

Les mots du discours dans les textes légaux et les chroniques. Parcours de l'ancien français au français moderne

Mathieu GOUX

Université de Caen Normandie

mathieu.goux@unicaen.fr

<https://orcid.org/0000-0003-4211-8309>

Pierre LARRIVÉE

Université de Caen Normandie

pierre.larrivee@unicaen.fr

<https://orcid.org/0000-0001-8447-7102>

Resumen

Nuestra contribución explora el uso de marcadores del discurso desde el francés antiguo hasta el francés moderno. El estudio se basa en corpus dotados de textos jurídicos (MICLE) y textos históricos (CHRONIQUES) elaborados en la Universidad de Caen. Mostramos que la representación diacrónica de los marcadores del discurso está relacionada con las propiedades textuales de estos dos géneros textuales, pero también con el espacio que se concede al discurso oral: los textos jurídicos aumentan este espacio, mientras que las crónicas lo disminuyen. También analizamos las estrategias para recuperar ocurrencias y mostramos que los analizadores sintácticos contemporáneos son capaces de identificar automáticamente los marcadores del discurso si los textos han sido transcritos con precisión, pero que la clasificación manual de los segmentos del discurso directo mejora significativamente la calidad de la identificación de los marcadores.

Palabras clave: marcadores del discurso, francés diacrónico, textos jurídicos, lingüística de corpus, herramientas de corpus

Résumé

Notre contribution explore l'emploi des marqueurs discursifs de l'ancien français au français moderne. L'étude s'appuie sur des corpus outillés de textes juridiques (MICLE) et historiques (CHRONIQUES) développés à l'université de Caen. Nous montrons que la représentation diachronique des marqueurs discursifs est en lien avec les propriétés textuelles de ces deux genres, mais aussi avec la place faite à l'oral : les textes juridiques l'augmentent, tandis que les chroniques la diminuent. Nous discutons également des stratégies de requêtes des occurrences et montrons que les analyseurs syntaxiques contemporains sont capables de repérer automatiquement les marqueurs discursifs si les textes ont été précisément transcrits, mais qu'un typage manuel des segments au discours direct améliore notablement la qualité des relevés.

Mots-clés : marqueurs discursifs, diachronie du français, textes juridiques, linguistique de corpus, corpus outillés

* Artículo recibido el 21/02/2025, aceptado el 10/08/2025.

Abstract

This paper explores the use of discourse markers from Old French to Modern French. The study is based on two corpora developed at the University of Caen (France): one composed of legal texts (MICLE), the other of historical texts (CHRONIQUES). We show that the diachronic representation of discourse markers is correlated to the textual properties of these two genres, but also to the place that they give to orality: as legal texts increased it, chronicles decrease it. We also discuss strategies for retrieving occurrences, and we conclude that contemporary syntactic parsers are capable of automatically identifying discourse markers, but only if the texts have been accurately transcribed. We also remark that manual typing of direct speech segments significantly improves the quality of the results.

Keywords: discourse markers, French diachrony, legal texts, corpus linguistics, corpus tools

1. Introduction

Notre contribution s'intéresse à la question des « mots du discours », ou des marqueurs discursifs, dans les textes légaux et historiques, sur une diachronie allant de l'ancienne langue française jusqu'à la période moderne. Cet ensemble linguistique, notamment rencontré dans le cadre des interactions orales spontanées, a fait l'objet d'un grand nombre d'études ces dernières années (Dostie, 2004 ; Fisher, 2006 ; Lefeuve et Dostie, 2017 ; entre autres), mais elle pose encore des questions de délimitation et de définition. Ces questions posent des difficultés d'autant plus grandes en diachronie, où les textes qui attestent de réels échanges de proximité sont rares. C'est pourquoi la recherche s'est davantage attachée à l'oral représenté (Rodríguez Somolinos, 2011 ; Glikman et Mazziotta, 2013 ; Guillot *et al.*, 2014 ; Lefeuve et Parussa, 2020 ; Dufter et Zepp-Zwirner, 2025), dont la fidélité aux pratiques réelles est difficile à évaluer sans point de comparaison avec des textes non-fictionnels, susceptibles d'être plus proches de la réalité discursive du temps.

Nous désirons participer à cette discussion en apportant l'éclairage que peuvent nous fournir les textes légaux et les chroniques de l'histoire du français quant à cette problématique. Le choix de deux genres différenciés se justifie en cela qu'ils pourraient documenter les mots du discours à des degrés différents, dans des contextes contrastés ; voire des formes diverses. Bien que ces textes transcrivent d'une façon particulière ces échanges oraux, et même si ces échanges subissent des transformations liées aux enjeux stylistiques ou culturels de ces genres textuels, ou encore des modifications dépendantes directement de la mise en texte de la parole de l'autre (Dourdy et Spacagno, 2020), l'origine historiquement située des textes juridiques et des chroniques nous permet de les considérer comme des témoins d'importance quant à l'étude de l'oral spontané en diachronie (Dourdy et Spacagno, 2020 ; Lefeuve et Parussa, 2020).

Il s'agira ainsi pour nous, dans cette contribution, d'analyser la façon dont les textes légaux et les chroniques historiques exhibent des traits de l'oralité (les « mots du

discours »), en se fondant sur une étude de corpus. Il s'agira à la fois de déterminer des mouvements d'évolution en diachronie longue, mais également d'éprouver plusieurs stratégies de repérage des marqueurs discursifs s'appuyant, notamment, sur une annotation automatisée des textes anciens. Les outils d'analyse morphosyntaxiques s'étant notablement améliorés ces dernières années (Grobol et Crabbé, 2021 ; Prévost *et al.*, 2024), nous avons voulu tester différentes méthodes pour relever automatiquement les occurrences pertinentes.

Dès lors, cette étude explore deux questions différentes : (i) d'une part, et au-delà des problèmes de définition des « mots du discours » sur lesquels nous reviendrons ci-après, nous désirons explorer une question méthodologique. Quelles sont les stratégies à notre disposition, avec l'outillage accessible aujourd'hui dans les corpus (lemmatisation, étiquetage morphosyntaxique, scores de dépendance, etc.), pour repérer ces mots du discours, et comment évaluer leur efficacité respective ? (ii) D'autre part, se servir de ces différentes stratégies pour proposer un panorama, en diachronie, de l'emploi des mots du discours dans les textes légaux et les textes historiques, et établir une chronologie du changement et des tendances d'emploi non seulement dans le temps, mais également entre les genres textuels eux-mêmes. Notamment, il a été observé que les textes juridiques de l'histoire du français révèlent plus tôt l'émergence de phénomènes morphosyntaxiques (Larrivée et Goux, 2024a ; Larrivée et Goux, 2024b) et textuels (Goux, 2022b), et nous désirons vérifier si ce caractère innovant se rencontre dans le domaine des marqueurs discursifs.

