

# Évolution des interrogatives partielles directes en français : le cas de *combien*

Christophe Benzitoun<sup>1,\*</sup>

<sup>1</sup>Université de Lorraine & CNRS ATILF

**Résumé.** Les interrogatives ont donné lieu à un grand nombre de travaux en français, tant en diachronie qu'en synchronie. Cependant, ces études ne sont pas toujours comparables à cause des données prises en compte ou des catégories retenues. C'est la raison pour laquelle nous nous proposons de relancer le travail de description grâce aux corpus désormais disponibles. Dans cet article, nous nous intéressons à l'interrogatif *combien* et nous montrons qu'il présente une grande diversité de fonctionnements eu égard aux types de données étudiées.

**Abstract. Evolution of direct interrogatives in French: the case of *combien*.** Interrogatives have been the subject of a great deal of work in French, both diachronically and synchronically. However, previous studies are not easy to compare because of data or categories used, which can be heterogeneous. This is why we propose to redo some descriptions thanks to corpora now available. In this article, we focus on *combien* and we show that it presents a great diversity of functioning with regard to the data studied.

## 1 Introduction

De nombreuses études ont été menées sur les interrogatives partielles en français depuis plus d'un siècle, tant en diachronie qu'en linguistique variationniste (Coveney, 2011). Les formes que peuvent prendre les interrogatives sont très diversifiées et il n'est pas évident d'expliquer les raisons de l'emploi d'un agencement plutôt qu'un autre. Ces travaux antérieurs ont toutefois montré une évolution des constructions interrogatives allant dans le sens d'une neutralisation progressive de l'inversion sujet-verbe et d'un déplacement du mot interrogatif de la position pré-verbale vers la position post-verbale (Foulet, 1921 ; Marchello-Nizia et al., 2020). Cependant, ces changements ne se produisent pas pour l'ensemble des mots interrogatifs. Par exemple, *pourquoi* et *comment* ont tendance à apparaître régulièrement en tête d'énoncé. Il existe donc des fonctionnements spécifiques qu'il serait utile de documenter avec précision. De plus, les changements ne se produisent pas à la même vitesse dans tous les types de texte. Pour toutes ces raisons, nous pensons qu'il est important de mener des études descriptives sur chaque interrogatif, en commençant

---

\* [Christophe.Benzitoun@univ-lorraine.fr](mailto:Christophe.Benzitoun@univ-lorraine.fr)

ici par *combien*, plutôt que de traiter l'ensemble des interrogatives de manière globale comme cela a souvent été fait.

Dans cette étude, nous nous limitons uniquement aux interrogatives partielles directes (*Combien ça coûte ?*), laissant de côté les interrogatives indirectes (*Je me demande combien ça coûte*). Ce que nous considérons comme des interrogatives, ce sont des constructions vérifiant certaines propriétés avant tout syntaxiques et sémantiques. Elles doivent comporter la proforme *combien* ainsi qu'une modalité interrogative. Par exemple, nous avons écarté les exemples du type suivant qui comportaient une valeur de haut degré (associée aux exclamatives) et non une modalité interrogative :

(1) **Combien** il avait dû se faire violence... (écrit, Frantext)

Formellement, cette tournure présente certaines contraintes de position. Par exemple, il n'est pas possible de déplacer l'interrogatif (Ex. 2) comme on pourrait le faire dans une interrogative (Ex. 3) :

(2) \*Il avait dû se faire violence **combien**...

(3) **Combien** tu me donnes ? (écrit, Frantext) / Tu me donnes **combien** ?

Dans un premier temps, nous ferons état d'études récentes sur corpus portant sur les interrogatives en français et nous détaillerons plus particulièrement le fonctionnement de *où* pour lequel nous disposons d'informations précises sur des données diversifiées. Dans un second temps, nous présenterons notre corpus d'étude ainsi que notre méthodologie avant de préciser nos résultats sur l'interrogatif *combien*. Pour finir, nous évoquerons les conséquences que l'on peut tirer de nos données concernant l'étude de l'évolution grammaticale de la langue française.

## 2 Quelques études antérieures sur corpus

La question de l'ordre des mots dans les interrogatives a été traitée dans de nombreuses études (Coveney, 2011). Mais c'est surtout à partir de la seconde moitié du XXe siècle que se sont développés les travaux systématiques sur corpus. Dans ce qui suit, nous donnons des indications pour l'ensemble des interrogatifs.

### 2.1 Panorama des interrogatives partielles directes

Dans sa synthèse de 2011 (version mise à jour en 2020), Coveney donne à voir des différences quantitatives importantes dans les choix formels des interrogatives partielles. Les interrogatives sont classées principalement en fonction de l'ordre des mots. Les observations de Coveney (2011) permettent de dresser un premier bilan : à l'oral, les structures sans inversion du sujet sont fréquemment utilisées dans des contextes spontanés alors que celles avec inversion se rencontrent majoritairement dans des situations formelles. Également, les fréquences d'emploi de certaines tournures semblent varier en fonction de la catégorie socioprofessionnelle du locuteur. Coveney a recensé des interrogatives avec complémentateur (*Combien qu'il gagne ?*) chez des ouvriers alors que celles-ci étaient totalement absentes chez les personnes appartenant à un milieu bourgeois. Mais tous les résultats présentés dans la synthèse de Coveney ne sont pas faciles à comparer, à cause de la classification des exemples, de l'origine sociale des locuteurs et des situations de parole : conversations, radio, télévision, etc.

Or, si l'on veut pouvoir observer l'évolution des tournures interrogatives en français, il est nécessaire de disposer de données comparables. Selon le médium considéré et la provenance des données sur lesquelles se sont basés les chercheurs, les résultats sont

sensiblement hétérogènes et les différences observées ne sont pas toujours possibles à interpréter. Même quand les données semblent appartenir à des situations proches, il arrive que l'on relève des divergences importantes.

Ci-dessous, nous croisons les résultats issus de trois études portant sur des données radiophoniques à deux époques différentes. Les catégories visibles dans le tableau correspondent aux interrogatives construites par antéposition de l'interrogatif sans inversion sujet-verbe (ANTE : *Où tu vas ?*), avec la particule *est-ce que* (ESK : *Où est-ce que tu vas ?*), avec inversion du sujet (INV : *Où vas-tu ?*) et en position postverbale (INSITU : *Tu vas où ?*).

