

Chaînes de référence et narration : analyse d'un corpus médiéval diachronique

Céline GUILLOT-BARBANCE

École Normale Supérieure de Lyon

celine.guillot@ens-lyon.fr

<https://orcid.org/0000-0002-9784-663X>

Resumen

Este estudio presenta una descripción de las cadenas de referencia (menciones sucesivas de un mismo objeto del discurso) en un corpus de textos narrativos escritos entre comienzos del siglo XII y finales del XV. Nos basamos en un corpus digitalizado y anotado mediante la plataforma de análisis TXM, así como en una serie de mediciones, relativas a la cobertura de los textos por las cadenas, el tamaño de las mismas, sus referentes y sus categorías gramaticales. Con la finalidad de ofrecer una aproximación a la posible función discursiva de las cadenas de referencia en diacronía, ponemos en relación los resultados de estas mediciones con las divisiones internas de los textos (unidades estructurales y discurso directo).

Palabras clave: cadena de referencia, diacronía, francés medieval, estructuración textual, narración.

Résumé

Cette étude vise à décrire les chaînes de référence (mentions successives d'une même entité de discours) dans un corpus de textes narratifs écrits entre le début du XII^e et la fin du XV^e siècle. On s'appuie pour cela sur un corpus numérique annoté et une série de mesures, qui portent sur la couverture des textes par les chaînes, la taille des chaînes, leurs référents, leurs catégories grammaticales, et qui sont réalisées grâce à la plateforme d'analyse TXM. Les résultats de ces mesures sont mis en relation avec les divisions internes des textes (unités structurales, discours direct) et la progression textuelle, afin de donner un aperçu de la fonction discursive possible des chaînes en diachronie.

Mots-clés : chaîne de référence, diachronie, français médiéval, structuration textuelle, narration.

Abstract

The study aims to describe reference chains (successive references to the same discourse entity) in a corpus of narratives written between the early twelfth and the late fifteenth century.

* Artículo recibido el 15/09/2020, aceptado el 30/03/2021.

The research is based on an annotated digital corpus and frequency indicators, which encompasses text coverage by chains, the lengths of the chains, their referents and grammatical categories. Frequencies are measured thanks to the textual analysis platform TXM. The results are linked to text internal division (structural units, direct discourse) and to textual progression in order to provide an overview of the discourse function that chains may have in diachrony.

Keywords: reference chain, diachrony, Medieval French, text structure, narration.

1. Introduction

La recherche présentée ici porte sur les chaînes de référence (mentions successives d'un même objet de discours) d'un corpus médiéval diachronique (début XII^e - fin XV^e siècles) composé de cinq textes narratifs. Elle se situe dans le prolongement de travaux antérieurs (voir notamment les numéros 195 de *Langages* et *Langue française* dirigés par Schnedecker *et al.*) et s'appuie sur le cadre théorique, la méthodologie et le corpus développés au sein du projet *Democrat* (*DEscription et MOdélisation des Chaînes de Référence : outils pour l'Annotation de corpus (en diachronie et en langues comparées) et le Traitement automatique*)¹. Notre étude complète les résultats précédents et fournit de nouvelles données, dans le but d'établir une typologie des chaînes de référence qui rende compte de leurs variations en diachronie et en fonction du type de texte.

L'un des apports de la présente recherche est de chercher à faire le lien avec une linguistique diachronique de l'écrit telle que définie par Combettes (2012). Dans cette perspective, on s'est notamment intéressé aux modes de progression thématique et de structuration des textes médiévaux dans leurs relations avec les chaînes. Mais l'étude de ces corrélations est particulièrement difficile à mener en raison de l'absence de cadre d'analyse unifié, de la multiplicité des paramètres à prendre en compte et du manque d'outils adaptés. On se contentera ici de présenter quelques pistes de recherche, en les étayant sur un volume de données relativement important (environ 50 000 mots) et en explorant les outils développés à cette fin dans la plateforme d'analyse textométrique TXM (*cf.* Heiden *et al.*, 2010 et *Textométrie*, 2007-2010).

2. Présentation du corpus et du cadre d'analyse

Les chaînes de référence (CR) se définissent comme la « suite des expressions d'un texte entre lesquelles l'interprétation construit une relation d'identité référentielle » (Corblin, 1985 : 174). L'exemple (1) permet de suivre les expressions co-

¹ Le projet *Democrat* (<https://www.lattice.cnrs.fr/democrat>) a bénéficié d'un financement de l'ANR (ANR-15-CE38-0008). Je tiens à remercier les membres du projet *Democrat* et mes collègues de l'IHRIM (Matthieu Decorde, Serge Heiden, Alexei Lavrentiev et Bénédicte Pincemin) qui développent la plateforme TXM et la *Base de français médiéval*, ainsi que Sophie Bordes, qui a annoté le corpus lors de son stage de M1.

référentes appartenant à deux CR, celle du roi Clodoveu (en rouge) et celle de la reine Baltalt (en bleu) :

(1) *Quant li rois Clodoveu et le roine Baltalt eurent tant esté emsamle qu'il eurent .ii. enfans, si vint en volenté au roi qu'il alast en pelerinaige en le sainte tere de Jherusalem. Il cela tant chele volenté a le roine que ele s'aperchut qu'il estoit en son cuer tormentés d'aucune secree pensee qu'ele ne savoit mie* (Vie de sainte Bathilde, version 1, p. 5).

Les expressions référentielles qui constituent chaque CR et se suivent à la manière de maillons qui s'enchaînent les uns aux autres seront appelées *mentions* dans le reste de l'étude. On parlera de CR dès lors que plus de deux mentions coréférentielles se suivent. Les mentions isolées ou reliées en paires, que les « notions d'anaphore et de coréférence suffisent amplement à décrire » (Schnecker et Landragin, 2014 : 4), n'entrent pas dans des chaînes.

L'un des objectifs du projet *Democrat* cité plus haut était de produire un corpus diachronique de français écrit (XII^e. - XXI^e s.) annoté en mentions et en chaînes. Ce corpus est composé d'extraits qui couvrent les 10 000 premiers mots (environ) de textes de différents genres textuels. Il comporte au total 689 000 mots et est accessible sous licence Creative Commons - Attribution - Partage dans les Mêmes Conditions 4.0 International. Les cinq œuvres sur lesquelles porte notre étude ont été sélectionnées à l'intérieur de cet ensemble et font partie des premiers textes médiévaux annotés dans le cadre du projet. L'annotation des CR est une tâche particulièrement complexe (pour plus de détails sur l'annotation avec la plateforme TXM, voir Guillot-Barbance et Qui-gnard, 2019 : §19 à 27). Cette annotation ayant été réalisée par des stagiaires, plusieurs vérifications ont été nécessaires avant de pouvoir exploiter le corpus, ce qui explique qu'il n'ait pas été possible d'utiliser davantage de textes. Il est certain que des erreurs d'annotation subsistent malgré ces précautions, mais on peut espérer qu'elles sont compensées et neutralisées grâce au volume de données traitées.

L'un des partis pris théoriques du projet *Democrat* a été d'annoter les CR indépendamment de la structure interne des textes (frontières de livres, chapitres, paragraphes, discours direct, etc.). Toutes les mentions coréférentes ont donc été rassemblées dans une seule et même chaîne continue. Bien qu'il soit discutable sur le plan théorique, ce choix permet de bien distinguer le suivi des chaînes et des divisions textuelles. Il facilite l'analyse de leurs interactions réciproques.

Le tableau 1 présente les extraits de textes analysés. Il permet de repérer des déséquilibres (pas de texte au XIV^e s., textes anciens en vers et textes récents en prose, genres textuels hétérogènes), dont certains sont inhérents aux données disponibles pour la période ancienne. Malgré ces imperfections, ce corpus a l'avantage de rassembler des textes qui sont tous narratifs sur une diachronie assez longue (début XII^e siècle - fin XV^e siècle) :

Texte ²	Auteur	Titre	Date	Forme	Genre	Nombre de mots ³
Roland	anonyme	<i>Chanson de Roland</i>	ca 1100	vers	épique	13 059
eneas1	anonyme	<i>Eneas</i>	ca 1155	vers	roman	11 647
SBath1	anonyme	<i>Vie de sainte Bathilde (Version I)</i>	2 ^e m. 13 ^e s.	prose	hagiographie	11 175
Jehpar	anonyme	<i>Roman de Jean de Paris</i>	1494	prose	roman	11 915
commyn1	Ph. de Commynes	<i>Mémoires</i>	ca. 1490-1505	prose	mémoires	11 268

Tableau 1. Présentation des textes du corpus.

Du point de vue structurel, les textes se répartissent en trois ensembles :

	Structure de niveau 1	Structure de niveau 2
Roland	112 laisses	-
eneas1	(initiales rubriquées)	-
SBath1	(initiales rubriquées)	-
Jehpar	25 titres	-
commyn1	livre 1	2 titres

Tableau 2. Divisions textuelles du corpus.

Comme toutes les chansons de geste anciennes, la *Chanson de Roland* se caractérise par sa structure en laisses assonancées et le passage annoté regroupe 112 laisses de tailles très inégales. Même si elle paraît répondre à des critères métriques et rythmiques plutôt que discursifs, cette structure est très présente dans l'œuvre.

Eneas (en vers) et la *Vie de sainte Bathilde* (en prose) se rapprochent par l'absence de divisions supérieures et intermédiaires. La difficulté d'accès aux sources primaires n'a pas permis de se faire une idée très précise de la structuration des manuscrits édités. Des initiales colorées semblent marquer dans les deux cas des unités de taille variable, mais les éditeurs n'en ayant pas rendu compte, ces éléments n'ont pu être exploités. On peut sans trop s'avancer dire qu'il « manque » à ces textes des unités graphiques intermédiaires (paragraphe, groupes de vers, etc.), dont le rôle comme palier de traitement de l'information a été souligné (Adam, 2018).

