

Les chaînes de référence dans les récits brefs en français : étude diachronique (XIII e – XVI e s.)

Vanessa Obry, Julie Glikman, Céline Guillot-Barbance, Bénédicte Pincemin

► **To cite this version:**

Vanessa Obry, Julie Glikman, Céline Guillot-Barbance, Bénédicte Pincemin. Les chaînes de référence dans les récits brefs en français : étude diachronique (XIII e – XVI e s.). Langue française, Armand Colin, 2017, 195, pp.91 - 110. hal-01598911

HAL Id: hal-01598911

<https://hal.archives-ouvertes.fr/hal-01598911>

Submitted on 30 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les chaînes de référence dans les récits brefs en français : étude diachronique (XIII^e – XVI^e s.)

Vanessa OBRY

Université de Haute-Alsace & Institut de recherche en Langues & Littératures Européennes – ILLE (EA 4363)

Julie GLIKMAN

Université de Strasbourg & Linguistique, Langues, Parole – LiLPa (EA 1339)

Céline GUILLOT-BARBANCE

ENS Lyon & Institut d'Histoire des Représentations des Idées dans les Modernités – IHRIM (UMR 5317)

Bénédictine PINCEMIN

CNRS & Institut d'Histoire des Représentations des Idées dans les Modernités – IHRIM (UMR 5317)

Résumé

Cet article s'intéresse aux chaînes de référence dans un corpus de textes narratifs brefs à tonalité comique composés entre le début du XIII^e et le milieu du XVI^e siècle. Seuls les référents animés humains ont été pris en compte dans l'analyse. Le corpus a été manuellement annoté grâce au logiciel ANALEC et analysé *via* la plateforme TXM. Nous décrivons les paramètres de variation des chaînes de référence en proposant un ensemble de mesures permettant la comparaison des résultats avec des études sur d'autres genres et d'autres périodes : la densité référentielle, le nombre de chaînes, leur longueur, le coefficient de stabilité référentielle ou la diversité des désignations des référents, la catégorie grammaticale des maillons et la nature du premier maillon. L'analyse outillée du corpus permet de dégager plusieurs spécificités et évolutions des chaînes de référence dans les genres brefs.

Mots –clés

chaîne de référence ; genre textuel ; récits brefs ; diachronie ; animés humains

Abstract

Coreference chains in French short narratives : a diachronic study (13th-16th c.)

This paper investigates coreference chains in a corpus of short humorous narrative texts in French from the early 13th to the mid-15th century. The study is restricted to human referents. Using ANALEC to manually encode our description, we analyze our data by means of the TXM software. The study establishes the variation parameters that allow for a comparison of our results with the results obtained in other studies: so-called 'referential density', the number of chains, their lengths, the referential stability coefficient or the variety in the denominations of the referents, the grammatical category of the links in the coreferential chains and the part-of-speech of the first mention. The analysis of the corpus by means of the tools mentioned above enables us to establish a number of specific characteristics and diachronic changes in coreferential chains in short texts

Key words

coreference chains ; textual genre ; short texts ; diachronics ; human referents.

1. Introduction

Les chaînes de référence (désormais CR) dans les états anciens du français et leur évolution dans les textes écrits nous sont très mal connues. Notre étude repose sur une approche diachronique et se centre sur les référents animés humains dans les genres narratifs brefs. Elle complète une recherche antérieure (Glikman, Guillot-Barbance & Oby 2014) qui comparait récits longs et brefs, textes en vers et en prose, ainsi que plusieurs genres textuels (roman, miracle, vie de saint, traduction biblique) dans un corpus synchronique de la fin du XII^e et du début du XIII^e siècle. Cette première analyse a conduit à mettre en relation les types de CR et d'expressions référentielles (désormais ER) entrant dans leur composition avec le degré de saillance locale et globale de chaque entité référentielle. Les textes narratifs longs, qui peuvent instancier de manière simultanée un grand nombre de référents humains et qui comportent plusieurs référents principaux, se démarquent assez nettement des récits dont la taille et le nombre de référents sont plus restreints. La présente étude est consacrée à cette seconde catégorie.

Notre recherche s'effectue dans le cadre du projet DEMOCRAT (Landragin 2016). Elle adopte la définition des CR mise en place dans le projet, à savoir la suite de toutes les expressions coréférentielles d'un texte. Le passage d'une unité textuelle à une autre (changement de paragraphe, passage au discours direct, etc.) et le choix des types d'ER (le nom propre, p. ex.) n'entraînent pas de rupture ni le démarrage d'une nouvelle CR. Ces choix paraissent les plus prudents s'agissant de textes anciens dont la structuration et la mise en page obéissent à des contraintes différentes de celles de l'imprimé moderne et compte tenu du caractère fragmentaire de nos connaissances sur la valeur sémantique des types d'ER dans les états de langue anciens (pronom zéro, nom propre, etc.). Nous avons en revanche considéré que la CR avait une limite inférieure et commençait à partir de trois mentions (ou maillons), les suites de deux mentions coréférentielles se laissant décrire au travers du mécanisme de l'anaphore (Schneidecker 1997).

Notre étude se base sur l'approche typologique, par genre textuel, développée par les membres du projet DEMOCRAT (voir notamment Schnedecker & Longo 2012, Schnedecker 2014 et ce volume). Elle vise à nous renseigner sur les caractéristiques typologiques des CR et sur leur évolution diachronique dans les récits brefs anciens, qui se transforment, eux aussi, au cours de la période étudiée.

2. Corpus et méthodologie de recherche

2.1. Traits génériques des récits brefs : continuité et variations diachroniques

Nous avons choisi d'analyser des textes relevant de genres narratifs brefs à tonalité comique, représentés par le fabliau au XIII^e siècle, puis par la nouvelle aux XV^e et XVI^e siècles. Notre corpus se compose ainsi de trois parties :

- deux fabliaux attribués à Jean Bodel et datant des premières années du XIII^e siècle : *Le Vilain de Bailleul* (désormais *Vilain*) et *Gombert et les deux clercs* (désormais *Gombert*) ;
- la première nouvelle du recueil anonyme des *Cent nouvelles nouvelles* (désormais *CNN*), composées vers 1460 ;
- les deux premières nouvelles des *Nouvelles recreations et joyeux devis* de Bonaventure des Périers (désormais *Des Périers*), datant de 1561¹.

Ces textes ont en commun la brièveté, la dimension narrative et une même tonalité, qui justifient leur regroupement. Leur taille relève du même ordre de grandeur, si on les compare notamment aux romans qui leur sont contemporains : *Vilain* et *Gombert* comportent respectivement 720 et 1 210 mots, la nouvelle des *CNN* 2 307 mots et les deux nouvelles de *Des Périers* font au total 2 115 mots, mais se rapprochent par leur longueur respective des fabliaux (899 et 1 216 mots), la première nouvelle se composant elle-même de trois histoires successives.