**Tableau 1.** Répartition des interrogatives partielles dans des émissions de radio

<i>Année</i>	<i>ANTE</i>	<i>ESK</i>	<i>INV</i>	<i>INSITU</i>	<i>Nombre d'interrogatives</i>
1960-1970 (Behnstedt, 1973)	10%	3%	62%	25%	4367
2004 (Thiberge, 2020)	21,3%	11,4%	58,8%	8,5%	211
2005-2009 (Hamlaoui, 2009) <sup>1</sup>		32%		47,3%	222

Behnstedt (1973) traite des emplois dans des émissions de radio enregistrées dans les années 1960 jusqu'à début 1970, tandis que Thiberge (2020) et Hamlaoui (2009) se sont concentrés sur l'année 2004 pour l'un et 2005-2009 pour l'autre. Le contexte d'une émission de radio implique souvent que les participants surveillent leur langage et prêtent attention aux tournures qu'ils emploient voire lisent un texte. On remarque dans l'étude de Behnstedt ainsi que dans celle de Thiberge que l'emploi majoritaire reste l'inversion (autour de 60%) à 40 ans d'intervalle. En revanche, la catégorie qui vient ensuite est l'insitu pour le premier et l'antéposition de l'interrogatif sans inversion du sujet pour le second.

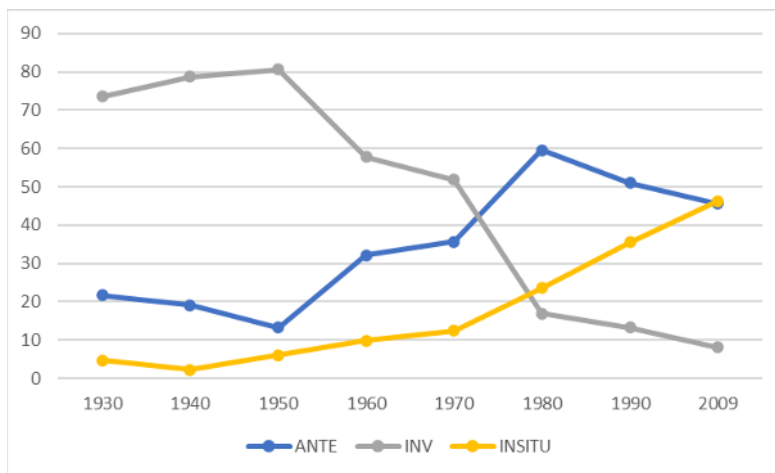
Dans l'article d'Hamlaoui (2009), on observe une forte progression de l'emploi de l'insitu qui passe à près de 50%. Pour les inversions, on ne sait pas si elles sont incluses dans les antépositions ou si l'autrice n'en a pas observé. En tout cas, leur fréquence est forcément nettement moins importante que celle relevée par Thiberge (2020). Ces différences de fréquence sont sans doute dues au degré de formalité. L'émission étudiée par Hamlaoui est *Là-bas si j'y suis* (France Inter) et plus particulièrement les interrogatives utilisées par l'une des reporters. Elle précise que cette reporter utilise un français parlé « non-standard » et « informel ». Thiberge, de son côté, a analysé des données issues de l'émission *Le téléphone sonne* sur France Inter, qui adopte un registre de langue plus soutenu.

Le fait de travailler à partir d'émissions de radio ne garantit donc pas d'avoir des données homogènes exhibant des productions linguistiques similaires. La ligne éditoriale des émissions ainsi que le degré de formalité des intervenants peuvent avoir une forte influence sur les résultats. La fréquence relativement faible (autour de 200), toutes interrogatives confondues, peut aussi engendrer des variations importantes. De même,

l'absence de catégorisations homogènes dans le classement des interrogatives rend les comparaisons difficiles.

Assiste-t-on à une forme de stabilité de l'inversion dans les émissions de radio, comme tendrait à le montrer la comparaison entre Behnstedt et Thiberge, accompagnée d'une augmentation de l'antéposition et une diminution de l'in-situ ? Doit-on, au contraire, conclure que l'in-situ progresse à 40 ans d'intervalle ? Ces travaux antérieurs ne permettent malheureusement pas de trancher. Il faut donc se tourner vers des publications ayant abordé l'évolution des interrogatives de manière unitaire. C'est le cas de la thèse de Farmer (2015).

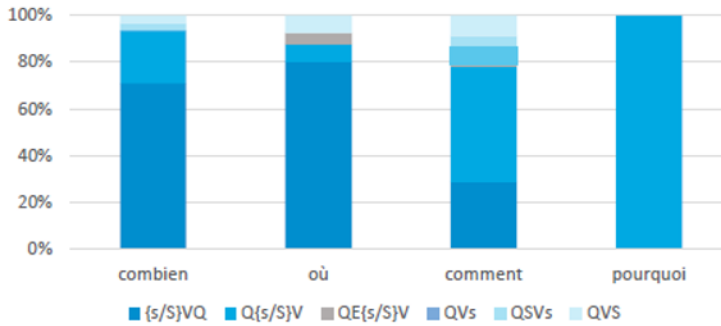
Dans la partie qui nous intéresse, Farmer (2015) se concentre sur les structures interrogatives partielles utilisées dans les dialogues de films français entre 1930 et 2009. Elle remarque une forte diminution de l'emploi de l'inversion au profit de l'antéposition et de l'in-situ. Alors que l'inversion était largement majoritaire par rapport à l'in-situ, les rôles se sont inversés depuis les années 1980. On voit également que l'in-situ est sans doute en train de prendre l'ascendant sur l'antéposition. On assisterait donc, au moins depuis les années 1950, à une évolution sensible de l'ordre des mots dans les interrogatives partielles (du moins dans les dialogues de films). Ses résultats sont basés sur 1231 occurrences d'interrogatives. La représentation graphique ci-dessous a été tracée par nos soins à partir des résultats présentés dans Farmer (2015). Nous en ferons de même pour les autres études dans la suite de cette partie (excepté pour le graphique de Reinhardt 2019, qui est extrait de l'étude originale).



**Fig. 1.** Les interrogatives partielles dans les dialogues de films [1930-2009]<sup>2</sup> (Farmer, 2015 : 387)

L'axe des ordonnées correspond aux pourcentages et l'axe des abscisses aux décennies. On voit clairement que le pourcentage d'inversion a été divisé par 4 en une trentaine d'années (1950-1980) pendant que celui de l'antéposition a été multiplié par près de 6 sur la même période. Quant au pourcentage de l'in-situ, il a été multiplié par 4 entre 1970 et le début des années 2000.

Cependant, il est important de tenir compte d'un paramètre, à savoir la fréquence respective de chaque interrogatif. Comme le montre le graphique ci-dessous extrait de Reinhardt (2019), il y a de grandes disparités de fonctionnement entre interrogatifs dans les émissions de télé-réalité. Alors que *où* et *combien* sont très majoritairement en position in-situ, *comment* est principalement antéposé et *pourquoi* ne connaît que la position initiale. On voit également que pour l'ensemble des interrogatifs, l'inversion du sujet est marginale, excepté pour *comment*.