Les deux textes du XV^e s. sont plus explicitement structurés. Le *Roman de Jehan de Paris* comporte des titres (non numérotés) qui délimitent des unités de petite taille (1 ou 2 pages de folio au maximum). Le texte de Commynes est organisé en unités plus importantes (livres), auxquelles s'ajoutent des titres intermédiaires (parfois, mais pas toujours, ajoutés par l'éditeur).

Tous les extraits annotés comportent par ailleurs des passages au discours direct. Le *Roland* se distingue par le fait que ces passages occupent la plus grosse partie de l'extrait (6838 mots dans le discours direct / 6221 mots en dehors), le texte de

² Dans tous les tableaux, on utilise les identifiants de la *Base de français médiéval* (d'où proviennent les extraits) pour les identifier.

³ Les ponctuations sont incluses dans le total.

Commynes par le fait qu'ils sont presque inexistantes (le discours indirect l'a plus ou moins remplacé). Les trois autres textes se situent entre ces deux extrêmes.

3. Caractéristiques globales des CR dans le corpus narratif

En nous appuyant sur une série d'indicateurs définis dans les études antérieures (voir notamment Glikman *et al.*, 2014 ; Obry *et al.*, 2017 ; Schnedecker, 2017) et implémentés dans la plateforme TXM (Heiden, 2019), nous tentons de caractériser les chaînes du corpus et des textes. La description porte sur les CR prises en bloc, sans tenir compte de leur progression au fil du texte.

3.1. Couverture des textes par les CR

Le tableau 3 permet d'apprécier la couverture des textes par les expressions référentielles. La colonne « Mentions » indique le nombre d'expressions annotées dans chaque extrait et la colonne « Référents » le nombre de référents différents correspondant à ces expressions. On observe que ces données sont relativement stables dans le corpus. Seul l'extrait le plus long, *Roland*, se démarque par un nombre de mentions et de référents un peu supérieur aux autres, et un autre texte, les *Mémoires* de Commynes, contient un peu moins d'expressions référentielles. Si l'on rapporte le nombre de mentions au nombre de mots de chaque extrait, les expressions référentielles couvrent entre 32% et 39% de la surface des textes du corpus⁴.

Les trois colonnes suivantes donnent une idée de la couverture des textes par les chaînes. On constate que les CR n'incluent qu'une faible proportion des référents discursifs (toujours moins de 10%), la très grande majorité d'entre eux n'étant pas assez importants pour être repris plus de deux fois. Mais les expressions référentielles entrant dans les chaînes occupent une part importante des textes, puisque plus de la moitié des mentions sont dans des CR. Du point de vue du matériau linguistique, les CR sont donc très présentes dans nos textes narratifs, même si beaucoup de référents trop évanescent leur échappent :

	Référents	Mentions	CR	Référents dans CR	Mentions dans CR
Roland	1873	5081	127	7%	65%
eneas1	1635	4220	107	6%	63%
SBath1	1703	4202	130	8%	61%
Jehpar	1634	4380	90	5%	64%
commyn1	1710	3659	148	9%	56%

Tableau 3. Référents, mentions et CR par texte.

Le tableau fait également apparaître quelques variations permettant de comparer les textes. *Jehan de Paris* contient peu de CR, ce qui explique le pourcentage assez

⁴ Ces chiffres sont approximatifs car ils ne tiennent pas compte du fait qu'une mention peut inclure un nombre de mots variable. Mais ils semblent indiquer que les expressions référentielles occupent environ un tiers des textes narratifs étudiés.

faible de référents qui leur sont intégrés. Le nombre de mentions incluses dans les CR étant au contraire élevé, on peut supposer que les chaînes sont relativement longues dans ce texte. *Commynes* se situe à l’opposé : c’est le texte dans lequel le nombre de CR est le plus élevé et le pourcentage de mentions le plus faible, ce qui implique des CR plus courtes qu’ailleurs. Les relevés suivants vont confirmer et préciser ces spécificités.

3.2. Typologie des CR : données comparatives sur la taille des CR

Le tableau 4 permet de comparer les longueurs des CR réparties en quatre grands types. Les chaînes courtes (3 ou 4 mentions) apparaissent comme étant toujours très nombreuses (entre 30 et 40% du total), ce qui confirme les relevés effectués sur des corpus très différents (voir notamment Guillot-Barbance et Quignard, 2019 : § 36). Les chaînes intermédiaires (entre 5 et 9 mentions) sont en nombre légèrement inférieur, sauf chez *Commynes* où elles sont particulièrement fréquentes. Les chaînes longues (entre 10 et 19 mentions) regroupent approximativement entre 10 et 20% des CR des textes. Quant aux chaînes très longues (à partir de 20 mentions), elles varient davantage en fonction des extraits :

	Nombre de CR par type de taille			
	3-4	entre 5 et 9	entre 10 et 19	à partir de 20
roland	42 (33%)	34 (27%)	24 (19%)	27 (21%)
eneas1	39 (36,5%)	26 (24%)	15 (14%)	27 (25,5%)
SBath1	51 (39%)	40 (31%)	23 (18%)	16 (12%)
jehpar	27 (30%)	25 (28%)	10 (11%)	28 (31%)
commyn1	46 (31%)	65 (44%)	20 (14%)	17 (11%)

Tableau 4 : Taille des CR par texte (nombre de mentions).

Le tableau confirme que les *Mémoires* de *Commynes* ont beaucoup de chaînes courtes et intermédiaires (75% du total) et peu de chaînes très longues. La *Vie de sainte Bathilde* s’en rapproche avec 70% des CR qui ont moins de 10 mentions et seulement 12% de CR de plus de 20 mentions. *Jean de Paris* est au contraire le texte dans lequel les CR très longues sont les plus nombreuses, du fait surtout d’un déficit en CR rassemblant entre 10 et 20 mentions.

Ces indications peuvent être complétées par un examen plus précis des chaînes les plus longues de chaque extrait :

	CR1	CR2	CR3	CR4	CR5	Total
roland	222	305	334	337	528	1726
eneas1	63	84	104	505	562	1318
SBath1	72	221	321	353	558	1525
jehpar	127	131	288	305	441	1292
commyn1	90	95	97	150	320	752

Tableau 5 : Taille des 5 CR les plus longues par texte (nombre de mentions).

Le tableau 5 révèle d'assez fortes disparités. *Roland* se distingue par la longueur de ses chaînes, longueur qui pourrait être favorisée par la taille supérieure de l'extrait. La position extrême de *Commynes* dans le corpus est confirmée, les chaînes longues dans ce texte étant nettement moins longues qu'ailleurs. *Eneas* est dominé par deux chaînes particulièrement étendues, ce qui est le cas aussi mais dans une moindre mesure de la *Vie de sainte Bathilde*, alors même que ce dernier texte a assez peu de CR très longues (voir tableau 4). Le *Roman de Jehan de Paris* se caractérise au contraire par des écarts de taille moins marqués et les CR les plus longues ne totalisent pas un nombre de mentions très élevé. Ce texte se signale en réalité par un nombre important de CR qui ont entre 20 et 100 mentions.

Ces indications sur les tailles des CR nous renseignent de manière indirecte sur les référents les plus proéminents des textes. *Eneas* a deux référents omniprésents (Didon et Enée), la *Vie de Sainte Bathilde* est dominée par les figures de Bathilde et son mari le roi Clodoveu (et leurs enfants), la *Chanson de Roland* par celle de Charlemagne. Les textes narratifs les plus récents du corpus ne sont pas caractérisés par la même suprématie d'un ou deux personnages centraux. *Jehan de Paris* met en scène plusieurs référents importants (Jehan de Paris, le roi d'Angleterre, le roi de France, le roi d'Espagne, les souverains d'Espagne, etc.). Quant à l'extrait de *Commynes*, s'il met bien en avant le duc Charles de Bourgogne, il ne lui accorde pas une place démesurée.

3.3. Typologie des CR : données comparatives sur les référents des CR

Comme on vient de le voir, ce sont les référents animés humains qui occupent une place centrale dans les textes narratifs. Dans notre corpus, plus de 50% des CR ont un référent animé humain et les deux textes dans lesquels la proportion est la plus forte sont les deux textes du XV^e siècle :

	Nombre de CR	Nombre de référents animés humains
roland	127	66 / 52%
eneas1	107	51 / 48%
SBath1	130	71 / 55%
jehpar	90	53 / 59%
commyn1	148	101 / 68%

Tableau 6 : Nombre de référents animés humains dans les CR

On constate également que les référents humains sont très majoritairement masculins. La seule œuvre qui se démarque sur ce point est la *Vie de sainte Bathilde*, l'un des deux personnages principaux étant la reine Bathilde.

