	99	10	

Tableau 4 Résultats des patrons TXM pour les clivées C'est X qu- et Il y a X qu- de plus de 7 mots.

Comme on l'a vu dans le tableau 3, nous avons extrait toutes les constructions dont la copule et *qu-* sont séparés par 7 à 20 mots. Notons que les patrons ci-dessus prennent en compte les deux pronoms relatifs (*que* et *qui*) dans le même patron. Même si nous estimons peu probable que l'élément clivé contienne plus de 20 mots, il faut souligner que cette étude ne prend pas en compte les clivées dont l'élément clivé compte plus de 20 mots.

Par ailleurs, il faut noter que le patron permettant de détecter les pseudo-clivées se distingue des autres patrons, et cela puisque l'élément clivé se trouve dans la deuxième partie de la clivée. En effet, la partie de la clivée qui est identifiée par le patron des pseudo-clivées est la première, pas la deuxième. Par conséquent, l'élément clivé est identifié dans le contexte droit. Ceci mentionné, la limite de ce patron est que la première partie ne peut pas contenir plus de 20 mots. Pour cette raison, nous avons vérifié à l'aide d'un patron TXM qu'aucune pseudo-clivée à relatif initial ne contienne une première partie de plus de 20 mots :

Nb	Construction	Patron	Total	VP
9	Pseudo-clivée (> 7 mots)	[frlemma="ce"][frlemma="que" "qui"][frpos!="SENT"]{20,40} [frlemma="ce"][frlemma="être" & frpos="VER.*"]	1	0

Tableau 5 Résultats des patrons TXM pour les pseudo-clivées avec un relatif initial de plus de 20 mots.

3.1.2 Analyse des faux positifs

Dans cette sous-partie, les faux positifs des premiers patrons seront décrits et exemplifiés. Pour chaque type de faux positifs, des améliorations possibles seront proposées, ainsi que des limites potentielles de TXM. Nous allons finir par discuter de quelques cas ambigus.

Qu- appartenant à une autre proposition ou phrase

Dans les patrons 1-4, aucune contrainte a été mise sur les mots de l'élément clivé ([]{0,7}). Ce choix a été fait pour être sûr de détecter toutes les clivées, même dans les cas des problèmes de ponctuation. Cependant, il a pu être constaté que les patrons doivent exclure les ponctuations de fin de phrase dans l'élément clivé.

- 1) **C'était** toujours le même résultat. Alors **que** Swan commença à s'impatienter [...] (CO-3e-2018-VTAC305-D1-R7-V1_N)
- 2) mais avec ce temps, **c'était** insensé ! Personne ne comprit ce **qui** leur était passé par la tête. (UN-M2-2020-UCL-D1-R40-V1_N)
- 3) **Il n'y avait** rien. Alors **qu'**il était en train de repartir dans leur chambre (CO-4e-2018-LSPJJRC-D1-R28-V1_N)

En revanche, l'élément clivé peut contenir d'autres types de ponctuations, tel que des parenthèses 4), des virgules 5) ou même des points de suspension 6):

- 4) mais, il y avait **une personne au comptoir (qui nous a finalement trouvés)** qui nous expliqua que tous les clients rentrent avant la tombée de la nuit par peur de voir des démons en particulier le démon E N D qui aurait disparu il y a 15 ans de cela. (CO-3e-2018-FSBJC6-D1-R9-V1_N)
- 5) C'est **l'un d'eux, Nicolas**, qui m'a tout expliqué [...](UN-M2-2021-UCL-D1-R64-V1_N)
- 6) Un instant et c'était **des projets, des rêves...toute une vie** qui s'écroulaient. (UN-M2-2021-UCL-D1-R69-V1_N)

En analysant les faux positifs de nos premiers patrons, nous observons également des occurrences où les mots qui précèdent *qu-* ne constituent pas à un seul syntagme ou une seule proposition subordonnée. Pour exemplifier, notre patron a identifié deux constructions dans l'exemple 7), à savoir *c'est faux et que* et *c'est des bêtises et sa mère lui prouve que*. Tout d'abord, nous notons que le *que* est une conjonction de subordination dans les deux cas. Deuxièmement, la phrase contient deux propositions principales coordonnées par *et*, dont la première contient deux subordonnées complétives (*que c'est faux* et *que c'est des bêtises*). Ainsi, le *que* du premier faux positif ne fait pas partie de la même subordonnée que *c'est faux*, et le *que* du deuxième faux positif ne fait pas partie de la même proposition que *c'est des bêtises*.

- 7) Il dit que **c'est** faux et **que c'est** des bêtises et sa mère lui prouve **que** les loups ne sont pas méchants de toute façon. (CO-4e-2018-LSPJRD-D1-R15-V1_N)

Pour améliorer nos patrons, nous ajoutons une contrainte sur le mot qui précède *qu-*. Ce mot ne peut pas être une conjonction de coordination. En revanche, nous gardons la possibilité d'avoir un verbe avant *que* dans les clivées prototypiques pour pouvoir identifier les clivées dont l'élément clivé est une subordonnée qui contient un verbe (voir 8)).

- 8) Je sais que **c'est** juste pour que je t'obéisse **que** tu as inventé cette histoire. (UN-M2-2020-UCL-D1-R80-V1_N)

Les faux positifs dont le relatif appartient à une autre proposition figurent également dans les patrons des clivées non-prototypiques :

- 9) **Il y avait** des bêtes et ça fait longtemps **que** je ne me suis pas amusé ! (CO-6e-2017-VTAC602-D1-R18-V1_N)
- 10) Il l'a rencontré **il y a** bientôt dix ans alors **que** les sites de rencontres commençaient seulement à devenir à peu près pratique (UN-M2-2021-UCL-D1-R63-V1_N)
- 11) **Il n'y avait** aucun bruit ce **qui** rendait l'atmosphère pesante. (UN-M2-2021-UCL-D1-R47-V1_N)
- 12) **ils avaient** foi, celui **qui** faisait brûler leur cœur d'une charité ineffable. (UN-M2-2021-UCL-D1-R24-V1_N)

Dans 10), *que* appartient à la locution *alors que*. Nous observons plusieurs occurrences parmi les faux positifs dont *que* appartient à une locution ou une négation (*parce que*, *pendant que*, *plus que*, *alors que*, *plutôt que*). Nous observons non seulement des faux positifs avec *que* mais aussi avec *qui*. L'usage du pronom relatif *ce qui* dans l'exemple 11) indique que l'antécédent est une proposition complète. Ainsi, la phrase n'a pas besoin de la partie qui suit le pronom relatif pour fonctionner. Dans l'amélioration des patrons, nous ajoutons une contrainte du mot qui précède *que* pour que celui-ci ne soit pas un des mots suivants : *parce*, *pendant*, *plus*, *plutôt*. Notons que nous n'y ajoutons pas *alors*, et cela puisque cela enlèverait les véritables clivées de type *c'est alors que*.

C'est X SCONJ

Comme évoqué ci-dessus, certaines occurrences données en sortie de nos patrons ne contiennent pas un pronom relatif, mais une conjonction de subordination :

- 13) **C'est** normal **que** le ciel soit noir à 18 heures. (EC-CM2-2016-SGLEA-D1-R2-V1_N)
 14) Un manoir, même repoussant, **c'était** toujours mieux **qu'**un petit appartement à Paris. (EC-CM2-2016-SGLEA-D1-R2-V1_N)

Même si l'étiquetage morphosyntaxique de Treetagger permet de limiter la catégorie grammaticale de *qui* ou *que* à un pronom relatif, nous avons pris la décision de ne pas nous servir de cette étiquetage dans nos patrons. Ce choix a été fait puisque la performance de Treetagger reste plutôt faible, surtout pour les données non-standard comme les écritures scolaires de jeunes élèves. En d'autres termes, il y a un risque que Treetagger attribue l'étiquette SCONJ (conjonction de subordination) même aux pronoms relatifs, ce qui résulterait dans plus de silence.

Constructions copulatives canoniques

Nos premiers résultats extraits de TXM contiennent des propositions relatives, à savoir des constructions copulatives canoniques (désormais CCC). Comme déjà soulevé dans la partie 1.3.3 Les constructions homonymes le pronom *ce* dans les CCC n'est pas vide de sens, mais entretient une relation anaphorique avec un élément déjà introduit ou connu dans le discours. De plus, le pronom relatif dans la deuxième partie ne précède pas le prédicatif de la phrase, mais il fait partie d'un syntagme nominal. Voici quelques exemples tirés de notre corpus. Nous soulignons l'élément dont le pronom du CCC fait référence selon nous :

- 15) Il se retourna en entendant ce grand bruit. **C'était** un éboulement de pierres **qui** allait droit sur eux. (EC-CE2-2016-VRX-D1-R1-V1)
 16) Tina avait vraiment peur de rentrer chez elle , car **c'était** la seule **qui** rentrait toute seule à pied (CO-3e-2018-FSBJC6-D1-R14-V1_N)

Treetagger ne nous permet pas d'identifier des relations de coréférence entre les mots du corpus. De plus, la version de corpus que nous utilisons n'est pas annoté en coréférence. Par conséquent, il n'y a pas d'amélioration du patron TXM qui permette de les éviter.

Nos résultats contiennent également une construction dont l'analyse demandait d'aller plus loin que de vérifier si le pronom *ce* avait un lien anaphorique avec un élément antécédent. En effet, dans 17), le pronom clitique *ce* au début de la construction n'a pas de référent anaphorique :

- 17) **C'est** l'histoire d'une jeune fille **qui** vivait avec sa famille dans une petite maison en Californie. (CO-3e-2018-VTAC305-D1-R9-V1_N)

À premier abord, cette construction pourrait être prise pour une clivée dont l'élément clivé est le syntagme nominal *l'histoire d'une jeune fille*, et cela puisque le pronom *ce* n'a pas de référent, comme dans le cas d'une clivée. Cependant, si *l'histoire d'une jeune fille* constituait un élément clivé, cela voudrait dire que la tête de ce syntagme, à savoir *l'histoire*, serait l'argument du verbe recteur *vivait*. Nous présumons que cela n'est pas le cas, le verbe étant plutôt associé à *une jeune fille*. Par conséquent, la proposition relative doit être analysée comme appartenant au syntagme nominal *une jeune fille*, qui a son tour appartient au syntagme prépositionnel (*d'une jeune fille*) dans le syntagme nominal (*l'histoire d'une jeune fille*). En d'autres termes, le pronom *ce* est le sujet, la copule est le verbe recteur et l'attribut est un syntagme nominal complexe qui contient une relative :

C'	est	l'histoire d'une jeune fille qui vivait avec sa famille dans une petite maison en Californie.
SUJET	VERB	ATTRIBUT

Notons que cette phrase pourrait être formulée avec un autre présentatif, tel que *voici* :

Voici l'histoire d'une jeune fille qui vivait avec sa famille dans une petite maison en Californie.

Selon Rouget et Salze (1986, 123), l'anaphore n'est pas toujours réalisé en discours, mais peut être inféré à l'aide de la situation d'énonciation. Cependant, dans le cas ci-dessus, l'anaphore est difficile à reconstruire.

Constructions existentielles ou locatives

Si les CCC étaient fréquentes dans les patrons pour détecter les clivées prototypiques, les patrons pour les clivées de type *il y a x qui* ont généré des constructions existentielles ou locatives :

- 18) **Il y a** régulièrement de petites secousses sismiques **qui**, parfois, font quelques dégâts matériels. (UN-M2-2020-UCL-D1-R115-V1_N)
- 19) À côté de la maison **il y avait** un garçon **qui** s'appelait Jean. (EC-CE2-2016-SGLEB-D1-R1-V1_N)

Comme dans le cas des propositions relatives, ce type de faux positifs ne peut pas être évité dans TXM.

Autres faux positifs

Cette sous-partie est consacrée aux faux positifs qui ne rentrent pas dans les catégories traitées jusqu'ici. Il s'agit avant tout des constructions non-prototypiques.

Quand les faux positifs de *IL Y A X que* ne peuvent pas être catégorisés comme une construction existentielle ou locative, il peut s'agir d'une construction temporelle, comme dans 20). Nous notons également que notre patron devrait inclure le pronom *il* au début pour ne pas identifier des occurrences telles que 21):

- 20) Ah ! **Il y a** bien longtemps **que** je n'avais pas ri comme ça (EC-CM2-2016-SGLEB-D1-R24-V1_N)
- 21) L'on raconte qu'un couple de traders **y avait vécu et que** ils s'adonnaient à des plaisirs malsains impliquant toutes sortes d'objets dangereux et d'animaux de diverses espèces (UN-M2-2021-UCL-D1-R26-V1_N)

La majorité des sorties du patron pour identifier les clivées de type *J'AI X qui* sont de type *pronom + verbe + SN avec relatif* :

- 22) **Elle** avait un frère **qui** était très jeune. (EC-CM1-2017-TBZX-D1-R1-V1_N)
- 23) **Ils** eurent tous un frisson **qui** leur parcourut le corps. (CO-6e-2016-PJPR5-D1-R15-V1_N)

Comme déjà soulevé dans 1.3.3 Les constructions homonymes, ce type de construction n'est pas équivalente à une clivée. Cependant, étant donné que la forme est homonyme à une clivée, aucune modification du patron TXM permet d'éviter ces constructions. Quoiqu'il en soit, ce type de clivée reste rare dans le corpus.

Cas ambigus

Avant de proposer des améliorations des patrons, il faut soulever une difficulté identifiée dans l'usage de TXM : les cas où l'élément clivé contient *que* ou *qui*. Les patrons qui finissent avec *que* ne sont appliqués que jusqu'à la première occurrence de *que*. De même, les patrons avec *qui* finissent au premier *qui*. Voici deux exemples où nous indiquons en gras les séquences qui ont été identifiées par le patron :

- 24) Mais **c'est depuis qu'**il est mort qu'il se passe des choses bizarres dans la maison (CO-4e-2018-LSPJJRC-D1-R11-V1_N)
- 25) Je sais que **c'est juste pour que** je t'obéisse que tu as inventé cette histoire. (UN-M2-2020-UCL-D1-R80-V1_N)

Bien que les séquences identifiées par TXM ne constituent pas des clivées complètes, elles font partie des clivées (*c'est depuis qu'il est mort que, c'est juste pour que je t'obéisse que*). Pour cette raison, les occurrences de ce type sont incluses parmi les vrais positifs. Cependant, nous estimons que ce problème pourrait être réglé à l'aide des scripts Python.

3.1.3 Amélioration des patrons

Suite à notre tri des faux positifs et des vrais positifs de nos premiers patrons, notre gold-standard a pu être établi. Nous le présentons dans le tableau ci-dessous :

Quantification du gold-standard	
Type de clivée	VP + FN
<i>C'est X que</i>	73
<i>C'est X qui</i>	47
<i>Il y a X que</i>	1
<i>Il y a X qui</i>	21
<i>Pronom avoir X qui</i>	2
<i>Pseudo-clivées</i>	17
TOTAL	161

Tableau 6 Quantification du gold-standard

L'étape suivante consistait à améliorer nos patrons. Nos améliorations concernent uniquement les constructions *c'est X que, c'est X qui* et *il y a X qui*. Cette limitation a été faite en raison du faible nombre de vrais positifs des patrons *il y a X que* et *j'ai X qui*. Dans le cas des pseudo-clivées, nous jugeons la performance de la première version du patron suffisamment bonne.

Les nouveaux patrons sont listés ci-dessous :

Nb	Construction	Patron	Total	VP ¹⁸	Précision
1	<i>C'est X que</i>	[frlemma="ce"][frlemma="être" & frpos="VER.*"][frpos!="SENT"]{0,15}[frlemma=".*" &! frpos="KON" &! frpos="SENT" &! frlemma="parce pendant plus plutôt ce"][frlemma="que"]	107	73	0.68
2	<i>C'est X qui</i>	[frlemma="ce"][frlemma="être" & frpos="VER.*"][frpos!="SENT"]{0,14}[frlemma!="ce" &! frpos="KON" &! frpos="SENT"][frlemma="qui"]	122	46	0.38
4	<i>Il y avoir X qui</i>	[frlemma="il"] [word="n"]{0,1}[frlemma="y"][frlemma="avoir"][frpos!="SENT"]{0,6}[frpos!="SENT" &! frlemma="ce"][frlemma="qui"]	35	21	0.60
TOTAL			264	144	

Tableau 7 Patrons TXM améliorés

Les modifications de chaque patron sont résumées ci-dessous :

C'est X que

- Pas de ponctuation de fin de phrase dans l'élément clivé
- *Que* n'est pas précédé par : *pendant, parce, plus, plutôt, ce* ou une conjonction de coordination
- Le nombre maximal de mots dans l'élément clivé est ajusté selon l'élément clivé le plus long

C'est X qui

- Pas de ponctuation de fin de phrase dans l'élément clivé
- *Que* n'est pas précédé par : *ce* ou une conjonction de coordination
- Le nombre maximal de mots dans l'élément clivé est ajusté selon l'élément clivé le plus long

¹⁸ VP = Vrais positifs

Il y a X qui

- La construction commence par le pronom *il*
- Il y a une négation optionnelle avant *y*
- Pas de ponctuation de fin de phrase dans l'élément clivé
- *Que* n'est pas précédé par : *ce*, un verbe ou une conjonction de coordination
- Le nombre maximal de mots dans l'élément clivé est ajusté selon l'élément clivé le plus long

Dans le tableau ci-dessous nous présentons les f-mesures des nouvelles versions de nos patrons. Les f-mesures ont été calculées en utilisant notre gold-standard comme référence. Les valeurs entre parenthèses indiquent les précisions de nos premiers patrons, à savoir les patrons montrés dans le tableau 3. L'amélioration de la précision est la plus remarquable pour les clivées de type *c'est X que*, dont la précision a augmenté de 16 points.

<i>Construction</i>	<i>Rappel</i>	<i>Précision</i>	<i>F-mesure</i>
<i>C'est X que</i>	1.00	0.68 (0.52)	0.81
<i>C'est X qui</i>	0.98	0.38 (0.36)	0.55
<i>Il y a X qui</i>	1.00	0.60 (0.57)	0.75

Tableau 8 F-mesures des patrons TXM améliorés

En consultant les F-mesures ci-dessus, il est à noter que le rappel du patron *c'est x qui* n'atteint pas 1, mais 0,98. Il s'agit d'un seul faux négatif, comme on le voit dans l'exemple :

- 26) **C'est** tout un tintamarre de métallique, une cavalcade sur une rue pavée **qui** lui emplissait les oreilles.
(UN-M2-2020-UCL-D1-R48-V1_N)

Étant donné que *pavée* est un participe passé et que notre patron ne prend pas en compte les occurrences avec un verbe avant *qui*, 26) n'a pas été identifiée. Néanmoins, nous constatons que ce faux négatif n'affecte pas la f-mesure de manière décisive.