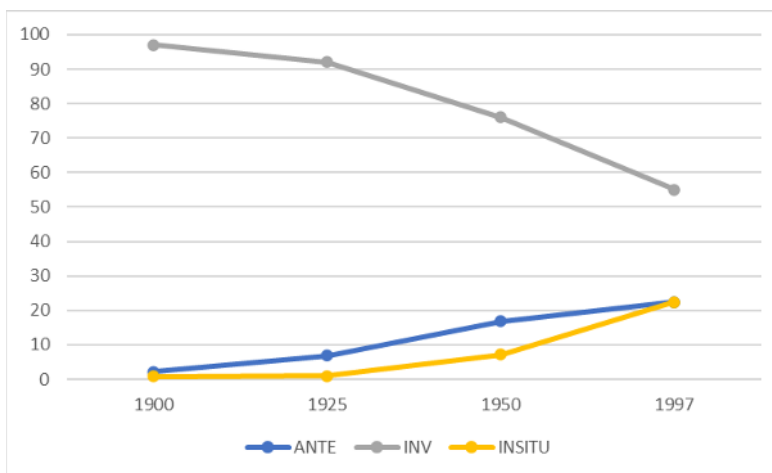


**Fig. 2.** Fonctionnement de 4 interrogatifs dans des émissions de télé-réalité [2016-2017] (Reinhardt, 2019 :159)<sup>3</sup>

On peut déduire de ce graphique que la fréquence respective des interrogatifs peut avoir une influence sur l'antéposition et la postposition. Par exemple, une plus grande proportion de *comment* et de *pourquoi* ferait mécaniquement augmenter le taux général d'antéposition. Cela illustre l'intérêt d'observer chaque interrogatif indépendamment pour mettre en évidence des dynamiques particulières, ce qu'ont déjà fait certains auteurs cités. Ci-dessous, nous prenons l'exemple de *où* en nous appuyant sur des travaux récents provenant de diverses sources de données afin d'en dessiner l'évolution.

## 2.2 Le cas de *où*

L'étude de Rossi-Gensane et al. (2021), portant sur 2731 occurrences, indique pour *où* une diminution de l'utilisation de l'inversion clitique et une légère et régulière augmentation de l'antéposition et de l'in-situ dans les dialogues de romans entre 1900 et 1997 (données extraites de Frantext). Dans les graphiques 3 et 4, les catégories des interrogatives sont moins détaillées que celles du graphique 2. Toutefois, les tendances sont tellement marquées que l'on peut conclure à une différence très nette entre les dialogues de romans en 1997 et la télé-réalité des années 2010.

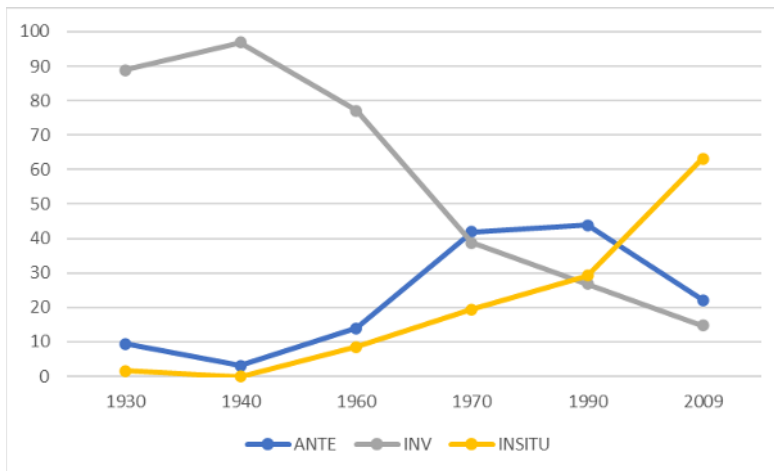


**Fig. 3.** Les interrogatives en *où* dans les dialogues de romans [1900-1997] (Rossi-Gensane et al., 2021)

En prolongeant le graphique précédent grâce à la partie de la thèse de Reinhardt (2019) portant sur les romans policiers publiés dans les années 2000, on observe que l'inversion se

maintient autour de 48% alors que l'in-situ augmente jusqu'à atteindre près de 40% et l'antéposition simple se situe autour de 10%. Il y a donc un léger repli de l'inversion du sujet et l'in-situ prend le pas sur l'antéposition dans les romans policiers des années 2000. À noter toutefois que ces derniers résultats prennent en compte l'intégralité du texte dans les romans policiers et pas juste les dialogues comme Rossi-Gensane et al. (2021). Une fois de plus, la comparabilité n'est pas parfaite.

Concernant *où*, Farmer (2015) observe des tendances similaires dans les dialogues de films, où l'inversion passe de près de 100% en 1940 à 15% en 2009. L'antéposition et l'in-situ, elles, progressent depuis les années 1940, et on remarque que l'augmentation de l'in-situ s'intensifie à partir des années 1990 passant de 30% cette décennie-là à plus de 60% en 2009. Ces résultats ont été obtenus à partir de 303 occurrences.



**Fig. 4.** Les interrogatives en *où* dans les dialogues de films [1930-2009] (Farmer, 2015)

En reprenant les exemples dans l'annexe de Thiberge (2020) émanant du corpus ESLO 2, l'in-situ est à près de 70% et l'inversion autour de 6% (seulement 3 occurrences dont 2 prononcées par un enseignant à l'école) et il n'y a qu'un seul exemple d'antéposition simple.

Des travaux récents permettent donc d'observer précisément l'évolution de *où* à partir de diverses sources de données et de montrer une convergence vers le même point, à savoir une augmentation de l'in-situ et une réduction de l'inversion du sujet. Mais cela ne se fait pas à la même vitesse dans tous les genres textuels. On voit que, dans la littérature par exemple, le changement est plus lent, alors qu'il est le plus avancé en français parlé de type conversationnel et dans les dialogues de films.

Cependant, la comparaison directe d'études distinctes n'est pas aisée étant donné que l'on dispose rarement d'informations concernant l'ensemble des paramètres. En effet, il existe de fortes disparités en fonction de la composition des corpus, des catégories distinguées (notamment pour *est-ce que/qui* où la quasi-totalité des exemples proviennent en réalité de *qu'est-ce que/qui* et qui sont parfois inclus parfois pas), du nombre d'occurrences traitées et des sous-ensembles à l'intérieur d'un même corpus (les parties narratives vs les discours directs dans les romans ; les différentes émissions de radio). Malgré la profusion d'études, il nous semble donc utile de reprendre le travail sur des bases homogènes.