3.4. Typologie des CR : données comparatives sur les catégories grammaticales des mentions

La fréquence des catégories grammaticales offre des données particulièrement utiles pour typer les CR et les textes. La figure 1 détaille ces fréquences par texte, que les expressions référentielles soient ou non dans des chaînes. Pour bien interpréter le

graphique, il faut tenir du compte du fait que Roland comporte plus de mentions que les autres textes et Commynes moins :

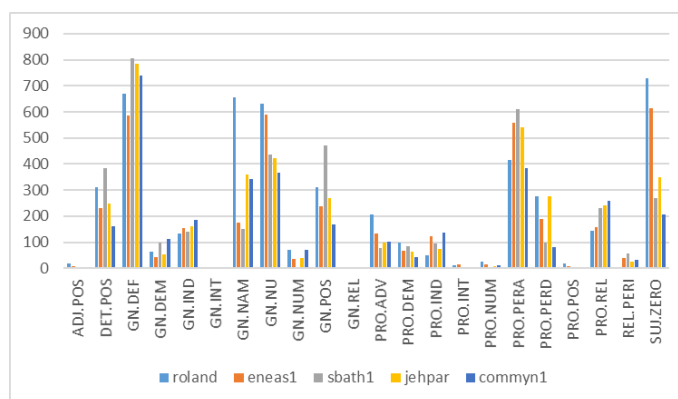


Figure 1. Catégories grammaticales des mentions par texte⁵.

Plusieurs évolutions diachroniques se repèrent à travers le graphique : la baisse du nombre de GN nus (sans déterminant), l'accroissement du nombre de GN indéfinis (qui les remplacent en partie), la chute très importante du sujet zéro, l'accroissement du nombre de pronoms relatifs (lié peut-être au développement d'une certaine complexité syntaxique).

La figure 2 reprend les mêmes données en éliminant les catégories peu fréquentes :

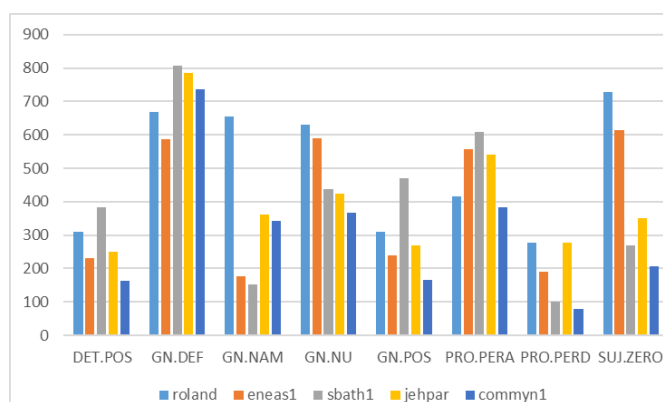


Figure 2. Principales catégories grammaticales des mentions par texte.

⁵ADJ.POS : adjectif possessif (*li miens amis*) ; DET.POS : déterminant possessif ; GN.DEF : groupe nominal défini ; GN.DEM : groupe nominal démonstratif ; GN.IND : groupe nominal indéfini ; GN.INT : groupe nominal interrogatif ; GN.NAM : nom propre ; GN.NU : groupe nominal sans déterminant ; GN.NUM : groupe nominal numéral ; GN.POS : groupe nominal possessif ; GN.REL : groupe nominal relatif ; PRO.ADV : *en/y* ; PRO.DEM : pronom démonstratif ; PRO.INT : pronom interrogatif ; PRO.NUM : pronom numéral ; PRO.PERA : pronom personnel anaphorique ; PRO.PERD : pronom personnel déictique ; PRO.POS : pronom possessif ; PRO.REL : pronom relatif ; REL.PERI : pronom relatif périphrastique (*celui qui*, etc.) ; SUJ.ZERO : pronom personnel sujet non exprimé.

Ce graphique simplifié fait bien apparaître la fréquence accrue à partir du XIII^e siècle et dans les textes en prose des GN définis et dans une moindre mesure du pronom personnel anaphorique (parallèle à la diminution du sujet zéro). Les noms propres, possessifs et pronoms déictiques ne suivent pas de tendance claire et leur emploi semble plus dépendant des particularités idiosyncrasiques de chaque texte (part du discours direct, liens de parenté entre les personnages, etc.). On verra plus loin que le nom propre et le GN défini peuvent se rapprocher du point de vue de discursif.

La comparaison des fréquences de ces catégories majeures selon que les mentions sont à l'extérieur des CR (figure 3) ou à l'intérieur (figure 4) apporte des précisions importantes :

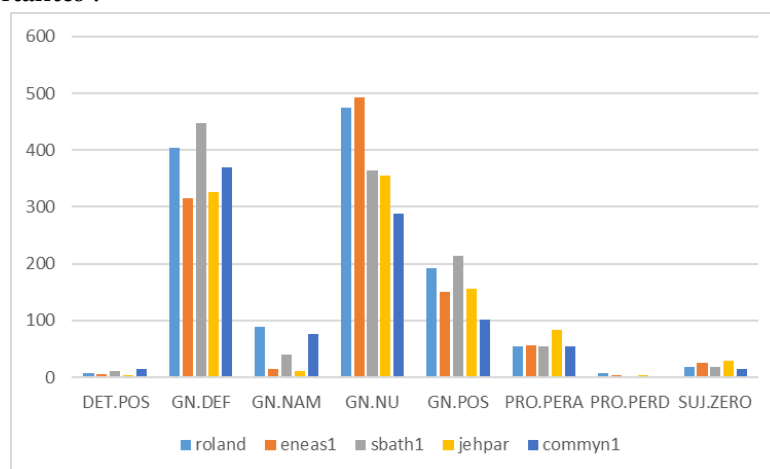


Figure 3. Principales catégories grammaticales des mentions hors CR.

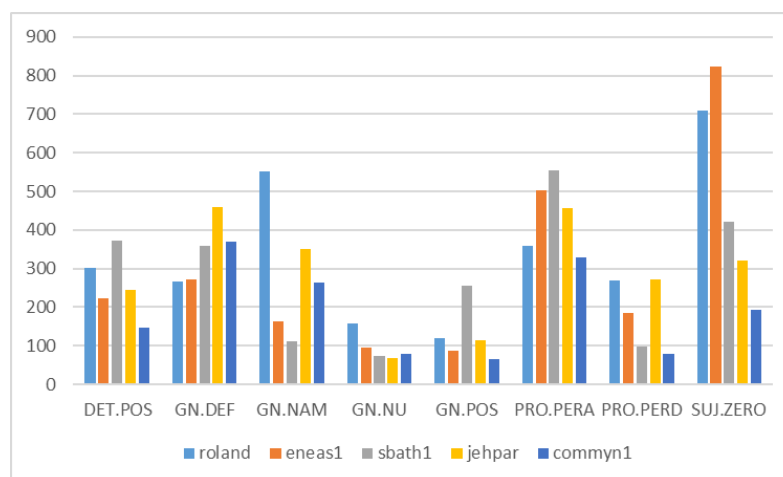


Figure 4. Principales catégories grammaticales des mentions dans les CR.

La différence la plus spectaculaire concerne les GN nus, qui apparaissent en proportion et en nombre beaucoup plus grand en dehors des CR, ce qui était prévisible car les référents de ces GN indéfinis ne sont généralement pas repris. À l'inverse, l'emploi de toutes les formes qui marquent la forte accessibilité référentielle (pronom sujet

zéro, pronom personnel et déterminant possessif) est presque exclusivement limité aux chaînes. Ce qui en revanche peut sembler plus surprenant est le maintien dans les CR d'un nombre important de marqueurs de faible ou moyenne accessibilité référentielle (Ariel, 1990).

Le GN défini étant l'un des moyens privilégiés d'introduire un nouveau référent qui peut disparaître aussitôt, il était prévisible qu'il soit plus fréquent dans les mentions hors CR. On observe toutefois qu'il est bien représenté dans les CR de tous les textes, et tout particulièrement dans les textes en prose à partir du XIII^e siècle. Le nom propre est surtout utilisé dans les CR de trois textes. Dans *Roland*, c'est même la deuxième catégorie en fréquence. Ce qui frappe surtout, c'est qu'il est toujours plus fréquent dans les CR qu'à l'extérieur. On peut voir là un effet de la saillance dont le porteur du nom est doté dans le récit (Schnecker, 1997), les référents nommés étant en général beaucoup réinstanciés, en particulier quand ils sont humains. Mais la place occupée par ces deux marqueurs mérite d'être soulignée et leur répétition dans les CR narratives devra être examinée.

La Figure 4 permet de préciser le profil de chaque texte du point de vue des chaînes. Deux des trois extraits les plus anciens se caractérisent par la fréquence des expressions pronominales, le sujet zéro dans *Eneas* et le pronom personnel anaphorique dans *Sainte Bathilde*. Dans ce dernier texte, le GN défini est relativement bien représenté aussi. *Roland* se rapproche de ces œuvres anciennes par la fréquence du pronom sujet zéro mais s'en distingue par la fréquence du nom propre. Les deux textes les plus récents sont ceux dans lesquels la part des expressions nominales (nom propre et GN défini) est la plus forte et la part des expressions pronominales la plus faible.

4. CR et (dis)continuité discursive dans le corpus narratif

Les chaînes de référence permettant le suivi des référents de discours, elles participent au marquage de la continuité discursive. On peut dès lors s'interroger sur la façon dont elles s'articulent avec la structure textuelle et s'intéresser à la façon dont elles se succèdent au fil du texte. On tentera dans une dernière section de donner quelques indications sur la structuration des parties narratives des textes à travers les chaînes.

4.1. CR et structure textuelle : la distinction narration / discours direct

Comme on l'a indiqué dans la section 1, les chaînes de référence ont été annotées de manière continue dans le corpus, sans qu'on tienne compte des frontières textuelles. Si ce point de vue peut se défendre dans la mesure où le texte écrit implique par définition une forme de linéarité et de séquentialité obligée, on peut se demander dans quelle mesure ces frontières influent sur les chaînes. Nous nous concentrerons dans cette section sur l'insertion du discours direct dans la narration et sur les interactions réciproques entre les unités discursives délimitées par ce biais et les chaînes.