La brièveté des textes est à relier au rapport qu'ils entretiennent à l'oralité. Les fabliaux sont conçus pour être diffusés oralement. Les nouvelles, ultérieures et regroupées en recueil, sont davantage dépendantes de la structure du livre, mais elles gardent néanmoins des traces de cette oralité : les *CNN*, écrites sur le modèle du *Décameron* de Boccace, enchâssent les nouvelles dans un récit cadre et chacune est racontée, oralement, par un personnage ; le texte de *Des Périers* porte aussi des marques d'oralité, notamment du fait des fréquentes interventions du narrateur.

L'homogénéité du corpus doit cependant être nuancée. Les fabliaux ont circulé indépendamment les uns des autres : les deux textes de notre corpus ne se trouvent pas toujours dans le même manuscrit, et ils ne se suivent pas. Les nouvelles appartiennent à un genre distinct qui se développe, en prose, à la fin du Moyen Âge et fait apparaître d'autres habitudes narratives. Le rôle du recueil dans l'émergence d'une conscience du genre de la nouvelle au Moyen Âge a bien été démontré (Labère 2006). Les *CNN*, premier recueil de nouvelles en français, en sont des témoins et *Des Périers*, à la Renaissance, s'inscrit dans cette lignée. L'intégration des nouvelles dans les recueils a notamment des conséquences sur la structuration des textes et sur le statut des titres. Il est cependant reconnu que la nouvelle de la fin du Moyen Âge est l'héritière directe des genres brefs précédents, dont le fabliau fait partie (Dubuis 1973, Labère 2006, Vélissariou 2012), et que la diffusion des fabliaux n'est pas étrangère à la structure du recueil (Collet, Gingras & Trachsler 2012). Une autre différence formelle réside dans l'opposition du vers et de la prose. Le passage du vers, employé dans le fabliau, à la prose des nouvelles correspond à l'évolution générale des genres narratifs au Moyen Âge.

2.2. Caractéristiques et structure des textes

2.2.1. Statut des titres

Tels qu'ils nous sont parvenus, les fabliaux sont précédés d'un titre. Mais celui-ci est souvent repris par l'éditeur moderne à une rubrique de manuscrit, située au début du texte ou à la fin, et variant d'un manuscrit à l'autre. Si cette rubrique peut programmer la lecture lorsqu'elle est présente au début de l'œuvre, il semble plus logique de considérer le titre comme extérieur au fabliau lui-même et ayant peu d'influence sur la composition des CR. Il ne permet pas, comme celui des portraits journalistiques étudiés par C. Schnedecker, de « prédire une partie du matériau lexical qui sera utilisé dans les chaînes de référence » (2005 : 93). Dans les *CNN*, les nouvelles sont numérotées et ne comportent pas de titre. Le début de la nouvelle, précédé de l'indication du personnage qui la raconte, coïncide avec le début de la prise de parole de ce dernier. Les nouvelles de *Des Périers* comportent, en revanche, des titres : *Des troys folz*, *Caillette*, *Triboulet* et *Polite* et *Du chancre, bassecontre de Saint Hilaire de Poitiers, qui accompagna les chanoines à leurs potages* : ceux-ci présentent une première fois les référents humains principaux et peuvent exercer une influence sur les CR.

2.2.2. Organisation narrative, types de référents et de situations

La brièveté des fabliaux et de la nouvelle des *CNN* implique une forte cohésion et une simplicité de la trame narrative, plus nette encore dans les textes du XIII^e siècle. Parmi les critères distinctifs des genres narratifs brefs médiévaux, P. Zumthor retient l'unité de l'événement narré : l'action est tendue vers sa résolution, la fin constitue la « pointe » du récit et la conclusion résout complètement l'intrigue (Zumthor, 1983 : 7-8). Le récit s'organise autour d'un ou de quelques événements centraux. Les personnages, en nombre restreint, se rattachent à des types traditionnels (le prêtre, le paysan, le bourgeois, l'épouse...) et ne sont pas toujours nommés.

Le texte de *Des Périers* se distingue parce qu'il contient quatre histoires dans deux nouvelles. La première nouvelle met en scène successivement chacun des personnages mentionnés dans le titre – *Caillette*, *Triboulet* et *Polite* – dans trois histoires distinctes. La seconde nouvelle réunit dans une même histoire les personnages également mentionnés dans le titre. Si les personnages centraux se

¹ Dans le recueil, il s'agit des deuxième et troisième textes numérotés, ils sont précédés d'une « Première Nouvelle en forme de préambule ».

rattachent eux aussi à des types – en particulier les fous de la première nouvelle –, ils sont cependant plus individualisés que ceux des fabliaux et des *CNN*. La présence du nom propre en constitue un indice important.

Nos textes se fondent sur les ressorts traditionnels du comique que sont la tromperie, l'adultère et le quiproquo : les personnages masquent leur identité, se font passer pour d'autres, font passer les autres pour coupables d'actes qu'ils ont eux-mêmes commis. La récurrence de ces situations, qui posent, à l'échelle de l'intrigue, des questions de référence, nous invitera à nous demander quel est l'impact de ce facteur d'ordre stylistique sur la construction des CR.

2.3. Méthodologie : annotation et analyses contrastives

Notre étude permet la comparaison avec d'autres genres médiévaux et avec des textes modernes de genre approchant. Nous nous intéresserons en particulier aux travaux portant sur les CR des faits divers (Schneedecker & Longo 2012, Schneedecker 2014). Il s'agit d'écrits narratifs également brefs, bien que les récits médiévaux soient un peu plus longs que les faits divers (en moyenne 208 mots) ; leur action est unique et simple, entièrement résolue à la fin du texte et met en scène un nombre restreint de référents humains rarement nommés ; les individus sont stéréotypés voire assignés à des « rôles thématiques » (« le petit voyou, le criminel par accident, la victime âgée, etc. », Schneedecker & Longo, *op. cit.* : 1960).

Dans notre corpus, nous avons repéré les ER et annoté leurs caractéristiques à l'aide du schéma d'annotation commun défini au sein du projet MC4 (Landragin & Schneedecker 2014 ; Mélanie-Becquet & Landragin 2014) avec le logiciel ANALEC (Landragin, Poibeau & Victorri, 2012). Toutes les mentions d'animés humains ont été annotées. La catégorie grammaticale de chaque expression a été encodée (GN défini, nom propre, pronom personnel, etc.). Les suites de maillons dépassant trois occurrences ont été reliées par des CR associées au nom du référent. La cohérence de l'annotation sur l'ensemble du corpus a fait l'objet d'une vérification croisée et de tests de cohérence.