Grâce à l'essor des corpus, de grandes quantités de données diversifiées sont désormais disponibles, aussi bien orales qu'écrites. Ces données nous offrent la possibilité de préciser les résultats obtenus dans les études antérieures, en contrastant différents registres et genres ainsi qu'en réalisant des études sur la grammaire de chaque interrogatif.

Nous allons maintenant présenter les corpus utilisés et la méthodologie que nous avons adoptée dans le cadre de notre propre travail sur les interrogatives. Comme nous l'avons précisé en introduction, nous allons nous intéresser uniquement aux interrogatives directes en *combien*.

## 3 Données et méthode

### 3.1 Présentation du corpus de travail

Nos données comportent à la fois des textes écrits et des transcriptions de français parlé. Elles proviennent de quatre sources : le corpus d'étude du français contemporain (CEFC), Frantext, les Enquêtes Sociolinguistiques à Orléans (ESLO) et l'Areneum Francogallicum Gallicum.

Le CEFC (Debaisieux & Benzitoun, 2020) est un corpus réalisé dans le cadre du projet ANR Orfeo (Outils et Recherches sur le Français Écrit et Oral). Il est le fruit de la récupération, la vérification et l'homogénéisation de données hétérogènes pour en faire un corpus d'étude. Il est composé des corpus suivants, qui ont été intégrés en totalité ou partiellement :

- Pour l'oral : CFPP2000, CLAPI, C-ORAL-ROM, CRFP, FLEURON, French Oral Narrative, OFROM, TCOF, TUFS et VALIBEL ;
- Pour l'écrit ; ANNODIS, Chambers-Rostand, CoMeRe, FRANTEXT, Est Républicain.

Le corpus est librement accessible sur Ortolang ou en passant par l'interface de recherche à l'adresse : <https://orfeo.ortolang.fr/>. Nous avons, pour cette étude, séparé les données en deux sous-corpus : un sous-corpus écrit et un sous-corpus oral. La partie écrite comporte des textes scientifiques, des écrits électroniques, de la littérature et de la presse. Elle fait environ 7 millions de mots (hors ponctuation). La partie orale, quant à elle, contient 3,5 millions de mots, environ 2500 locuteurs et des situations de parole variées.

Frantext (<https://www.frantext.fr/>) est une ressource textuelle comportant 5532 ouvrages, soit 262 millions de mots (en août 2021), et composée principalement d'œuvres littéraires. Les textes qui la composent s'étalent sur près d'un millier d'années. Nous n'avons ici retenu que les romans datant de deux périodes éloignées d'un siècle d'intervalle (1890-1910 et 1990-2010) afin d'observer l'évolution des interrogatives en *combien*. Ces deux sous-corpus comprennent respectivement 74 textes pour 4,5 millions de mots et 124 textes pour 9,9 millions de mots.

La troisième ressource utilisée dans le cadre de notre étude est le corpus ESLO. C'est le fruit d'un projet du Laboratoire Ligérien de Linguistique (UMR7270) à l'université d'Orléans. Ce corpus est composé d'enregistrements sonores recueillis entre 1968 et 1974 (ESLO1) d'une part, et à partir de 2008 (ESLO2) d'autre part (<http://eslo.huma-num.fr/>). L'ensemble fait environ 5 millions de mots. Nous n'avons pas travaillé sur la totalité d'ESLO, mais seulement sur une sous-partie nommée ESLO-MD (MD pour Micro-Diachronie) comportant des entretiens, des conférences et des repas. ESLO-MD fait un peu moins d'un million de mots et se répartit de la façon suivante : 450.000 mots provenant d'ESLO1 et 520.000 d'ESLO2 (Abouda & Skrovec, 2018). Il est quantitativement et qualitativement équilibré entre les deux périodes contenues dans ESLO, ce qui permet d'envisager des études sur la diachronie récente.

Enfin, la dernière ressource est l'Areneum Francogallicum Gallicum (AFG). Il s'agit d'un corpus provenant de textes du domaine .fr aspirés sur Internet entre 2013 et 2019. Nous nous sommes basés sur la partie la plus restreinte qui fait 100 millions de mots, sachant que le corpus complet fait près de 3 milliards de mots juste pour la France et 10 milliards pour la francophonie (Benko, 2014).

### 3.2 Extraction des exemples

Pour mener à bien nos analyses sur le CEFC et le corpus ESLO-MD, nous avons utilisé le logiciel de textométrie TXM (Heiden et al., 2010). Celui-ci nous a permis de générer les concordances en conservant les métadonnées que nous jugeons pertinentes pour l'exploitation ultérieure. Concernant la requête, nous avons simplement recherché l'ensemble des occurrences de la forme *combien* puis nous avons exporté les concordances vers Excel pour leur annotation manuelle avec les différentes catégories décrites ci-dessous.

La version du CEFC dont nous disposons pour l'exploitation par TXM ne permet pas d'afficher les identifiants des locuteurs directement dans les concordances afin de savoir où commence et où se termine un tour de parole (pour la partie orale). Ainsi, il a fallu vérifier manuellement dans le texte intégral une partie des exemples pour en connaître la délimitation exacte.

Pour Frantext, nous avons utilisé l'interface en ligne du moteur Allegro. Pour l'AFG, nous avons eu recours à l'instance de NoSketch Engine proposée à l'adresse : <http://unesco.uniba.sk/>. Concernant l'AFG, pour éviter de traiter un trop grand nombre d'exemples, nous avons effectué un tri aléatoire puis réduit à 500 occurrences.

Durant la phase d'annotation manuelle, nous avons repéré les interrogatives indirectes ainsi que les emplois à valeur de haut degré (voir ci-dessous) et les avons écartés de nos résultats finaux. Ce travail a été réalisé de manière collective dans le cadre d'un cours de Master 2 de sciences du langage<sup>4</sup>. De ce fait, les exemples ont été traités et annotés par différentes personnes et ont tous été vérifiés par l'enseignant.

La typologie retenue pour classer les interrogatives est présentée dans la partie suivante.

### 3.3 Les différentes catégories distinguées

Parmi les différents ordonnancements que nous avons rencontrés, nous avons choisi d'en distinguer uniquement 5 pour les raisons que nous explicitons plus loin. La première catégorie contient les antépositions simples (noté ANTE) dans lesquelles l'interrogatif est placé en tête d'énoncé et non suivi d'une inversion sujet-verbe.

- (4) **Combien** vous en avez en tout là ? (écrit, Frantext)  
 (5) **à combien** tu dis qu'il faut passer les vitesses (oral, ESLO)

Il peut arriver que la particule interrogative *est-ce que* suive l'interrogatif.