Les figures 5 à 13 permettent de mettre en évidence texte par texte les particularités des CR selon qu'elles se trouvent à l'intérieur ou à l'extérieur du discours direct⁶. La taille du discours direct étant très variable d'un texte à l'autre, seules les fréquences relatives sont indiquées dans les figures. Les textes sont présentés dans l'ordre chronologique.

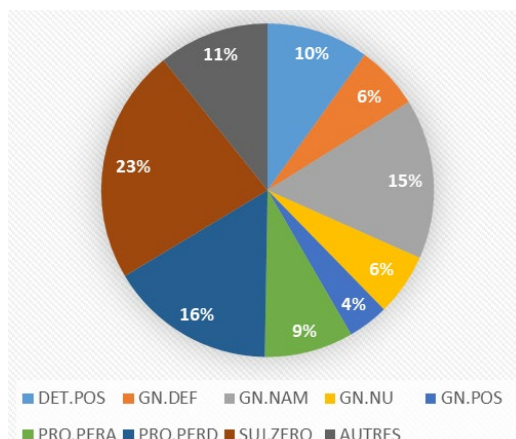


Figure 5. Catégories grammaticales des CR du discours direct dans Roland.

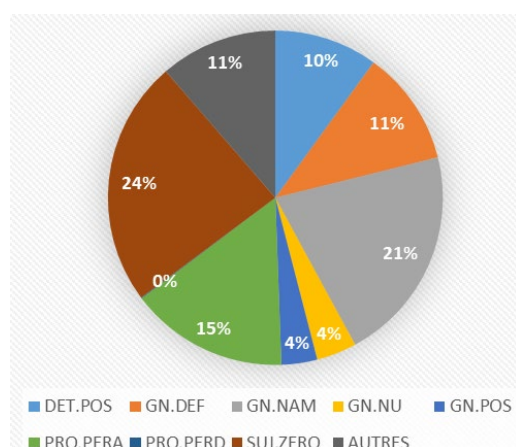


Figure 6. Catégories grammaticales des CR hors discours direct dans Roland.

La comparaison des parties du *Roland* permet de souligner la fréquence des pronoms déictiques dans le discours direct, ce qui est attendu, mais elle montre aussi la part très importante prise par le nom propre et le GN défini dans les CR extérieures au discours direct, puisque ces deux catégories occupent 32% du total des catégories, tandis que la somme du pronom sujet zéro et du pronom anaphorique ne dépasse pas 39%. Dans le discours direct les écarts sont plus marqués, le nom propre et le GN

⁶Les données présentées dans les figures ont été recueillies grâce à l'encodage du discours direct et à la fonctionnalité « création de sous-corpus » de la plateforme TXM (sous-corpus des parties au discours direct / sous-corpus des parties hors discours direct dans chaque texte).

défini représentant 21% de l'ensemble, les pronoms zéro et anaphorique 32% et les pronoms déictiques 16%. Si l'on distingue d'un côté les catégories grammaticales réservées aux référents les plus accessibles grâce au contexte (pronom sujet zéro, pronom personnel anaphorique et déictique, déterminant possessif) et de l'autre les catégories devant *a priori* marquer une moins grande accessibilité référentielle (nom nu, nom propre, GN défini et GN possessif), la tendance à privilégier les premières est plus forte dans le discours direct (58%) que dans le reste du texte (49%). Et il semble que ce soit surtout le nom propre et le GN défini qui soient renforcés dans les parties non oralisées.

Ces tendances se vérifient dans tous les autres textes du corpus. Les figures 7 et 8 permettent de constater que dans *Eneas* le GN défini et le nom propre occupent également une place plus importante à l'extérieur du discours direct (18%) qu'à l'intérieur (13%), même si les fréquences sont bien moindres que dans le *Roland* et si les pronoms sujets zéro et anaphorique dominent nettement (45% des CR dans le discours direct, 53% en dehors).

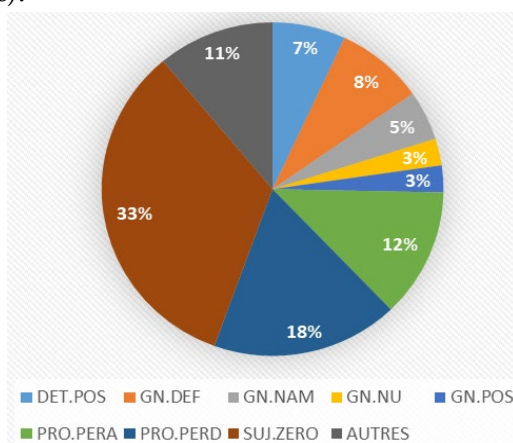


Figure 7. Catégories grammaticales des CR du discours direct dans *Eneas*.

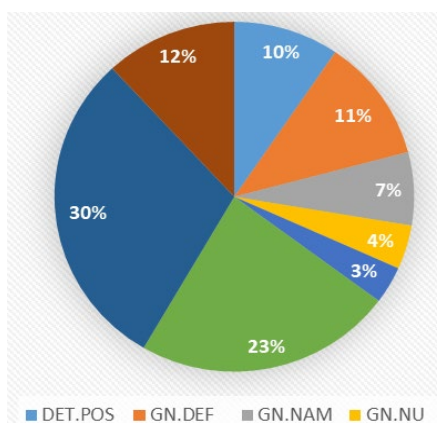


Figure 8. Catégories grammaticales des CR hors discours direct dans *Eneas*.

Dans la *Vie de sainte Bathilde* les écarts de fréquence sont particulièrement nets avec 20% de noms propres et GN définis en dehors du discours direct et seulement 5% dans le discours direct⁷. Comme dans les textes précédents les pronoms zéro et anaphoriques restent cependant très fréquents (27% des CR dans le discours direct, 40% en dehors) :

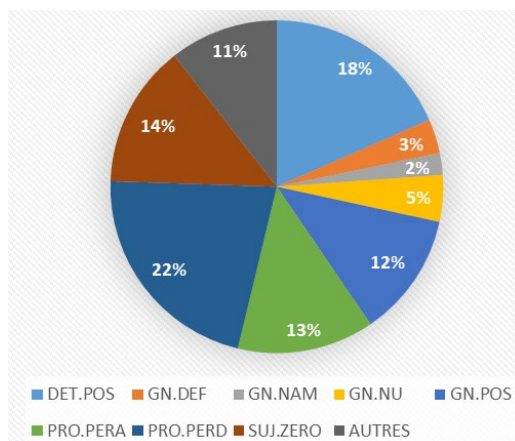


Figure 9. Catégories grammaticales des CR du discours direct dans *Sainte Bathilde*.

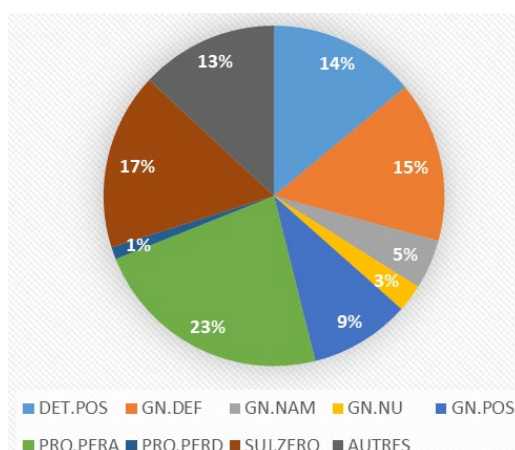


Figure 10. Catégories grammaticales des CR hors discours direct dans *Sainte Bathilde*.

On a déjà remarqué que les deux textes du XV^e siècle se distinguent par le fait que les catégories nominales y sont particulièrement fréquentes. En réalité, c'est surtout à l'extérieur du discours direct que leur sur-représentation est manifeste. Dans *Jehan de Paris*, 43% des occurrences hors discours direct sont des noms propres et GN définis, alors que les pronoms zéro et personnel ne représentent que 34% des occurrences. Dans le discours direct en revanche les deux catégories nominales ne totalisent plus que 14% des occurrences et les sujets zéro et pronoms anaphoriques sont plus utilisés qu'eux

⁷Il faut souligner que le discours direct est assez réduit dans ce texte (1653 mots), ce qui rend les données peu informatives.

(26%). On note par ailleurs que le nom propre est particulièrement utilisé dans les noms de pays (France, Angleterre, Espagne, etc.) qui entrent dans les titres de plusieurs personnages principaux désignés par des GN définis (le roi de France, le roi d'Angleterre, le roi d'Espagne etc.) :

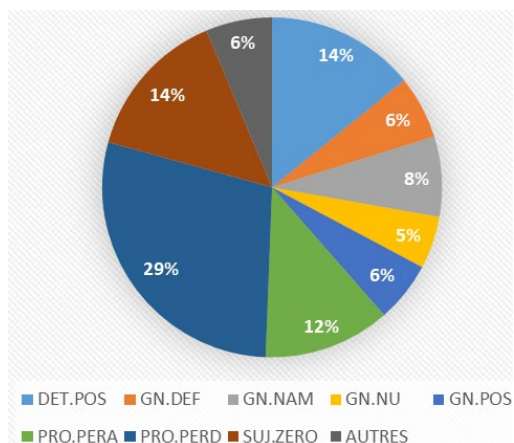


Figure 11. Catégories grammaticales des CR du discours direct dans *Jehan de Paris*.