Au fil des travaux précédents (cf. les études de Schneedecker citées en bibliographie), l'analyse des CR a permis d'identifier plusieurs paramètres de variation qui semblent pertinents pour une typologie des CR par genre. L'annotation tient compte de ces paramètres. Elle est homogène et exhaustive pour les animés humains, ce qui permet de quantifier les phénomènes observés. Un ensemble de six mesures, directement inspirées de travaux antérieurs, a été conçu pour caractériser les CR à partir des annotations et en fonction des paramètres de variation. Chaque mesure a été implémentée dans la plateforme TXM (Heiden, Magué & Pincemin 2010) sous la forme d'une macro dont l'utilisateur peut ajuster les paramètres. Ces développements ont été réalisés dans le cadre d'un chantier du projet DEMOCRAT qui consiste à intégrer les fonctionnalités d'ANALEC dans TXM. Les annotations réalisées dans ANALEC ont donc été importées dans TXM puis analysées par de nouveaux outils d'exploitation pour produire les mesures souhaitées.

Nous présentons dans les sections suivantes les paramètres de variation et les mesures associées, et nous analysons, pour chaque mesure, les résultats fournis par le corpus. Notre objectif est double : décrire la typologie des CR dans nos textes en la comparant à d'autres et progresser dans la définition d'une méthodologie adaptée à l'analyse des CR et reproductible.

3. Mesures générales : présence des ER et des CR dans les textes

3.1. Densité référentielle

La densité référentielle est calculée en divisant le nombre d'ER par le nombre de mots du texte. Les ER prises en compte sont de tout type, singletons, suites de deux expressions coréférentielles ou expressions entrant dans des CR de référents animés humains.

Nous faisons l'hypothèse que la densité référentielle dans les récits brefs anciens sera assez proche de celle des faits divers, dans lesquels 19% des mots sont des expressions référant à des animés humains (Schneedecker 2014) : il s'agit en effet dans les deux cas de récits brefs dans lesquels les référents animés humains occupent la place de premier plan.

Dans notre corpus, la mesure de la densité référentielle par texte est assez stable, autour de 19,6% pour Des Périers, 21,6% pour *Gombert* et 23,6% pour *Vilain* ; elle descend à 15,4% pour *CNN*, ce qui reste sensiblement supérieur à ce que l'on observe dans d'autres genres, tels que les 4% des portraits journalistiques (Schneedecker 2005). Cela confirme notre hypothèse de la proximité globale de nos textes avec les faits divers. Ces résultats montrent une certaine cohérence dans les genres narratifs brefs centrés sur des personnages.

Toutes les expressions référant à des animés humains n'étant pas incluses dans des CR d'au moins trois maillons, la mesure de la densité référentielle doit être mise en relation avec une mesure de la proportion d'ER hors CR. Un texte qui aurait une proportion d'ER hors CR élevée aurait un faible taux de référents suffisamment saillants pour être ré-instanciés dans des CR. Dans *CNN*, *Gombert* et *Vilain*, moins d'une dizaine d'ER n'entrent pas dans des CR (entre 2 à 5% des ER), ce qui indique que la majeure partie des référents sont saillants. Dans Des Périers au contraire, 15% des ER n'entrent pas dans des CR car plusieurs personnages (ou regroupements de personnages) sont mentionnés seulement une à deux fois.

3.2. Nombre de CR

Le nombre de CR renseigne sur le caractère mono- ou pluri-référentiel du texte. Les textes mono-référentiels comportent principalement une CR longue et à longue portée, centrée sur le personnage principal. Parmi les genres déjà étudiés, cela concerne surtout les portraits journalistiques, mais aussi près de 40% des faits divers. Dans la plupart des cas cependant, les textes comportent plusieurs CR et sont, à des degrés divers, pluri-référentiels.

La mono / pluri-référentialité se distingue par ailleurs du caractère mono-centré ou pluri-centré du texte, le nombre de référents centraux ou proéminents pouvant être variable et relativement indépendant du nombre de CR (un petit nombre de CR induit un nombre réduit de référents proéminents, mais l'inverse n'est pas nécessairement vrai). Nous verrons que la prééminence du référent est plus directement liée à la longueur des CR.

Cette mesure renseigne aussi sur le nombre de référents potentiellement concurrents et sur les risques d'ambiguïtés référentielles, ces risques augmentant avec le nombre de CR. Cette dernière donnée a elle-même des répercussions sur la composition des CR (voir *infra*).

Les textes étudiés sont pluri-référentiels et comportent un nombre assez réduit de personnages centraux, correspondant le plus souvent à des rôles et à des types prédéfinis. On peut supposer que, en ce qui concerne le nombre de CR, nos récits brefs se situent dans une zone moyenne. En effet, *CNN* compte 8 CR comme *Gombert*. *Vilain* en compte 5 (mais ce texte est aussi plus court que les autres). L'extrait de Des Périers comptabilise 22 CR. Ce nombre est lié au fait que l'extrait comporte 4 récits ; l'ordre de grandeur est ainsi le même. Par ailleurs, dans tous les textes, le regroupement de plusieurs référents ou le changement de point de vue (jeu de quiproquos) créent de nouvelles CR.

Ce nombre de CR par texte est relativement élevé quand on le compare aux CR des faits divers, où 40% des textes sont centrés sur un personnage et un tiers sur deux, et aux CR des recettes de cuisine et des introductions de textes encyclopédiques, qui comptent en moyenne 3 ou 4 CR par textes (Schneidecker 2014). Dans tous les cas, la taille du texte entre également en ligne de compte.

3.3. Longueur des CR

3.3.1. Définitions et hypothèses

La longueur moyenne des CR est calculée à partir du nombre moyen de maillons contenus dans une CR. Des disparités très fortes ont été observées entre les genres textuels, les textes informatifs (textes de loi et rapports publics) ayant une moyenne de 3 maillons, les textes narratifs (romans) ayant des CR beaucoup plus longues. La longueur des CR (brute comme moyenne) est aussi à considérer en rapport avec la longueur du texte, les textes brefs ayant par définition des CR courtes.