3.1.4 Annotation des vrais positifs

Les résultats des requêtes ont été extraits en format .csv où ils ont subi plusieurs étapes d'annotation. Dans la première étape d'annotation, nous avons séparé les constructions homonymes et erronées des constructions clivées (voir 3.1.2 Analyse des faux positifs).

Dans un deuxième temps, nous avons approfondi notre analyse en établissant les niveaux d'annotation suivants :

Métadonnées

- Niveau scolaire (école primaire, collège, université)
- Classe (CE2, CM1, CM2, 6^e, 4^e, 3^e, M2)

Informations linguistiques

- Nombre de mots-formes dans le texte
- Fréquence relative des clivées/texte
- Fonction syntaxique de l'élément clivé
- Constituant syntaxique de l'élément clivé
- Nombre de mots-formes dans l'élément clivé
- Type de clivée (*C'est X que*, *C'est X qui*, *il y a X qui*, *il y a que*, pronom avoir *qui*, pseudo-clivée)

Les métadonnées, à savoir le niveau scolaire et la classe, permettent de faire une analyse contrastive des fréquences des différents types de clivées. Le nombre de mots-formes par texte permet de calculer la fréquence relative des clivées. Le nombre de mots-formes par texte a été calculé à l'aide d'un script Python. Le script Python segmentait les textes normalisés bruts donnés en entrée pour donner le nombre

de mots-formes par texte en sortie. La décision a été prise de se fonder sur les mots-formes pour assurer la cohérence avec TXM qui traite les ponctuations comme des mots. La fréquence relative, qui correspond au nombre de clivées par texte divisé par le nombre de mots-formes par texte, a été calculée automatiquement à l'aide d'un autre script Python.

Nous avons annoté la fonction syntaxique et le constituant syntaxique de l'élément clivé manuellement afin de pouvoir décrire la nature des constructions employées. En ce qui concerne l'annotation des constituants syntaxiques, nous employons ceux cités dans la partie 1.3.2 L'élément clivé. Cependant, nous avons fait quelques adaptations de ces catégories. Premièrement, les catégories *Adverbe* et *Adjectif* sont nommées *syntagme adverbial* et *syntagme adjectival* dans ce travail, et cela pour mieux représenter la nature du constituant en question. Deuxièmement, nous ne faisons pas la distinction entre de différents types de subordinées dans le corpus comme dans l'article de Rouquier (2018). Finalement, l'annotation a fait apparaître la nécessité d'ajouter une catégorie supplémentaire, à savoir le gérondif :

27) C'est **en sortant de chez elle** qu'elle se retournait pour contempler sa maison. (UN-M2-2020-UCL-D1-R30-V1_N)

Lors de l'annotation des constituants syntaxiques, nous avons rencontrés certaines difficultés de catégorisation. Par exemple, bien que les segments *pour ça* et *comme ça* sont tous les deux des locutions adverbiales, nous avons annoté *pour ça* comme un SP, et cela puisque le pronom est précédé d'une préposition. En revanche, la locution *comme ça* a été annoté comme un adverbe, étant donné son usage adverbial. Nous sommes conscientes de la possibilité d'annoter les deux locutions comme des pronoms accentués.

Dans la mesure où les constituants syntaxiques potentiels d'une clivée dépendent du type de clivée, nous faisons la distinction des différents types. Le tableau ci-dessous illustre les catégories potentielles de chaque type :

C'est X qu-	Il y a X qu-	Pronom avoir X qui	Pseudo-clivée
<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué • Syntagme prépositionnel • Adverbe • Syntagme adjectival • Subordonnée • Verbe à l'infinitif • Gérondif 	<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué • Syntagme prépositionnel (avec négation) 	<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué 	<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué • Syntagme prépositionnel • Subordonnée

Tableau 9 Constituant syntaxique de l'élément clivé

Concernant le nombre de mots-formes dans l'élément clivé, nous avons effectué ce comptage à la main, en prenant en compte les signes de ponctuation comme la virgule ou la parenthèse. Nous tenons à préciser que nous considérons que l'élément clivé d'une pseudo-clivée s'arrête à la fin de la phrase. Ainsi, l'élément clivé de la pseudo-clivée ci-dessous contient 16 mots.

28) Ce qu'il aimait le plus c'est **engloutir les enfants lors des grandes pluies afin de les piéger à jamais dans le sol**. (UN-M2-2021-UCL-D1-R63-V1_N)

Nous ne prenons pas en compte la ponctuation de fin de phrase puisqu'elle ne serait pas incluse dans une version clivée prototypique de la phrase :

29) *c'est engloutir les enfants lors des grandes pluies afin de les piéger à jamais dans le sol que...!***c'est engloutir les enfants lors des grandes pluies afin de les piéger à jamais dans le sol_que....*

3.2 Deuxième étape : modèle d'extraction des clivées avec Python

Dans cette section, nous présenterons un modèle d'extraction des clivées. Il s'agit d'un script Python permettant d'extraire des constructions clivées. Ce dernier est fondé sur les relations syntaxiques entre les mots d'une construction clivée. Les relations syntaxiques sont identifiées dans notre corpus à l'aide d'un post-traitement effectués avec le parser Stanza. Dans ce qui suit, nous commencerons par décrire ce dernier. Ensuite, l'élaboration de notre script Python sera explicitée. Nous évaluerons la performance de notre modèle en comparant la f-mesure à celles de nos meilleurs patrons TXM. Cette section terminera par une analyse des faux positifs et des faux négatifs identifiés dans les sorties de notre modèle.

3.2.1 Post-traitement des données avec Stanza

L'annotation générée par Treetagger est limitée aux lemmes et aux parties du discours. Cela veut dire que les relations syntaxiques entre les constituants ne sont pas prises en compte. Comme nous l'avons déjà évoqué dans la partie Qu- appartenant à une autre proposition ou phrase, le manque d'annotation des relations syntaxiques rend la tâche de détection des clivées plus compliquée. Dans une tentative de résoudre ce problème, nous procédons par un post-traitement des données avec un tagger plus avancé, à savoir Stanza (Qi et al. 2020). Stanza donne à chaque mot d'une phrase :

- Un « id » indiquant sa position dans la phrase
- Sa forme
- Son lemme
- Sa catégorie grammaticale
- De l'information morphologique (pour un verbe : modalité, nombre, personne, temps, fini/non-fini)
- La position du gouverneur syntaxique du mot
- La relation de dépendance avec le gouverneur
- Sa position en termes de caractères

L'annotation de Stanza se fonde sur les dépendances universelles¹⁹ (désormais UD). L'UD est un projet international collaboratif qui vise à réduire la variation interlinguistique dans les annotations syntaxiques. Le schéma d'annotation de l'UD est basé sur plusieurs jeux d'étiquettes, notamment *Stanford dependencies* (de Marneffe et al. 2014) pour les dépendances syntaxiques, *Google universal* (Petrov et al. 2012) pour les étiquettes morphosyntaxiques et *Intersect Interlingua* (Zeman 2008) pour les étiquettes morphologiques. Dans le cadre du projet UD, des banques de corpus arborés ont été créés pour plus de 100 langues, dont le français.

En ce qui concerne les dépendances syntaxiques, l'UD propose 37 étiquettes de base (de Marneffe et al. 2021). De plus, pour certaines langues, des étiquettes spécifiques ont été rajoutées afin de mieux pouvoir traiter des particularités syntaxiques de la langue. Pour le français, la catégorie *advcl:cleft* a été rajoutée au 37 étiquettes de base afin de pouvoir distinguer les clivées des subordinées (De Marneffe et al. 2019). La relation *advcl:cleft* correspond à la relation entre la tête de l'élément clivé et le verbe recteur de la phrase. Les exemples ci-dessous, tirés du site d'UD, illustrent comment les clivées sont identifiées à l'aide de la relation *advcl:cleft* :

¹⁹ <https://universaldependencies.org/>



Figure 4 : Exemples de l'étiquette *advcl:cleft* du site de dépendances universelles françaises (<https://universaldependencies.org/fr/dep/advcl-cleft.html>)

Comme le montrent les exemples d'UD, la relation *advcl:cleft* est conçue pour identifier tous types de clivées, pas seulement les clivées prototypiques. Il est pourtant à noter que, dans les exemples annotés avec les dépendances universelles, l'occurrence *il y a longtemps que j'habite ici* a été analysée comme une clivée. Néanmoins, dans notre analyse, ce type de construction n'est pas inclus puisqu'il est considéré comme une locution.

Dans les exemples de la figure 4, *advcl:cleft* partage le même gouverneur que le constituant avec l'étiquette *nsubj:expl*. Nous remarquons pourtant que cette relation n'est pas présente dans nos données. À la place, il s'agit de la relation *expl:subj*, comme dans cet exemple tiré d'UD :

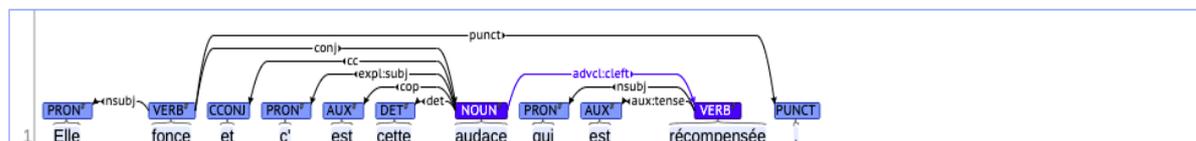


Figure 5 : Exemple de l'étiquette *expl:subj* du site de dépendances universelles française (https://universaldependencies.org/treebanks/fr_gsd/fr_gsd-dep-advcl-cleft.html)

L'étiquette *expl:subj* est utilisée pour les sujets explétifs, à savoir les sujets qui ne portent pas de rôle sémantique (de Marneff et al. 2021), comme c'est le cas pour le pronom en tête d'une clivée. Ainsi, une séquence de mots dont le premier mot porte le rôle de sujet explétif (*expl:subj*) et le dernier le rôle de clivée (*advcl:cleft*) constitue une construction clivée. Dans cette structure, l'élément clivé est la

séquence de mots qui commence après le mot avec l'étiquette *cop* et qui continue jusqu'à un *qui* ou un *que* gouverné par un constituant *advcl:cleft*.

Il est possible de constater que le post-traitement des données avec Stanza rend possible une autre façon d'extraire les constructions clivées. Cette dernière sera explicitée dans la partie suivante.

3.2.2 Extraction avec un script Python

Comme déjà évoqué dans la partie précédente, les annotations de Stanza permettent d'identifier les clivées à partir des relations syntaxiques entre les constituants. Dans le but d'améliorer la performance de la méthode pour extraire les clivées du corpus, un script Python a été développé, désormais appelé *Python1*.

Avant d'élaborer *Python1*, nous avons dû passer par plusieurs étapes. Afin de pouvoir traiter le corpus phrase par phrase, la première étape consistait à modifier les sorties de Stanza pour avoir le format CONLLU. À la différence de nos premières sorties Stanza, le format CONLLU sépare chaque phrase par une ligne vide, ce qui facilite le traitement au niveau de la phrase. Une fois les fichiers modifiés, nous avons procédé à la création d'un script Python permettant d'extraire toutes les phrases contenant la relation *advcl:cleft*. Cette étape a été importante pour pouvoir avoir une idée du rappel de l'annotation. Nous avons observé que les sorties de cette extraction ne contenaient aucune construction clivée de type non-prototypique (*pronom avoir X qui, il y a X qu-* ou *pseudo-clivée*). L'absence des clivées non-prototypiques dans ce premier script nous a amené à nous limiter aux constructions prototypiques dans la conception du *Python1*.

Python1 a une structure imbriquée. Dans un traitement phrase par phrase, nous extrayons celles qui correspondent aux conditions suivantes :

- La phrase a un élément *root* qui gouverne un constituant *advcl:cleft* et un constituant *expl:subj*.
- Le lemme *qui* ou *que* est gouverné par le constituant *advcl:cleft*.
- Le constituant *expl:subj* correspond au lemme *ce*.

3.2.3 Performance des scripts Python

Le script *Python1* a donné 85 occurrences en sortie, dont 24 étaient des constructions avec *qui* et 61 des constructions avec *que*. En comparant ces sorties avec notre *gold-standard* pour les clivées prototypiques, nous observons que, parmi les 24 occurrences de type *c'est X qui*, 14 occurrences constituaient de véritables clivées (de vrais positifs). Parmi les 61 occurrences de type *c'est X que*, 49 occurrences correspondaient à de vrais positifs. Le tableau ci-dessous permet de comparer la performance des dernières versions des patrons TXM avec le script *Python1* :

Méthode	Total	VP	Rappel	Précision	F-mesure
TXM <i>c'est X qui</i>	122	46	0.98	0.38	0.55
<i>Python1 qui</i>	24	14	0.30	0.58	0.40
TXM <i>c'est X que</i>	107	73	1.00	0.68	0.81
<i>Python1 que</i>	61	49	0.67	0.80	0.73

Tableau 10 F-mesures des derniers patrons TXM et du script python

Comme le montre le tableau 10, la précision du script Python est plus élevée que la précision des patrons TXM. Néanmoins, le rappel est plus élevé pour les patrons TXM que pour les scripts Python. Les f-mesures des scripts Python sont plus basses pour les clivées avec *qui* que les clivées avec *que*. Il est à noter que les rappels des scripts Python sont plus faibles que la précision dans les deux cas. Cela signifie que, même si l'extraction des clivées avec Python réussit mieux à distinguer les vrais positifs des faux

positifs que les patrons TXM, elle réussit moins bien à identifier tous les vrais positifs. Autrement dit, les sorties des scripts Python contiennent plus de silence.

3.2.4 Analyse du bruit et du silence dans le script Python

Afin de mieux comprendre la performance faible du script Python, nous avons observé le bruit et le silence plus en détail. Le bruit des sorties du *Python1* peuvent être classé en trois types. Le premier type correspond à des constructions homonymes. Plus précisément, il s'agit exclusivement des propositions relatives sans fonction prédicative :

- 1) Tina avait vraiment peur de rentrer chez elle , car **c'était** la seule **qui** rentrait toute seule à pied (CO-3e-2018-FSBJC6-D1-R14-V1_N)
- 2) C'est l'histoire d'un garçon qui s'appelait Finn le petit humain, **c'était** un grand aventurier **qui** battait n'importe quel monstre grâce à sa Vitesse (CO-6e-2016-VTAC603-D1-R3-V1_N)
- 3) S'il y a bien une chose de plus dans ce lycée **ce sont** les cours de développement de magie **qui** nous apprennent à mieux la maîtriser. (CO-3e-2018-FSBJC6-D1-R9-V1_N)
- 4) Quand tout à coup quelqu'un sonna à la porte. **C'était** la voisine **qui** habitait la maison à le fond de la rue (CO-3e-2016-VTAC305-D1-R25-V1_N)

Ce type de bruit montre que les relations syntaxiques d'UD ne suffisent pas pour distinguer les clivées des constructions homonymes. Dans les occurrences ci-dessus, nous observons deux problèmes liés à l'annotation faite par Stanza. Premièrement, le verbe qui est annoté comme la réaction d'une clivée fait en effet partie d'une subordonnée ordinaire. Deuxièmement, le pronom *ce* n'est pas sémantiquement vide comme l'implique l'étiquette *expl:subj*. Dans les exemples ci-dessus, *ce* réfère à un élément du contexte antérieur.

Le deuxième type de bruit identifié correspond à des constructions interrogatives partielles ou exclamatives avec *qu'est-ce que* ou *qu'est-ce qui* :

- 5) Qu'est-ce qu'il y a ! (EC-CE2-2018-TFPPX-D1-R12-V1_N)
- 6) Qu'est-ce qu'il y a ! dit Paul à Tom (EC-CE2-2018-TFPPX-D1-R12-V1_N)
- 7) Mais qu'est-ce que tu racontes, chérie ? (UN-M2-2020-UCL-D1-R48-V1_N)
- 8) Qu'est-ce que vous me voulez ! (CO-6e-2016-PJPR1-D1-R18-V1_N)
- 9) Qu'est-ce que l'on attend ? (CO-6e-2018-VTAC605-D1-R16-V1_N)
- 10) Qu'est-ce qu'elle fut heureuse ! (UN-M2-2021-UCL-D1-R50-V1_N)
- 11) Qu'est-ce qui s'est passé ? (EC-CM1-2015-TFGLX-D1-R11-V1_N)
- 12) Qu'est-ce que ces jumeaux étaient intrépides. (UN-M2-2020-UCL-D1-R90-V1_N)
- 13) Qu'est-ce qui avait bien pu provoquer un tel boucan ? (UN-M2-2020-UCL-D1-R64-V1_N)
- 14) Mais qu'est-ce qu'elle était grande ! (EC-CM2-2016-MQRVX-D1-R13-V1_N)
- 15) Qu'est-ce que tu fais là ? (CO-6e-2018-VTAC605-D1-R9-V1_N)

Même si la présence des phrases interrogatives parmi les clivées peut sembler étonnant, elle n'est pas involontaire de la part des auteurs du schéma UD du français (de Marneffe et al. 2019). Selon ces derniers, les interrogatives partielles peuvent être interprétées comme des clivées puisqu'elles peuvent être paraphrasées (16)). De plus, comme le constate Obenauer (1981), la question peut être formulée sans mise en relief (17)).

- 16) **C'est** quoi **que** tu me racontes ?
- 17) Que me racontes-tu ?

À notre connaissance, les exclamatives n'ont pas été interprétées comme des clivées. Cependant, en suivant la même logique que pour les interrogatives, nous prétendons qu'elles peuvent être exprimées sans mise en relief :

18) Mais qu'elle était grande !

Étant donné que les interrogatives et les exclamatives avec *qu'est-ce qu-* sont rarement prises en compte dans les études sur les clivées, nous les laissons de côté dans cette étude. Nous retiendrons toutefois l'intérêt de ces constructions pour des études ultérieures.

Le troisième type de bruit correspond en effet à des véritables constructions clivées qui n'ont pas été identifiées dans notre *gold-standard* puisqu'elles contiennent une négation.