Notons néanmoins que la notion de *genre textuel* pose d'importantes difficultés méthodologiques, notamment liées à l'hétérogénéité des textes (*cf.* Ayres-Bennett, 2025), hétérogénéité qui pourrait amener à considérer une découpe plus fine en séquence textuelle. Il est cependant difficile d'annoter automatiquement un texte particulier en séquences sur la base d'indices explicites, comme l'attribution d'une étiquette particulière (narratif, descriptif, explicatif, etc.) à une séquence dépend tant de critères linguistiques que d'une impression laissée à la lecture. Aussi, l'entrée générique, aussi imparfaite serait-elle, permet d'orienter l'analyse de prime abord, en attendant de pouvoir affiner les observations en la complétant d'une analyse séquentielle. Comme on le verra de plus *infra*, procéder de la sorte permet également d'identifier des emplois stratégiques de certains marqueurs discursifs susceptibles d'orienter la découpe des textes en séquences. L'analyse séquentielle n'est donc pas entièrement absente de notre démarche, même si elle ne détermine pas, en amont, les données que nous avons collectées.

Nous avons organisé notre contribution en trois parties, pour cette étude qui se veut pilote quant au repérage (semi-)automatisé des mots du discours dans les corpus outillés. Dans un premier temps, nous reviendrons sur la définition délicate des marqueurs discursifs, en retenant les propriétés morphologiques, syntaxiques, sémantiques et pragmatiques qui ont conditionné la recherche de nos occurrences. Ensuite, nous présenterons les corpus sur lesquels nous nous sommes appuyés pour effectuer cette

recherche, les requêtes informatiques utilisées pour ce faire et les difficultés rencontrées pour collecter les occurrences en fonction de la stratégie adoptée. La troisième et dernière partie sera consacrée à l'analyse des occurrences et à l'évaluation desdites stratégies.

2. Définition

La définition des « mots du discours », ou « marqueurs discursifs », fait encore l'objet de nombreuses discussions, et il est difficile, semble-t-il, d'avoir une définition arrêtée sur les propriétés de ces objets polymorphes qui peuvent prendre la forme d'interjections (*ben, eh, oh*), d'outils de gestion de la structuration du discours (*euuh*), de formes à valeur phatique (*tu vois, tu sais, comprends bien, faut voir, allons*), de connecteurs (*donc, alors, mais, et, là*), etc. La liste extensive des occurrences susceptibles de rentrer dans ce paradigme complique les analyses, et oblige à centrer les définitions retenues (cf. Dostie et Pusch, 2007 ; Glikman *et al.*, 2019).

Pour notre travail, nous nous sommes fondés sur les travaux d'H. L. Andersen (2007), de G. Dostie (2004) et de C. Bolly (2012) pour identifier une série de traits qui permettent d'isoler ces objets complexes. Ainsi, pour Dostie (2004 : 43-45), les marqueurs discursifs se définissent selon trois paramètres principaux : (i) ils ne contribuent pas au contenu propositionnel, ou à la prédication de l'énoncé ; (ii) ils sont multifonctionnels dans le cadre des interactions verbales, et peuvent endosser plusieurs types de relation sémantiques ou argumentatives ; (iii) ils ont un rôle interactionnel, voire textuel, et agissent au niveau de l'articulation de l'interaction ou de sa connexité. Andersen (2007 : 13-14) ajoute à ces paramètres trois autres traits qui nous ont semblé pertinents, notamment dans le cadre d'une recherche des marqueurs discursifs dans un corpus outillé : elle observe effectivement que ces marqueurs se caractérisent aussi par (iv) leur invariabilité morphologique ; par (v) leur position relativement libre dans l'énoncé et, enfin, par (vi) leur sens subjectif ou intersubjectif. C. Bolly (2012 : 147) résume, dans un court paragraphe, les contours de cette famille spécifique :

Les marqueurs de discours sont des unités indexicales, qui guident l'interlocuteur / scripteur sur la façon dont il faut organiser et manipuler l'information conceptuelle. Leur sémantisme peut être vu comme le résultat d'un processus de déplacement par abstraction / aréférenciation.

Ou encore, si nous reprenons la *GGHF*, ces marqueurs « se caractérisent par leur optionalité syntaxique, leur position très souvent initiale, leur réduction phonologique, leur capacité à articuler les plans discursifs local et global, leur sous-détermination sémantique et, enfin, leur fonction cohésive » (Badiou-Monferran et Marchello-Nizia, 2020 : 1637). Ces critères sont utiles et pertinents mais leur généralité amène, parfois, à inclure dans cette famille des éléments qui peuvent s'y rattacher mais dont le rôle, tant sémantique que textuel, semble pourtant fondamentalement distinct de celui des marqueurs discursifs. Notamment, les arguments mis en avant par Andersen *supra*

semblent autant décrire ces marqueurs que la plupart des membres de la famille des adverbes et il est vrai que, morpho-syntaxiquement, la frontière entre ces paradigmes est ténue. Quant aux paramètres identifiés par Dostie, ils semblent communs avec la famille des connecteurs ou des mots de liaison (Goux, 2021 ; Lehmann, 2023 : 125-126) qui, du reste, sont souvent analysés comme des adverbes ou des locutions adverbiales. Il est vrai, cependant, que certains mots-outils comme *donc* ou *alors* seront analysés selon les contextes tant comme des « mots du discours » que comme des connecteurs (cf. Riegel *et al.*, 2021 : 1052). Partant, et même si nous voulions exclure de notre recherche les mots de liaison puisque relevant, selon nous, d'une autre démarche et d'un autre dispositif interactionnel, il semble que seule une analyse de détail, prenant en compte l'intégralité des informations contextuelles et situationnelles, permette d'identifier correctement les marqueurs discursifs. Le rôle de la discussion étant notamment de mettre en avant les enjeux de l'hétérogénéité de la catégorie des marqueurs discours, nous ne développerons pas plus en amont les liens entre connecteurs et pragmatification, qui mériteraient une analyse à part. Certaines occurrences, quoi qu'il en soit, seront nécessairement ambiguës et indécidables : c'est une conséquence inévitable de ce flottement terminologique et de l'absence de véritables tests opérationnels permettant de repérer sans équivoque ces objets.

Mais au-delà de ces enjeux épistémologiques, ces difficultés ont des effets concrets pour la recherche : comment, dans ces conditions particulières, pouvons-nous faire une étude sur corpus, et à partir de quels paramètres (morphologiques, syntaxiques, sémantiques, pragmatiques...) ? Et, le cas échéant, sommes-nous obligés d'annoter manuellement les occurrences pour en faire un relevé systématique ? Notre objectif effectivement, dans cette contribution, n'est pas seulement d'identifier des évolutions et des changements en diachronie, mais aussi d'éprouver des méthodes de collecte d'occurrences qui permettent de mener cette identification. En nous fondant ainsi non seulement sur les paramètres mis en avant par la recherche, mais aussi sur les propriétés et les dimensions des corpus d'étude sélectionnés, que nous présenterons ci-après, nous avons choisi d'explorer trois grandes familles de stratégies :

– En premier lieu, un repérage entièrement automatisé, fondé sur l'annotation morphosyntaxique des corpus en parties de discours d'une part, en fonction syntaxique en dépendances de l'autre. Comme précisé *supra*, les marqueurs discursifs sont souvent rapprochés, voire confondus avec les adverbes qui sont assez bien repérés par les analyseurs (*parsers*) par leur invariabilité morphologique. De même, leur non-participation au contenu propositionnel se traduit par des relations de dépendance spécifiques que les analyseurs savent identifier avec beaucoup de régularité.