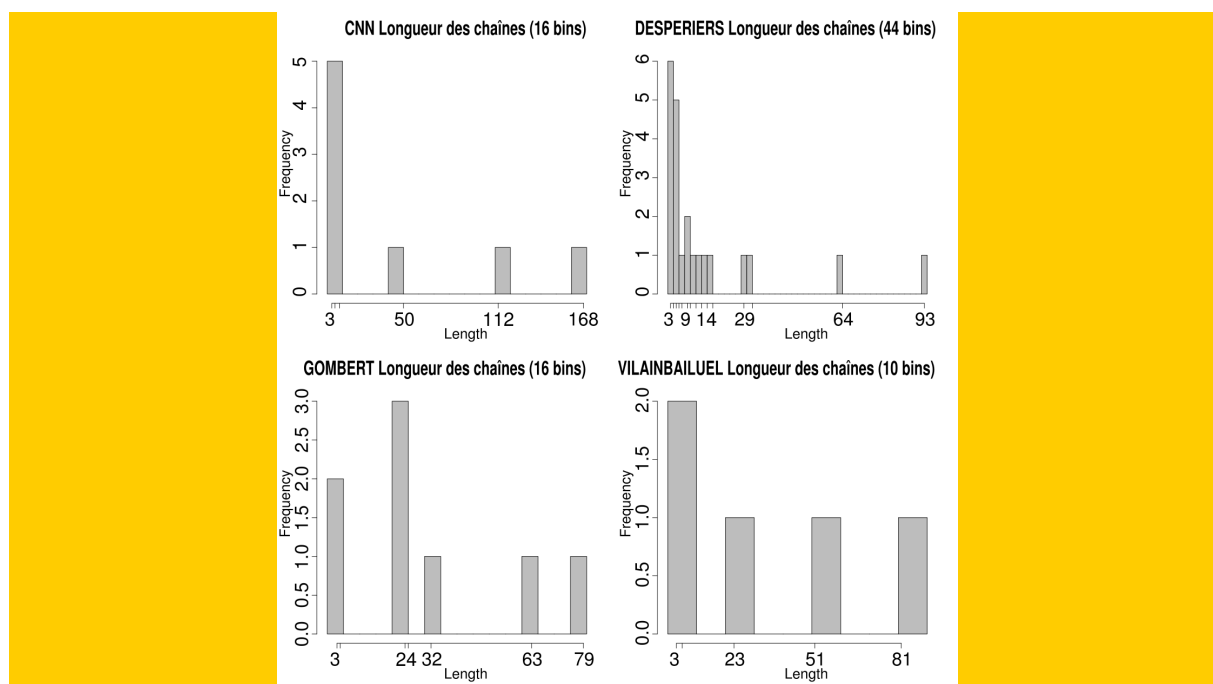
La réduction de la longueur des CR à un chiffre moyen ne rend pas compte de la répartition des tailles de CR au sein d'un texte, notamment du fait qu'il y a beaucoup de CR courtes et peu de CR longues. Elle ne dit rien non plus de la taille maximale et minimale des CR, ni de la longueur la plus fréquente dans le texte. Dans les introductions des articles encyclopédiques, par exemple, la longueur moyenne est assez faible, mais la longueur maximale est élevée, ce qui est normal pour des textes centrés du début à la fin sur le référent-thème de l'article. Il serait intéressant de voir si la taille de la plupart des CR est plutôt brève, comme dans les faits divers, où plus de 60% des CR comptent entre 3 et 6 maillons.

3.3.2. Analyse : longueur moyenne des CR et fréquence des CR brèves et des CR longues

La longueur moyenne des CR dans notre corpus, du fait de nos choix d'annotation, ne nous donne peut-être pas un résultat directement comparable avec les résultats des autres études et montre, en outre, un déséquilibre entre nos textes. La longueur moyenne est en effet de 44,25 maillons pour *CNN* alors que Des Périers a une moyenne de 16,09 maillons. Dans *Gombert* et *Vilain*, les longueurs moyennes sont assez proches avec, respectivement, 31,87 et 32,6 maillons. La différence extrême entre *CNN* et Des Périers s'explique assez facilement du fait que, pour une taille équivalente, *CNN* constitue un récit unique autour des mêmes personnages tandis que Des Périers comporte quatre récits différents.

Il nous a semblé plus intéressant de considérer l'ensemble des longueurs observées. Cela peut se représenter à l'aide de graphiques montrant les différentes longueurs de CR présentes dans un texte (sur l'axe horizontal) et, pour chacune de ces longueurs, le nombre de CR correspondant (sur l'axe vertical). Les diagrammes (Fig.1) ont été construits en groupant les longueurs par intervalles (bins) : par exemple, le premier bâton donne l'effectif global des CR de longueur entre 1 et 5, puis le bâton suivant celles de longueur 6 à 10, etc. La taille des intervalles est ajustée pour chaque graphique – elle peut donc différer d'un graphique à l'autre – mais il s'agit toujours d'un découpage régulier et continu de l'intervalle total des longueurs observées. L'abscisse indiquée sous le bâton le plus à droite renseigne sur la taille de la CR la plus longue du texte.

Figure 1. Histogrammes de la longueur des CR pour chacun des textes



D'après ces données, la moitié des CR d'un texte ont une longueur courte de 3 à 6 maillons (sur les graphiques, bâtons groupés et hauts à gauche) et les autres CR sont de longueurs inégales et dispersées (bâtons plus écartés, de hauteur 1, au milieu et à droite). Les textes sont globalement centrés autour de trois référents principaux, les CR courtes étant parfois liées à des regroupements de personnages (p. ex., dans *CNN*, la femme et l'amant). *Gombert* est cependant différent des autres textes, avec deux chaînes très longues, quatre CR relativement longues et de longueurs proches les unes des autres, et enfin deux CR courtes. Le texte est en effet un huis clos centré autour de cinq personnages principaux (Gombert, sa femme, sa fille et les deux clercs), les chaînes correspondant à un personnage ou à un regroupement de personnages (ex. les deux clercs pris ensemble ou séparément).

D'autres disparités concernent plutôt la façon dont les maillons des CR se répartissent dans le texte. Tel est le cas, en particulier, de la persistance du référent (Givón 1983) : un référent persistant est caractérisé par une CR à la fois longue, dense et continue / régulière. La mesure de la longueur des CR pourrait donc être complétée par une mesure de la régularité d'apparition du référent dans le texte et de la portée (Schneidecker 2014), ou taille de l'empan de texte sur laquelle la CR s'étend.

4. Coefficient de stabilité référentielle et diversité des CR

4.1. Définitions et hypothèses

Le coefficient de stabilité renseigne sur la diversité des redénominations dans les CR. Cette mesure s'inspire d'une étude de M. Perret (2000) et s'obtient en divisant, pour un référent donné, le nombre total d'ER nominales (GN et noms propres) par le nombre de ses désignations nominales différentes. Pour être pleinement opératoire, la mesure du coefficient de stabilité doit neutraliser les variations formelles non pertinentes (en particulier, pour la période médiévale, les variations casuelles) et s'appuyer sur les lemmes, et non sur les formes des expressions référentielles nominales². Plus le coefficient de stabilité est fort, moins les désignations nominales sont variées et plus les CR peuvent sembler monotones. Initialement forgée pour rendre compte de la pauvreté de la variation lexicale dans les œuvres médiévales, cette mesure s'est en réalité avérée fort utile pour contraster des genres modernes : les textes informatifs (comme les introductions des articles encyclopédiques) et les recettes de cuisine se distinguent par un taux de stabilité très élevé, les portraits journalistiques sont souvent à l'opposé.