19) **Ce n'est** qu'à le petit matin **qu'**il eut le courage de raconter à tout le village ce qu'il avait vu (UN-M2-2020-UCL-D1-R108-V1_N)

20) **Ce n'est** que vers 1h de le matin **que** des lumières bleues apparurent sur les murs de la gare. UN-M2-2020-UCL-D1-R6-V1_N

Ce type de « bruit »²⁰ ne constitue que 2 occurrences des sorties du *Python1*. Néanmoins, afin de vérifier que Python a bien identifié toutes les clivées prototypiques avec négation, nous avons conçu un patron TXM supplémentaire²¹. Les résultats de notre patron TXM supplémentaire montrent que le script *Python1* a pu identifier 2 sur 7 clivées de ce type. Nous estimons que l'absence de ces 7 occurrences n'affecteront pas considérablement notre analyse statistique. Pour cette raison, nous ne les incluons pas dans cette étude. Bien que ces occurrences ne soient pas incluses dans l'analyse, elles prouvent qu'il y a du potentiel d'améliorer le *gold-standard* en incluant les clivées prototypiques avec négation.

En ce qui concerne le rappel du *Python1*, nous avons observé un échantillon de quelques faux négatifs des deux types de constructions (*c'est X qui* et *c'est X que*). L'analyse du silence montre que le verbe du relatif est interprété soit comme la tête d'une proposition subordonnée relative, soit la tête d'une subordonnée complétive (d'un verbe ou d'un adjectif)²². En employant la nomenclature d'UD, il s'agit de la relation *acl:relcl* et *ccomp*. Voici deux exemples d'annotation tirés du corpus (Figure 6 et 7), suivis de deux exemples tirés du site d'UD (Figure 8 et 9):

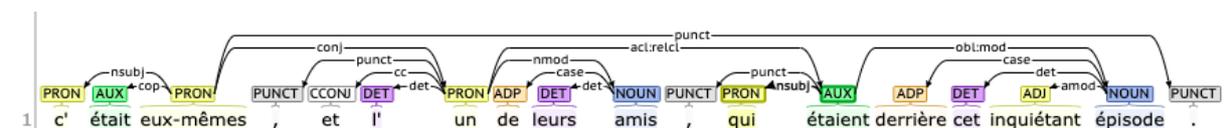


Figure 6 : Annotation par Stanza d'une phrase tirée de UN-M2-2018-TUTJ2-D1-R1-V1_N

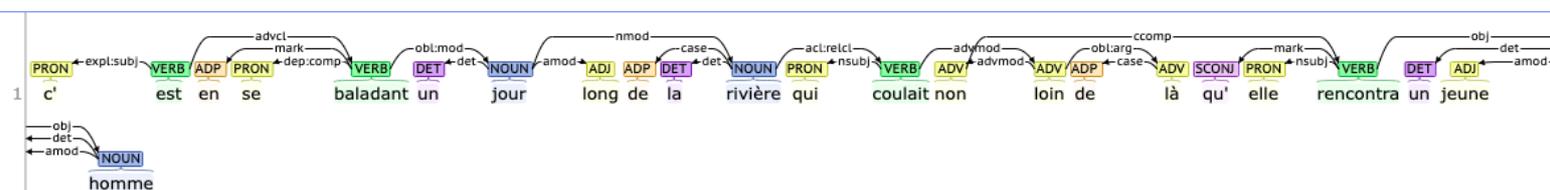


Figure 7 : Annotation par Stanza d'une phrase tirée de UN-M2-2020-UCL-D1-R56-V1_N

²⁰ Nous continuons d'utiliser le terme de bruit dans ce passage même s'il s'agit des véritables clivées, et cela dû à leur absence dans notre gold-standard.

²¹ Le patron TXM supplémentaire est le suivant : [frlemma="ce"][word="n"][frlemma="être" & frpos="VER.*"][frpos!="SENT"] {0,15} [frlemma="*" & ! frpos="KON" & ! frpos="SENT" & ! frlemma="parce|pendant|plus|plutôt|ce"][frlemma="que|qui"]

²² De tels compléments figurent par exemple dans le discours rapporté.

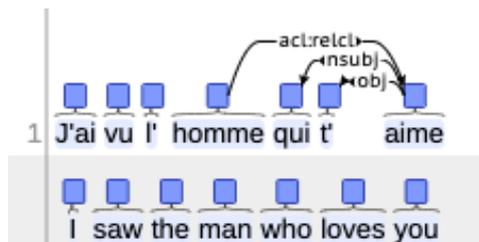


Figure 8 : Exemple de l'étiquette `acl:relcl` du site de dépendances universelles française (<https://universaldependencies.org/fr/dep/acl-relcl.html>)

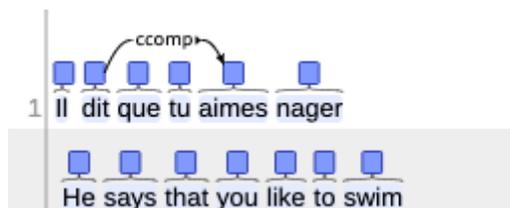


Figure 9 : Exemple de l'étiquette `ccomp` du site de dépendances universelles française (<https://universaldependencies.org/fr/dep/ccomp.html>)

Dans la figure 6, le verbe *étaient* est analysé comme la tête d'une subordonnée relative gouvernée par un groupe nominal. Notons que le pronom *ce* n'occupe pas le rôle du sujet explétif dans l'exemple, mais du sujet ordinaire. Dans la figure 7, le verbe *rencontra* est analysé comme la tête d'une subordonnée complétive.

Par ailleurs, il est à noter que l'annotation faite par Stanza ne gère pas mieux que les patrons TXM les clivées dont l'élément clivé contient une subordonnée (cf. Cas ambigus). Dans la figure 10, une relation clivée est identifiée entre *qu'il est mort* (gouverneur : *mort*) et *depuis*, tandis que notre analyse manuelle identifie la relation clivée entre *qu'il se passe des choses bizarres* (gouverneur : *se passe*) et *depuis qu'il est mort* (gouverneur : *depuis*). Nous présumons que l'interprétation erronée de Stanza peut être expliquée par le manque d'une analyse de la locution *depuis que*.

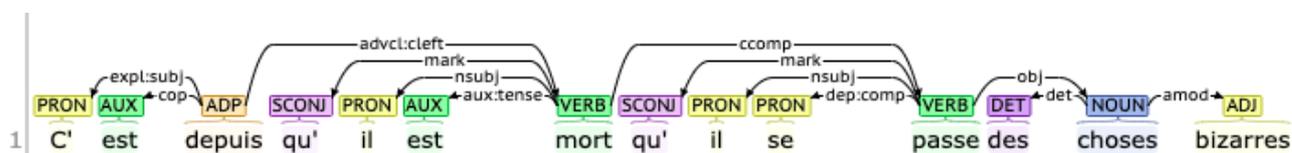


Figure 10 : Annotation par Stanza d'une phrase tirée de CO-4e-2018-LSPJJRC-D1-R11-V1_N

Dans la figure 11, l'analyse devient complexe due au fait que la clivée est imbriquée dans une subordonnée complétive gouvernée par *je sais*. En effet, le verbe *inventé* pourrait être considéré aussi bien comme le gouverneur d'une subordonnée complétive à *je sais*, que le gouverneur d'une relation clivée avec *juste pour que je t'obéisse*.

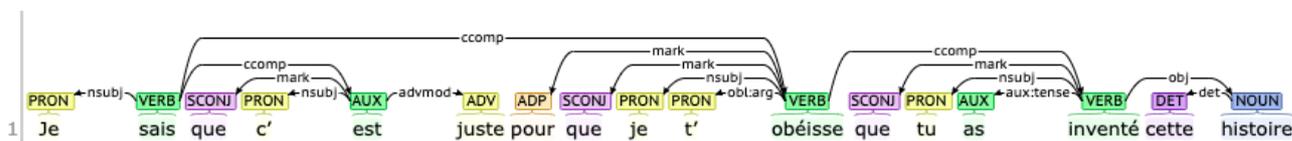


Figure 11 : Annotation par Stanza d'une phrase tirée de UN-M2-2020-UCL-D1-R80-V1

Bien évidemment, le silence du *Python1* peut également s'expliquer par la nature non-standard de nos données. Comme nous l'avons soulevé dans la partie 2.2.2 Normalisation, les textes n'ont pas été normalisés au niveau de la ponctuation. Des phrases longues et chevauchées peuvent effectivement perturber l'analyse syntaxique faite par les outils. Quoiqu'il en soit, il est possible de constater que les performances relativement faibles des méthodes proposées dans ce travail démontrent la difficulté d'identifier les constructions clivées automatiquement. Plus généralement, la démarche que nous présentons ici nous semble bien illustrer l'enjeu que représente le traitement de données non-standard.

PARTIE 4 – Analyse descriptive des clivées

L'analyse qui suivra dans cette partie est fondée sur le *gold-standard* établi dans la partie précédente, reprise ci-dessous :

Type de clivée	VP + FN
<i>C'est X que</i>	73
<i>C'est X qui</i>	47
<i>Il y a X que</i>	1
<i>Il y a X qui</i>	21
<i>Pronom avoir X qui</i>	2
<i>Pseudo-clivées</i>	17
TOTAL	161

Tableau 6 Quantification du gold-standard

La première sous-partie sera consacrée à une description globale des clivées, à savoir leur répartition dans le corpus étudié. Dans la deuxième sous-partie, l'usage des clivées sera décrit à l'aide des résultats de l'annotation manuelle. Nous nous concentrerons sur trois caractéristiques de l'élément clivé : sa fonction syntaxique, le type de constituant syntaxique faisant l'objet du clivage et la longueur de la clivée (en nombre de mots-formes). La dernière sous-partie dans cette section sera consacrée à une analyse qualitative de quelques clivées fréquentes, traitées dans une étude de Roubaud et Sabio (2018).

4.1 Description globale

Dans le corpus, composé de 613 textes, il y a au total 161 clivées réparties dans 124 textes. Par rapport au nombre de mots-formes dans le corpus, cela correspond à une fréquence relative d'environ 0.09%. Parmi les 124 textes contenant des clivées, 21 des textes proviennent de l'école primaire, 35 proviennent du collège et 68 à l'université. Le tableau 11 ci-dessous donne un aperçu de l'ensemble des clivées identifiées dans le corpus, à savoir notre *gold-standard* (cf le tableau 6).

Niveau Scolaire	Classe	Nb de textes	Nb de mots-formes	Nb de clivées	Nb de textes avec clivées	Fréquence relative (nb de clivées/nb de mots-formes)
École primaire	CE2	55	6495	8	7	0,12%
	CM1	53	6467	4	4	0,06%
	CM2	59	11275	14	10	0,12%
Collège	6e	114	21868	18	14	0,08%
	4e	47	13576	10	8	0,07%
	3e	54	17808	18	13	0,10%
Université	M2	231	102577	89	68	0,09%
TOTAL		613	180066	161	124	0,09%

Tableau 11 Description de la distribution des clivées

Comme le montre le diagramme ci-dessous (cf. la figure 12), aucun texte ne contient plus de 4 clivées, et la majorité des textes qui contiennent des clivées en contiennent une seule. Néanmoins, le plus souvent, le texte ne contient pas de clivée.

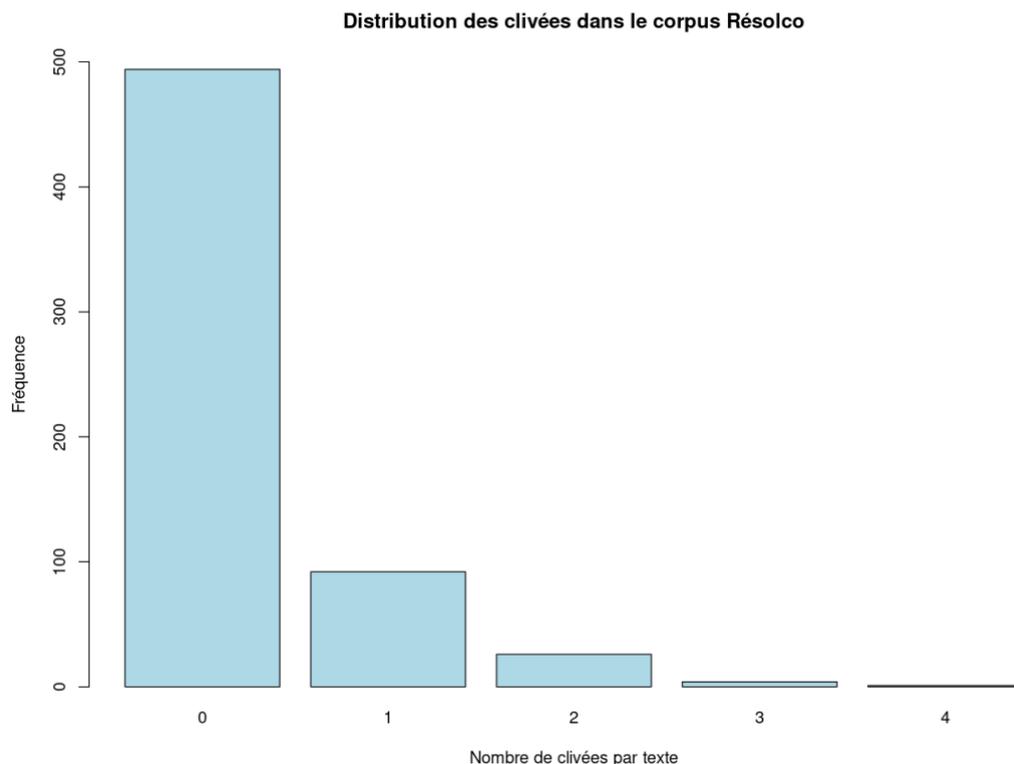


Figure 12 : Distribution de nombre de clivées par texte

Parmi les 161 clivées, une grande majorité sont des clivées prototypiques (75%). En revanche, les clivées de type *il y avoir X que* et *pronom avoir X qui* sont extrêmement rares dans le corpus, avec respectivement 1 et 2 occurrences. La distribution des différents types de clivées est illustrée dans le diagramme de Pareto ci-dessous (cf. la figure 13). L'axe y à gauche indiquent la fréquence absolue tandis que l'axe y à droite indiquent la fréquence relative des constructions. La courbe liée par des points indique l'accumulation de la fréquence relative des constructions. Il est à noter que l'accumulation des fréquences relatives des constructions les plus fréquemment employées, à savoir *c'est X que* et *c'est X qui*, constitue 75% des clivées. Les clivées prototypiques sont donc largement majoritaires.

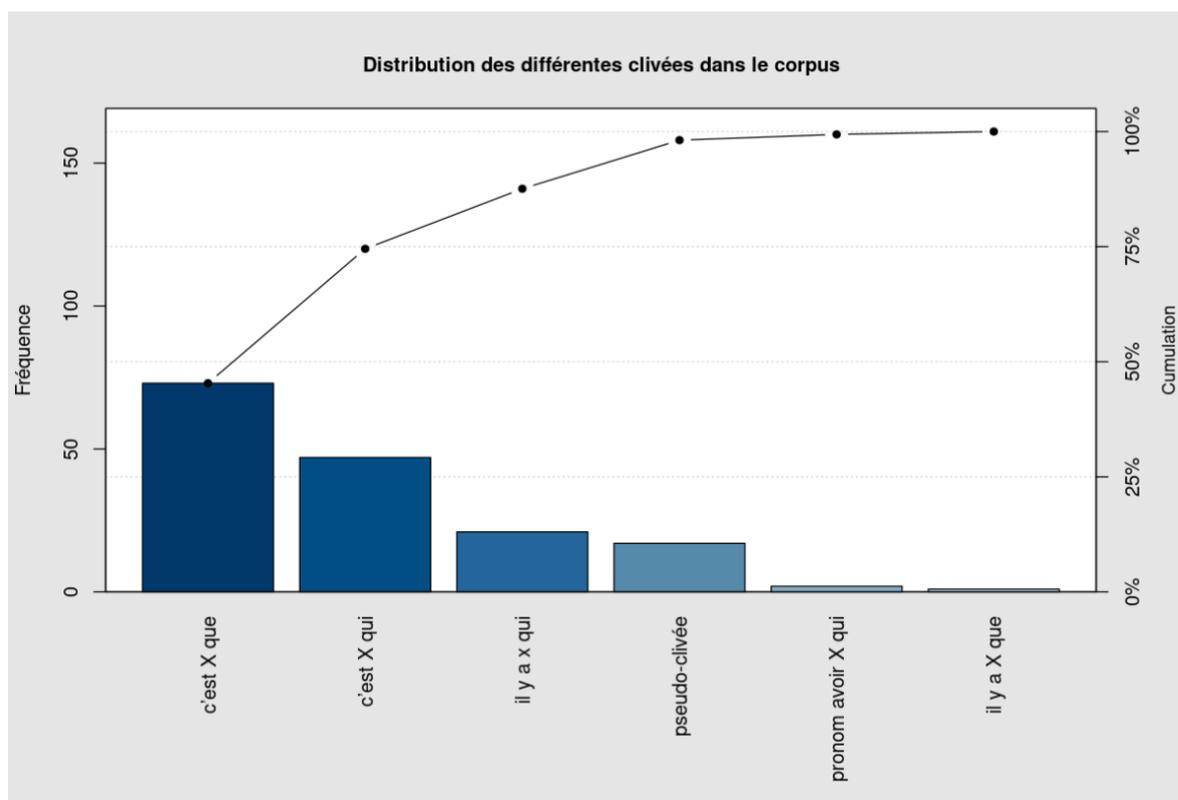


Figure 13 : Distribution des différents types de clivées

Une étude plus détaillée des occurrences des clivées a permis d'identifier un cas où deux types de clivées se chevauchent. En effet, dans 1), l'élément clivé de la pseudo-clivée est lui-même une clivée de type *il y a X qui*. Ce cas particulier est compté dans notre étude comme deux occurrences de clivées, une occurrence d'une pseudo-clivée et une occurrence d'une clivée de type *il y a X qui*.

- 1) Ce qu'ils ignoraient c'est qu'il y avait une vieille sorcière qui y habitait (EC-CM2-2016-SGLEA-D1-R22-V1_N)

4.2 Résultats de l'annotation

Dans cette partie, nous décrivons les constructions clivées identifiées dans l'étape de l'annotation manuelle des données. Dans ce qui suit, la fonction syntaxique de l'élément clivé sera analysée, ainsi que le type de constituant faisant l'objet du clivage et la longueur de la clivée. Ensuite, afin d'approfondir notre analyse, quelques constructions clivées récurrentes, antérieurement étudiées par Roubaud et Sabio (2018), seront observées plus précisément.