- (6) **combien de lettres est-ce que** vous écrivez en moyenne (oral, ESLO)

Mais étant donné qu'il y avait fort peu d'exemples de *combien est-ce que*, nous les avons reversés dans une catégorie nommée AUTRES.

Nous avons retenu la classe des interrogatives in-situ (noté INSITU). Il s'agit d'exemples où l'interrogatif se trouve après le verbe.

- (7) ton quartier tu le définirais **comment** (oral, CEFC)  
 (8) vous avez **combien d'enfants** (oral, ESLO)

Nous avons identifié trois types d'inversion sujet-verbe, parmi lesquelles se trouvent les inversions du pronom clitique sujet (noté INVCLIT).

- (9) **Combien d'année** allait-elle passer ainsi, à attendre la vieillesse ? (écrit, Frantext)  
 (10) **combien** en possédez-vous environ (oral, ESLO)



Les inversions complexes, quant à elles, fonctionnent de la même manière que les inversions clitiques mais elles comportent en plus le sujet lexical entre l'interrogatif et le verbe.

(11) **Combien de temps le Marquis espérait-il pouvoir se « raccrocher » en sûreté ?** (écrit, Frantext)

Cela étant, les cas d'inversion complexe étant rares dans les données que nous avons étudiées et très proches syntaxiquement des inversions clitiques, nous avons décidé de les regrouper avec ces dernières.

Le troisième type d'inversion est représenté par les inversions stylistiques du sujet (ou nominales), constructions dans lesquelles le sujet, placé après le verbe, est un groupe nominal et pas un pronom clitique sujet (noté INVNOM).

(12) **Combien font deux et deux ?** (écrit, Frantext)

(13) **Combien de temps dure ce genre d'état ?** (écrit, Frantext)

Pour ce qui concerne les occurrences que nous avons écartées, on peut mentionner les *combien* à valeur de haut degré, déjà évoqués plus haut.

(14) **Tâche ô combien ingrate, mais socialement des plus utiles.** (écrit, Frantext)

Nous avons également mis de côté les interrogatives inachevées et n'avons conservé qu'une seule occurrence dans les cas où il y avait une répétition de l'interrogatif. Les interrogatives indirectes ont subi le même sort.

(15) **tu vois plus à combien tu roules** (oral, ESLO)

de même que les occurrences ne dépendant pas d'un verbe.

(16) **combien d'articles d'OGM de combien d'encyclopédies papier ou d'ouvrages de référence utilisent cet exemple dans un texte de synthèse de quelques pages ?** (écrit, CEFC)

Par ailleurs, nous ne prenons pas en compte les mots interrogatifs dont la position dans la phrase ne varie pas, puisque c'est précisément cette mobilité du mot interrogatif que nous étudions. C'est le cas des interrogatifs sujets (Ex. 17), mais aussi des interrogatifs « seuls » (Ex. 18), qui ne sont pas accolés à une structure verbale.

(17) **Combien de mineurs sont restés au fond ?** (écrit, Frantext)

(18) **alors combien d'heures** (oral, ESLO)

Enfin, nous avons rencontré des interrogatives avec le complémenteur *que* après l'interrogatif *combien*. Ces formes étant très rares, elles seront classées dans la catégorie AUTRES avec les exemples en *combien est-ce que*.

(19) **combien qu'il nous avait dit qu'il se faisait** (oral, ESLO)

De même, nous avons classé dans la catégorie AUTRES les clivées.

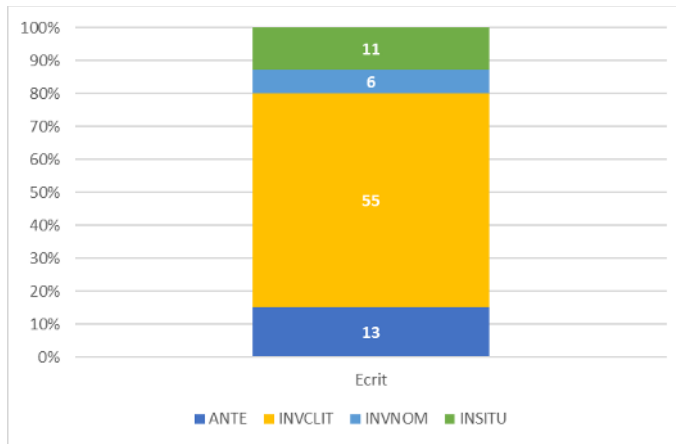
(20) **c'est dans combien d'années qu'on aura la soupe** (oral, CEFC)

En résumé, nous avons annoté notre corpus avec les catégories suivantes :

- Antéposition de *combien* sans inversion du sujet (ANTE)
- Inversion du sujet clitique et inversion complexe (INVCLIT)
- Inversion stylistique ou nominale du sujet (INVNOM)
- Postposition de *combien* après le verbe (INSITU)
- *Combien* suivi de *est-ce que*, de *que* ou dans une clivée (AUTRES)

## 4 Résultats

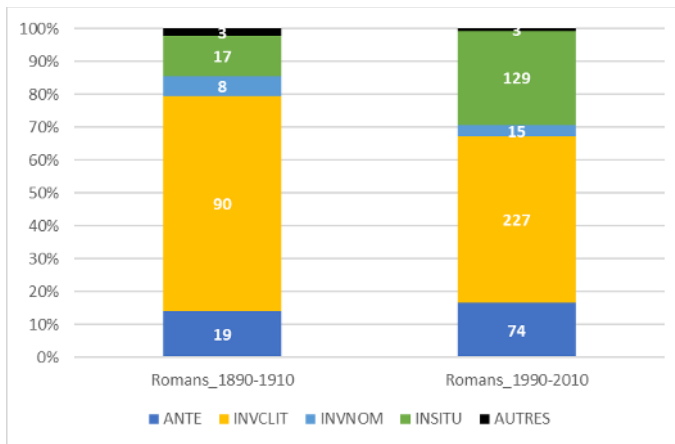
Nous allons commencer la présentation de nos résultats par la partie écrite du CEFC (les nombres au centre du graphique correspondent au nombre d'occurrences de *combien* dans chaque catégorie).



**Fig. 5.** Les interrogatives en *combien* dans le CEFC (partie écrite)

On voit que plus de 60% des occurrences sont composées de formes interrogatives avec inversion clitique du sujet. Nous précisons que l'inversion complexe représente 15 occurrences sur les 55 exemples classés dans cette catégorie. Mais ce résultat donne une image déformée de l'usage de *combien* en français contemporain écrit. En effet, dans le CEFC, les textes littéraires, provenant tous de la base Frantext, s'échelonnent de 1890 à 1921 pour des questions de droits d'auteur. Or, 50 occurrences sur les 85 émanent de cette partie du corpus. Nous avons donc dû recourir à d'autres sources de données pour avoir une image plus fidèle du français contemporain écrit. Nous en profitons pour étudier l'évolution des interrogatives en *combien* dans la littérature à un siècle d'intervalle.