Une mesure de diversité (Pincemin & Malrieu 2014) consistant plus simplement à dénombrer le nombre de désignations différentes d'un référent serait en fait plus robuste que ce coefficient, qui semble vouloir normaliser la mesure en la rapportant à la longueur de la CR. Or, on peut s'attendre linguistiquement à ce que la diversité des désignations ne soit pas proportionnelle à la longueur de la CR : d'une part, l'effet de variation peut être bien présent, malgré des reprises de mentions à l'identique, si celles-ci sont à certaine distance dans la linéarité du texte et que d'autres formes de désignations se sont intercalées ; d'autre part, les possibilités de redénomination ne sont pas infinies en pratique au sens où, sauf genre textuel particulier, les nouvelles formes finissent par être perçues comme trop lourdes ou trop recherchées. Par conséquent, cette mesure de la stabilité aura tendance à croître avec la longueur de la CR (elle-même liée au genre textuel). Dans l'article de M. Perret (2000), les résultats et l'interprétation auraient été très probablement perturbés si l'on avait pris en compte tous les personnages pour les textes longs (romans), et non seulement les deux ou trois principaux : des personnages secondaires auraient pu être détectés comme de plus faible stabilité que les personnages principaux. Nous proposons donc de tester cette nouvelle mesure (diversité des désignations des référents) aux côtés du coefficient de stabilité référentielle sur les CR les plus longues des textes, pour une première confirmation de cette intuition sur nos données.

L'étude de M. Perret montrait déjà un coefficient de stabilité relativement bas dans le texte des *Cent nouvelles nouvelles*, ce coefficient allant de 1 à 2,5 suivant les personnages (par comparaison, les romans en prose des XIV^e et XV^e s. ont un coefficient qui va de 32 à 5,16 ; ce qui peut s'expliquer en partie par leur longueur supérieure). Elle constatait aussi que le personnage central des récits (le héros éponyme des romans, la femme dans les nouvelles) est le référent dont le taux de stabilité est souvent le plus faible (parmi les CR les plus longues). Notre étude permet de vérifier ces observations.

4.2. Analyse

Les fabliaux ne se conforment pas au modèle étudié par M. Perret dans les *CNN*. Le personnage féminin du *Vilain* présente un coefficient de stabilité équivalent à celui des deux référents masculins centraux, son mari et le prêtre (respectivement 1,6, 1,3 et 1,6), cela correspond à un très petit nombre de formes et les expressions se limitent à l'emploi du nom propre *Erme* et à deux substantifs noyaux de GN, *fame* et *dame*. On aurait pourtant pu supposer que davantage de variation serait permise par l'absence de concurrence avec un autre référent féminin, outre l'habitude du genre qui accorde au personnage féminin un mode de référence varié. Le mari trompé, personnage pour lequel la CR est la plus longue, est lui aussi désigné par très peu d'expressions nominales : les noyaux des GN sont : *vilain*, rendant compte de son statut social, et un terme relationnel, *frere*. Le prêtre est, quant à lui, instancié par deux synonymes désignant sa fonction, *prestre* et *chapelain*. La faiblesse du coefficient de stabilité ne traduit pas des variations de redénomination mais tient à la brièveté des CR. Elle masque en réalité la monotonie des CR. Les personnages sont désignés par des expressions rappelant une propriété immuable, en particulier leur statut social. L'exemple de *Gombert* confirme que le coefficient de stabilité ne rend pas compte à lui seul de la diversité des CR. Le référent le plus fréquemment instancié, Gombert, est celui pour lequel on trouve le plus de formes différentes (13). Le coefficient de 1,3 le concernant est certes inférieur à celui des CR référant à son épouse (1,5), à sa fille (2) ou encore au groupe des deux clercs (1,5). Mais le coefficient est le même pour chacune des CR désignant les deux clercs séparément, qui sont pourtant beaucoup moins variées (3 formes différentes seulement). Pour Gombert, on relève le nom propre seul, le nom propre modifié (*sire Gombert* et *dant Gombert*) et les GN *uns vilains*, *li vilains*, *vostre pere*, *l'oste* et *ses ostes*. La variété est donc tempérée par la répétition de substantifs dont le sens rappelle les qualités stables et permanentes du personnage (statut social et filiation). Seule une qualité non permanente, liée au rôle dans l'intrigue, est convoquée

² Cela comprend les têtes lexicales mais aussi leurs déterminations. Notre corpus n'étant pas lemmatisé, nous n'avons pu réaliser cette mesure sur les lemmes.

dans *li ostes* ; il ne s'agit pourtant pas d'un statut qui évolue au cours du texte. Les deux clercs sont quant à eux rarement désignés par des GN, et les deux CR font apparaître seulement deux syntagmes identiques : *li clers* et *ses compains*. Le choix de ces GN révèle une absence de véritable soin à distinguer les référents potentiellement concurrents.

Dans *CNN*, le personnage féminin est celui pour lequel le coefficient de stabilité est le plus faible (1) mais, parmi les 112 mentions du référent, seules 3 sont des GN, dont les noyaux sont le nom *femme* et un synonyme à valeur péjorative, *gouge* (la mesure de diversité est de 3). Au mari trompé, référent de la CR la plus longue de la nouvelle, correspond un coefficient de stabilité inférieur à celui qui concerne son voisin, le bourgeois (1,25 pour le premier, 1,9 pour le second). La CR désignant le bourgeois a une diversité de 7 mais varie autour de deux substantifs seulement, *bourgeois* et *voisin*. Seule l'expression *son amy nouvel* fait apparaître une qualité transitoire du personnage. Il en est de même pour les formes référant au mari trompé (diversité de 8) : s'opposent ainsi *un bon compaignon, nostre compaignon, le bon compaignon, son mary*, d'une part, et *le povvre desolé*, d'autre part, qui rend compte d'un jugement subjectif du narrateur. La diversité, mieux que le coefficient de stabilité, traduit le fait que les personnages masculins centraux voient légèrement varier leur désignation, qui tient compte ponctuellement de leur situation : on pourrait y lire le signe d'une plus grande individuation du personnage de la nouvelle par rapport à celui du fabliau, puisque quelques ER rendent compte de ses évolutions, il s'émancipe un peu du type auquel il appartient.