– Deuxièmement, un repérage mixte que nous avons qualifié de « semi-automatisé » et qui combine ces annotations morphosyntaxiques faites automatiquement avec un typage manuel des documents explorés. Notamment, nous avons ajouté une strate complémentaire d'information énonciative dans les textes pour identifier les

segments au discours direct. Cette couche d'annotation supplémentaire n'a pas été faite spécifiquement dans le cadre d'une analyse des marqueurs discursifs en diachronie, mais nous nous sommes rendu compte qu'elle pouvait être exploitée pour cette analyse.

– Enfin, un repérage entièrement manuel fondé sur une recherche ciblée de marqueurs discursifs fréquents (termes d'adresse, adverbes en *-ment*, interjections, etc.) et par un parcours systématique des sites préférentiels d'emploi (ou considérés comme tels) de ces marqueurs : en priorité, le début ou la fin des « phrases¹ » et le début ou l'alternance des tours de parole dans les segments au discours direct. Bien que ce repérage ait été construit à partir de listes fermées d'éléments, établies préalablement, nous nous sommes néanmoins appuyés sur certaines annotations automatisées comme la lemmatisation ; mais cette automatisation n'a pas orienté nos recherches.

Nous avons comparé ces trois stratégies, en tâchant de trouver particulièrement les occurrences qui n'avaient pas été isolées par les précédentes. Ainsi, nous avons tout d'abord commencé par les recherches automatisées, avant de faire des recherches mixtes puis entièrement manuelles, ce qui nous a permis de vérifier l'efficacité de l'outil informatique dans la collecte des occurrences.

3. Corpus et requêtes

3.1. Corpus d'étude

Pour cette analyse, nous nous sommes appuyés sur deux corpus développés au laboratoire CRISCO, à l'Université de Caen Normandie : MICLE, composé d'un ensemble de textes légaux (procès et styles de procédure) du XIII^e au XVII^e siècles et issus de la région normande (*cf.* tableaux 1 et 2), et CHRONIQUES (*cf.* tableau 3), composé de chroniques historiques du XII^e au XIX^e siècles et issus de la région normande également. Ces corpus ont été annotés en parties du discours selon trois standards (Universal Dependencies [UD]², PRESTO³ et UPENN⁴), ont été lemmatisés et ont été annotés syntaxiquement selon le formalisme *Universal Dependencies*. L'annotation a été faite par l'analyseur HOPS (« an HOnest Parser of Sentences⁵ »), qui a été développé et déployé dans le cadre du corpus PROFITEROLE (Grobol & Crabbé 2021 ; Prévost

¹ Bien évidemment, la notion de « phrase », ainsi que sa délimitation, sont hautement problématiques avant la période moderne. Les découps opérées dans les corpus MICLE et CHRONIQUES, que nous avons sondés pour cette contribution, ont répondu à des contraintes scientifiques particulières s'appuyant tant sur la réalité des documents anciens (et notamment les informations typographiques, mais aussi sur le sens des énoncés) que sur les fonctionnalités des logiciels d'annotation exploités. Nous renvoyons à Goux (2024a : § 1.2) pour un argumentaire détaillé de ces choix pour le corpus MICLE (et, par extension, pour le corpus CHRONIQUES, qui a suivi les mêmes démarches).

² <https://universaldependencies.org> (consulté le 12 septembre 2024).

³ <https://presto.ens-lyon.fr> (consulté le 12 septembre 2024).

⁴ <https://www.ling.upenn.edu/hist-corpora/annotation/index.html> (consulté le 12 septembre 2024).

⁵ <https://github.com/hopsparser/hopsparser> (consulté le 12 septembre 2024).

et al. 2024), avec une correction manuelle des annotations (au moins pour les fonctions nucléaires sujet et objet direct, intégralement pour les parties du discours et des lemmes). Les corpus sont accessibles sur le portail TXM-CRISCO et le site CriscoHT⁶, qui proposent une présentation complète de ces corpus et de ses textes, à laquelle nous renvoyons. Nous renvoyons également à Goux (2024a) pour une description du corpus MICLE, les règles de sa constitution ayant également été suivies pour le corpus CHRONIQUES⁷.

TITRE COURT	DATE	LANGUE	NOMBRE DE TOKENS
<i>Assises de Normandie</i>	1207	Ancien français	4 766
<i>Anglo-Norman Year Books 1292</i>	1292	Anglo-normand	33 557
<i>Atirements et Jugiés</i>	1314	Ancien français	15 232
<i>Anglo-Norman Year Books 1340</i>	1340	Anglo-normand	49 203
Style et usage de l'échiquier de Normandie	1425	Moyen français	71 903
<i>Procès de Jeanne d'Arc</i>	1431	Moyen français	41 967
<i>Réhabilitation de Jeanne d'Arc</i>	1450	Moyen français	8 716
<i>Procès de Sorcellerie (Guernesey)</i>	1563	Moyen français	12 190
<i>Style de Procédure Terrien</i>	1578	Moyen français	40 794
<i>Procès Bavent</i>	1643	Français classique	3 343
<i>Confession Bavent</i>	1652	Français classique	33 390

Tableau 1. Corpus MICLE Noyau

TITRE COURT	DATE	LANGUE	NOMBRE DE TOKENS
<i>Roche-Guyon</i>	1502	Moyen français	41 149
<i>Style Rouillé</i>	1539	Moyen français	35 134
<i>Fille possédée</i>	1591	Moyen français	30 773

Tableau 2. Corpus MICLE Satellite

TITRE COURT	DATE	LANGUE	NOMBRE DE TOKENS
<i>History of Kings of England</i>	1190	Anglo-normand	5 798
<i>Croniques de Normendie</i>	1223-1453	Ancien français	45 958
<i>Histoire des Ducs de Normandie</i>	13 ^e s.	Ancien français	64 911
<i>Chronique normande du 14^e siècle</i>	1373	Ancien français	54 299
<i>Abbrégé de l'histoire de la Normandie</i>	1665	Français classique	58 185

⁶ <https://txm-crisco.huma-num.fr>, <https://criscoht.unicaen.fr> (consulté le 12 septembre 2024). On trouve sur ces sites les notices bibliographiques complètes des ouvrages listés dans les tableaux 1 à 3.

⁷ Les chronolectes indiqués dans les tableaux suivent la répartition conventionnelle qui fait s'arrêter l'Ancien français en 1400 et le Moyen français en 1600. Cette découpe est encore d'usage pour l'encodage des fichiers XML selon les standards de la TEI, même si elle ne correspond plus à l'avancée de la recherche sur la périodisation de la langue française. Elle s'installe néanmoins dans une tradition numérique qui assure l'interopérabilité des corpus historiques entre eux. Voir (<https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>) (consulté le 16 juin 2025) pour la liste des codes concernés, notamment « fro » (Old French [842-ca. 1400]) et « frm » (Middle French [ca. 1400-1600]).