4.2.1 L'élément clivé : fonction syntaxique

Nous rappelons que l'élément clivé peut avoir des fonctions syntaxiques variées : sujet, complément direct, complément indirect ou circonstant. Les résultats de notre annotation montrent que l'élément clivé est le plus souvent un sujet ou un circonstant. Ainsi, il est possible de constater que nos résultats sont conformes aux études antérieures (cf. 1.3.2 L'élément clivé). La figure 14 ci-dessous illustre la distribution des fonctions syntaxiques de l'élément clivé dans le corpus. Comme dans la figure 13 il s'agit d'un diagramme de Pareto dont l'axe y à gauche indique la fréquence absolue, et l'axe y à droite indique le pourcentage du total. Le diagramme 14 montre que les sujets constituent en peu moins de 50% des clivées analysées, tandis que les circonstants constituent environ 40%. Les compléments indirects ne constituent que 2 occurrences.

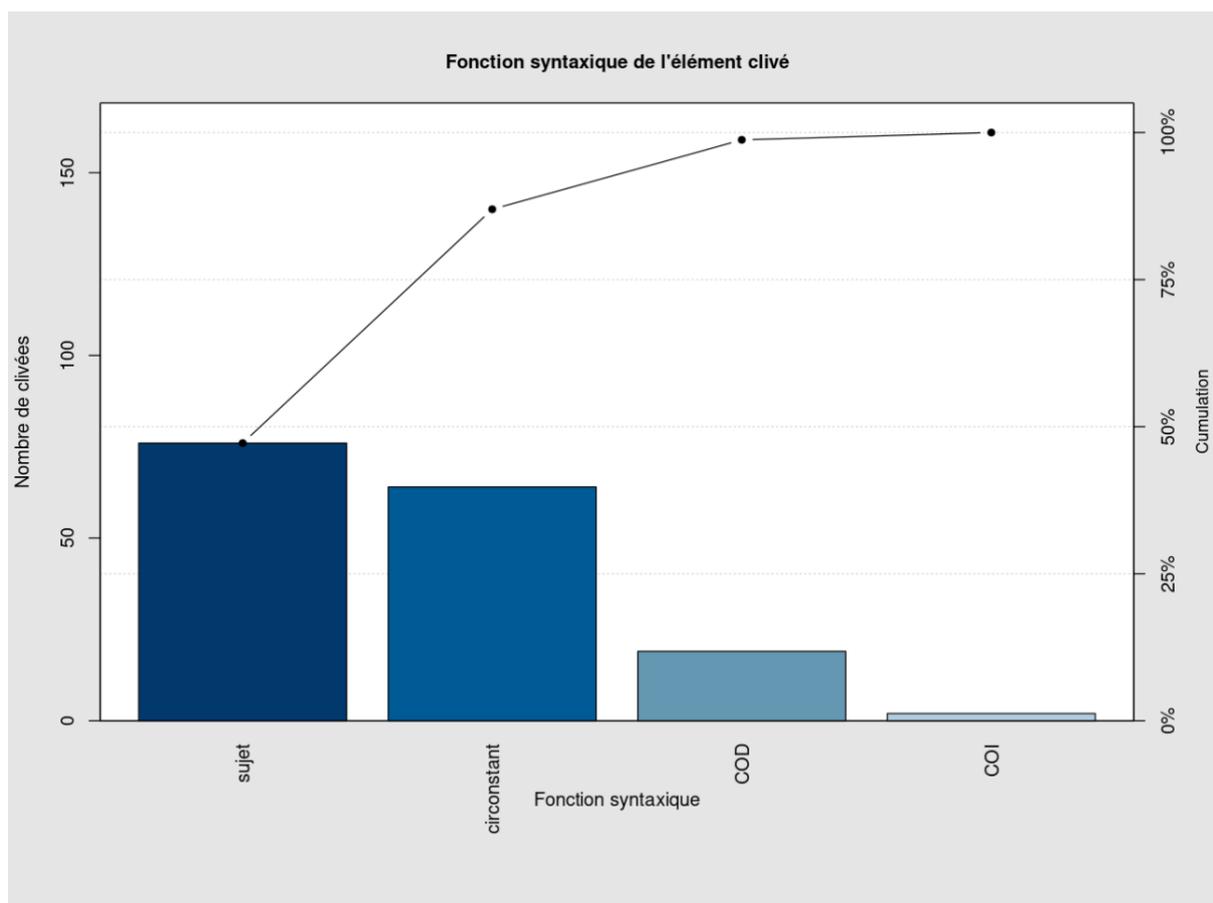


Figure 14 : Distribution des fonctions syntaxiques de l'élément clivé

Bien évidemment, les fonctions syntaxiques de l'élément clivé sont liées au type de construction, les constructions avec *qui* impliquant la présence d'un sujet. Nous donnons ci-dessous un exemple de chaque type de construction présent dans notre corpus pour chaque fonction syntaxique :

Sujet

- 2) J'ai **tous mes amis** qui sont ici [...] (EC-CM2-2016-SGLEB-D1-R12-V1_N)
- 3) Il y a **encore le monstre** qui rode. (CO-6e-2017-VTAC602-D1-R21-V1_N)
- 4) Mais ce qui était inédit, c'est **qu'elle les gardait chez elle** [...] (UN-M2-2021-UCL-D1-R19-V1_N)
- 5) À sa mort, c'est **son fils** qui hérita du manoir. (EC-CM1-2017-TBZX-D1-R5-V1_N)

Circonstant

- 6) C'est **comme ça** que le mois dernier j'ai fini par atterrir dans un quartier tranquille de la banlieue d'Avalon avec ma mère et mon petit frère. (UN-M2-2021-UCL-D1-R25-V1)

COD

- 7) Ce qu'elle ignorait, c'est **que son petit frère, Gilles, l'avait entendue ouvrir la fenêtre** (UN-M2-2021-UCL-D1-R75-V1_N)
- 8) C'est **comme ça** qu'ils avaient pris l'habitude de nommer ce futur petit appartement [...] (UN-M2-2020-UCL-D1-R71-V1_N)
- 9) Et, il n'y avait **pas que la maison** qu'elle avait fait sienne (UN-M2-2020-UCL-D1-R81-V1_N)

COI

- 10) C'est **à lui** que l'aventure arriva (mauvaise aventure). (CO-6e-2016-PJPR1-D1-R21-V1_N)

Comme l'ont déjà évoqué Roubaud et Sabio (2018, 5), les clivées avec *comme ça* peuvent porter soit le rôle de circonstant, soit le rôle d'un complément d'objet direct (voir 6) et 8)).

4.2.2 L'élément clivé : constituant syntaxique

En observant la courbe du diagramme de pareto ci-dessous (cf. la figure 15), on constate que les syntagmes nominaux, accompagnés par les adverbes et les syntagmes prépositionnels, constituent plus de 75% des clivées identifiées. Le diagramme montre également que les constituants les plus rares des clivées sont les verbes à l'infinitif, les noms propres et les gérondifs. La fréquence élevée des syntagmes nominaux peut être expliquée par le fait que l'élément clivé est le plus souvent sujet (cf. la figure 14).

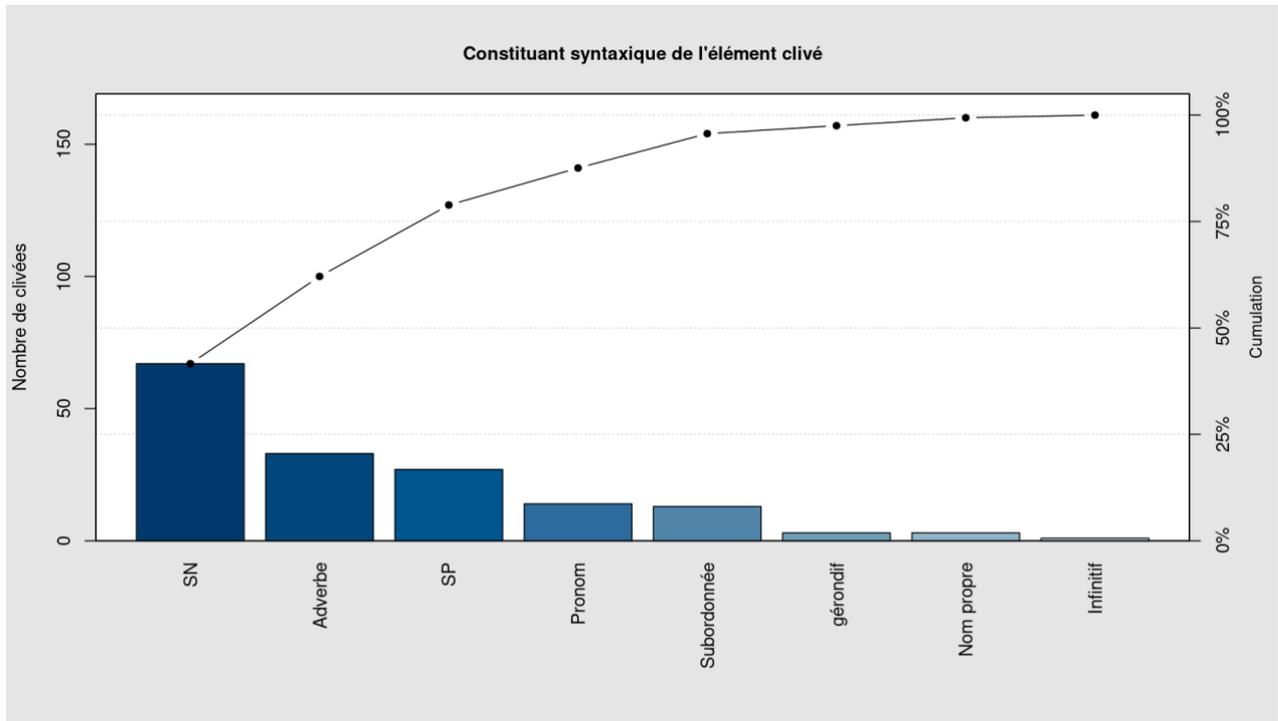


Figure 15 : Distribution des constituants syntaxiques de l'élément clivé

Afin d'avoir une meilleure compréhension de la distribution des constituants syntaxiques, nous rappelons que le type de clivée détermine les constituants syntaxiques qui peuvent se trouver dans l'élément clivé :

C'est X qu-	Il y a X qu-	Pronom avoir X qui	Pseudo-clivée
<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué • Syntagme adverbial • Syntagme adjectival • Syntagme prépositionnel • Subordonnée • Verbe à l'infinitif • Gérondif 	<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué • Syntagme prépositionnel (avec négation) 	<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué 	<ul style="list-style-type: none"> • Syntagme nominal • Nom propre • Pronom accentué • Syntagme prépositionnel • Subordonnée

Tableau 9 Constituant syntaxique de l'élément clivé

Comme le montre la figure 16 ci-dessous, les syntagmes nominaux sont présents dans tous les types de clivées. Nous notons que les clivées avec *avoir* ne sont employées qu'avec des syntagmes nominaux autres que des pronoms accentués et des noms propres. Ces derniers figurent uniquement dans les clivées prototypiques. L'analyse montre que l'élément clivé dans les clivées prototypiques est constitué de presque tous les constituants syntaxiques qui figurent dans la liste (cf. le tableau 9). Les seuls constituants syntaxiques qui ne sont pas présents sont les adjectifs et les infinitifs. Les constituants

syntaxiques qui ont la fréquence la plus élevée dans les clivées prototypiques sont les adverbes 11) et les syntagmes prépositionnels 12). Notons que le syntagme prépositionnel dans 12) est une locution adverbiale qui porte le même rôle sémantique que l’adverbe dans 11).

- 11) il se retourna en entendant ce grand bruit. Et c’est là qu’il a vu une personne ! (EC-CE2-2016-SGLEB-D1-R1-V1_N)
 12) C’est à ce moment-là qu’il tombe nez à nez avec la meute de loups qui le fixait de le regard (CO-3e-2018-FSBJC6-D1-R6-V1)

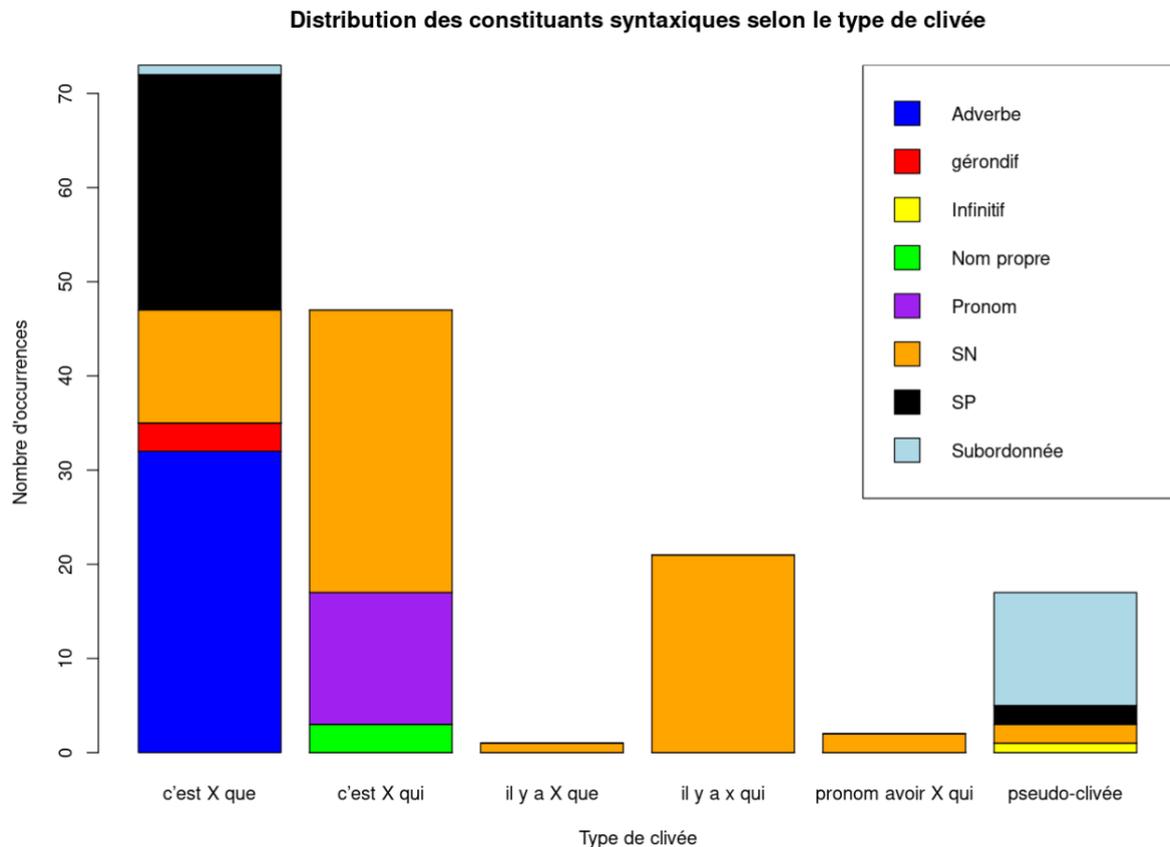


Figure 16 Distribution des constituants syntaxiques selon le type de clivée

Il est à noter que toutes les clivées dont l’élément clivé est un syntagme adverbial correspondent aux constructions clivées cohésives que nous allons traiter en détail sous

4.2.4 Comparaison avec l’étude de Roubaud et Sabio (2018)(*c’est comme ça/ainsi/alors/là que*).

En ce qui concerne la distribution des constituants syntaxiques selon la classe scolaire, l’élément clivé recouvre une plus grande diversité de constituants dans les textes rédigés par les étudiants de master. En effet, les pronoms accentués, l’infinitif et les gérondifs figurent uniquement dans les textes rédigés par les locuteurs à l’université, comme le montre la figure 17.

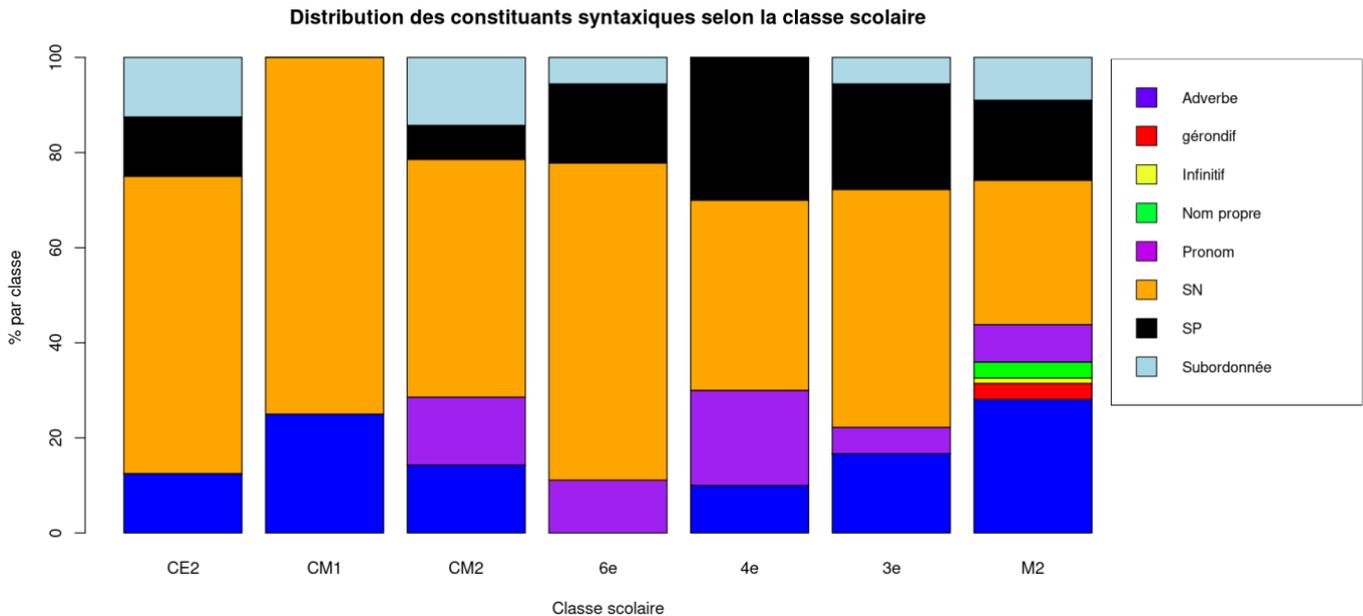


Figure 17 Distribution des constituants syntaxiques selon la classe scolaire

4.2.3 L'élément clivé : nombre de mots

Dans l'extraction des clivées (cf. 3.1), nous avons souligné le fait que l'élément clivé contient rarement plus de 7 mots, l'exception toutefois faite des pseudo-clivées, qui, dans notre étude, contiennent plus de 13 mots dans 50% des cas. Le tableau ci-dessous rassemble les valeurs qui nous permettent d'interpréter la distribution de la variable de la longueur de l'élément clivé. Le tableau montre que les longueurs des clivées sont plus ou moins regroupées autour de la moyenne, sauf les pseudo-clivées qui ont une longueur plus variable.