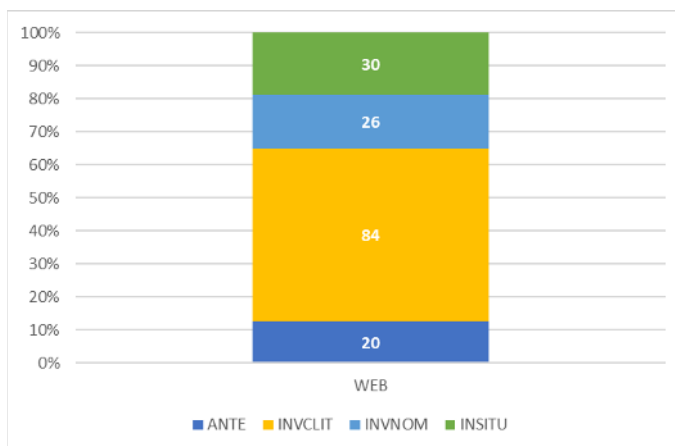
Dans le graphique ci-dessous obtenu à partir de la base Frantext, nous comparons les romans à deux époques. On voit que les romans durant la période 1890-1910 sont très proches des résultats trouvés dans le CEFC (écrit). On observe une multiplication par deux de la part des in-situ dans les résultats et une régression de l'inversion clitique à l'époque contemporaine. Les exemples en *combien que* et *combien est-ce que* sont extrêmement peu fréquents. Il n'y en a que 6 en tout.



**Fig. 6.** Les interrogatives en *combien* dans Frantext (limitation aux romans 1890-1910 / 1990-2010)

Dans le sous-corpus correspondant aux « écrits non planifiés » (sms, forums, tweets et wiki issus du corpus CoMeRe, un sous-corpus du CEFC), il y a 3 occurrences d'antéposition simple, 9 occurrences d'in-situ, 9 occurrences d'inversion clitique et 1 occurrence d'inversion nominale. Il y a donc une proportion équivalente entre l'inversion et l'in-situ. Dans Guryev (2017), les résultats sont encore plus spectaculaires (quoique basés sur un très faible nombre d'occurrences) : 7 in-situ pour 1 inversion dans un corpus de SMS suisses.

Dans l'AFG, en revanche, les résultats sont très différents de ceux relevés dans les autres écrits électroniques (voir paragraphe précédent).



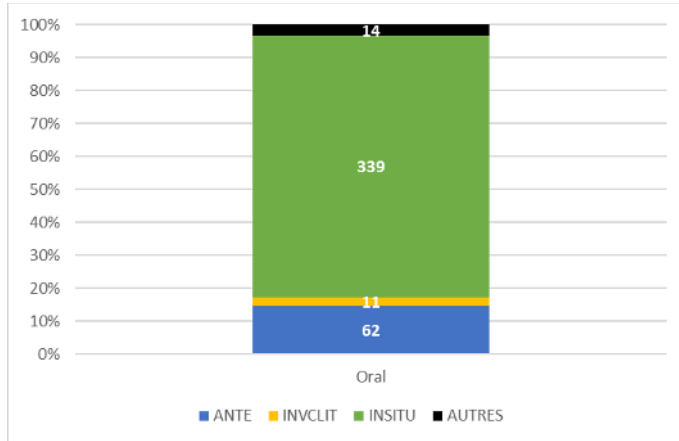
**Fig. 7.** Les interrogatives en *combien* dans l'AFG (partie Araneum Gallicum Minus)

De façon assez inattendue à nos yeux, ces données donnent à voir une langue un peu plus soutenue que les romans contemporains, du moins pour ce qui est des interrogatives en *combien*. Il y a moins d'in-situ, autant d'inversions clitiques et plus d'inversions nominales du sujet. Mais la composition précise d'une ressource issue du web est difficilement contrôlable et peut avoir une incidence importante sur les résultats.

En tout cas, on voit que les résultats fluctuent d'un genre textuel à un autre, avec une tendance marquée à la réduction de l'inversion et à l'augmentation de l'in-situ dans les romans, une répartition proche des romans contemporains dans les écrits électroniques de l'AFG et un recours plus important à l'in-situ sur les forums de discussion ou dans les

tweets. Toutefois, l'inversion sujet-verbe semble se maintenir à un niveau élevé dans la plupart des productions écrites. Mais les données dont nous disposons pour la presse et les textes scientifiques ne comportaient pas assez d'exemples pour étudier *combien*. Un travail complémentaire sera nécessaire au moins pour ces deux genres.

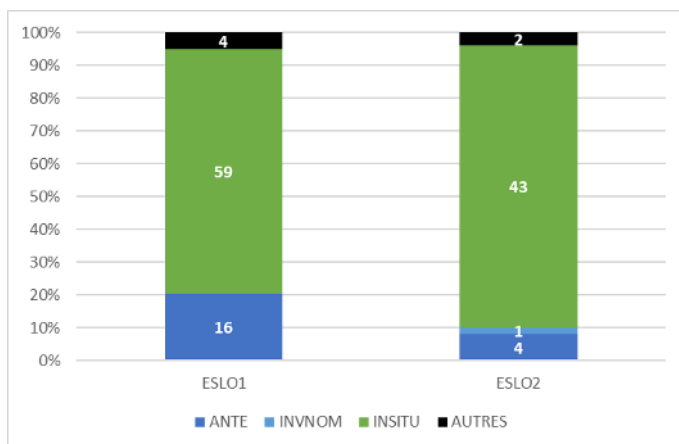
À présent, nous allons exposer nos résultats issus de la partie orale du CEFC.



**Fig. 8.** Les interrogatives en *combien* dans le CEFC (partie orale)

Dans ce sous-corpus, une majorité écrasante (80%) est en position in-situ et les inversions sont quasi-inexistantes (11 en tout). 7 sont produites en Belgique, et 8 sont utilisées dans un contexte professionnel. De plus, il n'y a aucune occurrence d'inversion complexe. Et il y a seulement 3 occurrences de *combien est-ce que*, 4 de clivées et 7 de *combien que*. Ces dernières ont toutes été produites en Lorraine, ce qui illustre peut-être une variante régionale (mais nous verrons plus loin qu'on en trouve aussi dans d'autres régions). Globalement, on retrouve une répartition proche de celle observée dans les émissions de télé-réalité (voir Figure 2 ci-dessus).

Pour avoir une idée de l'évolution du fonctionnement des interrogatives en *combien* en français parlé, nous présentons maintenant les résultats obtenus dans le corpus ESLO-MD. Pour rappel, ESLO1 contient des enregistrements faits dans les années 1970 et ESLO2 à partir de 2008.