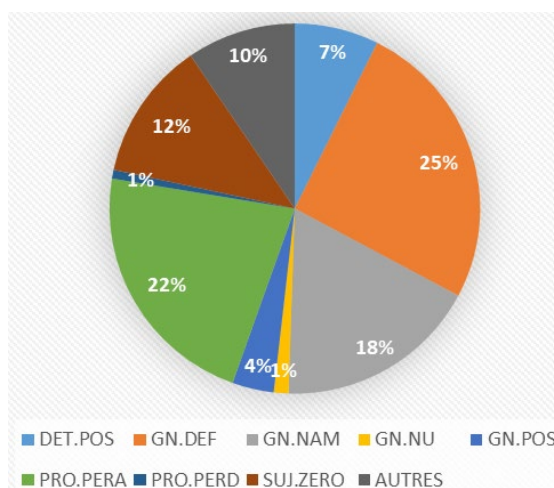


Figure 12. Catégories grammaticales des CR hors discours direct dans *Jehan de Paris*.

Le texte de Comynes comportant très peu de discours direct (265 mots), on se contentera ici de donner les proportions dans les parties non oralisées pour confirmer la sur-représentation du nom propre et du GN défini (34%) par rapport aux pronoms zéro et anaphorique (28%) :

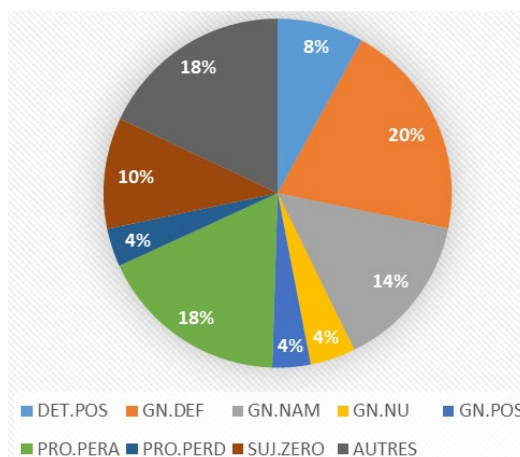


Figure 13. Catégories grammaticales des CR hors discours direct dans Commynes.

Ces données, qui semblent converger dans tous les textes du corpus, rejoignent les résultats d'autres recherches utilisant des outils statistiques pour contraster la fréquence des catégories grammaticales selon qu'on se situe dans le discours direct ou en dehors. Ces précédents travaux (Guillot *et al.*, 2015 ; Guillot-Barbance *et al.*, 2017) avaient déjà mis en évidence la sur-représentation du nom et ses déterminants dans les parties non oralisées des textes et un déficit corollaire des catégories liées au nom dans le discours direct. Notre étude apporte des informations supplémentaires en montrant que dans les textes narratifs ce sont surtout le nom propre et le GN défini qui sont concernés et que cette tendance se vérifie qu'on tienne compte des chaînes ou de toutes les mentions dans les textes. Pour des questions de place nous n'avons pas ajouté les graphiques correspondant aux mentions hors CR, mais dans ce cas aussi la proportion de noms propres et GN définis est plus forte à l'extérieur de l'oral représenté. Le GN défini dominant toujours davantage parmi les mentions hors CR que dans les chaînes, il faudrait une analyse statistique plus poussée pour voir si ces écarts de fréquence sont plus ou moins marqués selon qu'on se trouve dans le discours direct ou en dehors. Une telle analyse pourrait conforter le constat que ce qui est remarquable c'est bien la fréquence du GN défini et du nom propre dans les chaînes (par rapport aux autres mentions) à l'intérieur des parties non oralisées des textes (par rapport aux passages au discours direct).

4.2. Progression des CR au fil des textes

Le corpus annoté permet l'exploitation de plusieurs outils de visualisation disponibles dans la plateforme TXM, parmi lesquels la fonctionnalité « Progression ». Le graphique de progression présenté dans la Figure 14 permet de suivre au fil d'*Eneas* chaque mention des référents listés à droite et associés à une couleur. Chaque marche du graphique (axe vertical) correspond à une occurrence du référent dans le texte (axe horizontal) et les paliers marquent l'absence d'occurrence, c'est à dire une interruption de la chaîne qui peut éventuellement redémarrer plus loin.

Les expressions référentielles étant très nombreuses dans les textes, le graphique de progression d'un extrait entier aurait été totalement illisible. Chaque extrait de 10 000 mots a donc été divisé en tranches de 500 mots et les graphiques de progression ont été produits sur ces extraits⁸. On a pris en compte tous les référents entrant dans une chaîne à l'échelle de l'extrait complet, même si le nombre d'occurrences d'un élément était inférieur à 3 dans la portion de 500 mots examinée. La limite de 500 mots est totalement arbitraire et ne tient pas compte des divisions internes au texte mais elle a l'avantage d'être constante et de favoriser la comparaison. Elle a le désavantage de tronquer les chaînes, qui se retrouvent alors d'un graphique à l'autre :

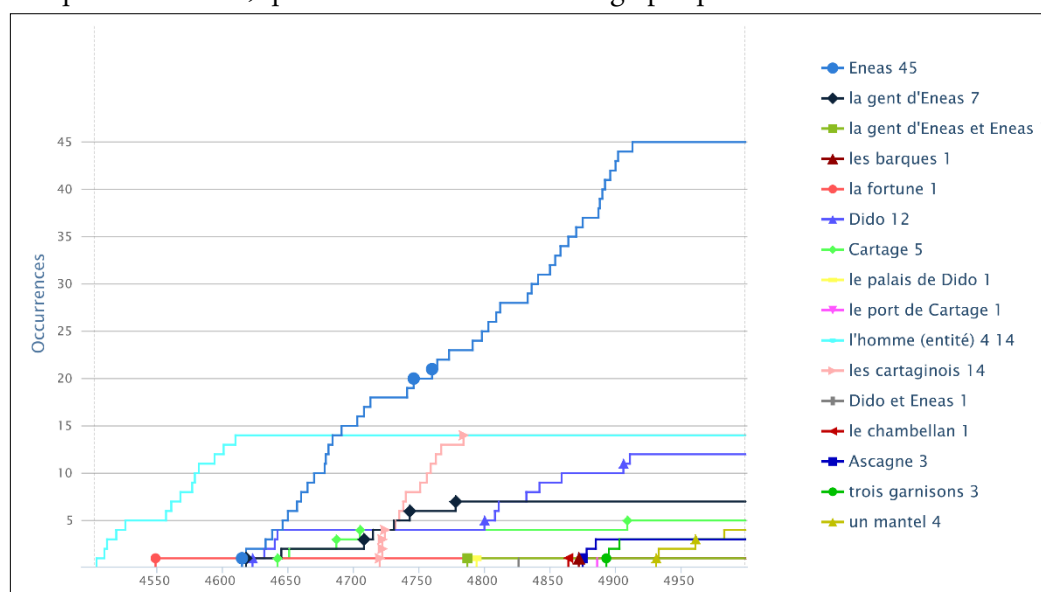


Figure 14. Graphique de progression des CR dans *Eneas* (mots 4500 à 5000).

Cet exemple de progression est assez représentatif de ce qu'on rencontre dans le corpus. Il fait bien apparaître la succession des chaînes et la domination d'une longue chaîne, dans le cas présent celle d'Enée qui se poursuit de manière constante des mots 4620 à 4920 environ. De manière très visible, les chaînes se différencient par leur longueur et la distance qui sépare leurs mentions successives, avec des paliers plus ou moins longs, plus ou moins nombreux. Les mentions isolées qui n'entrent dans aucune chaîne ne décollent pas de l'axe représentant les mots du texte. Les formes géométriques colorées (ronds, losanges et triangles) correspondent aux occurrences des expressions à tête lexicale (GN).

Ce type de graphique permet de visualiser de manière immédiate ce que Givón (1983) appelle des « topiques continus ». Ce sont les référents proéminents dont les chaînes sont relativement longues sur une portion de texte et dont les mentions sont très rapprochées sur cette portion. Mais il n'est pas très facile de distinguer clairement

⁸Nous avons analysé les graphiques de progression de 14 extraits de 500 mots dans chaque texte.

les textes du corpus en comparant visuellement tous les graphiques de progression. Deux textes qu'on a rapprochés plus haut du fait qu'ils présentent une très forte proportion de chaînes courtes et intermédiaires illustrent toutefois deux cas de figure opposés. Les extraits de la *Vie de sainte Bathilde* (Figure 15) se caractérisent le plus souvent par un assez petit nombre de chaînes longues et surtout par une ou deux chaînes très longues et compactes, dont les mentions sont à faible distance les unes des autres. Les extraits des *Mémoires de Commynes* (Figure 16) ont beaucoup de chaînes courtes et intermédiaires (entre 3 et 9 mentions) qui interfèrent avec des chaînes longues qui sont beaucoup moins longues et dont les mentions sont plus espacées ou denses sur une distance très limitée. Quant aux trois autres textes, ils se situent dans une zone intermédiaire entre ces deux extrêmes :

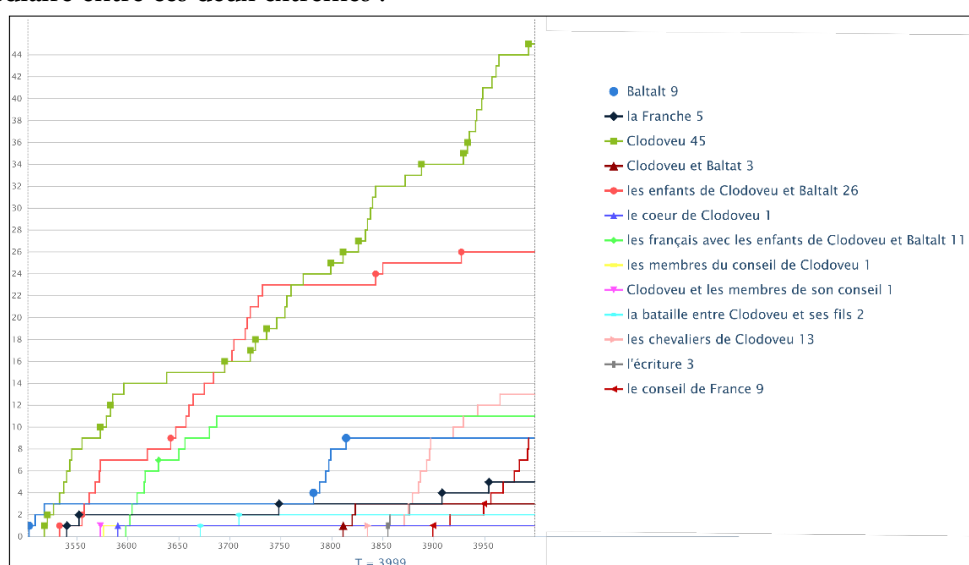


Figure 15. Graphique de progression des CR dans *Sainte Bathilde* (mots 3500 à 4000).