Type de clivée	Min	Max	Q1	Médiane	Q3	Moyenne	Écart-type
<i>C'est X que</i>	1	16	1	2	3	2.64	2.43
<i>C'est X qui</i>	1	15	1	2	4.5	3.47	3.26
<i>Il y a X que</i>	5	5	5	5	5	5	-
<i>Il y a X qui</i>	2	11	2	2	3	2.90	1.95
<i>Pronom avoir X qui</i>	3	3	3	3	3	3	0
<i>Pseudo-clivée</i>	2	27	8	13	16	12.53	6.96

Tableau 12 Description quantitative des différents types de clivées (valeur minimale, maximale, moyenne et écart-type)

Parmi les clivées les plus longues, nous trouvons des gérondifs 13) ou des SN complexes 14)-15). Dans le cas des pseudo-clivées, les éléments les plus longs sont les subordonnées, comme on le voit dans l'exemple 16).

- 13) C'est **en se promenant un jour le long de la rivière qui coulait non loin de là** qu'elle rencontra un jeune homme. (UN-M2-2020-UCL-D1-R56-V1_N)
- 14) C'est **une mère au foyer se nommant Stephanie et ses 2 enfants ; Kyllian et Ilan** qui vécurent cette expérience. (CO-4e-2018-LSPJRD-D1-R15-V1_N)
- 15) mais, il y avait **une personne au comptoir (qui nous a finalement trouvés)** qui nous expliqua que tous les clients rentrent avant la tombée de la nuit par peur de voir des démons en particulier le démon E N D qui aurait disparu il y a 15 ans de cela. (CO-3e-2018-FSBJC6-D1-R9-V1_N)
- 16) Pourtant, ce qu'ils ne savent pas, c'est **que Wilfried le bûcheron leur avait sauvé la vie cette nuit-là en tuant, avec sa hache, le crocodile qui les prenait pour son prochain repas...** (UN-M2-2020-UCL-D1-R58-V1)

En examinant la distribution des longueurs des clivées selon la classe scolaire dans le tableau 13, nous observons que, même si les éléments clivés de plus de 10 mots-formes restent rares, de telles clivées sont inexistantes à l'école primaire. Nous notons également que toutes les clivées de plus de 15 mots sont présentes dans les textes rédigés par des étudiants de master.

Classe scolaire	Nb de clivées					
	1-5 mots-formes	6-10 mots-formes	11-15 mots formes	16-20 mots-formes	21-25 mots-formes	25-30 mots-formes
CE2	7	1	0	0	0	0
CM1	4	0	0	0	0	0
CM2	12	2	0	0	0	0
6e	16	2	0	0	0	0
4e	8	1	1	0	0	0
3e	16	1	1	0	0	0
M2	65	13	3	4	3	1
TOTAL	128	20	5	4	3	1

Tableau 13 Fréquence absolue des clivées selon la longueur de l'élément clivés

4.2.4 Comparaison avec l'étude de Roubaud et Sabio (2018)

Comme nous l'avons évoqué dans l'état de l'art (1.3.6 Le statut des clivées à l'oral et à l'écrit), Roubaud et Sabio (2018) ont publié une étude où certaines clivées fréquentes ont été quantifiées dans un corpus oral et un corpus écrit. Les auteurs ont observé que les constructions *c'est comme ça que*, *c'est pour ça que* et *c'est là que* avaient une fréquence élevée à l'oral par rapport à l'écrit. En revanche, les constructions *c'est alors que*, *c'est ainsi que* et *c'est comme cela que* étaient plus fréquentes à l'écrit. Pour cette raison, nous avons observé si et comment ces constructions sont employées dans les textes du corpus. Pour faciliter la lecture et pour mieux distinguer les deux catégories des clivées, la première catégorie de clivées sera nommée *clivée de l'immédiat* et la deuxième *clivée de la distance*.

Toutes les constructions analysées par Roubaud et Sabio (2018) figurent bel et bien dans nos données. Dans ce qui suit, nous allons commencer par présenter la distribution des clivées de l'immédiat pour ensuite présenter la distribution des clivées de la distance.

Par rapport aux 161 clivées dans le corpus, les clivées de l'immédiat constituent environ 12% des clivées (11,8%) correspondant à un total de 19 occurrences. Le tableau 14 ci-dessous montre la fréquence absolue des clivées de l'immédiat par classe scolaire. Dans la mesure où le nombre de clivées par classe est très variable, la fréquence brute est accompagnée par la fréquence relative selon le nombre de clivées par classe. Le tableau 14 montre que la fréquence relative des clivées de l'immédiat est la plus élevée dans les textes rédigés par les élèves en 3^e au collège. En revanche, les textes rédigés par les élèves en CM1 et en 4^e ne contiennent aucune clivée de ce type. La clivée *c'est là que* est la plus fréquemment employée (5,6%), or la clivée *c'est pour ça que* a une distribution plus équilibrée dans les classes : cette clivée est employée au moins une fois dans 5 classes sur 7 tandis que la construction *c'est là que* ne figure que dans 3 classes sur 7. La clivée *c'est comme ça que* est employée le plus rarement. De plus, on note qu'elle ne s'emploie que par les étudiants de master.

Classe	Nb de clivées	Clivées de l'immédiat			TOTAL /CLASSE
		<i>C'est pour ça que</i> Nb (%)	<i>C'est comme ça que</i> Nb (%)	<i>C'est là que</i> Nb (%)	
CE2	8	1 (12.5%)	0 (0.0%)	1 (12.5%)	2 (26%)
CM1	4	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
CM2	14	1 (7.1%)	0 (0.0%)	0 (0.0%)	1 (7.1%)
6e	18	1 (5.6%)	0 (0.0%)	0 (0.0%)	1 (5.6%)
4e	10	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
3e	18	3 (16.7%)	0 (0.0%)	2 (11.1%)	5 (27.8%)
M2	89	1 (1.1%)	3 (3.4%)	6 (6.7%)	10 (11.2%)
TOTAL	161	7 (4.3%)	3 (1.9%)	9 (5.6%)	19 (11.8%)

Tableau 14 Distribution des clivées de l'immédiat par classe scolaire

Le tableau 15 ci-dessous montre la distribution des clivées de la distance. Ces dernières constituent 12,4% du total des clivées, ce qui correspond à 20 occurrences. La clivée de type *c'est alors que* est la plus fréquente (8,1%), et elle est la seule parmi les clivées de la distance à figurer dans plusieurs classes scolaires, à savoir dans 4 sur 7 classes (CM1, CM2, 3^e et M2). Il est possible de noter que les locuteurs à l'université sont les seuls à employer toutes les trois constructions. Si l'on compare les fréquences relatives des différentes classes, on note que les clivées de la distance sont les plus fréquentes en CM1. Ceci dit, il faut toutefois noter que les textes rédigés par les élèves en CM1 ne contiennent que 4 clivées au total.

Clivées de la distance								
Classe	Nb de clivées	<i>C'est alors que</i> Nb (%)		<i>C'est ainsi que</i> Nb (%)		<i>C'est comme cela que</i> Nb (%)		TOTAL /CLASSE
CE2	8	0	(0.0%)	0	(0.0%)	0	(0.0%)	0 (0.0%)
CM1	4	1	(25.0%)	0	(0.0%)	0	(0.0%)	1 (25.0%)
CM2	14	2	(14.3%)	0	(0.0%)	0	(0.0%)	2 (14.3%)
6 ^e	18	0	(0.0%)	0	(0.0%)	0	(0.0%)	0 (0.0%)
4 ^e	10	0	(0.0%)	0	(0.0%)	0	(0.0%)	0 (0.0%)
3 ^e	18	2	(11.1%)	0	(0.0%)	0	(0.0%)	2 (11.1%)
M2	89	8	(9.0%)	6	(6.7%)	1	(1.1%)	15 (16.9%)
TOTAL	161	13	(8.1%)	6	(3.7%)	1	(0.6%)	20 (12.4%)

Tableau 15 Distribution des clivées de la distance par classe scolaire

Revenons maintenant aux quatre occurrences de la clivée *c'est comme ça/cela que*. En effet, cette construction est traitée plus en détail par Roubaud et Sabio (2018). Les auteurs distinguent deux emplois de *c'est comme ça que* : la clivée contrastive et la fonction discursive de connecteur. Ce dernier usage, qui n'est pas classé comme une véritable clivée par les auteurs, peut avoir plusieurs fonctions : introduction d'une conséquence, introduction d'un trait saillant ou introduction d'une illustration. Selon Roubaud et Sabio (2018), la fonction discursive la plus fréquente est celle qui sert à introduire une conséquence des faits antérieurement mentionnés. Selon les auteurs, cette dernière fonction « se rencontre essentiellement dans les productions narratives » (Roubaud et Sabio 2018, 12).

Dans notre annotation, la distinction entre la clivée et la fonction discursive de *c'est comme ça que* n'a pas été faite préalablement, et cela puisque nous considérons que la fonction discursive de connecteur peut être interprétée comme une clivée cohésive, au même titre que Karssenberget Lamiroy (2017) identifie la clivée suivante comme une clivée cohésive :

- 17) En quarante-deux, les bombardements! Et **c'est là que** beaucoup de gens sont partis de Toulon. (Karssenberget Lamiroy 2017)

Autrement dit, nous considérons que la structure informationnelle de la fonction de connecteur correspond aux clivées dont la deuxième partie contient de l'information nouvelle tandis que la première partie renvoie au contexte antérieur (cf. 1.3.5 Analyse pragmatique et informationnelle).

Afin d'approfondir notre analyse, nous faisons une analyse qualitative des clivées de type *c'est comme ça que*. Nous incluons, comme le font Roubaud et Sabio (2018), la construction *c'est comme cela que*. Selon Roubaud et Sabio (2018, 15), la clivée contrastive peut être réécrite avec la locution *de cette manière-là* tandis que la fonction discursive de connecteur peut être réécrite avec un des équivalents sémantiques suivants : *en conséquence*, *si bien que*, *résultat :*, *ce qui fait que* (Roubaud et Sabio 2018, 12). Les autrices définissent également quatre propriétés qui permettent de distinguer les véritables clivées de la fonction discursive de connecteur :

Clivée contrastive

- Comme ça* peut être remplacé par un autre élément lexical
- Comme ça* peut être contrasté avec un autre élément du même type
- Comme ça* peut être modifié par un adverbe
- Comme ça* peut être instancié dans d'autres dispositifs, par exemple avec *il y a*

Pour contextualiser les propriétés listées dans a)-d) ci-dessus, nous utilisons un exemple utilisé par les autrices (Roubaud et Sabio 2018, 6) :

L1 : et vous parlez l'arabe marocain ou l'arabe standard
L2 : l'arabe marocain et standard les deux
L1 : les deux vous avez étudié donc à l'école ou
L2 : non non non c'est à force de parler avec avec mes oncles mes mes tantes et
J'ai un oncle qui est prof euh qui est professeur au Maroc des fois il m'apprend
des petits mots des on discute on essaie de d'avoir des conversations et il
m'écoute parler il me reprend il me corrige puis **c'est comme ça que** j'ai appris
mais je sais pas l'écrire

L'exemple ci-dessus permet d'illustrer les quatre propriétés de la manière suivante :

- a) *Comme ça* peut être remplacé par un autre élément lexical : c'est *en parlant arabe avec ma famille* que j'ai appris
- b) *Comme ça* peut être contrasté avec un autre élément du même type : ce *n'est pas comme ça* que j'ai appris *mais en suivant des cours du soir*
- c) *Comme ça* peut être modifié par un adverbe : c'est *uniquement / surtout* comme ça que j'ai appris l'arabe
- d) *Comme ça* peut être instancié dans d'autres dispositifs, par exemple avec *il y a : il n'y a que comme ça* que j'ai appris

Nous nous sommes appuyées sur les propriétés et les définitions décrites ci-dessus dans notre analyse des clivées identifiées dans le corpus. Ainsi, notre analyse des quatre occurrences du corpus est la suivante :

Clivée contrastive

- 18) Cette pendule faisait vibrer les murs de la demeure de madame Lise. **C'est comme ça qu'**on l'appelait dans le quartier, tout le monde la connaissait (UN-M2-2018-TUTJ2-D1-R8-V1)
- 19) Même les jumeaux le savaient : dès qu'elle aurait récolté l'argent nécessaire, ils plieraient bagage afin de partir vivre dans « un monde meilleur ». **C'est comme ça qu'**ils avaient pris l'habitude de nommer ce futur petit appartement situé en centre-ville, idéalement proche d'une école et du travail de leur mère. (UN-M2-2020-UCL-D1-R71-V1_N)
- 20) On a jamais retrouvé la botte de son frère, ce qui confirme l'existence de ce monstre. Depuis cette aventure, les enfants ne sortent plus la nuit. **C'est comme cela qu'**elle finit son histoire alors que les deux têtes brunes disparaissent déjà sous les couvertures pour ne pas risquer d'apercevoir le monstre par la fenêtre. (UN-M2-2021-UCL-D1-R63-V1_N)

Fonction discursive/ clivée

- 21) Au final, l'inconvénient principal de cette séparation, c'est le déménagement : devoir retrouver une nouvelle maison, de nouvelles habitudes, me recréer un nouveau chez moi. Je vous passe les détails...mais **c'est comme ça que** le mois dernier j'ai fini par atterrir dans un quartier tranquille de la banlieue d'Avalon avec ma mère et mon petit-frère. (UN-M2-2021-UCL-D1-R25-V1_N)

Selon notre analyse, 3 occurrences sur 4 peuvent être interprétées comme des constructions clivées. Nous défendons notre interprétation en renvoyant à la réécriture possible des constructions clivées avec *de cette manière-là* :

- 22) On l'appelait **de cette manière-là**
- 23) Ils avaient pris l'habitude de nommer ce futur petit appartement **de cette manière-là**
- 24) Elle finit son histoire **de cette manière-là**

En revanche, nous prétendons que l'exemple 21) peut être interprété soit comme une clivée, soit comme un connecteur discursif. L'arrivée dans un quartier tranquille de la banlieue peut être jugé comme la conséquence d'une suite d'événements (la séparation suivie du déménagement). Pourtant, il est également possible de dire que *c'est en déménageant que le personnage se retrouve dans un quartier*

tranquille de la banlieue. Autrement dit, *comme ça* peut être remplacé par un autre élément lexical, ce qui est plutôt en faveur de la présence d'une clivée.

La construction *c'est pour ça que* est plus fréquente chez les locuteurs au collège et à l'école primaire qu'à l'université. En observant les textes des locuteurs à l'université, nous avons remarqué que d'autres constructions sont employées pour exprimer le lien entre une cause et un effet, telles que *voilà pourquoi* et *c'est pourquoi*.

Notons également que la construction *c'est là que* a une fréquence relative plus élevée chez les collégiens et les élèves à l'école primaire que chez les étudiants à l'université. Nous observons que d'autres clivées sont employées avec le même sens, à savoir *c'est à ce moment(-là) que* et *c'est à cet instant que*. De même, la locution *à ce moment-là* et l'adverbe *là* sont employés sans mise en relief :

25) **A ce moment-là**, Violette se leva pour aller chercher du popcorn (UN-M2-2020-UCL-D1-R63-V1_N)

26) **Là**, il aperçut une grosse voiture de Mr WILSON, le voisin (CO-3e-2018-FSBJC6-D1-R5-V1)

Les nombreuses manières d'exprimer un lien logique (cause-effet) ou de donner un cadre temporel que nous avons mentionnées ci-dessus ne sont pas quantifiées dans le cadre de ce travail. Néanmoins, nous prétendons que ces dernières méritent d'être examinées d'avantage dans une étude ultérieure.

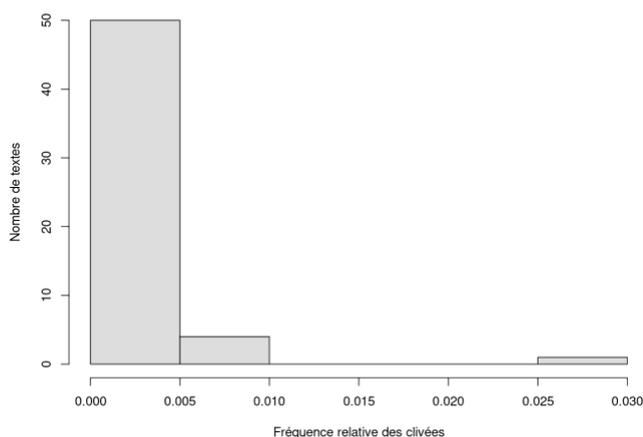
PARTIE 5 – Analyse contrastive des clivées

Ce chapitre est consacré à une analyse inférentielle des clivées. Dans un premier temps, nous examinerons la fréquence relative des clivées par classe scolaire et par niveau de scolarité, sans pour autant faire une distinction des différents sous-types de clivées. Dans un second temps, nous analyserons les différents types de clivées de plus près.

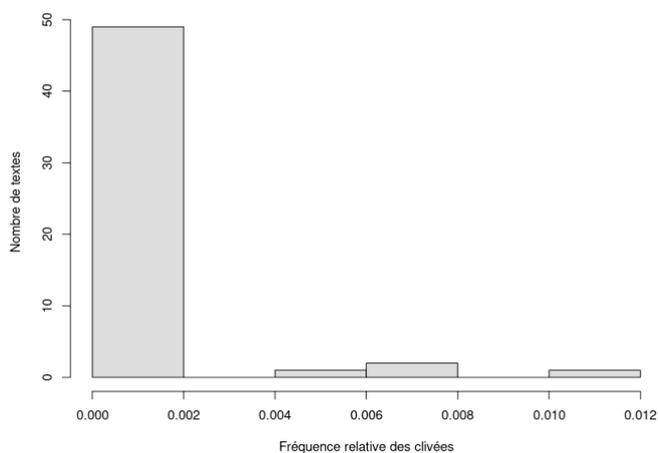
5.1 Fréquence relative des clivées selon le niveau de scolarité

Nous rappelons qu'un de nos objectifs est d'examiner si la fréquence des clivées diminue au fur et à mesure que le niveau de littératie augmente. Le niveau de littératie correspond dans cette étude au niveau de scolarité (école primaire à l'université). Pour rajouter encore plus de précision, nous examinons également la variable de la classe scolaire (CE2 à M2). Afin de choisir un test statistique convenable, nous avons d'abord observé la distribution de la variable de la fréquence relative des clivées. Les histogrammes ci-dessous montrent la distribution de la fréquence relative selon la classe scolaire. Les histogrammes montrent clairement qu'il s'agit d'une variable non-paramétrique, puisque la distribution n'est dans aucun cas normale. Ceci dit, la distribution n'est pas étonnante : nous examinons une construction assez peu fréquente, présente dans 124 textes sur 613. Par conséquent, dans la plupart de textes, la fréquence relative correspond à 0.