Chez Des Périers, certaines expressions nominales témoignent aussi de la subjectivité du narrateur, avec une valeur plus ou moins ironique des adjectifs : *ce sage homme Caillette, tant d'honnestes jeunes gens, le povvre chancre*. Si les référents secondaires sont instanciés par des expressions qui varient peu, la diversité touche de manière inégale les personnages centraux des deux nouvelles. La CR référant à Triboulet, le premier des trois fous de la première nouvelle, semble la plus stable (coefficient : 4) et ce résultat correspond à quatre occurrences du nom propre (diversité : 1). On trouve un taux de variation légèrement plus important pour Caillette, avec le nom propre, *le povvre Caillette* et *ce sage homme* (coefficient : 2, diversité : 4). Enfin, Polite est désigné à deux reprises seulement par *un autre fol nommé Polite*, et par son nom. La référence au chancre, protagoniste de la seconde nouvelle, donne lieu à plus de variations (coefficient : 1,3 ; diversité : 6). Il est désigné comme *un chancre, ce Bassecontre, le povvre chancre* ou encore *ce fol*. Ce dernier personnage rejoint le modèle des protagonistes masculins des *CNN* alors que, au contraire, la présence du nom propre, tout en rendant saillant le référent, semble figer le contenu de la CR.

Malgré une tendance, partagée par l'ensemble des récits, à une grande homogénéité des CR, dont rendent compte à la fois la mesure de la diversité des expressions référentielles et le sémantisme des noms noyaux des GN, on observe des variations plus importantes dans la nouvelle que dans le fabliau, ce qui rend compte de l'évolution du statut du personnage dans ces genres.

5. Composition des CR

5.1. Catégorie grammaticale des maillons

5.1.1. Définitions et hypothèses

La fréquence des catégories grammaticales dans les CR est probablement le paramètre qui présente les variations les plus fortes et qui donne les informations les plus riches sur la typologie des CR. Il renseigne notamment sur le degré d'homogénéité des CR et, à ce titre, il doit être mis en relation avec le paramètre de la stabilité référentielle.

La mesure des catégories grammaticales permet aussi d'étudier la part relative des expressions marquant une accessibilité référentielle faible, moyenne ou élevée. La distribution de ces catégories dépend du caractère mono- ou pluri-référentiel du texte et des risques d'ambiguïté ; les marques de faible accessibilité étant moins attendues en cas de compétition référentielle.

5.1.2. Marqueurs de haute et de faible accessibilité

Tableau 1. Répartition des catégories grammaticales représentées dans les CR

	Fabliaux		CNN		Des Périers	
Nom Propre	19	4,5%	0	-	13	3,7%
N ou GN sans dét	11	2,6%	6	1,7%	9	2,5%
GN Indéfini	3	0,7%	2	0,6%	4	1,1%
Pron Indéfini	9	2,2%	0		19	5,4%
Pron Num	2	0,5%	0		0	
GN Défini	43	10,3%	30	8,5%	22	8,7%
GN Possessif	11	2,6%	33	9,3%	7	2%
GN Démonstratif	0		1	0,3%	6	1,7%
Pron Démonstratif	21	5%	5	1,4%	1	0,3%
Pron possessif	2	0,5%	0		3	0,8%
Dét Possessif	35	8,4%	95	26,8%	33	9,3%
Adjectif Possessif	1	0,2%	0		0	
Pron Pers Anaphorique	95	22,7%	99	28%	136	38,4%
Pron adverbial	1	0,2%	0		1	0,3%
Pron Pers Déictique	44	10,5%	34	9,6%	71	20,1%

Pron Relatif	10	2,4%	8	2,3%	17	4,8%
Sujet non exprimé d'un impératif	11	2,6%	10	2,8%	0	
Sujet non exprimé d'un V conjugué à un autre mode	100	23,9%	31	8,9%	12	3,4%

L'ordre de présentation des catégories grammaticales dans le tableau 1 s'inspire de l'échelle d'accessibilité définie par M. Ariel (Ariel 1990), qui met en relation le choix des ER et la plus ou moins grande activation des référents. Dans tous les textes, c'est une expression marquant une accessibilité haute qui est majoritaire : 23,9% des maillons sont des verbes conjugués sans sujet exprimé dans les fabliaux, 28% et 38,4% des pronoms personnels dans *CNN* et Des Périers respectivement. Le tableau 2 regroupe les expressions en deux catégories, d'après M. Ariel (1990) et C. Schnecker et L. Longo (2012) : nous distinguons, d'une part, les noms propres, les noms ou GN sans déterminants, les GN et pronoms indéfinis, les GN définis, les GN et pronoms démonstratifs et possessifs, en tant que marques de faible ou moyenne accessibilité et, d'autre part, les marques de haute accessibilité que sont les déterminants possessifs, pronoms personnels, adverbiaux et relatifs et les verbes conjugués sans sujet exprimé. On exclut de ce calcul les impératifs et les pronoms personnels déictiques figurant en discours direct.

Tableau 2. Répartition des marques de faible, moyenne et haute accessibilité

	Fabliaux		CNN		Des Périers	
Marques de faible et moyenne accessibilité	117	32,20%	77	24,80%	81	28,60%
Marques de haute accessibilité	246	67,80%	233	75,20%	202	71,40%

La part globalement importante des marques de haute accessibilité est à relier à la relative brièveté des textes, qui implique la persistance des référents et un nombre restreint de personnages centraux. Même si l'on inclut dans ces décomptes l'ensemble des expressions référentielles renvoyant à des animés humains, cette proportion varie très peu. Les différences entre les trois parties de notre corpus s'expliquent surtout par la longueur des unités narratives. L'extrait des *CNN*, constitué d'une seule nouvelle, présente les mêmes personnages du début à la fin, ce qui explique la proportion légèrement supérieure de marques de haute accessibilité. Dans les deux fabliaux et dans Des Périers, le plus grand nombre de CR implique l'introduction de nouveaux référents et la présence plus importante de marques de faible accessibilité.

5.1.3. Variations des catégories représentées

L'une des principales différences entre les textes du corpus est la diminution de la part des verbes conjugués dont le sujet pronominal n'est pas exprimé (23,9% des maillons dans les fabliaux ; 8,9% dans les *CNN* ; 3,4% chez Des Périers). Cela n'a rien d'étonnant et est à relier à l'augmentation générale de l'expression du sujet pronominal en moyen français et dans les états de langue plus tardifs.

De nouveaux types d'expressions apparaissent dans les nouvelles, tels que le déterminant *ledit* dans les *CNN* ou le relatif *lequel* dans les deux textes les plus récents, mais ces éléments ne font pas varier sensiblement la répartition des catégories d'expressions.

Parmi les marques de faible accessibilité, le nom propre a un rôle qui varie selon les textes, et les données statistiques n'en rendent que partiellement compte. Son emploi est toujours limité à quelques personnages centraux. Parmi les fabliaux, seuls le protagoniste et sa femme sont nommés dans *Gombert*, mais la plupart des emplois du nom propre concernent Gombert. Dans le *Vilain*, le personnage principal est anonyme et seule son épouse, *dame Erme*, porte un nom assez peu employé. Dans les *CNN*, tous les personnages sont anonymes. Dans Des Périers, l'ensemble des emplois se répartit dans la première nouvelle (les trois noms sont ceux des trois fous éponymes) tandis que les personnages de la seconde sont anonymes. Les hommes sont donc plus volontiers nommés que les femmes (mais ils sont plus nombreux) et les noms des hommes sont plus employés : la référence aux femmes dépend des expressions désignant leur mari, avec un recours fréquent au déterminant possessif (*sa femme*).