<i>Essai sur l'histoire de Normandie</i>	1766	Français classique	33 818
<i>Histoire de la Normandie ancienne</i>	1862	Français moderne	50 549

Tableau 3. Corpus CHRONIQUES

3.2. Stratégies de requêtage

Pour repérer les marqueurs discursifs à partir de l'annotation de ces textes, plusieurs stratégies sont alors disponibles. Comme la recherche des adverbes en eux-mêmes jetterait un filet trop large pour nos besoins, nous avons sélectionné dans un premier temps les interjections ([udpos="INTJ"]⁸). La documentation du site UD donne une description et quelques exemples de l'emploi de cette étiquette dans le jeu qui a été appliqué au corpus :

An interjection is a word that is used most often as an exclamation or part of an exclamation. It typically expresses an emotional reaction, is not syntactically related to other accompanying expressions, and may include a combination of sounds not otherwise found in the language⁹ (<https://universaldependencies.org/u/pos/INTJ.html>).

Compte tenu de leur participation à la dimension émotionnelle de l'interaction, et de leur analyse indépendante des relations syntaxiques traditionnelles, il nous a semblé que les interjections composent pour nous les meilleures candidates pour cette recherche.

Une autre stratégie de collecte a nécessité de chercher les mots qui, toujours selon les principes du système UD, ne participent pas, sur le plan syntaxique, à la structure de la phrase analysée. Deux fonctions en particulier sont dédiées à ces relations, et nous ont semblé intéressantes : *discourse* d'une part, *vocative* de l'autre ([function="discourse|vocative"]).

[Discourse] is used for interjections and other discourse particles elements, which are not clearly linker to the structure of the sentence except in an expressive way¹⁰ (<https://universaldependencies.org/u/dep/discourse.html>, consulté le 12 septembre 2024).

⁸ Les requêtes CQL sont celles qui ont été utilisées via le portail *TXM-Crisco* (voir note précédente) pour collecter les occurrences.

⁹ « Une interjection est un mot employé généralement comme exclamation ou partie d'une exclamation. Elle exprime généralement une réaction émotionnelle, elle n'est pas syntaxiquement liée aux autres expressions qui l'accompagnent et elle peut inclure une combinaison de sons que l'on ne trouve nulle part ailleurs dans la langue concernée » (notre traduction).

¹⁰ « La relation *discourse* est utilisée pour les interjections et les autres marqueurs discursifs qui ne sont pas clairement liés à la structure de la phrase, sauf sur le plan émotionnel » (notre traduction).

The *vocative* relation is used to mark a dialogue participant addressed¹¹ (<https://universaldependencies.org/u/dep/vocative.-html>, consulté le 12 septembre 2024).

La fonction *discourse* renvoie ainsi à des « fonctions discursives », associée automatiquement par l'analyseur à des mots qui ne sont pas des dépendants, directement ou indirectement, de la fonction *root* (« racine », typiquement le verbe principal ou le substantif des phrases nominales) de la phrase analysée. Cette déconnexion syntaxique étant l'un des traits donnés aux marqueurs discursifs, *discourse* est un des meilleurs candidats pour la collecte des occurrences. La fonction *vocative* est, de ce point de vue syntaxique, proche de *discourse*, mais s'en distingue par ses propriétés morphologiques (la fonction porte sur des noms ou des pronoms, alors que *discourse* est associée à des interjections ou des adverbes, avec une distribution détachée). Bien que les vocatifs ne soient pas généralement inclus dans la famille des marqueurs discursifs, nous avons choisi de collecter les occurrences, ne serait-ce parce que les vocatifs sont généralement employés dans des contextes de dialogues qui sont, par essence, susceptibles d'illustrer des marqueurs discursifs. Ils sont donc susceptibles de nous aiguiller vers des énoncés pertinents pour cette problématique, quand bien même ne seraient-ils pas, *stricto sensu*, des mots du discours.

Cette première stratégie de requêtage, par parties de discours et par fonction syntaxique en dépendances, nous a permis de collecter un certain nombre d'occurrences à partir desquelles nous avons manuellement supprimé les faux-positifs et les erreurs résiduelles d'étiquetage. On notera que ces analyses automatiques dépendent évidemment, en amont, des choix de transcription observés lors de la préparation des corpus et notamment de la ponctuation et la découpe en « phrase », ou en unités de traitement. C'est le cas, bien entendu, pour n'importe quelle recherche faite sur les grands corpus outillés, mais compte tenu des difficultés propres aux textes antérieurs à l'invention de l'imprimerie ou de la première modernité (Parussa 2023 ; Goux, 2024a ; Goux, 2024b ; Goux et Pinzin, à paraître), il convient d'être particulièrement attentif à ces difficultés.

Pour contourner ces difficultés, d'autres stratégies ont été considérées aux côtés de cette recherche entièrement automatisée. En s'appuyant cette fois-ci sur la transcription des données, leur annotation en parties du discours et la lemmatisation, nous avons également recherché les adverbes en *-ment* ([udpos="ADV" & word="*.ment"]) et les mots situés en tête de phrase ([n="1"]), ainsi que les groupes de mots situés en tête de phrase et séparés du reste de celle-ci par un signe de ponctuation ([n="1"]+[word=",";"]). Là encore, nous avons parcouru manuellement les occurrences pour éliminer le bruit, et nous avons comparé ces résultats aux précédents.

¹¹ « La relation de dépendance *vocative* est utilisée pour marquer l'adresse à un participant d'un dialogue » (notre traduction).

Enfin, dans la mesure où il a été identifié les segments au discours direct dans les textes du corpus MICLE, nous avons parcouru manuellement ces segments pour identifier les éventuelles occurrences de marqueurs discursifs qui n'auraient pas été collectées par les autres méthodes (`<q_direct="True">[word=".*"]+</q>`)¹². Là encore, nous avons comparé ces résultats avec les précédents afin d'évaluer l'efficacité de cette recherche.

4. Relevés et analyses

Nous allons à présent analyser les occurrences relevées dans les corpus MICLE et CHRONIQUES, en distinguant les trois stratégies présentées plus haut (automatique, par partie du discours et fonction syntaxique ; semi-automatique, par recherche croisée des adverbes en *-ment* et des groupes en tête de phrase ; manuelle, par exploration des segments au discours direct).

4.1. Recherches automatiques

Les recherches automatiques sur les interjections nous ont permis de collecter quelques occurrences. Pour MICLE, ces interjections n'apparaissent qu'à compter du XV^e siècle, et se retrouvent surtout dans le type particulier du genre légal qu'est le témoignage, notamment l'ensemble de dépositions qu'est la *Réhabilitation de Jeanne d'Arc* qui revient, avec un dispositif narratif particulier, sur le procès de Jeanne d'Arc, et *Confession Bavent*, récit-témoignage des événements subis par Madeleine Bavent (cf. tableau 4, exemples 1 et 2¹³).