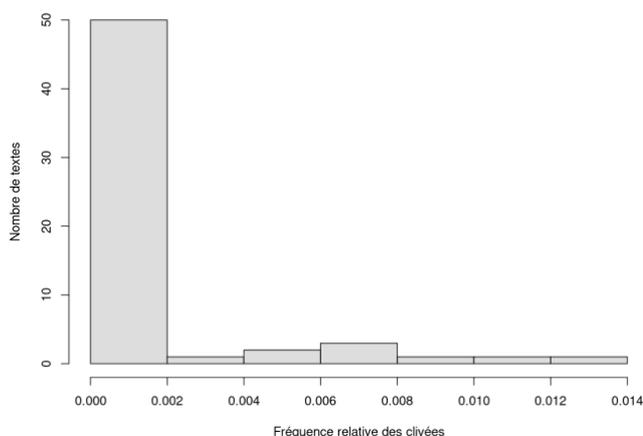
Histogramme de la fréquence relative des clivées dans des textes rédigés en CE2



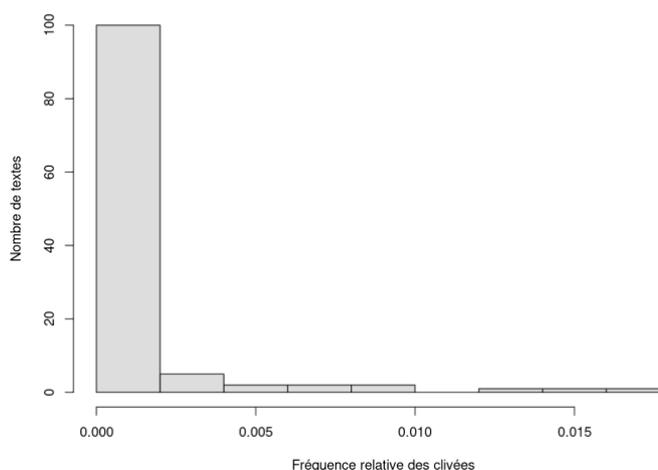
Histogramme de la fréquence relative des clivées dans des textes rédigés en CM1



Histogramme de la fréquence relative des clivées dans des textes rédigés en CM2



Histogramme de la fréquence relative des clivées dans des textes rédigés en 6e



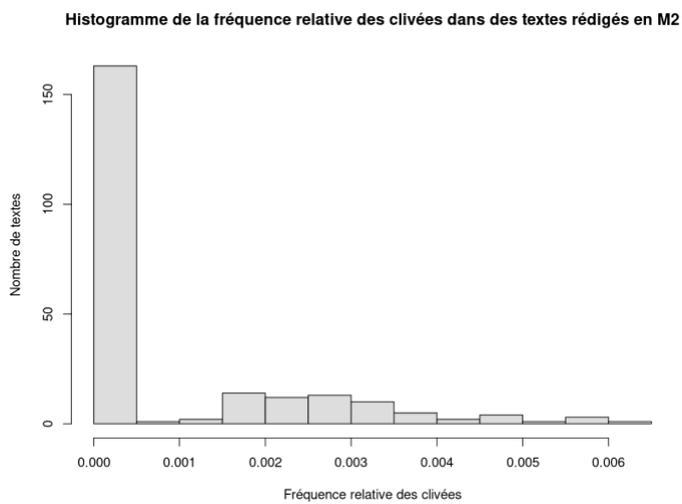
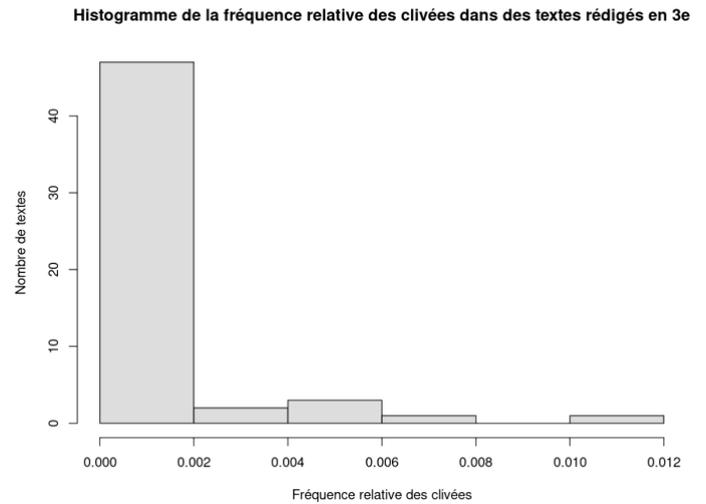
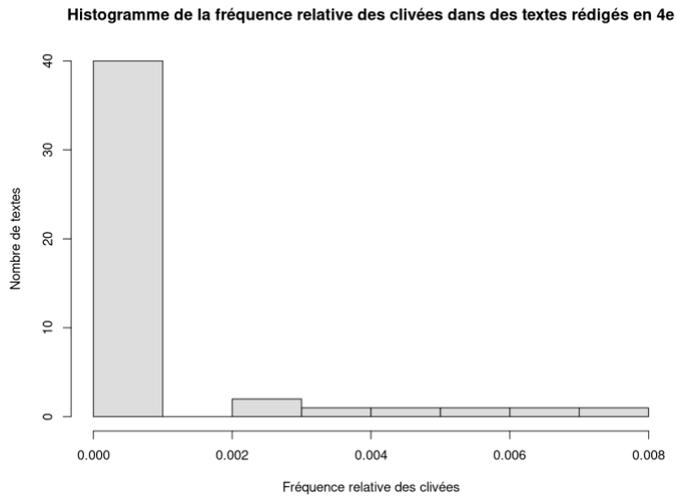


Figure 18 Histogrammes de la fréquence relative des clivées selon la classe scolaire

Pour plus facilement pouvoir comparer la distribution dans les différentes classes, nous observons également les boîtes à moustache illustrées dans la figure 19. Les boîtes à moustaches montrent que la médiane de la fréquence relative est de 0 pour toutes les classes scolaires. Seuls les textes produits par des étudiants de master ont une valeur du troisième quartile supérieure à 0. Pour les classes scolaires restantes, toutes les fréquences relatives supérieures à 0 sont considérées comme des valeurs extrêmes, comme le montrent les points au-dessus de l'extrémité supérieure.

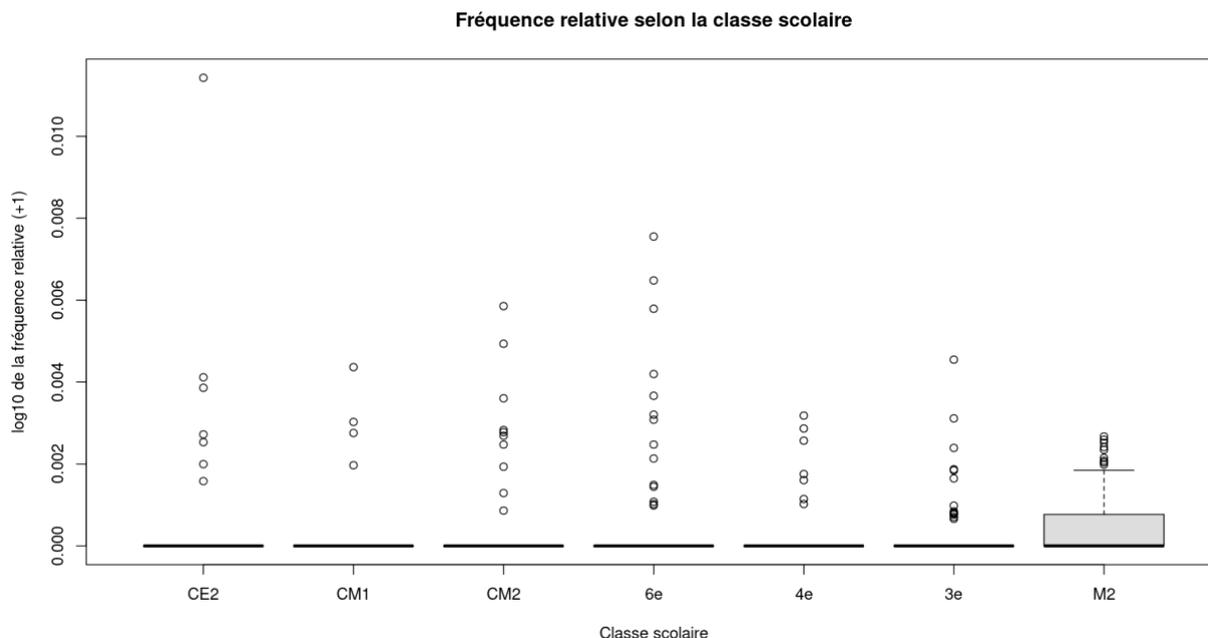


Figure 19 Boîtes à moustaches de la fréquence relative des clivées selon la classe scolaire

La dispersion de la fréquence relative des clivées du corpus est également illustrée dans le tableau 16 ci-dessous. En comparant la moyenne et l'écart-type de chaque classe, on constate que les fréquences relatives ne sont pas regroupées autour de la moyenne, mais elles sont extrêmement dispersées : dans plusieurs cas l'écart-type est même plus élevé que la moyenne.

Classe scolaire	Min	Max	1Q	Médiane	3Q	Moyenne	Écart-type
CE2	0	0.027	0	0	0	0.0019	0.004
CM1	0	0.0101	0	0	0	0.0005	0.002
CM2	0	0.014	0	0	0	0.0011	0.003
6e	0	0.018	0	0	0	0.0009	0.003
4e	0	0.007	0	0	0	0.0007	0.002
3e	0	0.0105	0	0	0	0.00089	0.002
M2	0	0.006	0	0	0.002	0.00085	0.001

Tableau 16 Répartition de la fréquence relative des clivées selon la classe scolaire

Les boîtes à moustaches dans la figure 19 et les valeurs dans le tableau 16 montrent clairement que la variable de la fréquence relative n'a pas une distribution normale dans le corpus. Il s'agit en d'autres termes d'une variable non-paramétrique. Par conséquent, nous avons choisi d'utiliser le test de corrélation de Spearman. Ce test convient aux données non-paramétriques pour observer s'il y a une corrélation linéaire entre deux variables. Dans notre cas, nous souhaitons observer si la fréquence relative des clivées diminue au fur et à mesure que le niveau d'étude augmente, ce qui se traduit en une corrélation linéaire négative entre les deux variables.

Dans ce travail, nous considérons que les résultats sont significatifs si la valeur-p est inférieure à 0,05. Une corrélation linéaire négative correspond à un coefficient de corrélation de Spearman de -1 et une corrélation linéaire positive correspond à un coefficient de corrélation de Spearman de 1. Si le coefficient de corrélation de Spearman est 0, cela veut dire qu'il n'existe aucune corrélation linéaire entre les variables. Ainsi, pour confirmer notre hypothèse concernant la corrélation entre la variable de la fréquence des clivées et le niveau de scolarité, il faut un coefficient de corrélation de Spearman de près de -1 et une valeur-p inférieure à 0,05.

Nous précisons que nous ne pouvons pas nous appuyer sur la fréquence relative par classe telle qu'elle est indiquée dans le tableau 10, et cela puisque nos données contiennent un nombre considérable de valeurs extrêmes. À la place, nous considérons la fréquence relative par individu (c'est-à-dire par texte).

Par ailleurs, le test de corrélation de Spearman prend en compte tous les individus sauf les individus aberrants, ce qui veut dire que l'on ne peut pas s'appuyer sur la moyenne d'un groupe.

Les résultats de la corrélation de Spearman figurent dans le tableau 13 ci-dessous.

	<i>Corrélation de Spearman</i>	
	Valeur-p	Coefficient de corrélation
<i>Fréquence relative Niveau de scolarité</i>	0.0003	0.15
<i>Fréquence relative Classe scolaire</i>	0.0001	0.16

Tableau 17 Corrélation de Spearman pour les variables de fréquence relative et le niveau scolaire/ class scolaire

Les résultats présentés dans le tableau 13 ci-dessus sont significatifs, aussi bien en observant la classe scolaire que le niveau scolaire, et cela puisque la valeur-p ne dépasse pas le seuil de 0,05. Cependant, le coefficient de corrélation est de 0,15 et de 0,16, ce qui est très faible. En effet, on peut même noter que le coefficient de corrélation est plus près de 1 que de -1. Autrement dit, si l'on veut indiquer une direction de la corrélation entre les deux variables, il s'agit plutôt d'une corrélation linéaire *légèrement* positive, bien que l'on soit loin d'une corrélation parfaite.

5.2 Type de clivée selon le niveau de scolarité

Dans la partie précédente, nous avons constaté que les clivées ne s'emploient pas de plus en plus au fur et à mesure que le niveau de scolarité augmente. Le coefficient de corrélation de Spearman étant de 0.15 et 0.16, la corrélation est plus au moins inexistante. Avec une valeur-p de 0.003 et 0.001, les résultats peuvent être considérés significatifs. Il faut pourtant souligner que ces résultats concernent *l'ensemble des clivées*. Cela veut dire qu'aucune distinction a été faite entre les différents sous-types de clivées. Dans ce qui suit, nous allons examiner si l'emploi d'un certain type de clivée dépend du niveau de scolarité du locuteur. Pour ce faire, des tests chi-deux seront effectués. En traitant un sous-type à la fois, nous indiquons pour chaque texte avec une clivée, s'il y a présence (OUI) ou absence (NON) du sous-type en question. Notons que nous nous intéressons uniquement aux textes avec présence des clivées dans cette partie, pas à la totalité des textes comme dans la partie précédente. Par conséquent, la fréquence relative n'est plus utilisée dans cette partie.

Nous ne traiterons pas les clivées de type *pronom avoir X qui* séparément de cette partie de l'analyse. Par ailleurs, aucune distinction ne sera faite entre les différents types de pronoms relatifs : *C'est X que* et *C'est X qui* seront traités ensemble, de la même manière que *Il y a X que* et *Il y a X qui*.

D'abord, une observation des tableaux d'effectifs observés et attendus pour chaque type de clivée a permis de constater que certains effectifs attendus étaient inférieurs à 5 à cause de la faible fréquence des clivées de chaque sous-type par niveau de scolarité. Les tableaux d'effectifs observés et attendus sont présentés dans le tableau 18 ci-dessous. Les valeurs inférieures à 5 sont mises en gras :

		<i>Effectifs observés</i>			<i>Effectifs attendus</i>		
		<i>Clivée prototypique</i>					
<i>Présence dans le texte</i>		École primaire	Collège	Université	École primaire	Collège	Université
<i>NON</i>		9	9	10	4.74	7.90	15.35
<i>OUI</i>		12	26	58	16.26	27.10	52.65
		<i>Clivée il y a X qu-</i>					
<i>Présence dans le texte</i>		École primaire	Collège	Université	École primaire	Collège	Université
<i>NON</i>		25	14	66	17.78	29.64	57.58
<i>OUI</i>		10	7	2	3.22	5.36	10.42

Présence dans le texte	Pseudo-clivée			École primaire	Collège	Université
	École primaire	Collège	Université			
NON	18	33	57	18.29	30.48	59.22
OUI	3	2	11	2.71	4.51	8.77

Tableau 18 Tableau d'effectifs observés et attendus selon le type de clivée analysée

Les effectifs attendus inférieurs à 5 dans le tableau 18 nous ont obligés de regrouper les niveaux de scolarités afin d'avoir des résultats statistiquement fiables. Dans les résultats présentés dans ce qui suit, les textes rédigés par des locuteurs à l'école primaire et au collège sont traités comme une seule catégorie : *l'enseignement obligatoire*. Nous justifions ce regroupement par le fait que le français, et par extension l'écriture en français, est enseigné de façon explicite à l'école primaire et au collège, mais pas au deuxième cycle de l'université. De plus, les diagrammes d'association générés à partir des tableaux d'effectifs montrent clairement que les proportions de ces deux niveaux de scolarité se rapprochent.

Dans ce qui suit, nous présenterons les résultats du test chi-deux par type de clivée. Comme dans la partie précédente, nous considérons que le seuil de signification est 0,05. Cela veut dire qu'il y a dépendance entre les deux variables si la valeur-p est inférieure à ce dernier. S'il y a dépendance, nous nous intéresserons également à la taille d'effet de la dépendance. La taille d'effet sera calculée avec le coefficient V de Cramer. Le coefficient V de Cramer permet de calculer, pour un tableau de contingence, le rapport entre le chi-deux et la valeur maximale possible pour le chi-deux pour un tableau de cette taille et le même effectif total. Le coefficient peut être situé entre 0 et 1. Si la valeur V de Cramer est 1, cela veut dire qu'il s'agit d'une association parfaite : les variables sont associées à 100%. Au contraire, si la valeur V de Cramer correspond à 0, cela indique que les variables ne sont pas du tout associées.

5.2.1 Clivées prototypiques

Le tableau 20 ci-dessous montre les effectifs observés de chaque catégorie ainsi que les proportions. En observant le tableau, il est possible de constater que la majorité des textes avec clivées contiennent des clivées prototypiques, et cela dans tous les niveaux de scolarité. Parmi les clivées produites par les locuteurs à l'école primaire et au collège, 68% sont prototypiques, tandis que 85% des clivées sont prototypiques chez les locuteurs à l'université.

Texte avec clivée prototypique	L'enseignement obligatoire	Université
NON	18 (32%)	10 (15%)
OUI	38 (68%)	58 (85%)
TOTAL	56 (100%)	68 (100%)

Tableau 19 Tableau d'effectifs observés et les proportions pour les clivées prototypiques

Le test de chi-deux confirme que les deux variables sont dépendantes, la valeur-p étant inférieure au seuil de signification. Cela veut dire qu'il est plus probable de trouver un texte avec une clivée dans un texte rédigé par un étudiant de master que par un élève à l'école primaire ou au collège.