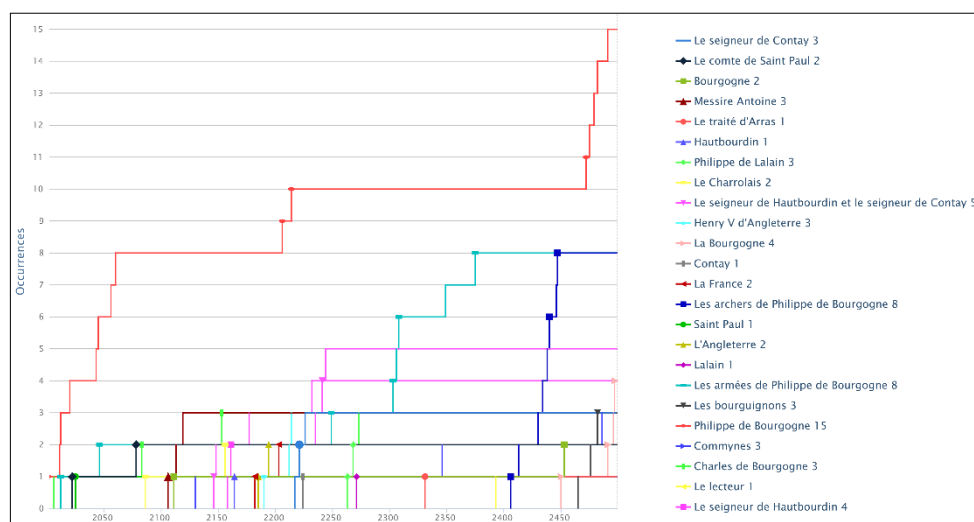


Figure 16. Graphique de progression des CR dans *Commynes* (mots 2000 à 2500).

4.3. Structuration des parties narratives du corpus à travers les chaînes

Dans ses nombreux travaux sur la progression thématique, B. Combettes (notamment 1983 et 2012) a souligné l'importance dans la prose narrative médiévale de la progression à thème constant, centrée sur un référent humain, sujet syntaxique des énoncés, à travers lequel les autres éléments sont généralement introduits. Ce type de progression fait d'un référent saillant et continu l'élément organisateur de l'unité narrative et produit un effet de resserrement grâce à des unités très cohésives et rapprochées :

C'est la progression à thème constant, *caractéristique de la narration* dans la prose narrative en ancien français, qui assure, si l'on peut dire, la dépendance du second plan et sa faible importance, ne serait-ce que du point de vue quantitatif. *Ce type de progression permet en effet à un référent saillant, d'ordinaire « personnage principal », d'apparaître comme thème et comme sujet syntaxique dans des énoncés successifs, et de servir ainsi de relais dans le maintien de la cohérence du passage.* C'est par l'intermédiaire de cet actant, par l'emploi de procès de perception ou d'action que vont se trouver introduits les référents nouveaux, supports éventuels d'une description, d'un commentaire. *Le syntagme sujet n'étant pas obligatoirement exprimé, la première place de la proposition se trouve d'ordinaire remplie par un marqueur qui renforce le lien contextuel (si, or, lors ...) et accentue l'impression d'une cohérence « étroite », resserrée* (Combettes, 2012 : 7, les italiques sont nôtres).

Ce constat nous semble pouvoir être étendu aux œuvres versifiées. En revanche, il paraît moins bien convenir aux textes les plus récents du corpus.

Nous commencerons notre enquête par *Eneas*. Comme l'a illustré la Figure 14, ce texte est marqué par la succession de CR diverses, qui peuvent dans certains cas s'interrompre et reprendre. Le (re)démarrage de chaque chaîne de taille moyenne ou longue se fait le plus souvent par la (ré)introduction d'un référent humain en proposition principale et en fonction syntaxique sujet *via* le nom propre ou un GN défini, en général en début ou fin de vers. Ce référent saillant domine un segment de discours par ses reprises successives, même si un ou plusieurs autres référents peuvent également apparaître et même s'il n'est pas nécessairement le sujet constant de tous les énoncés qui suivent. C'est surtout ce mode d'introduction d'un personnage nouveau ou qui revient sur le devant de la scène qui semble lui conférer une saillance particulière jusqu'à ce qu'un nouveau personnage, introduit de la même façon, vienne prendre sa place.

(2) *Li Greu* prenent les citeains, / nus n'eschapot d'entre *lor* mains / ne l'esteüst morir a honte [...] / Tote ert la vile mise en cendre. / Ocis i fu *li reis Prianz* / o *sa* fame, o *ses* anfanz : / unc ne fu mais tant grant ocise. / *Menelaus* a vanjance prise : / toz *fist* les murs aplanoier / por le tort fait de *sa* moillier. / A une part de

la cité / tint Eneas une erité, / de la vile bien grant partie. / Quant il a cele noise oïe, / si regarda vers lo donjon / et vit la grant destrucion ; / n'est mervoille s'il ot peor. / Venus la deesse d'amor, / qui est sa mere, li a noncié / que Troïen sont trebuchié ; / li deu an ont pris lor vanjance. / Comanda li, sanz demorance / s'en tort, ainz quel prengnent li Gré, / et ce li comandent li dé / que il aut la contree querre / dunt Dardanus vint an la terre, / qui fonda de Troie les murs. / Eneas n'estoit pas seürs (Eneas, v. 11-42).

Ce schéma nous semble remplir une fonction discursive essentielle dans une œuvre dépourvue de structuration interne explicite autre que la segmentation en vers. Ce mode de progression centré sur un référent mis en relief permet de créer des unités intermédiaires, de petite taille, entre le niveau global du texte et le niveau très local des unités syntaxiques et métriques. Bien qu'*Eneas* ait une proportion relativement faible de noms propres et de GN définis, ces derniers semblent donc jouer un rôle discursif de premier plan à l'intérieur des chaînes et des parties narratives du texte.

La *Vie de Sainte Bathilde* est moins marquée par la succession rapide de chaînes différentes au fil du discours. Le texte présente une très grande quantité de CR qui disparaissent très vite et quelques CR très longues, notamment celles de Clodoveu, Bathilde et leurs enfants. Les CR très longues, même si elles sont continues, peuvent également être scandées par des reprises nominales. Bien qu'ici la nomination d'un référent sujet marque moins souvent le passage d'un élément central à un autre, nous pensons qu'elle continue d'avoir pour effet de marquer le début d'une nouvelle unité narrative en réaffirmant la saillance d'un personnage qui peut rester le même. Il est possible que le caractère figé et répétitif des reprises lexicales (avec les N *roi* et *roïne* notamment) renforce leur fonction de balisage. Les redénominations s'accompagnent souvent aussi d'indications temporelles (en romain dans l'exemple), en particulier de subordinées en *quant*, qui accentuent le passage à une nouvelle séquence narrative. La reprise nominale peut alors se trouver dans la subordinée antéposée au verbe principal :

(3) *Li enfant qui eurent receu cele sentence par le jugement le roïne leur mere le souterrent a mout bone patience et se recorderent qu'il l'avoient bien deservi car il ne demoura mie en iaus qu'il n'ocheissent leur pere et prièrent Nostre Seigneur mout humlement qu'il ne fussent pour che pechié livré a pardurable torment et abandonnerent leurs cuers et leurs cors au servige leur createur en vegilles et en orisons et en aumosnes et en mout de maniere d'astinence de leurs cors et en tele maniere demorerent une pieche del tans en le maison leur pere. Li rois Clodoveu qui resgardeoit ses enfans qui tous jours se seoient et nule fois ne se levoient si en eut mout grant pitié et vint un jour a le roïne et li dist : « Ha ! Dame, comment porrons nos mais toute no vie souffrir et endurer a veir devant nos le fouleure de nos enfans ? [...] » Quant le roïne aperchut le grant debonnaireté le roi son seigneur et le naturel pitié qu'il avoit de ses enfans, si le commencha mout debonnairement a conforter et dist : « Beneois soit Nostres Sires, li Dex Israel, qui a visité*

*et racaté son pueple et qui le grant orguel de nos enfans a mué en si grant humilité [...] » Après ches coses fu **cele sainte roïne**⁹ en assiduelx orisons vers Nostre Seigneur qu'il par se grant misericorde leur demoustrast chou qu'il li plairoit qu'il fesissent de leurs enfans. Ne demora gaires après ces coses que **la roïne** vint a ses enfans et parla a iaus en cheste maniere : « Ha ! Comme est bons li Diex Israel a chiaus qui sont de droit cuer [...] » (Vie de sainte Bathilde, pp. 19-20).*