TEXTE	NOMBRE D'OCCURRENCES
<i>Procès Jeanne d'Arc</i> (1431)	1
<i>Réhabilitation de Jeanne d'Arc</i> (1450)	7
<i>Sorcellerie Guernesey</i> (1563)	3
<i>Confession Bavent</i> (1652)	5

Tableau 4. Interjections dans le corpus MICLE

(1) « *Ha*, noble maison de France, qui as tousjours esté protectrice de la foy, as-tu esté ainsi abusée, de te adhérer à une hérétique et scismatique ! » (*Réhabilitation*, 1450).

(2) Il me parle aussi-tost, & me dit : *Hé bien*, mon cœur, t'avois-je pas dit avec verité, que nous nous verrions encore aujourd'huy ? (*Confession Bavent*, 1652).

¹² À la date de rédaction de l'article, tous les textes n'ont pas fait l'objet d'un typage des segments au discours direct. Pour MICLE, l'*Anglo-Norman Year Books* de 1340, *Roche-Guyon* (1502), *Sorcellerie Guernesey* (1563) et *Confession Bavent* (1652) n'ont pas été typés. Pour CHRONIQUES, seule l'*Histoire des ducs de Normandie* (XIII^e siècle) a été typée.

¹³ Dans les exemples, les italiques mettent en lumière les formes commentées.

Dans CHRONIQUES, nous trouvons davantage d'occurrences à partir du XIII^e siècle, même si elles demeurent rares ; en revanche, nous n'en trouvons plus après le XV^e siècle et le nombre d'occurrences baisse régulièrement (cf. Tableau 5, exemples 3 et 4).

TEXTE	NOMBRE D'OCCURRENCES
<i>Histoire des Ducs</i> (13 ^e s.)	7
<i>Cronique de Normendie</i> (13 ^e -15 ^e s.)	5
<i>Chronique Normande</i> (1373)	3
<i>Histoire de la Normandie ancienne</i> (1862)	1

Tableau 5. Interjections dans le corpus CHRONIQUES

(3) *Ha*, Robiert ! dist li rois, je les connois trop bien (*Histoire des Ducs*, 13^e s.)

(4) Donc fust cryé « *Ho* » de par le roy de France (*Cronique de Normendie*, 13^e-15^e s.¹⁴)

Le faible nombre d'occurrences nous empêche de proposer des généralités. Nous savons cependant, en nous fondant sur des études précédentes sur le genre juridique (Marnette, 2006 ; Goux, 2022a), que celui-ci établit progressivement et tardivement le discours direct comme le témoignage du « vrai » à partir d'une situation qui considérerait le discours indirect comme seul garant de la véracité des témoignages : ces résultats iraient dans ce sens. Au contraire, les chroniques se distingueraient par un détachement de la parole rapportée au profit d'un récit dépersonnalisé, qui marquerait peut-être la recherche d'une plus grande objectivité.

La recherche au moyen de la fonction *discourse* donne des résultats différents. Cela est dû au fait que les interjections ne sont pas toujours annotées en *discourse* mais parfois en *root*, car il s'agit du seul mot de la phrase ainsi analysée et que toutes les phrases doivent, selon ce schéma d'annotation et par défaut, avoir une et une seule fonction *root*. Nous trouvons ainsi toujours des occurrences dès le XIII^e siècle dans MICLE, mais sans évolution notable (cf. Tableau 6), et certes davantage dans CHRONIQUES, mais le texte du XIII^e siècle monopolise quasiment toutes les occurrences relevées (cf. Tableau 7, exemples 5 et 6).

TEXTE	NOMBRE D'OCCURRENCES
<i>Anglo-Norman Year Books</i> (1292)	1
<i>Anglo-Norman Year Books</i> (1340)	1
<i>Procès de Jeanne d'Arc</i> (1431)	3
<i>Réhabilitation de Jeanne d'Arc</i> (1450)	1
<i>Sorcellerie Guernesey</i> (1563)	4
<i>Fille possédée</i> (1591)	1
<i>Confession Bavent</i> (1652)	3

Tableau 6. Fonction *discourse* dans le corpus MICLE

¹⁴ Notons que (4) est un exemple de discours direct cité par la narration, alors que (3) est une séquence entièrement au discours direct.

TEXTE	NOMBRE D'OCCURRENCES
<i>Histoire des Ducs</i> (13 ^e s.)	16
<i>Cronique de Normandie</i> (13 ^e -15 ^e s.)	1
<i>Histoire de la Normandie ancienne</i> (1862)	1

Tableau 7. Fonction *discourse* dans le corpus CHRONIQUES

(5) [...] ilz aperceurent ung batteau au bec a la chiepvre don elles furent joyeuses, disant : « *Dieu mercy*, voecy noz gentz venir ! » (*Sorcellerie Guernesey*, 1563)

(6) Or alés, dist elle, au duc, si verrés quel droit il vos en fera (*Histoire des Ducs*, 13^e s.)

Pour ce dernier exemple, nous pouvons nous interroger sur le statut de *or*, qui a certes un rôle de connecteur consécutif, que le DMF glose par « à présent donc, vraiment, à l'évidence », mais dont les fonctions s'orientent également du côté des marqueurs discursifs. Il renforce, toujours selon le DMF, l'assertion, l'injonction ou le souhait, l'exclamation, etc. Comme nous le soulignons en introduction, un certain nombre d'outils participent de ces différents plans d'interprétation de l'énoncé, et nous n'avons pas voulu exclure *a priori* ces termes qui sont, par nature, polyfonctionnels.

Quoi qu'il en soit, ces données interrogent, mais elles nous semblent surtout liées à la préparation des textes du corpus. Notamment, et afin d'avoir une annotation en dépendances exploitable, les textes anciens ont été souvent « sur-segmentés » en phrases distinctes alors que la réalité langagière du temps était sans doute bien plus complexe. L'objectif initial des corpus, qui portait sur des enjeux syntaxiques et typologiques liés à l'évolution de la structure V2 de l'ancienne langue française (Poletto *et al.*, 2024), rentre ainsi en tension avec la réalité des marqueurs discursifs qui demandent à être envisagés au-delà d'un cadre phrastique et syntaxique. Partant, quand bien même l'analyseur reconnaîtrait-il efficacement les marqueurs discursifs par l'intermédiaire de la fonction *discourse*, la préparation initiale du texte influence beaucoup leur repérage effectif¹⁵.

Concernant la fonction *vocative*, nous trouvons davantage d'occurrences, et en nombre suffisant pour observer une décroissance marquée dans le corpus MICLE (cf. Tableau 8).

TEXTE	NOMBRE D'OCCURRENCES
<i>Anglo-Norman Year Books</i> (1292)	175
<i>Anglo-Norman Year Books</i> (1340)	33
<i>Style et usage de l'Échiquier de Normandie</i> (1425)	13
<i>Procès de Jeanne d'Arc</i> (1431)	19
<i>Réhabilitation de Jeanne d'Arc</i> (1450)	16
<i>Roche-Guyon</i> (1502)	5

¹⁵ Nous renvoyons ici à la discussion de Goux (2024b : §7-13) sur l'annotation des corpus en diachronie, et les problèmes engendrés par leur spécialisation.