Test de chi-deux de Pearson	
χ^2	4.39
Degré de liberté	1
Value-p	0.04

Tableau 20 Test de chi-deux de Pearson, clivées prototypiques et niveau de scolarité

Il faut toutefois noter que la valeur V de Cramer, à savoir la taille d'effet de cette dépendance, est de 0.19. Cette valeur pourrait être interprétée comme une relation faible, étant donné qu'une association complète correspond à 1, et l'absence d'une association correspond à 0.

5.2.2 Il y a X qu-

Si les clivées prototypiques sont plus probables à trouver dans les textes écrits par des locuteurs à l'université, les clivées de type *il y a X qu-* semblent être plus probables à trouver dans des textes écrits par des locuteurs à l'école primaire et au collège. 30% des clivées produites par des collégiens et des élèves à l'école primaire sont de type *il y a X qu-*, tandis que ce même type de clivée ne constitue que 3% des clivées produites par des étudiants de master. Le tableau 21 montre les effectifs observés de même que les proportions par sous-groupe.

<i>Texte avec clivée il y a X qu-</i>	<i>Enseignement obligatoire</i>		<i>Université</i>	
<i>NON</i>	39	(70%)	66	(97%)
<i>OUI</i>	17	(30%)	2	(3%)
TOTAL	56	(100%)	68	(100%)

Tableau 21 Tableau d'effectifs observés et les proportions pour les clivées IL Y A X qu-

Le test de chi-deux confirme que les deux variables sont fortement dépendantes, la valeur-p étant extrêmement basse comparée au seuil de signification de 0,05.

<i>Test de chi-deux de Pearson</i>	
X^2	15.74
Degré de liberté	1
Value-p	7.267e-05

Tableau 22 Test de chi-deux de Pearson, clivées IL Y A X QU- et niveau de scolarité

La taille d'effet de cette association est de 0,36. Il s'agit en d'autres termes d'une association plus forte entre le niveau de scolarité et les clivées *il y a X qu-* qu'entre le niveau de scolarité et les clivées prototypiques. Ceci dit, la taille d'effet demeure moyennement faible.

5.2.3 Pseudo-clivées

Le tableau de proportions ci-dessous indique que tous les niveaux de scolarité favorisent d'autres types de clivées que les pseudo-clivées. Il est pourtant possible de constater que les pseudo-clivées sont plus rares dans les productions à l'école primaire et au collège qu'à l'université.

<i>Texte avec pseudo-clivée</i>	<i>Enseignement obligatoire</i>		<i>Université</i>	
<i>NON</i>	51	(91%)	57	(84%)
<i>OUI</i>	5	(9%)	11	(16%)
TOTAL	56	(100%)	68	(100%)

Tableau 23 Tableau de d'effectifs observés et les proportions pour les pseudo-clivées

Comme il est démontré dans le test de chi-deux ci-dessous, l'association entre la présence d'un texte avec une pseudo-clivée et le niveau de scolarité n'est pas suffisamment forte pour que les résultats soient significatifs, la valeur-p étant de 0,35.

<i>Test de chi-deux de Pearson</i>	
X^2	0.86
Degré de liberté	1
Value-p	0.35

Tableau 24 Test de chi-deux de Pearson, pseudo-clivées et niveau de scolarité

Vu que la valeur-p du test chi-deux est supérieur au seuil de signification (0,05), nous n'avons pas calculé le coefficient de V de Cramer.

PARTIE 6 – Discussion

Dans ce qui suit, les résultats de notre étude seront discutés. Nous commencerons par aborder les observations faites dans l'étape de l'extraction des clivées. Ensuite, l'usage des clivées sera discuté. Nous finirons par traiter les résultats de notre analyse inférentielle.

6.1 La détection semi-automatique des constructions clivées : une tâche complexe

Nous rappelons qu'un des objectifs de cette étude était de nature plutôt méthodologique, l'objectif étant de proposer une méthode permettant d'extraire semi-automatiquement des constructions clivées et disloquées d'un corpus écrit. Étant donné la quantité des données clivées à annoter et analyser, nous avons toutefois pris la décision de nous limiter à ces dernières, pour traiter les dislocations dans une étude ultérieure.

Une première extraction des clivées a été faite avec l'outil TXM, et cela puisqu'il est doté d'une interface maniable et d'une multitude de fonctionnalités. Nos premiers patrons TXM qui avaient pour but d'identifier toutes les clivées du corpus, avaient naturellement une faible précision. Après avoir étudié de plus près les faux positifs de nos patrons, il a pu être constaté que le bruit ne contenait pas uniquement des constructions avec une forme de surface identique à celle des clivées, mais également d'autres types de faux positifs. Parmi ces derniers se trouvent les cas où l'élément entre la copule et la relative ne constitue pas un seul syntagme. Les formes homonymes étaient difficilement évitées par un patron formulé en CQL, nous avons toutefois réussi à améliorer la précision des patrons en rajoutant des contraintes pour éviter de détecter les autres types de faux positifs. Cependant, la précision reste finalement assez faible dans la mesure où TXM ne dispose pas de moyen d'identifier les syntagmes d'un texte.

Il est également important de soulever que les améliorations de nos patrons ont été effectuées en fonction du bruit du corpus en question, et que le même patron pourrait avoir une plus mauvaise précision ou un plus faible rappel s'il était appliqué à un autre corpus. Par exemple, la clivée dans 1) ne serait pas détectée par notre deuxième patron TXM, et cela puisque le premier *que* est précédé de *pendant*. Notons aussi que nos patrons améliorés ne prennent pas en compte les clivées prototypiques dont l'élément clivé est une subordonnée qui commence par *que*. Ainsi, une clivée de type 2) ne serait pas identifiée à l'aide de notre patron amélioré. Pourtant, comme le constatent Abeillé et Godard (2021, 1958), les subordonnées complétives ont plutôt tendance à figurer dans des pseudo-clivées que dans des clivées prototypiques. Un exemple d'une telle pseudo-clivée est donné en 3).

- 1) C'est **pendant qu'on calcule, qu'on analyse les pour et les contre**, que la vie passe, et qu'il ne se passe rien (Marc Lévy²³)
- 2) C'est **qu'il puisse faire naître en nous-même l'infinie diversité** qui donne du prix à nos plus précieux rouleaux de soie. (Rouquier, 2018)
- 3) Ce qui nous donne du prix à nos plus précieux rouleaux de soie, c'est **qu'il puisse faire naître en nous-même l'infinie diversité**.

Il faut également souligner la difficulté liée au seuil de mots de l'élément clivé. À notre connaissance, il n'y a pas d'expression en CQL permettant de contourner le seuil de mots d'un élément potentiellement vide. Étant donné qu'il est impossible de savoir le nombre maximum d'une clivée, nous nous sommes limitées à 20 mots. Ce seuil a ensuite été modifié pour chaque patron en fonction de la clivée la plus longue détectée afin de diminuer le bruit autant que possible. Bien évidemment, cela pourrait affecter le rappel si l'on l'utilise avec un autre corpus. Dans la mesure où nous nous sommes limitées à un seuil de 20 mots, nous n'avons pas pu identifier la clivée ci-dessous, trouvée par hasard dans le corpus. Nous présumons toutefois que de telles occurrences sont rares.

²³ Citation repérée sur le site du Figaro 1 juin 2022 : <http://evene.lefigaro.fr/citation/pendant-calcule-analyse-contre-vie-passe-passe-rien-42715.php>

- 4) **C'est** avec une grande impatience, et non sans une pointe d'appréhension (sa grand-mère, connue pour son esprit critique, mettait un point d'honneur à toujours livrer des avis objectifs, même à son petit-fils adoré), **que** Charles parcourut la distance qui séparait son habitation en centre-ville du domicile de son ancêtre. (UN-M2-2021-UCL-D1-R41-V1)

La difficulté d'identification des syntagmes avec un patron TXM nous a amenée à faire un post-traitement des données avec Stanza. Les annotations de Stanza nous ont permis d'évaluer la performance du modèle d'annotation d'UD français sur des données non-standard en élaborant un script Python. Notre script, qui extrayait les clivées à l'aide de la relation syntaxique entre le verbe de la relative et le gouverneur de l'élément clivé, n'a pas été plus performant que les patrons TXM au niveau du rappel. Cependant, la précision du script était meilleure que la précision des patrons TXM, et cela puisque nous n'avions plus la difficulté de distinguer des syntagmes. Quoiqu'il en soit, il est possible de constater que la nature non-standard des données entrave la performance des annotations de Stanza. Comme nous l'avons soulevé dans 3.2.2 Extraction avec un script Python, la relation syntaxique des clivées (*advcl:cleft*) est censée être identifiée pour tous les types de clivées traités dans cette étude²⁴, or les seules clivées identifiées sont des clivées prototypiques. Nous supposons que les phrases longues et complexes, parfois avec une syntaxe erronée, ont affecté la performance de l'annotation du parser.

Notre script Python constitue une tentative pertinente d'utiliser les dépendances syntaxiques proposées par UD dans le but d'extraire des clivées, cependant nous estimons qu'il y a des manières plus efficaces d'utiliser les données annotées par Stanza. Une méthode plus efficace pourrait consister à identifier les séquences suivantes : *CE (NE) ÊTRE + syntagme + que/qui*, *IL (NE) Y AVOIR + syntagme + que/qui*, *PRONOM + AVOIR + syntagme + qui*. Une telle extraction ne reposerait pas sur le type de relations syntaxiques entre les éléments, ce qui limiterait le silence.

Même si l'amélioration de l'extraction mentionnée ci-dessus permettrait de limiter le bruit, cette méthode n'éviterait pas d'identifier les constructions homonymes. En considérant qu'une différence cruciale entre une relative « ordinaire » et une clivée prototypique est que le pronom *ce* a une fonction anaphorique dans cette première, nous estimons qu'une résolution de la coréférence aurait pu être appliquée aux données afin de pouvoir distinguer ces deux constructions de façon automatique. Cependant, la résolution de la coréférence ne permettrait pas de distinguer tous types de présentatives des clivées. En plus, comme le constatent Karssenberget Lamiroy (2017), les clivées cohésives (*c'est ainsi que, c'est alors que...*) peuvent être interprétées comme ayant un lien coréférentiel avec le contexte antérieur. Avec une résolution de la coréférence, ces clivées risqueraient d'être exclues.

Quoiqu'il en soit, il faut soulever, comme le font les auteurs de GGF (Abeillé et Godard 2021, 117), que les présentatives sont également bien attestées à l'oral spontané. Par conséquent, la prise en compte des présentatives dans l'étude des constructions caractéristiques de l'immédiat n'est pas à considérer comme un souci méthodologique.

6.2 L'usage des clivées

La fréquence relative des clivées dans le corpus correspond à 0.09%. Cette fréquence, qui peut apparaître basse, est similaire aux fréquences attestées dans les statistiques pour la relation syntaxique des clivées (*advcl:cleft*) dans les corpus arborés de l'UD :

²⁴ Hormis les pseudo-clivées qui sont identifiées comme des dislocations.

<i>Nom du corpus</i>	<i>Type de corpus</i>	<i>% advcl:cleft</i>
<i>UD_French_GSD</i>	Blogs et pages Wikipédia	0.05%
<i>UD_French_ParisStories</i>	Narration orale	0.08%
<i>UD_French_Sequoia</i>	Domaine médical, textes journalistiques, débat parlementaire, pages Wikipédia	0.03%
<i>UD_French_Rhapsodie</i>	Corpus oral	0.18%

Tableau 25 Comparaison des fréquences des clivées dans des corpus annotés avec les jeux d'étiquettes d'UD français (<https://universaldependencies.org/fr/dep/index.html>)

Notons que la fréquence est la plus élevée dans les corpus purement oraux (Rhapsodie et ParisStories), ce qui corrobore le fait que les clivées appartiennent aux traits caractéristiques de l'immédiat. Par ailleurs, il est intéressant de noter que la fréquence des clivées dans la narration orale est similaire à la fréquence des clivées dans Résolco. Nous reviendrons à la question de la narration plus tard. Néanmoins, vu la faible performance du modèle d'annotation d'UD appliqué à notre corpus, il est possible de questionner la fiabilité des statistiques présentées dans le tableau 25.

Peu surprenant, la majorité des clivées identifiées sont des clivées prototypiques. Les clivées les moins fréquentes sont les clivées de type « j'ai mon ami qui habite à Paris » ou « il n'y a que cette voiture que je conduis ». Il faut cependant noter que notre étude n'inclut pas tous types de clivées. Nous n'avons pas traité des clivées prototypiques avec négation et des clivées de type *c'est de X dont* et *c'est X où*. De même, nous n'avons pas inclus les pseudo-clivées dont la première partie n'est pas une relative. Il serait intéressant d'élargir l'échantillon des clivées analysées dans une étude ultérieure.

Nos résultats par rapport au rôle syntaxique de l'élément clivé sont conformes aux études antérieures : les sujets et les circonstants sont les fonctions syntaxiques les plus souvent clivées dans le texte. La fréquence élevée des sujets est peu surprenante, étant donné qu'ils peuvent figurer dans tous les types de clivées analysées et que les constructions avec *qui* sont majoritaires.

Nous avons noté que les éléments clivés du corpus constituent le plus fréquemment des syntagmes nominaux. Ce résultat peut être expliqué par le fait que tous types de clivées peuvent avoir un syntagme nominal dans l'élément clivé. Nos résultats montrent également que les adverbes qui portent le rôle de circonstant sont exclusivement ceux traités par Roubaud et Sabio (2018). Leur fréquence élevée dans plusieurs types de textes prouve leur importance discursive et expliquent les nombreuses études portant sur ce type de clivée (Roubaud et Sabio 2015, 2018 ; Karssenberget et Lamiroy 2017).

Concernant la longueur de l'élément clivé, exception faite des pseudo-clivées, toutes les clivées ont tendance à ne pas compter plus de 5 mots. En plus, il a été observé que les clivées les plus longues sont exclusivement produites par des locuteurs à l'université. Nous estimons que l'absence des clivées plus longues que 15 mots chez les élèves au collège et à l'école primaire peut s'expliquer par le fait que celles-ci constituent exclusivement des phrases complexes, souvent avec des subordinées. Étant donné qu'il a déjà été suggéré que la fréquence des phrases complexes augmente avec l'âge (Boivin et al. 2017), ce n'est pas surprenant si les locuteurs à l'université produisent des clivées plus longues et complexes que les locuteurs à l'école primaire.

6.3 La corrélation entre la fréquence des clivées et le niveau de scolarité

Une de nos hypothèses était que la fréquence des clivées diminue selon que le niveau de scolarité augmente. Afin de vérifier cette hypothèse, des tests de corrélation de Spearman ont été effectués, aussi bien avec les classes scolaires qu'avec les niveaux scolaires. Les résultats de ces tests n'ont pas confirmé l'hypothèse. En effet, aucune corrélation a été prouvée entre le niveau de scolarité et la fréquence des clivées. Cependant, nos tests statistiques bivariés qualitatifs semblent suggérer que parmi toutes les clivées identifiées, il est plus probable de trouver des clivées prototypiques dans les écrits scolaires produits à l'université, que dans les autres niveaux scolaires analysés. En revanche, les clivées avec *il y*

a semblent être plus probables à trouver dans les écrits scolaires produits à l'école primaire et au collège²⁵.

Quoi qu'il en soit, nous trouvons la distribution des clivées de type *il y a X qu-* intéressante en considérant les résultats de l'étude de Karssenberget Lahousse (2015) qui montrent clairement que ce type de clivée est plus fréquent à l'oral et à l'écrit informel qu'à l'écrit plus formel. Le fait que nos résultats indiquent que les clivées de ce type sont plus probables à trouver dans des textes rédigés à l'école qu'à l'université, pourrait signaler que les locuteurs à l'université sont plus conscients des différences entre les usages de l'immédiat et la distance. Néanmoins, la liaison entre le niveau de scolarité et l'emploi des clivées *il y a X qu-* pourrait être étudiée davantage, par exemple en se demandant s'il y a une corrélation linéaire entre la fréquence relative de ce type de construction et le niveau de scolarité. Il serait également intéressant de compléter ces résultats par une étude d'acceptabilité pour savoir si les clivées de type *il y a X qu-* sont jugées moins acceptables par les locuteurs à l'université que par les locuteurs aux niveaux scolaires plus bas.

Nous estimons que le manque d'une corrélation entre la fréquence relative des clivées et le niveau scolaire peut s'expliquer par le type de texte analysé, à savoir des textes narratifs. Dans ce qui suit, nous insistons sur le rôle que joue le type de texte dans nos résultats.