Nos textes se rapprochent des faits divers en raison de la part restreinte des marques de faible accessibilité référentielle, mais ils n'accordent pas le même privilège aux marques d'accessibilité moyenne : même en prenant en compte les ER hors CR pour tous les référents humains, la proportion de GN est bien moindre (55% de GN définis dans les faits divers).

5.1.4. Types de référents et catégories des ER

Si l'on s'intéresse à la représentation des différentes catégories d'expressions au sein de chacune des CR, les proportions analysées *supra* ne varient que faiblement. En revanche, la façon dont les différents textes instancient les référents les plus fréquemment mentionnés (*i.e.* dans les CR les plus longues) fait apparaître des types de référents qui semblent déterminer, en partie, la composition des CR et qui ne varient que très peu d'un texte à l'autre.

Les personnages nommés s'opposent bien sûr aux personnages anonymes. Lorsque le nom propre est présent, il occupe une part relativement constante de la CR (13 mentions sur 79 dans *Gombert*, soit 14,4% ; dans Des Périers, 6 mentions sur 32 pour *Caillette*, soit 18,8%, 4 mentions sur 29 pour *Triboulet*, soit 13,8%, 3 mentions sur 14 pour *Polite*, soit 21,4%). Dans ces quatre cas, les marques de haute accessibilité (verbes sans sujet exprimé et pronoms personnels) correspondent à environ la moitié des expressions. Lorsque les personnages sont anonymes, la part des marques de haute accessibilité est sensiblement la même, mais aucune catégorie grammaticale ne vient, à elle seule, compenser l'absence du nom propre et la CR est plus variée.

La comparaison des personnages féminins et des personnages masculins montre combien la concurrence référentielle influe sur le choix des expressions. La CR qui réfère à l'épouse du *Vilain* compte 35,3% de sujets non exprimés (23,9% dans les deux fabliaux), et celle qui réfère à l'épouse des *CNN* 32% de pronoms personnels (23% pour l'ensemble de la nouvelle). Les figures féminines centrales sont moins sujettes aux ambiguïtés référentielles que les référents masculins, qui sont toujours au moins deux.

La concurrence référentielle est également exploitée dans les passages relevant du quiproquo. Ce procédé, caractéristique du genre narratif bref comique, est favorisé par le caractère prototypique des personnages centraux et par la stabilité des CR. Dans *Gombert*, la femme s'adresse au clerc qu'elle prend pour son mari en employant le nom propre de son mari :

1. Et li clers ne s'oublia pas:
Avec la dame ala couchier; [...]
Or a Gombert bone mesnie,
Mout le maintent de male pile.
« Sire Gombert, fet dame Gille [...] » (*Gombert*, v. 114-122)
'Le clerc, lui, ne perdit pas de temps : il alla coucher avec la dame [...]. Maître Gombert a de bons serviteurs : il n'y vont pas de main morte avec lui ! « Messire Gombert, fit dame Gille [...] »'

En récit, le GN *li clers* et le nom *Gombert* distinguent les référents. La reprise du nom en discours direct permet au quiproquo de fonctionner car le comique repose sur la perception, par le lecteur ou auditeur, de l'absence de coréférence entre les mentions du même nom.

5.2. Début des CR

5.2.1. Définitions

La fréquence des catégories grammaticales en position de premier maillon peut varier du tout au tout. C'est tantôt le nom propre qui domine, tantôt le GN indéfini, tantôt le GN défini. Le choix du premier maillon donne des indications sur les procédés mis en œuvre dans le texte pour donner, ou pas, une prééminence ou saillance particulière au référent nouvellement introduit. L'étude des premiers maillons doit donc différencier les référents prééminents des autres référents du texte.

En français moderne, le nom propre est un bon indicateur de la saillance du référent dans la suite du texte. Dans les textes médiévaux, A. Carlier (2001) a observé déjà qu'une des conditions d'emploi de l'article indéfini était la reprise ultérieure du référent. À sa fonction numérale d'origine s'ajoute alors une fonction discursive

qui consiste à signaler à l'allocataire / lecteur l'introduction d'un nouveau référent dans l'univers du texte et à rendre ce référent nouvellement introduit disponible pour une reprise anaphorique. (Carlier, 2001 : 76)

Le « processus de renforcement pragmatique » qui sous-tend l'emploi de *un* comme article peut se traduire par une mise en saillance du référent nouvellement introduit. C'est ce que semblent suggérer les premiers maillons des CR des personnages centraux de miracles (Glikman, Guillot-Barbance & Obry 2014).

5.2.2. Analyse par texte : l'évolution du rôle du GN indéfini

Dans les fabliaux, le personnage du mari trompé est introduit par un GN indéfini, qu'il soit nommé ou non dans le récit :

2. Dont avint il, ce dist mon mestre,
C'uns vilains a Bailluel manoit;
Formenz et terres *ahanoit*,
N'*estoit* useriers ne changiere. (*Vilain*, v. 2-5)
'Il se trouva, comme le dit mon maître qu'un paysan demeurait à Bailleul. Il travaillait ses terres et son blé, il n'était ni usurier ni banquier.'

Le personnage éponyme de *Gombert* est introduit de la même façon, *un vilein*, et l'emploi de l'indéfini est réservé à ces personnages, à l'exception de *un petit enfant*, mentionné dans *Gombert*, pour introduire ce nouveau personnage (l'enfant ne joue pas un rôle prépondérant dans la suite de l'intrigue, mais son berceau constitue l'un des ressorts de la farce). Il semble conférer une saillance au référent, repris ensuite par des expressions de haute accessibilité, et indiquer l'introduction d'un référent central : les deux personnages concernés sont ceux pour lesquels la CR est la plus longue. À partir de ce personnage introduit par l'indéfini se construit un univers de référence. Ainsi, l'épouse du vilain est introduite dans les deux textes par un GN possessif *sa fame* et, dans *Gombert*, on trouve le même type d'expression pour *sa fille*. Le ou les personnages masculins concurrents sont introduits par d'autres types de GN. Dans *Vilain*, le GN défini *le chapelain* rattache le personnage à un type, celui du prêtre luxurieux, et indique aussi une notoriété limitée à l'univers défini par la référence géographique à Bailleul. Dans *Gombert*, c'est le pluriel *deus clers* qui introduit les autres référents masculins, l'expression étant suivie par des pronoms indéfinis, *l'uns* et *l'autre*, qui opèrent une extraction dans le groupe et donnent ensuite lieu à deux CR distinctes. On note donc une forme de dépendance des principales CR les unes par rapport aux autres. Les autres référents, essentiellement des groupes mentionnés localement et le narrateur, sont introduits par des expressions de différentes catégories, sans spécialisation notable. Dans notre étude antérieure (Glikman, Guillot-Barbance & Obry 2014), le référent central introduit par l'indéfini semblait bénéficier d'une saillance globale, qui permettait l'emploi de marques de haute accessibilité, même quand il n'était pas saillant localement. Dans nos fabliaux, le choix des ER reste lié au contexte référentiel et à la concurrence possible entre les référents humains masculins. En revanche, l'usage du nom propre dans *Gombert* semble conforter le statut spécifique du référent introduit par l'indéfini. Après l'introduction citée *supra*, une vingtaine de vers se déroule sans évoquer le vilain, puis on lit :

3. Tieus biens com *frere Gonbers* ot
Orent la nuit asez *si* oste (*Gombert*, v. 30-31)
'Tous les biens de frère Gombert furent cette nuit-là à la disposition de ses hôtes.'