<i>Sorcellerie Guernesey</i> (1563)	4
<i>Fille possédée</i> (1591)	7
<i>Confession Bavent</i> (1652)	21

Tableau 8. Fonction *vocative* dans le corpus MICLE

Il convient cependant d'observer deux emplois du vocatif dans le corpus. Dans les *Anglo-Norman Year Books* et dans les *Style et usage de l'Échiquier*, le vocatif se trouve dans des mises en scène de situation judiciaire, qui précisent le contenu et la forme des tours de paroles exigés par le dispositif législatif (7 et 8). Dans la mesure où la parole judiciaire est, lors de ces échanges procéduriers, performative, le respect de cette mise en scène doit être total : aussi, ces termes d'adresse ont tant un rôle interactionnel, en rappelant effectivement l'identité de la personne dépositaire du pouvoir judiciaire, qu'un rôle procédural, comme c'est l'exactitude des mots prononcés selon le style de procédure qui a valeur légale.

(7) [(E nekedent yl respundy ke par la meyn Estevene. Tyl-tone.)] *Sire*, joe vus dirroi coment yl fut seisi (*Anglo-Norman Year Books*, 1292).

(8) [...] celui pour qui le jugié seroit venu doit requérir et dire ainsi : *Sire juge*, je vous requier... (*Style et usage de l'Échiquier*, 1425).

Nous pouvons nous demander également, en (7), si *Sire* en adresse s'accompagne d'une désémantisation au regard de sa signification de base. Il nous semble néanmoins que le terme, malgré sa valeur de marqueur, nous renseigne cependant sur le statut de l'interlocuteur, *Estevene Paunsevot* dans l'extrait. Nous retrouvons là encore une poly-fonctionnalité du terme d'adresse, qui agit à plusieurs niveaux d'interprétation.

Mais si l'on excepte ces trois textes, les autres emplois du vocatif se situent au sein de témoignages et d'échanges dialogués et sont d'ailleurs souvent des insultes (9 et 10), plutôt que des titres ou des formules de politesse. La question se pose de savoir si ces échanges sont effectivement des paroles rapportées au plus proche de leur réalité énonciative, ou s'ils ont subi des réécritures les éloignant de cette réalité. On notera, quoi qu'il en soit, que ces vocatifs témoignent de l'évolution marquée de l'écriture juridique pour ces témoignages et, comme nous l'évoquions précédemment, qu'elle délaisse le style indirect qui avait cours jusques là (Goux, 2022a), au profit d'une écriture imitant l'oral du temps.

(9) [...] ledit d'Estouteville respondit : « je te quiers, *ribault* », et ce disant le rua par terre (*Roche-Guyon*, 1502).

(10) [...] il dyst a sadite fille : « Ha, *meschante putain*, va querir ton enfant et le monstre es voisynes ! » (*Sorcellerie Guernesey*, 1563).

La même chose s’observe dans le corpus CHRONIQUES, où l’on constate une tendance à la baisse (cf. Tableau 9). Cette fois-ci, les occurrences renvoient surtout à des titres nobiliaires et des noms propres, conformément aux événements relatés qui concernent de hauts personnages et non des particuliers (11 et 12). En revanche, et au regard des styles de procédure, il s’agit bien ici dans les chroniques de rapporter des échanges oralisés « réels », ou non procéduriers ; mais il se pose ici encore la question de la réalité des interactions rapportées surtout qu’au regard des témoignages judiciaires, ce ne sont généralement pas des témoins directs qui rapportent ces paroles.

TEXTE	NOMBRE D'OCCURRENCES
<i>Histoire des Ducs de Normandie</i> (13 ^e s.)	58
<i>Cronique de Normendie</i> (13 ^e -15 ^e s.)	3
<i>Chronique Normande</i> (1373)	4
<i>Essai sur l'histoire de Normandie</i> (1766)	1
<i>Histoire de la Normandie ancienne</i> (1862)	10

Tableau 9. Fonction *vocative* dans le corpus CHRONIQUES

(11) Lors parla Robers de Biethune, et lor dist : « *Segneur*, quels chose est chou vous loés au conte ? » (*Histoire des Ducs*, 13^e s.).

(12) [...] et le roy lui dist soutivement : « *Dame*, vous m’avez en convenant que ce sera cellé [...] » (*Chronique Normande*, 1373).

Ces emplois du vocatif, toutefois, s’écroulent dans le corpus passé le premier texte. Cela va de pair avec une diminution régulière des segments au discours direct dans le genre de la chronique qui ira de plus en plus, et à l’inverse du genre juridique, vers le style indirect. Les occurrences, un peu plus nombreuses, de vocatifs et de discours direct dans le texte le plus tardif du corpus, *Histoire de la Normandie ancienne*, montrent d’ailleurs que ces segments ne sont là que pour dramatiser ou mettre en scène les événements rapportés, en affectant un parler médiévalisant (ou, du moins, un parler imitant la façon dont on se figurait la langue médiévale au XIX^e siècle). Ainsi, en faisant parler l’Évêque de Paris, Gozlin, à la fin du IX^e siècle, l’auteur met en scène ce discours (13) :

(13) L’évêque [...] parla à Sigefroi en ces termes : « *Seigneur*, l’empereur Charles, dont vous connaissez la puissance, m’a chargé, et a chargé tous ceux qui sont dans ces murs, de la garde de cette place [...] » (*Histoire de la Normandie ancienne*, 1862).

Il y a donc des points de rencontre, sur l’emploi du vocatif, entre textes juridiques et textes historiques ; mais leur rôle et, partant, la place des marqueurs discursifs dans les textes, répondent à des logiques textuelles et communicationnelles différentes, entre théâtralisation de la parole historique, établissement d’un guide de procédure ou mise à l’écrit d’un témoignage dans le cadre d’une affaire judiciaire. Une étude systématique de ces tours de parole, par exemple en reprenant la méthodologie proposée par Vermander (2020) et inspirée de l’analyse de la conversation et des grammaires de

l'oral, en comparaison du travail qu'il a déjà mené sur le corpus des farces, serait cependant à conduire pour affiner l'analyse.

4.2. Recherches semi-automatisées et manuelles

Au regard de la recherche automatisée des occurrences, la stratégie semi-automatisée a été, en revanche, peu concluante. Les termes d'adresse ont toujours été correctement identifiés en *vocative* par l'analyseur, et aucun mot en tête de phrase, adverbe en *-ment* ou autre, ne s'est ajouté à notre collecte précédente. Les occurrences de marqueurs discursifs mettant en jeu des adverbes sont toujours employées, dans nos corpus, en association d'une interjection ou d'une autre partie du discours annotée comme *discourse* et, donc, avaient déjà été relevées par la recherche automatisée (14).

(14) *Ouy, vrayment*, il sert au procez (*Procès de Jeanne d'Arc*, 1431).