**Fig. 9.** Les interrogatives en *combien* dans le corpus ESLO-MD

Dans ESLO2, on observe une répartition proche de celle constatée dans le CEFC (partie orale). Il y a une augmentation assez sensible de l'in-situ entre ESLO1 et ESLO2, qui s'explique principalement par une diminution de l'antéposition. À noter que dans ESLO1, il y avait un nombre important d'inversions clitiques (31 occ.) et de renforcement par *est-ce que* (20 occ.). Cependant, nous les avons supprimés car cela émanait de la lecture (ou de la mémorisation) des questionnaires liés à l'enquête. Sans cette suppression, on aurait pu avoir l'illusion d'une disparition de l'inversion et de *combien est-ce que* à quelques décennies d'intervalle seulement. On peut toutefois en conclure que la forme des questions posées par les enquêteurs sont différentes entre ESLO1 et ESLO2. Pour finir, les 6 occurrences notées « AUTRES » correspondent à la forme *combien que*, ce qui démontre que cette variante ne se trouve pas qu'en Lorraine.

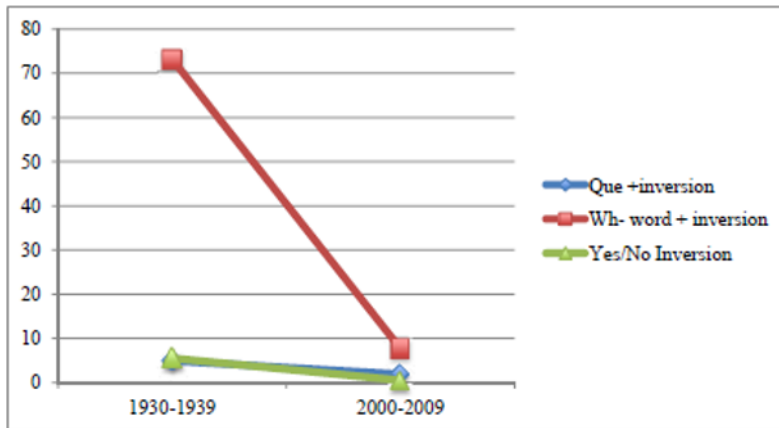
## 5 Conclusion

Dans ce travail, nous avons mis en évidence des écarts importants, en français contemporain, en ce qui concerne le fonctionnement des interrogatives en *combien* en fonction des genres textuels. À l'écrit, l'inversion sujet-verbe se trouve à un niveau élevé, généralement majoritaire, mais la quantité de données dont nous disposons ne nous a pas permis une étude approfondie en dehors des textes littéraires et de ceux issus d'internet. Et en nous appuyant sur des données diachroniques issues de romans, nous avons montré, à un siècle d'écart, que la tendance était à la diminution de l'inversion sujet-verbe.

Du côté du français parlé, cette même tournure est presque totalement absente. D'un point de vue diachronique, nous avons relevé quelques exemples dans les années 1970, mais tous étaient des questions posées par les enquêteurs et vraisemblablement préparées ou lues. Pour ce qui est des inversions sujet-verbe produites à l'oral spontané, on peut donc faire l'hypothèse qu'elles étaient déjà très peu fréquentes dans les années 1970.

En définitive, il y aurait une tendance générale en français à la neutralisation de l'inversion sujet-verbe dans les interrogatives. Cette évolution se produirait à des vitesses différentes en fonction des genres et du médium. Ainsi, certains genres seraient plus conservateurs que d'autres mais suivraient, malgré tout, les mêmes tendances. Ce phénomène a très bien été montré par McLaughlin (2021) dans son étude historique sur la presse française, en prenant l'exemple, entre autres, de la montée du clitique et des temps verbaux. Dans la littérature et la presse, des changements linguistiques ont eu lieu mais à plusieurs décennies d'écart.

Pour *combien* comme pour *où*, on a ainsi montré qu'il existait un parcours allant vers la réduction des structures à inversion du pronom clitique sujet. Cette tendance semble généralisable à l'ensemble des interrogatives, comme l'illustre le graphique ci-dessous extrait de Farmer (2015). Ce graphique a été obtenu en étudiant des dialogues de films. Il représente l'évolution du pourcentage d'emploi des structures à inversion du sujet par rapport aux autres agencements possibles pour les interrogatives totales, les interrogatives en *que/quoi* et les interrogatives partielles. On voit que les interrogatives partielles rejoignent, sur ce chemin, les interrogatives totales et les interrogatives partielles en *que/quoi*, à plusieurs décennies d'écart.



**Fig. 10.** L'évolution des interrogatives avec inversion du sujet dans des dialogues de films (Farmer, 2015 : 467)

Cela montre qu'il existe des trajectoires grammaticales pour lesquelles l'alignement se fait à des vitesses différentes en fonction des items impliqués.

Du côté de l'in-situ, nous avons observé une augmentation dans les romans et à l'oral. En français parlé contemporain, la fréquence de *combien* en position in-situ est de l'ordre de 80%. Dans les romans, elle se situe autour de 30%. Et il y a une quasi-absence généralisée du recours à *est-ce que*.

Une fois ce panorama dessiné, il nous reste encore à tenter d'expliquer pourquoi l'inversion du sujet se maintient à un haut niveau dans certains genres mais pas en français parlé spontané. Comment rendre compte de ce phénomène ? Nous faisons l'hypothèse que l'écrit est plus perméable aux discours normatifs omniprésents dans la société française. Par exemple, dans la 9<sup>e</sup> édition du Dictionnaire de l'Académie française, on peut lire :

*Une proposition interrogative directe exige l'inversion du sujet et du verbe.*  
(Entrée *Interrogatif*, 9<sup>e</sup> édition du Dictionnaire)<sup>5</sup>

On retrouve une conception proche dans les manuels scolaires, où l'inversion est associée à l'écrit et au langage soutenu :

***L'interrogation à l'oral et en langage familier se traduit souvent par la seule intonation*** : Tu viens demain ? *Le langage courant introduit est-ce que au début de la question. Le langage soutenu et l'écrit nécessitent l'inversion du sujet.*  
(*Le Jardin des Lettres 5<sup>e</sup>*, Magnard)

Ainsi, tous les élèves de France apprennent que les interrogatives doivent se construire avec une inversion du sujet. Et on peut constater, en écoutant des stations de radio comme *France Inter* ou *France Culture*, les efforts que font les journalistes pour respecter cette convention. C'est une des marques de langage soutenu et quiconque s'en écarte sur les antennes d'une radio publique s'expose à des remarques du type :

*Amis journalistes, généralement la forme correcte de l'interrogation se fait avec l'inversion du sujet ? C'est un peu déprimant de vous entendre dire "comment vous l'entendez ?" au lieu de "comment l'entendez-vous ?"*  
(message du 9/3/2022 lu sur le site de la médiatrice de Radio France)

Cette attitude puriste est parfaitement illustrée par la planche de bande-dessinée ci-dessous extraite d'un volume de Gaston Lagaffe des années 1970.