Un des intérêts de cette étude était d'analyser l'usage des clivées dans les écrits scolaires. Il est toutefois important de se rendre compte du fait que les écrits scolaires analysés font partie d'un type textuel spécifique : la narration. Comme le constatent Koch et Oesterreicher (2001), la narration constitue un cas particulier dans le continuum de l'oral et de l'écrit. Selon les auteurs, le genre narratif présuppose des conditions de l'écrit, telles que le détachement actionnel, situationnel et référentiel et une fixation thématique (cf. le tableau 1). Nous ajoutons qu'il s'agit d'une communication préparée, séparée au niveau spatio-temporel avec le moment de lecture. En même temps, le narrateur veut donner l'impression que les événements de la narration se déroulent devant les yeux du lecteur. Pour ce faire, il est courant d'insérer des dialogues et du discours rapporté. Nous avons observé plusieurs cas où les clivées figurent dans de tels contextes :

- 5) Un des frères dit : « **c'est pour ça qu'** il y avait du bruit hier. (EC-CM2-2016-SGLEA-D1-R7-V1_N)
- 6) « Vite **c'est papa qui** nous court après il faut se cacher » (EC-CM2-2016-SGLEB-D1-R20-V1_N)
- 7) – **C'est vous qui** avez renversé toutes les poteries d'Herphie ! (EC-CM2-2016-SGLEB-D1-R24-V1_N)

Nous pensons que le discours rapporté est à interpréter comme des productions imbriquées avec le but de refléter l'immédiat communicatif. Une autre stratégie employée dans la narration est le fait d'inviter le lecteur dans l'univers fictionnel en s'adressant au lecteur au sein du texte, comme le fait le narrateur dans les exemples ci-dessous tirés de nos données :

- 8) Je vous passe les détails...mais **c'est comme ça que** le mois dernier j'ai fini par atterrir dans un quartier tranquille de la banlieue d'Avalon avec ma mère et mon petit-frère. (UN-M2-2021-UCL-D1-R25-V1_N)
- 9) **C'est leur histoire que** je vais vous raconter (CO-6e-2016-VTAC603-D1-R11-V1_N)

En raison de la particularité du genre narratif, une attention particulière devrait être prêtée au contexte dans lequel les clivées figurent dans les productions écrites. Dans la mesure où notre extraction des clivées ne permet pas de prendre en compte les textes en entier, une quantification des contextes des clivées n'a pas été possible dans le cadre de ce travail. Cependant, nous envisageons une suite de ce travail où le contexte des clivées sera examiné de plus près. La prise en compte du contexte permettrait également une analyse au niveau de la structure informationnelle des clivées (cf. 1.3.5 Analyse pragmatique et informationnelle). Une telle analyse serait pertinente afin de confirmer les résultats de Karssenberget Lahousse (2015) concernant les structures informationnelles favorisées par les différents types de clivées, ainsi que d'observer comment les différentes structures informationnelles sont utilisées dans le genre narratif. Les résultats de l'analyse de Karssenberget Lahousse (2015) indiquent que les

²⁵ Nous soulignons que ces résultats ne prennent pas en compte tous les textes du corpus, mais uniquement les textes contenant au moins une clivée.

clivées prototypiques sont les plus souvent employées avec la structure informationnelle focus–arrière-fond tandis que les clivées avec *il y a* sont les plus souvent employées avec un tout-focus. Cela veut dire que les clivées prototypiques sont le plus souvent employées pour introduire un nouveau référent dans l'élément clivé, tandis que les clivées avec *il y a* donnent de la nouvelle information dans les deux parties de la construction. Cependant, l'analyse de Karssenberget Lahousse (2015) ne concerne pas de textes narratifs, mais des textes journalistiques, des entretiens et des forums de discussion. Nous nous demandons si les clivées cohésives ne sont pas plus fréquentes dans le genre narratif, comme déjà suggéré dans l'étude de Roubaud et Sabio (2018). Nous soulignons que nos résultats de l'annotation montrent que deux clivées cohésives qui servent à organiser la chronologie, à savoir *c'est alors que* et *c'est là que*, constituent presque 14% des clivées du corpus. À cette même catégorie de clivées s'ajoutent d'autres clivées identifiées dans le corpus, telles que *c'est à cet instant que* et *c'est à ce moment-là que*. Nous sommes certaines que l'étude des structures informationnelles des clivées du corpus pourrait contribuer à une meilleure compréhension de l'usage des clivées cohésives traitées dans la section 4.2.4 Comparaison avec l'étude de Roubaud et Sabio (2018).

Conclusion

Le point de départ dans ce travail était d'étudier les dislocations et les clivées dans les écrits scolaires, ces deux constructions étant typiques de l'immédiat communicatif, couramment appelées des *constructions de l'oral*. Notre objectif était non seulement de *décrire* comment ces dernières sont employées dans des écrits scolaires, mais aussi de les *comparer* selon le niveau de scolarité. Hormis ces deux objectifs, notre but était également d'explorer des méthodes possibles pour détecter ces types de constructions de façon semi-automatique.

Quelques écarts ont néanmoins été faits par rapports aux objectifs de départ. Tout d'abord, le travail chronophage d'annotation manuelle des clivées extraites ainsi que la richesse de nos données nous ont amenés à exclure les dislocations de l'analyse. Nous avons également pris la décision de ne pas prendre en compte tous les types de clivées, notamment tous les sous-types de pseudo-clivées et les clivées dont l'élément clivé compte plus de 20 mots-formes (à l'exception des pseudo-clivées).

La partie méthodologique de notre étude a démontré la difficulté de détecter les clivées de façon automatique, et cela surtout parce que l'interprétation des constructions homonymes nécessite une compréhension discursive et pragmatique que les outils n'ont pas.

Notre analyse des clivées identifiées dans le corpus a confirmé l'hypothèse que l'élément clivé est le plus souvent un sujet ou un circonstant, et cela avec une grande majorité. Nous avons également pu soulever le fait que la plupart des éléments clivés sont des syntagmes nominaux, des adverbes ou des syntagmes prépositionnels. Les clivées employées par les étudiants de master sont plus variées au niveau du constituant syntaxique de l'élément clivé, de même qu'au niveau de la longueur ce dernier. Pourtant, la plupart des clivées de corpus ne comportent pas plus de 5 mots.

Quant à la corrélation entre la fréquence des clivées et le niveau de scolarité, les résultats ne sont pas conformes à notre hypothèse. Il semble ne pas y avoir de corrélation entre les deux variables. Nous expliquons le manque de corrélation par la nature narrative des données analysées. Notre observation de la dépendance entre les différents types de clivées et le niveau scolaire est plus féconde, mais d'autres études seraient nécessaires afin de mieux expliquer le lien entre le niveau d'enseignement et l'usage de certaines clivées.

La richesse de notre sujet donne lieu à plusieurs pistes à suivre. Du point de vue méthodologique, nous sommes certaines que la performance de l'extraction automatique des clivées pourrait être améliorée, surtout en explorant davantage de manières de manipuler les données étiquetées par Stanza. De plus, notre analyse descriptive pourrait être complétée par une analyse approfondie du contexte des clivées, en mettant en avant le genre narratif du corpus. En analysant le contexte, il serait intéressant et pertinent d'observer la structure informationnelle des clivées de plus près.

Pour finir, ce travail contribue à l'analyse des clivées dans les écrits scolaires, aussi bien au niveau quantitatif que qualitatif. De plus, ce travail a permis d'évaluer deux modèles d'extraction des clivées, une tâche qui a été très peu documentée jusqu'à maintenant. Nous constatons que, bien que le projet É:CALM, dont Résolco fait partie, touche à sa fin avec la clôture en juin 2022, l'analyse et la manipulation des écrits scolaires demeure un enjeu important dans le domaine de la linguistique outillée et en TAL.

Bibliographie

- Abeillé, A., et Godard, D. (2021). *La grande grammaire du français* (Vols. 1–2). ACTES SUD.
- Apothéloz, D. et Roubaud, M.-N. (2018). Constructions pseudo-clivées, *Encyclopédie grammaticale du français*.
- Béguelin, M.-J., Denervaud, M. et Jespersen, J. (eds) (1988). *Écrire en français : cohésion textuelle et apprentissage de l'expression écrite*. Delachaux et Niestlé.
- Béguelin, M.-J. (dir.). (2000). *De la phrase aux énoncés : grammaire scolaire et descriptions linguistiques*. De Boeck-Duculot.
- Berrendonner, A. (2021). Constructions disloquées, *Encyclopédie Grammaticale du Français*.
- Blanche-Benveniste, C. Bilger, M., Rouget, C., Van Den Eynde, K., Mertens, P. et Willems, D. (1997). *Le français parlé : études grammaticales* (Sciences du langage). Editions du Centre national de la recherche scientifique.
- Blanche-Benveniste, C. (1997b). *Approches de la langue parlée en français*. Gap: Orphys.
- Blanche-Benveniste, C. (2002). Auxiliaires et degrés de « verbalité », *Syntaxe et sémantique*, 3, 75-97.
- Blasco-Dulbecco, M. (1999). *Les dislocations en français contemporain étude syntaxique* (Les français parlés - textes et études 1). Paris : H Champion.
- Boivin, M.-C., Roussel, K., Pinsonneault, R. (2017). Phrases complexes et maturité syntaxique : une comparaison entre des écrits d'élèves de 13 et 16 ans. *Lidil* [en ligne], 55.
- Carter-Thomas, S. (2009). The French c'est-cleft : function and frequency. *La linguistique systémique fonctionnelle et la langue française*, 127-157.
- Conti, V.(2010). La construction en avoir SN qui SV (« j'ai ma copine qui habite à Paris ») : une forme de dispositif clivé ?, *Linx*, 62-63, 63-87.
- De Cat, C. (2007). *French dislocation syntax, interpretation, acquisition* (Oxford studies in theoretical linguistics 17).
- Declerck, R. (1988). *Studies on copular sentences, clefts and pseudo-clefts*. Leuven : Leuven university press.
- De Marneffe, M-C, Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J. et Manning, C. (2014). Universal Stanford Dependencies: A cross-linguistic typology. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 4585–4592.
- De Marneffe, M-C., Guillaume, B., Perrier, G. (2019). Conversion et améliorations de corpus du français annotés en Universal Dependencies. *Revue TAL, ATALA (Association pour le Traitement Automatique des Langues)*, 60 (2), 71–95.
- De Marneffe, M-C., Manning, C., Nivre, J., Zeman, D. (2021). Universal Dependencies. *Computational Linguistics 2021*, 47 (2), 255–308.
- Gadet, F. (1997). *Le français ordinaire* (2e édition revue et augmentée ed., U Série Linguistique). Paris : Armand Colin.

- Gundel, J. K. (1988). Universals of topic comment structure. Dans M. Hammond, Edith A. Moravcsik et J. Wirth (eds), *Studies in Syntactic Typology*. Amsterdam/Philadelphia : John Benjamins.
- Heiden, S., Magué, J-P. et Pincemin, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement. *10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, 1021–1032.
- Karssenber, L. (2017). French 'il y a' clefts, existential sentences and the focus-marking hypothesis. *Journal of French Language Studies*, 27(3), 405–430.
- Karssenber, L. (2018). Il y a toujours un chameau qui tombe. Les multiples fonctions des clivées en *il y a*. *SHS Web of Conferences*, 46.
- Karssenber, L. et Lahousse, K. (2015). Two types of clefts in French: information structure, existential presupposition and syntax. *27 February 2015 IGG 41 Perugia*.
- Karssenber, L. et Lamiroy, B. (2017). C'est ainsi que: Grammaticalisation ou lexicalisation ou les deux à la fois ? *Journal of French Language Studies*, 27(2), 161–185.
- Katz, S. (1997). *The syntactic and pragmatic properties of the c'est-cleft construction*. Thèse de doctorat. Université du Texas à Austin.
- Koch, P. et Oesterreicher, W. (2001). 62. Gesprochene Sprache und geschriebene Sprache / Langage parlé et langage écrit. Dans G. Holtus, M. Metzeltin & C. Schmitt (Ed.), *Band I/2 Methodologie (Sprache in der Gesellschaft / Sprache und Klassifikation / Datensammlung und -verarbeitung)* (584–627). Berlin, New York: Max Niemeyer Verlag.
- Lahousse, K. (2022, 24 mai). *Les clivées françaises et leur acquisition : le rôle du contraste*, [présentation d'un conférencier invité]. Université Catholique de Leuven.
- Lambrecht, K. (2001a). Dislocation. Dans M. Haspelmath, E. König, W. Oesterreicher et W. Raible (eds), *Language Typology and Language Universals : An International Handbook* (Vol. 2., 1050–1078). Berlin, New York : Walter de Gruyter.
- Lambrecht, K. (2001b). A framework for the analysis of cleft constructions. *Linguistics*, 39(3), 463–516.
- Lambrecht, K. (2004). Un système pour l'analyse de la structure informationnelle des phrases. L'exemple des constructions clivées. Dans J. Fernandez-Vest, S. Shirley-Thomas (dir.), *Structure Informationnelle et Particules Énonciatives* (Grammaire & cognition 1-2). Paris Budapest Torino: L'Harmattan.
- Nølke, H., (1998). Il est beau le lavabo, il est laid le bidet, Pourquoi disloquer le sujet ?, Dans M. Forsgren, D. Jonasson, H. Kronning (éds), *Preedicatio, assertion, information, Actes du colloque d'Uppsala en linguistique française, 6-9 juin 1996*, *Studia Romanica Upsaliensia* 56, 385–393.
- Marchello-Nizia, C., Combettes, B., Prévost, S., & Scheer, T. (2020). *Grande Grammaire Historique du Français (GGHF)*, Berlin etc. : De Gruyter Mouton.
- Müller Blaser, G. (2007). La construction pseudo-clivée dans l'organisation d'activités complexes : questions de portée. *Cahiers de praxématique*, 48, 213–238.

- Perera, K. (1986). Grammatical differentiation between speech and writing in children aged 8-12. Dans: A. Wilkinson (éd), *The Writing of Writing* (90-108), OUP: Milton Keynes.
- Petrov, S., Das, D. et McDonald, R. (2012). A universal part-of-speech tagset. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, 2089–2096.
- Prince, E. F. (1978). A Comparison of Wh-Clefts and it-Clefts in Discourse. *Language*, 54(4), 883–906.
- Riegel, M., Pellat, J., et Rioul, R. (2009). *Grammaire méthodique du français* (4e édition entièrement revue "Quadrige" ed., Quadrige manuels). Paris : Presses universitaires de France.
- Riou, É., Hemforth, B. (2015). Dislocation clitique de l'objet à gauche en français écrit, *Discours*, 16.
- Roubaud, M-N. et Sabio, F. (2015). Les clivées en C'EST LÀ QUE, C'EST LÀ OÙ : structures et usages en français moderne. *Repères DoRiF, Recherches sur la syntaxe verbale en français et en italien. Hommage à Claire Blanche-Benveniste*, 6.
- Roubaud, M. N., et Sabio, F. (2018). C'est comme ça que j'ai perdu mon papa ! Les constructions en *c'est comme ça que* en français parlé et écrit. *SHS Web of Conferences*, 46.
- Rouget, C. et Salze, L. (1986). *C'est...qui, C'est...que*, le jeu des quatre familles. *Recherches sur le Français parlé*, 7, 117-139.
- Roulet, E. (2001). De la linguistique de la langue à l'analyse du discours. Dans: E. Roulet, L. Fillietaz & A. Grobet, *Un modèle et un instrument d'analyse de l'organisation du discours* (11-26), Berne: Peter Lang.
- Rouquier, M. (2018). Les constructions clivées. *Encyclopédie grammaticale du français*.
- Rouquier, M. (2020). *Constructions clivées*. Dans C. Marchello-Nizia, B. Combettes, S. Prévost & T. Scheer (dir.), *Grande Grammaire Historique du Français (GGHF)*, Boston: De Gruyter Mouton.
- Obenauer, H-G. (1981). Le principe des catégories vides et la syntaxe des interrogations complètes. *Langue française*, 52, 100-118.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J. et Manning, C. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *Association for Computational Linguistics (ACL) Proceedings of the System Demonstrations*, 101–108.
- Zeman, D. (2008). Reusable Tagset Conversion Using Tagset Drivers. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, 213–218.

Annexes

Annexe 1 – Extrait de l’annotation manuelle des sorties de TXM

ID	Classe	Niveau scolaire	Contexte à gauche	Pivot	Contexte à droite	Type	FS ²⁶	CS ²⁷	Nb de mots
CO-3e-2018-FSBJC6-D1-R4-V1_N	3e	COLLEGE	sortir quand le soleil se couche, il	y a une loi qui	a été votée, depuis un événement tragique. Il y a	il y a X qui	sujet	SN	2
CO-3e-2018-FSBJC6-D1-R6-V1_N	3e	COLLEGE	se retourna en entendant ce grand bruit et	c'est à ce moment - là qu'	il tombe nez à nez avec la meute de loups qui le	c'est X que	circonstant	SP	3
CO-3e-2018-FSBJC6-D1-R6-V1_N	3e	COLLEGE	disait que la nuit dans les bois il	y avait des loups qui	se déplaçaient en meute et chassaient tous ceux qui s'approchaient d'	il y a x qui	sujet	SN	2
CO-3e-2018-FSBJC6-D1-R9-V1_N	3e	COLLEGE	il n'y avait personne, mais il	y avait une personne au comptoir (qui	nous a finalement trouvés) qui nous expliqua que tous les clients	il y a x qui	sujet	SN	11
CO-3e-2018-VTAC305-D1-R1-V1_N	3e	COLLEGE	. Elle habitait cette maison depuis longtemps.	C'était là que	, avant leur décès, ses parents l'avaient élevée, et	c'est X que	circonstant	Adverbe	1
CO-3e-2018-VTAC305-D1-R1-V1_N	3e	COLLEGE	avaient élevée, et elle avait décidé que	ce serait là qu'	elle élèverait ses enfants, Viktor et Frank. Cette semaine était	c'est X que	circonstant	Adverbe	1
CO-3e-2018-VTAC305-D1-R2-V1_N	3e	COLLEGE	sortir qu'une seule fois ce mois,	c'est sûrement pour ça aussi que	le jardin n'est pas entretenu. continua Lucie ». Jules	c'est X que	circonstant	SP	4
CO-3e-2018-VTAC305-D1-R2-V1_N	3e	COLLEGE	cette maison depuis longtemps et depuis son décès	c'est Pauline sa fille qui	vit ici, elle reste dans la maison tout le temps,	c'est X qui	sujet	SN	3
CO-3e-2018-VTAC305-D1-R4-V1_N	3e	COLLEGE	nos bonbons. Dans le cas contraire,	c'est vous qui	nous donnerez les vôtres. Je me suis dit que cela en	c'est X qui	sujet	Pronom	1

²⁶ FS = Fonction syntaxique

²⁷ CS= Constituant syntaxique

Annexe 2 – Extrait du tableau quantitatif

ID copie	Nb_mots-formes	Nb_clivees	Freq_relative	Classe	Niveau scolaire
CO-3e-2016-VTAC305-D1-R1-V1_N	157	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R10-V1_N	181	1	0.0055248618784530384	7	2
CO-3e-2016-VTAC305-D1-R11-V1_N	156	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R12-V1_N	103	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R13-V1_N	176	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R14-V1_N	140	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R15-V1_N	221	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R16-V1_N	215	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R17-V1_N	263	1	0.0038022813688212928	7	2
CO-3e-2016-VTAC305-D1-R18-V1_N	455	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R19-V1_N	226	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R2-V1_N	140	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R20-V1_N	266	0	0.0	7	2
CO-3e-2016-VTAC305-D1-R21-V1_N	314	0	0.0	7	2

Déclaration sur l'honneur de non-plagiat

Je soussignée,

Nom, prénom : LIDEN, Maria
Régulièrement inscrite à l'Université de Toulouse II Jean Jaurès
N° étudiant : 22101323
Année universitaire : 2021-2022

certifie que le document joint à la présente déclaration est un travail original, que je n'ai ni recopié ni utilisé des idées ou des formulations tirées d'un ouvrage, article ou mémoire, en version imprimée ou électronique, sans mentionner précisément leur origine et que les citations intégrales sont signalées entre guillemets.

Fait à : Toulouse
Le : 13 juin 2022

Signature :