Concernant, enfin, l'analyse des segments au discours direct, celle-ci a donné des résultats plus probants. Le décompte a considéré les segments continus, c'est-à-dire sans passage à un autre type de discours au sein d'une même phrase. Même si le relevé des marqueurs discursifs et des vocatifs lui-même n'a pas différé de nos recherches précédentes, nous avons pu reporter ces résultats pour calculer leur fréquence d'emploi, en décomptant les segments au discours direct dans lesquels nous trouvions ces mots du discours et ces vocatifs, reportés sur le nombre total de segments. Bien que quatre textes n'aient pas été annotés, le relevé laisse voir cependant des évolutions intéressantes (*cf.* Tableau 10).

TEXTE	NOMBRE DE SEGMENTS CONTINUS AU DISCOURS DIRECT	SEGMENTS AVEC MARQUEURS DISCURSIFS ET/OU VOCATIFS
<i>Anglo-Norman Year Books</i> (1292)	773	144 (18,6%)
<i>Anglo-Norman Year Books</i> (1340)	Non annoté	/
<i>Style et usage de l'Échiquier de Normandie</i> (1425)	0	0
<i>Procès de Jeanne d'Arc</i> (1431)	439	26 (5,9%)
<i>Réhabilitation de Jeanne d'Arc</i> (1450)	69	26 (37,7%)
<i>Roche-Guyon</i> (1502)	Non annoté	/
<i>Sorcellerie Guernesey</i> (1563)	Non annoté	/
<i>Fille possédée</i> (1591)	19	13 (68,4%)
<i>Confession Bavent</i> (1652)	Non annoté	/

Tableau 10. Fréquence des marqueurs discursifs et des vocatifs dans les segments au discours direct du corpus MICLE

On note ainsi une augmentation progressive de la fréquence des marqueurs discursifs et des vocatifs dans les segments au discours direct, qui semble aller avec la

volonté de donner une couleur réaliste aux propos rapportés. La comparaison entre le texte de 1431 et celui de 1450 est assez marquante : il s'agit effectivement de la même affaire judiciaire, et des mêmes faits relatés, mais le texte de 1450, à la dimension littéraire plus marquée, emploie fréquemment davantage de marqueurs discursifs et de vocatifs dans ses segments au discours direct que le texte de 1431. Ce paramètre peut, à ce que nous croyons, donner des clés d'identification de sous-genres au sein des grands domaines de l'écrit, en distinguant parmi le domaine juridique des styles de procédures, des témoignages ou des récits d'événements judiciaires, sur la base d'indices explicites comme la densité de ces marqueurs discursifs. Du moins, il peut indiquer des évolutions dans la façon dont ils considèrent la place du discours direct dans leur économie, et la relation entre ces segments au discours direct et les marqueurs discursifs dans une perspective de théâtralisation ou de dramatisation de la parole rapportée¹⁶. Ces éléments pourraient permettre d'effectuer un repérage (semi-)automatisé des séquences textuelles, ou du moins d'initier des propositions de découpage pour affiner l'analyse.

5. Conclusions

Cette étude pilote nous amène à deux résultats principaux :

– D'une part, que les analyseurs comme HOPS sont efficaces pour identifier les marqueurs discursifs, et qu'une recherche automatisée des occurrences par parties du discours ou par fonction syntaxique en dépendances permet de collecter les occurrences pertinentes. Il faut néanmoins apporter deux bémols à ce résultat :

1. La préparation initiale des textes, et notamment la découpe du texte en « phrases » et la qualité de la transcription au regard du respect des ponctuations, est déterminante en ce qu'elle oriente très notablement la sortie de l'analyseur. Plus encore, il conviendrait peut-être d'envisager une découpe alternative des unités textuelles, complémentaire à une division syntaxique en phrase des données, et d'entraîner des modèles spécifiques à l'identification des marqueurs discursifs pour éliminer le bruit que nous avons rencontré lors de nos recherches.

2. Le typage des segments au discours direct permet d'étudier plus en profondeur l'évolution de l'emploi des marqueurs discursifs. Bien que ce typage manuel soit particulièrement long, et qu'il ajoute beaucoup de travail lors de la phase d'annotation des textes, son association avec la recherche automatisée des occurrences par parties du discours ou par fonctions syntaxiques nous semble être la meilleure stratégie pour la production de données de référence, et l'établissement d'une chronologie du changement. Plus encore, une découpe des textes en séquences, que ce soit selon un paramètre énonciatif du type discours direct / indirect, ou selon d'autres indices comme la densité

¹⁶ Comme dit précédemment, seule l'*Histoire des Ducs* du corpus CHRONIQUES a été typée : on y a dénombré 119 segments au discours direct et 13 avec un marqueur discursif ou un vocatif, soit 10,9%.

ou la fréquence des marqueurs discursifs, permettrait d'affiner notamment l'analyse en diachronie.

– D'autre part, et pour ce qui est de la dimension générique, nos relevés témoignent d'une différence entre les textes juridiques et les textes historiques. Ainsi, les textes du corpus MICLE emploient les marqueurs discursifs dans une perspective de théâtralisation ou de dramatisation de la parole juridique, dans les styles de procédure ou les témoignages, et ce alors que le style indirect demeure, du moins jusqu'à la période classique, le « style de la loi ». D'une façon opposée, dans les textes historiques du corpus CHRONIQUES, l'effet de vraisemblance est recherché à date ancienne tandis qu'il disparaît progressivement au profit d'une parole dépersonnalisée¹⁷.

Les perspectives de recherche, suite à ces résultats et ces observations préliminaires, s'orientent dès lors vers la consolidation et l'exploration de ces deux pistes d'étude. Du point de vue de la diachronie des marqueurs discursifs, l'analyse doit se poursuivre à d'autres genres textuels et à d'autres périodes de l'histoire de la langue, pour établir des schémas d'évolution. Nous mettons notamment l'accent sur le fait que cette recherche ne doit pas s'enquérir exclusivement des mécanismes de grammaticalisation de ces marqueurs, mais aussi de la fréquence d'emploi de ceux-ci et de leurs enjeux communicationnels, textuels, sémantiques et pragmatiques dans l'histoire de la langue. Enfin, du point de vue méthodologique, il faut encore affiner les procédures d'annotation et d'analyse, en trouvant l'équilibre entre une annotation automatisée exploitant tous les avantages des analyseurs, et une annotation manuelle rendant mieux compte de la réalité linguistique des textes du temps.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ANDERSEN, Hanne Leth (2007) : « Marqueurs discursifs propositionnels », *Langue française*, 154, 13-28.
- AYRES-BENNETT, Wendy (2025) : « Tracing change through different text types and genres: successes and challenges », *Journal of French Language Studies*, 35, p. e5. DOI : <https://doi.org/10.1017/S095926952500002X>
- BADIOU-MONFERRAN, Claire & Christiane MARCHELLO-NIZIA (2020) : « Les connecteurs », in C. Marchello-Nizia, B. Combettes, S. Prévost & T. Scheer (éd.), *Grande grammaire historique du français*, Berlin, De Gruyter Mouton, 1637-1638.