On remarque, dans les corpus, que la plupart des locuteurs parviennent, au moins à l'écrit, à faire l'effort d'utiliser les formes correctes dans des situations où l'on se doit de surveiller son langage, même quand cela contrevient à une tendance forte des usages grammaticaux du français contemporain. Le poids de la norme impose l'emploi d'un grand nombre de formes qui s'intègrent de manière plus ou moins harmonieuse à la grammaire première (Blanche-Benveniste, 1990). Cette intégration, à l'oral, ne semble pas évidente dans la grammaire de nombreux locuteurs sans doute à cause d'un changement linguistique bien avancé (du moins en français européen). L'inversion du sujet clitique dans les interrogatives tend de plus en plus à être un archaïsme, comme l'absence de montée du clitique (*je le veux voir*) ou l'usage exclusif du *ne* de négation (*je ne sais*).

En définitive, il existe deux forces qui tirent en sens inverse en français contemporain. L'une pousse vers l'extension aux interrogatives de l'ordre SVO et l'autre vers la conservation d'un ordre distinct entre interrogatives et déclaratives. Les corpus diversifiés en genre montrent clairement cette tension à l'œuvre en exhibant différents stades. Certaines productions écrites formelles recourent plus volontiers aux formes normatives traditionnelles alors que l'oral conversationnel, possédant un temps de planification réduit, est plus enclin à l'emploi de tournures bien intégrées au système. Ainsi, l'emploi des interrogatives à inversion du sujet clitique n'est sans doute pas automatisé pour la plupart des locuteurs contemporains. Cela nécessite un calcul supplémentaire et une forme de conscientisation rendue possible par le support écrit mais complexe à systématiser à l'oral.

À une même époque, il existerait donc des connaissances grammaticales ayant des statuts différents : certaines resteraient en usage principalement du fait de la pression normative (grammaire seconde) et d'autres seraient acquises plus tôt dans le développement langagier, sans passer par un apprentissage explicite (grammaire première). Dans nos futurs travaux, nous verrons si l'on peut généraliser cette hypothèse à l'ensemble des interrogatifs et mettre en œuvre un protocole permettant de vérifier cette hypothèse.

## Références bibliographiques

- Abouda, L. & Skrovec, M. (2018). Pour une micro-diachronie de l'oral : le corpus ESLO-MD. Congrès Mondial de Linguistique Française.
- Benko, V. (2014). Aranea: Yet Another Family of (Comparable) Web Corpora. In Petr Sojka, Aleš Horák, Ivan Kopeček and Karel Pala (Eds.): *Text, Speech and Dialogue. 17th International Conference, Brno, Czech Republic, September 8-12, 2014. Proceedings. LNCS 8655*. Springer International Publishing Switzerland, 2014. pp. 257-264.
- Behnstedt, P. (1973). Viens-tu? Est-ce que tu viens? Tu viens? Formen und Strukturen des direkten Fragesatzes im Französischen. *Tübinger Beiträge zur Linguistik*.

- Blanche-Benveniste, C. (1990). Grammaire première et grammaire seconde : l'exemple de *en*. *Recherches sur le français parlé*, 10, pp. 51-73.
- Debaisieux, J.-M. & Benzitoun, C. (dir.) (2020). *Orféo : un corpus et une plateforme pour l'étude du français contemporain*. *Langages*, n°219, Armand Colin.
- Coveney, A. (2011). L'interrogation directe. *Encyclopédie Grammaticale du Français*. [version mise à jour en 2020]
- Farmer, K. L. (2015). *Sociopragmatic variation in yes/no and wh-interrogatives in hexagonal French : A real time study of French films from 1930 to 2009*. Phd Thesis. Indiana University.
- Foulet, L. (1921). Comment ont évolué les formes de l'interrogation. *Romania*, tome 47 n°186-187, pp. 243-348.
- Guryev, A. (2017). *Les formes des interrogatives dans le Corpus suisse de SMS en français. Étude multidimensionnelle*. Thèse de doctorat, Universités de Neuchâtel et Paris 3 Sorbonne nouvelle.
- Hamlaoui, F. (2009). A prosodic study of wh-questions in French natural discourse. *LangUE Proceedings*.
- Heiden Serge, Magué Jean-Philippe, & Pincemin Bénédicte. (2010). TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement. In *JADT 2010 : 10th International Conference on the Statistical Analysis of Textual Data*. Rome, Italie.
- Marchello-Nizia, C., Combettes, B., Prévost, S. & Scheer, T. (2020). *Grande Grammaire Historique du Français (GGHF)*. Mouton De Gruyter (2 volumes).
- McLaughlin, M. (2021). *La presse française historique Histoire d'un genre et histoire de la langue*. Classiques Garnier.
- Reinhardt, J. (2019). *Regularity and variation in french direct interrogatives: The morphosyntax and intonation of question forms in reality TV shows, audio, books and teaching materials*. Phd thesis. Université de Bielefeld.
- Rossi-Gensane, N., Ursi, B., Acosta Cárdena, LF. (2021). Les structures interrogatives directes partielles fondées sur où dans les dialogues de romans français du XXe siècle. *Journal of French Languages Studies*, 31(2), pp. 1-23.
- Thiberge, G. (2020). *Structure syntaxique des interrogatives entre optionnalité linguistique et contraintes cognitives*. Thèse de doctorat. Université Paris 7 Diderot.

<sup>1</sup> Hamlaoui (2009) ne distingue pas différents types d'antéposition de l'interrogatif et elle ajoute une catégorie « other » qui regroupe tous les cas qui ne sont ni des antépositions, ni des postpositions et qui compte pour 20,7% du total.

<sup>2</sup> Dans le graphique, sont exclues les interrogatives en *que/quoi*.

<sup>3</sup> {s/S}VQ correspond à l'in-situ avec un sujet pronominal (s) ou nominal (S). Q{s/S}V est l'antéposition simple, QE{s/S}V les structures avec *est-ce que*, QV{s/S} l'inversion et enfin QSVs l'inversion complexe (par exemple *Comment Max fait-il ?*).

<sup>4</sup> Je tiens à remercier les étudiantes et l'étudiant du Master 2 Sciences du langage de l'université de Lorraine à Nancy.

<sup>5</sup> C'est nous qui soulignons.