Roland se distingue des deux œuvres précédentes par la prégnance du discours direct et par le fait que l'alternance des tours de parole participe à l'organisation discursive. Cette alternance est marquée par un système d'éléments délimitant le début et la fin des prises de parole, un même énoncé pouvant marquer à la fois la fin du tour qui précède et le début du tour qui suit (*dient Franceis, respunt Rollant* dans l'exemple 4). Ces éléments, que nous avons appelés annonces internes dans une étude précédente (Guillot *et al.*, 2018 : 284) et qui semblent assez spécifiques aux chansons de geste, se conjuguent aux annonces (au début des séquences, *ço dist Rollant*) et aux incisives (au milieu des séquences) pour segmenter les dialogues. On a constaté déjà la fréquence remarquable du nom propre dans ces différents types d'énoncés (Guillot *et al.*, 2018 : 291-292)¹⁰. Toujours en fonction sujet, le nom propre permet d'indiquer de manière explicite l'identité du locuteur des paroles rapportées et peut initier une chaîne qui se poursuit dans le discours direct à travers des pronoms déictiques et des sujets zéro :

(4) « *Ço dist **Rollant** : « Ço ert Guenes, **mis** parastre. » / **Dient Franceis** : « Car il le poet ben faire ; / Se lui lessez, n'i trametrez plus saive. » / **E li quens Guenes** en fut mult anguisables. / De **sun** col getet **ses** grandes pels de martre / **E est remés** en **sun** blialt de palie. / Vairs **out** les oilz e mult fier lu visage ; / **Gent out** le cors e les costez **out** larges ; / Tant par **fut** bels tuit **si** per l'en esguardent. / **Dist a Rollant** : « Tut fol, pur quei t'esrages ? / Ço set hom ben que **jo** sui tis parastres, / Si as juget qu'a Marsiliun en **alge**. / Se Deus ço dunet que **jo** de la repaire, / **Jo** t'en muvra une si grant contraire / **Ki** durerat a trestut tun edage. » / **Respunt Rollant** : « Orgoill **oi** e folage / Ço set hom ben, n'ai cure de manace ; / Mais saives hom, il deit faire message : / Si li reis voelt, prez **sui** por vus le **face** ! » (Chanson de Roland, v. 277-295).*

Ce type d'organisation discursive nous semble se conjuguer à la progression à thème constant évoquée ci-dessus (voir la CR de Guenes dans l'exemple). Le nom propre et dans une moindre mesure le GN défini peuvent donc servir dans ce texte

⁹Le GN démonstratif semble jouer un rôle proche de celui des GN définis, mais il faudrait naturellement une étude plus poussée pour préciser la fonction des démonstratifs dans le texte. Ce qui favorise le rapprochement est le fait que le GN apparaît en fonction sujet, qu'il entraîne la redénomination du référent et qu'il suit un complément temporel.

¹⁰L'étude de 2018 portait plus spécifiquement sur les incisives mais la fréquence du nom propre caractérise également les annonces.

aussi à mettre au premier plan l'élément organisateur du récit¹¹. Cette segmentation narrative interne semble parallèle à la division du texte en laisses, les laisses pouvant s'organiser autour d'un nombre très variable de référents alternativement saillants.

On a souligné plus haut le fait que le *Roman de Jean de Paris* comporte un grand nombre de personnages importants. Comme c'est généralement le cas dans les autres textes, les référents des GN définis sont soit des éléments qui disparaissent très vite (mentions hors chaînes), soit au contraire les personnages les plus présents dans le récit. Ce qui distingue assez nettement ce texte des autres est le fait que certaines chaînes ont plus de GN définis que de pronoms personnels anaphoriques et sujets zéro. L'examen des sept chaînes maximales montre que c'est le cas du roi d'Angleterre, du roi et de la reine d'Espagne. Le héros éponyme du récit et les souverains d'Espagne échappent à cette règle. Dans les chaînes du roi et de la reine de France, le nombre de GN définis est inférieur à la somme des pronoms mais reste exceptionnellement élevé :

	Jehan de Paris	Roi d'Angleterre	Roi de France	Roi d'Espagne	Souverains d'Espagne	Reine de France	Reine d'Espagne
DET.POS	66	30	29	15	14	14	12
GN.DEF	17	86	74	39	12	18	22
GN.DEM	3	0	0	0	0	1	1
GN.IND	2	0	1	0	0	0	0
GN.NAM	52	0	0	1	0	0	0
GN.NU	8	11	8	6	3	2	1
GN.POS	18	4	2	9	3	5	3
PRO.ADV	1	0	0	1	0	0	0
PRO.DEM	1	0	0	0	0	0	0
PRO.PERA	114	52	41	13	20	10	11
PRO.PERD	50	37	45	13	19	5	7
PRO.POS	2	1	0	0	0	0	0
PRO.REL	9	11	8	4	3	6	3
SUJ.ZERO	43	31	39	15	33	9	4

Tableau 7. Fréquence des catégories grammaticales des 7 CR les plus longues de *Jehan de Paris*.

Le récit mettant en scène plusieurs rois et reines qui sont alternativement saillants et qu'il s'agit de bien distinguer, on peut faire l'hypothèse que l'omniprésence des expressions nominales répond principalement à la volonté de réduire les risques d'ambiguïté. Lorsque le GN défini *le roi*, sans autre précision, est utilisé, c'est en général le roi de France qui est visé et le GN n'est pas réellement ambigu. Bien qu'il soit incomplet, il remplit la même fonction. Dans l'exemple 5, le titre (en romain) permet

¹¹ On peut hésiter à analyser les éléments permettant de baliser les propos rapportés comme faisant partie du discours direct lui-même (c'est la thèse de Marchello-Nizia, 2012) ou de la narration. Dans l'encodage du corpus, ces éléments ont été exclus du discours direct. Leur fonction organisationnelle en revanche ne fait pas de doute.

d'interpréter les deux premières occurrences du GN *le roi*, les suivantes renvoient toutes au roi de France :

(5) Comment **le roy de France** print congié *du roy d'Espagne* et de la royne, et comment le peuple du pays acompaignerent **le roy** grant chemin. *Toutesfoiz, pour abregier, le roy se despartit d'Espagne a grans pleurs et lamentations du roy, de la royne et de ceulx du pays, qui l'accompaignerent grant espace de temps, et donna le roy d'Espagne de grans et riches dons au roy et aux barons et chevaliers, tellement qu'il n'y avoit en toute l'armee petit ne grant qui ne s'en louast, et qui ne tint le roy d'Espagne a vaillant et puissant roy. Si firent tant par leurs journees qu'ilz vindrent a Paris ou ilz furent moult joyeusement et honorablement receuz, et dura la feste dix jours. Puis print chascun congié du roy qui bien les envoya joyeux et contans en leurs maisons* (*Roman de Jehan de Paris*, p. 16).

Ces éléments tendent à montrer que la progression thématique et la distribution des catégories grammaticales n'obéissent pas aux mêmes contraintes que dans les textes plus anciens. La structuration du texte en unités intermédiaires dotées d'un titre est à mettre en relation avec le fait que les expressions nominales n'y sont pas investies du même rôle discursif, la fonction de désambiguïisation semblant passer au premier plan, alors même que le nombre de référents humains est particulièrement important dans le récit. La fréquence particulière du GN défini est quant à elle surtout liée à la nature des référents les plus centraux (rois et reines) et à l'importance des titres.

Les *Mémoires* de Commynes ayant un nombre exceptionnellement élevé de chaînes courtes et intermédiaires, très peu de chaînes vraiment longues et un nombre particulièrement élevé de référents humains de sexe masculin, les risques d'ambiguïté référentielle sont amplifiés. Les tendances observées précédemment dans les chaînes les plus longues sont présentes mais sous forme atténuée : la chaîne du roi de France a plus de GN définis et noms propres (63) que de pronoms (48), mais les fréquences sont presque identiques dans celle du duc Philippe de Bourgogne (23 et 22 occurrences) et les rapports sont inversés dans la chaîne de Charles de Bourgogne, la plus longue de tout l'extrait (59 / 139 occurrences). Ce qui frappe surtout à la lecture de ce texte, c'est la profusion de référents instanciés à faible distance et la fréquence de l'anaphorique *ledit*. S'il est probable qu'il vise essentiellement à éviter les confusions entre entités concurrentes, sa multiplication devant les N *duc* et *comte*, parfois employés seuls, a pour conséquence qu'il remplit de manière assez imparfaite sa fonction désambiguïsante :

(6) *Ledit évesque* de Narbonne feit ce messaige au roy, quand il fut de retour, comme vous entendrez cy après. Ces parolles engendrèrent grand hayne *dudict conte de Charroloys* au roy, avec ce qu'il n'y avoit guères que le roy avoit rachapté les villes de dessus la rivière de Somme, comme Amyens, Abeville, Saint Quentin et autres, baillées par le roy Charles [Viiè] *audict duc Philippe de Bourgogne* par le traictié qui fut fait à Arras, pour en joyr par luy et ses hoirs masles, au rachapt de quatre cens mille escus. *Toutesfois, ledit duc* se trouvant en sa vieillesse, furent conduyctz tous ses affaires par monsr de Crouy et de Chimay, frères, et autres de

leur maison : reprint son argent du roy et restitua lesdictes terres, dont ledict conte, son filz, fut fort troublé, car c'estoient les frontières et limytes de leurs seigneuries, et y perdoient beaucoup de subjectz, bonnes gens pour la guerre (Philippe de Comynnes, *Mémoires*, t.1, p. 8-9).