¹⁷ Notons que le genre de la chronique historique, qui se développe notamment au XV^e siècle, est étroitement lié au moyen français : on peut donc s'attendre à ce que des innovations linguistiques apparaissent dans ce genre textuel à ce moment-là du temps. Nous remercions les évaluateurs d'avoir attiré notre attention sur cet élément.

- BOLLY, Catherine (2012) : « Du verbe de perception visuelle au marqueur parenthétique ‘tu vois’ : Grammaticalisation et changement linguistique », *Journal of French Language Studies*, 22 : 2, 143-164.
- DOSTIE, Gaétane (2004) : *Pragmaticalisation et marqueurs discursifs*. Bruxelles, Duculot.
- DOSTIE, Gaétane & Claus D. PUSCH (2007) : « Présentation. Les marqueurs discursifs. Sens et variation », *Langue française*, 154, 3-12.
- DOURDY, Laura-Maï & Michela SPACAGNO (2020) : « Donner la parole aux interrogés : une étude de l’oral représenté dans les comptes rendus de procès médiévaux aux XIV^e et XV^e siècles », *Langages*, 217, 119-132.
- DUFTER, Andreas & Susanne ZEPP-ZWIRNER (2025) : *La langue du peuple dans la littérature : Le français populaire mis en texte, de l’âge classique à nos jours*. Berlin, De Gruyter.
- FISCHER, Kerstin (2006) : *Approaches to discourse particles*. Oxford, Elsevier.
- GLIKMAN Julie & Nicolas MAZZIOTTA, (2013) : « Représentation de l’oral et structures syntaxiques dans la prose de la *Queste del saint Graal* (1225-1230) », in D. Lagorgette et P. Larrivé (éd.), *Représentation du sens linguistique*, Chambéry, Éditions de l’Université de Savoie, 69-90.
- GLIKMAN, Julie ; Gabriella PARUSSA & Richard WALTEREIT (2019) : « Les marqueurs du discours en diachronie du français : nouvelles perspectives », *Studia Linguistica Romanica*, 2.
- GOUX, Mathieu (2022a) : « Dire le droit et dire le vrai. La polyphonie dans la Coutume de Normandie », *Studia Linguistica Romanica*, 7, 1-20.
- GOUX, Mathieu (2022b) : « L’évolution des cadres de discours dans la Coutume de Normandie », *Studia Linguistica Romanica*, 8, 116-130.
- GOUX, Mathieu (2023) : « *D’une part et d’autre part* dans les coutumiers normands : entre corrélation et théâtralisation du dispositif juridique », in M. Goux et P. Mounier (dir.), *La Corrélation en diachronie longue (1450-1800). Phrase, texte et discours*. Paris, Honoré Champion, 155-172.
- GOUX, Mathieu (2024a) : « Enjeux des corpus bilingues en diachronie longue : l’exemple du projet MICLE », *Corpus*, 25. DOI : <https://doi.org/10.4000/corpus.8468>
- GOUX, Mathieu (2024b) : « De très grands corpus pour l’étude diachronique du français : annotations et informations métalinguistiques et paratextes », *Humanités numériques*, 9. DOI : <https://doi.org/10.4000/11wmv>
- GOUX, Mathieu & Francesco PINZIN (à paraître) : « Challenges of a Multilingual Corpus (Old French/Old Venetian): The example of the MICLE Project », in *Venise et la France. Similitudes, spécificités, interrelations*. Florence, Franco Cesati Editore.
- GROBOL, Loïc & Benoît CRABBÉ (2021) : « Analyse en dépendances du français avec des plongements contextualisés », *Actes de la 28^e conférence sur le traitement automatique des langues naturelles*. URL : <https://hal.archives-ouvertes.fr/hal-03223424>
- GUILLOT, Céline ; Alexei LAVRENTIEV ; Bénédicte PINCEMIN & Serge HEIDEN (2014) : « Le discours direct au Moyen Âge : vers une définition et une méthodologie d’analyse »,

- in D. Lagorgette & P. Larrivé (éd.), *Actes du colloque Représentation du Sens Linguistique V*. Chambéry, Éditions de l'Université de Savoie, 17-41.
- LARRIVÉE, Pierre & Mathieu GOUX (2024a) : « Croiser les corpus calibrés pour faire l'histoire de la langue : le cas de l'antéposition stylistique de l'infinitif et du participe », *Corpus*, 25. DOI : <https://doi.org/10.4000/corpus.8373>
- LARRIVÉE, Pierre & Mathieu GOUX (2024b) : « The evolution of bare nouns in the history of French. The view from calibrated corpora », *Journal of French Language Studies*. DOI: <https://doi.org/10.1017/S0959269524000061>
- LARRIVÉE, Pierre ; Cecilia POLETO ; Francesco PINZIN & Mathieu GOUX (2024) : « Asymmetry as a general cue for V2 (loss) », *Isogloss. Open Journal of Romance Linguistics*, 10: 7. DOI : <https://doi.org/10.5565/rev/isogloss.410>
- LEFEUVRE, Florence & Gaétane DOSTIE [éd.] (2017) : *Lexique, grammaire, discours. Les marqueurs discursifs*. Paris, Honoré Champion.
- LEFEUVRE, Florence & Gabriella PARUSSA (2020) : « L'oral représenté en diachronie et en synchronie : une voie d'accès à l'oral spontané ? », *Langages*, 217, 9-21.
- LEHMANN, Sabine (2023) : « Au commencement était le texte ou les marqueurs discursifs propositionnels comme traces de textualisation », in M. Saiz-Sanchez et S. Gomez-Jordana Ferrary (éd.), *Études de sémantique et pragmatique en synchronie et diachronie*. Chambéry, Presses Universitaires Savoie Mont-Blanc, 123-143.
- MARNETTE, Sophie (2006) : « La signalisation du discours rapporté en français médiéval », *Langue française*, 149, 31-47.
- PARUSSA, Gabriella (2023) : « Pragmatique historique et ponctuation : le meilleur usage de mauvaises données ? », in M. Saiz-Sanchez & S. Gomez-Jordana Ferrary (éd.), *Études de sémantique et pragmatique en synchronie et diachronie*, Chambéry, Presses Universitaires Savoie Mont-Blanc, 187-207.
- PRÉVOST, Sophie ; Loïc GROBOL ; Mathieu DEHOUC ; Alexei LAVRENTIEV & Serge HEIDEN (2024) : « Profiterole : un corpus morpho-syntaxique et syntaxique de français médiéval », *Corpus*, 25. DOI : <https://doi.org/10.4000/corpus.8538>
- RIEGEL, Martin ; Jean-Christophe PELLAT & René RIOUL (2021) : *Grammaire méthodique du français*. Paris, Presses Universitaires de France.
- RODRÍGUEZ SOMOLINOS, Amalia (2011) : « Présentation : les marqueurs du discours – approches contrastives », *Langages*, 184, 3-12.
- VERMANDER, Pierre (2020) : « Analyse de conversation et documents littéraires médiévaux », *Langage*, 217, 71-86.