L'exemple 6 montre que la progression à thème continu, centrée sur un référent dominant un segment discursif, n'est pas habituelle dans ce texte. Il semble que les GN définis et noms propres, qu'ils soient ou non précédés de *ledit*, répondent plus à une exigence d'identification référentielle explicite qu'à une fonction de structuration discursive.

5. Conclusion

L'étude des chaînes de référence, comme de tout phénomène textuel, pose de grands défis au linguiste, tant les facteurs mis en jeu sont multiples (facteurs morphologiques, syntaxiques, sémantiques, discursifs et cognitifs). Et même si les corpus numériques décuplent les possibilités d'analyse, il reste à inventer une méthodologie permettant de mettre au jour les éléments les plus pertinents, les patrons récurrents, les associations à privilégier.

Bien que la recherche présentée ici soit encore très partielle, elle s'appuie sur un cadre utilisé dans d'autres études, qui permet de comparer les résultats et d'affiner la méthode. L'annotation d'un corpus composé d'extraits de taille proche permet par exemple de faire apparaître des constantes dans la longueur des chaînes, les chaînes courtes (3 à 4 mentions) étant toujours les plus nombreuses dans tous les textes. Les différences ressortent elles aussi de manière très nette, les fréquences de ces chaînes courtes pouvant varier dans des proportions importantes selon les types de textes et les périodes (entre 30 et 40% des CR dans notre corpus, un pourcentage plus élevé dans une étude portant sur les *Essais sur la peinture* de Diderot).

Grâce aux outils développés dans la plateforme TXM, la mise en œuvre de ce cadre d'analyse a permis de mettre en évidence plusieurs spécificités de notre corpus narratif : large couverture des textes par les expressions référentielles (mentions) et les chaînes, importance des référents animés humains, fréquence du sujet zéro dans les textes les plus anciens, baisse des GN nus dans les textes les plus récents, etc. Les résultats qui nous ont semblé les plus marquants concernent la fréquence et le rôle des marques de faible et moyenne accessibilité référentielle (nom propre et GN définis) à l'intérieur des chaînes. Leur importance dans les parties narratives (hors discours direct) a été soulignée et mise en relation avec les fonctions référentielle et discursive qu'elles peuvent assurer dans la narration. Des hypothèses ont ainsi émergé sur les relations possibles entre les catégories grammaticales composant les chaînes (marques de forte ou faible accessibilité référentielle), la façon dont ces chaînes progressent (topiques plus ou moins continus), les modes d'enchaînements thématiques (progression à thème contant ou non) et la structuration interne du matériau discursif en unités supérieures

et intermédiaires. Des corrélations ont ainsi pu être établies entre le matériau des chaînes et l'évolution des modes de structuration textuelle, l'écrit développant au fil du temps un système de divisions internes plus explicite et plus élaboré (livres, chapitres, titres etc.). Au terme de l'étude, un début de diachronie semble se dessiner, que le développement de la linguistique textuelle diachronique que nous appelons de nos vœux pourra éventuellement confirmer.

RÉFÉRENCES BIBLIOGRAPHIQUES

a) Corpus

BFM = *Base de français médiéval* : Lyon, École Normale Supérieure. URL : <http://bfm.ens-lyon.fr/>.

Chanson de Roland (1972) : Édition de Gérard Moignet, Paris, Bordas. Publié en ligne par la BFM. URL : <http://catalog.bfm-corpus.org/roland>. Dernière révision le 24/02/2014.

COMMYNES, Philippe de (1924) : *Mémoires*. Édition de Joseph Calmette, Paris, Belles Lettres, t.1. Publié en ligne par la BFM. URL : <http://catalog.bfm-corpus.org/commyn1>. Dernière révision le 21/04/2016.

Eneas (1925) : Édition de Jean-Jacques Salverda de Grave, Paris, Champion, t.1. Publié en ligne par la BFM. URL : <http://catalog.bfm-corpus.org/eneas1>. Dernière révision le 30/11/2011.

Roman de Jean de Paris (1923) : Édition d'Edith Wickersheimer, Paris, Champion. Publié en ligne par la BFM. URL : <http://catalog.bfm-corpus.org/jehpar>. Dernière révision le 30/05/2000.

Vie de sainte Bathilde (Version I) (1996) : Édition d'Anders Bengtsson, Lund, Lund University Press. Publié en ligne par la BFM. URL : <http://catalog.bfm-corpus.org/SBath1>. Dernière révision le 30/05/2016.

b) Études

ADAM, Jean-Michel (2018) : *Le paragraphe : entre phrases et texte*. Paris, Armand Colin.

ARIEL, Mira (1990) : *Accessing Noun-Phrase Antecedents*. London, Routledge.

COMBETTES, Bernard (1983) : *Pour une grammaire textuelle : la progression thématique*. Bruxelles et Paris, De Boeck & Duculot.

COMBETTES, Bernard (2012) : « Linguistique textuelle et diachronie », in *Actes du 3^e Congrès Mondial de Linguistique Française*. EDP Sciences, 3-10. DOI : <https://doi.org/10.1051/shsconf/20120100344>

CORBLIN, Francis (1985) : « Remarques sur la notion d'anaphore ». *Revue québécoise de linguistique*, 15: 1, 173-195.

- Democrat (DEscription et MOdélisation des Chaînes de Référence : outils pour l'Annotation de corpus (en diachronie et en langues comparées) et le Traitement automatique*. URL : <https://www.ortolang.fr/market/corpora/democrat/v1.1>
- GIVÓN, Talmy (1983) : « Topic Continuity in Discourse : an Introduction », in Talmy Givón (éd.), *Topic Continuity in Discourse. A Quantitative Cross-Language Study*. Amsterdam/Philadelphia, John Benjamins Publishing Company, 5-41.
- GLIKMAN, Julie; Céline GUILLOT-BARBANCE & Vanessa OBRY (2014) : « Les chaînes de référence dans un corpus de textes narratifs médiévaux : traits généraux et facteurs de variation ». *Langages*, 195, 43-60.
- GUILLOT, Céline; Serge HEIDEN, Alexei LAVRENTIEV & Bénédicte PINCEMIN (2015) : « L'oral représenté dans un corpus de français médiéval (9^e-15^e) : approche contrastive et outillée de la variation diasystémique », in Kerstin Jeppesen Kragh & Jan Lindschouw (éds), *Les variations diasystémiques et leurs interdépendances dans les langues romanes. Actes du Colloque DIA II à Copenhague (19-21 nov. 2012)*. Strasbourg, Éditions de linguistique et de philologie, 15-27.
- GUILLOT-BARBANCE, Céline & Matthieu QUIGNARD (2019) : « Chaînes de référence et structure textuelle dans les *Essais sur la peinture* de Diderot ». *Discours*, 25, DOI : <https://doi.org/10.4000/discours.10421>
- GUILLOT-BARBANCE, Céline; Alexei LAVRENTIEV; Serge HEIDEN & Bénédicte PINCEMIN (2018) : « Diachronie de l'oral représenté : délimitation et segmentation interne du dialogue (IX^e-XV^e siècle) », in Wendy Ayres-Bennett et al. (éds), *Nouvelles voies d'accès au changement linguistique. Actes du colloque de la SIFD*, Paris, Classiques Garnier, 279-296.
- GUILLOT-BARBANCE, Céline, Bénédicte PINCEMIN & Alexei LAVRENTIEV (2017) : « Représentation de l'oral en français médiéval et genres textuels ». *Langages*, 208, 53-68.
- HEIDEN, Serge (2019) : *Manuel de TXM, Extension Annotation URS (Unité-Relation-Schéma) version 1.0*. DOI : <http://doi.org/10.5281/zenodo.3267345>.
- HEIDEN, Serge ; Jean-Philippe MAGUE & Bénédicte PINCEMIN (2010) : « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », in Sergio Bolasco, Isabella Chiari & Luca Giuliano (éds), *Proc. of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, Roma, Edizioni Universitarie di Lettere Economia Diritto, vol. 2, 1021-1032. URL : <https://doi.org/10.1051/shsconf/20120100344>.
- MARCHELLO-NIZIA, Christiane (2012) : « L'oral représenté : un accès construit à une face cachée des langues 'mortes' », in Céline Guillot et al. (éds), *Le changement en français. Études de linguistique diachronique*. Bern, Berlin et Bruxelles, Peter Lang, 247-264.
- OBRY, Vanessa; Julie GLIKMAN; Céline GUILLOT & Bénédicte PINCEMIN (2017) : « Les chaînes de référence dans les récits brefs en français : étude diachronique (XIII^e - XVI^e s.) ». *Langue française*, 195, 91-110.
- SCHNEDECKER, Catherine (1997) : *Nom propre et chaînes de référence*. Paris, Klincksieck.

- SCHNEDECKER, Catherine (2017) : « Les chaînes de référence : une configuration d'indices pour distinguer et identifier les genres textuels ». *Langue française*, 195, 53-72.
- SCHNEDECKER, Catherine & Frédéric LANDRAGIN (2014) : « Les chaînes de référence : présentation ». *Langages*, 195, 3-22.
- SCHNEDECKER, Catherine, GLIKMAN, Julie et Frédéric LANDRAGIN [éds] (2017) : *Langue française*, 195 [*Les chaînes de référence en corpus*].
- Textométrie* (2007-2010) : *Fédération des recherches et développements en textométrie autour de la création d'une plateforme logicielle ouverte*. Lyon, ENS. URL : <http://textometrie.ens-lyon.fr>.