La reprise de celui qui était désigné comme *un vilain* par un nom propre, sans passer par le processus de nomination, implique une forme de persistance du référent. La mention du nom propre, dans ce contexte, renforce d'autant la saillance du référent principal.

Dans *CNN*, les personnages masculins principaux, le bourgeois et le mari trompé, sont introduits par des GN indéfinis, suivis d'un pronom relatif de reprise :

4. En la ville de Valenciennes eut nagueres **ung notable bourgeois**, en son temps receveur de Haynau, **lequel** entre les aultres fut renommé de large et discrete prudence. Et entre **ses** loables vertuz celle de liberalité ne fut par la maindre, car par icelle **vint** en la grace des princes, seigneurs et aultres gens de tous estaz. (*CNN*, p. 23)

5. et de fait avoit une petite posterne vis a vis de laquelle demouroit **ung bon compaignon, qui tresbelle femme et gente** avoit et encores en meilleur point. (*CNN*, p. 23)

Les autres référents représentés dans cette nouvelle sont très peu nombreux et ne sont pas introduits par des GN indéfinis : on trouve un pronom personnel pour le groupe formé par la femme et le bourgeois, des GN définis désignant des groupes, et le déterminant possessif *nostre* pour le narrateur.

Dans les nouvelles de Des Périers, l'emploi du GN indéfini en première mention n'est plus réservé à un type de référent en particulier. Parmi les référents centraux, deux des protagonistes nommés de la première nouvelle sont introduits directement par leur nom, Caillette et Triboulet. Le nom propre en début de CR, qui répète le titre de la nouvelle, assure la saillance du référent, ensuite instancié par des marques de haute accessibilité. Le troisième fou de cette première nouvelle est pourtant introduit différemment : « Il y avoit **un autre fol** nommé Polite, qui estoit à un abbé de Bourgueil » (Des Périers, p. 12). Ce choix s'explique, sans doute, par l'éloignement de cette troisième histoire par rapport au titre de la nouvelle. Mais il s'apparente aussi à l'introduction, par un GN indéfini, du personnage central, anonyme, de la deuxième nouvelle :

6. En l'église Saint Hilaire de Poitiers y eut jadis **un chantre qui** servoit de bassecontre: **lequel** par ce qu'il estoit bon compaignon, et qu'il beuvoit bien, ainsi que voluntiers font telles gens, estoit bien venu entre les chanoines. (Des Périers, p. 12)

On retrouve le modèle de début de CR représenté dans *CNN*, mais cette structure n'est plus réservée aux personnages centraux. Les autres personnages qui ne sont pas nommés sont désignés par leur qualité ou leur profession et introduits par des GN définis ou indéfinis : *un des seigneurs de Cour, un maistre, un abbé de Bourgueil*.

Le rôle du GN indéfini dans l'introduction du référent évolue. Dans les fabliaux, l'indéfini, réservé aux personnages centraux, semble conférer une saillance au référent. Dans les *CNN*, l'indéfini est toujours réservé aux référents centraux et est inclus dans des suites d'expressions qui contribuent à la saillance plus ou moins forte du référent. Chez Des Périers, l'indéfini n'est plus lié à un type de référent et le rôle du nom propre comme indicateur de saillance dans la suite du récit semble s'affirmer.

6. Conclusion

Notre étude a permis de mettre en lumière plusieurs spécificités des CR des récits brefs anciens, parmi lesquelles certaines se trouvent aussi dans les faits divers modernes (densité référentielle des animés humains, nombre faible et taille réduite des CR, fréquence des noms relationnels) et d'autres sont plus typiques des états anciens du français et peuvent être relevés autant dans les genres brefs que dans les récits longs (fréquence du sujet zéro, stabilité et monotonie fortes ; voir Glikman, Guillot-Barbance, & Obry 2014). Notre analyse montre, par ailleurs, une évolution diachronique dans la composition des CR : le statut d'indicateur de saillance élevée d'abord attaché à l'article indéfini semble s'atténuer au XVI^e siècle et le nom propre paraît le remplacer dans ce rôle. De nouvelles expressions, encore rares, se développent à partir du XV^e siècle (*ledit, lequel*). Cette évolution serait à confirmer avec d'autres textes des mêmes périodes.

Ce qui ressort en outre de notre étude, c'est l'impact du type de référent, central ou non, sur les CR. Ce paramètre, même s'il n'est pas nécessairement mis en avant de manière explicite dans les travaux sur les CR, est toujours primordial, et ce sont les chaînes les plus longues, celles qui correspondent aux référents centraux, qui sont les plus étudiées et qui permettent les contrastes d'un genre à l'autre. Dans le cas présent, les changements qui affectent ces CR caractérisent plus largement l'évolution des récits brefs et l'individuation croissante des personnages importants (marquée notamment par la progression de la fréquence du nom propre en début de CR).

Notre étude visait également à rendre explicite et reproductible une méthodologie d'analyse des CR grâce au développement de mesures aptes à évaluer leur variation. La définition de ces mesures se fonde sur plusieurs études antérieures, mais lorsque cela nous a paru nécessaire, nous avons proposé des mesures complémentaires ou rectifié les définitions existantes. D'autres calculs demanderaient à être intégrés à l'analyse globale des CR (la mesure de la portée par exemple). La réflexion méthodologique s'est faite en lien avec le développement des outils permettant leur mise en œuvre et montre le lien étroit qui unit les choix de représentation numérique des textes et des annotations, d'une part, et des outils d'extraction et de calcul, d'autre part.

Remerciements : ce travail a été réalisé avec le soutien de l'ANR dans le cadre du projet DEMOCRAT – ANR-15-CE38-0008 – et grâce à Serge Heiden et Matthieu Decorde (ENS de Lyon, UMR 5317 IHRIM) pour l'implémentation des mesures dans TXM.

Références

Éditions des textes du corpus :

- Cent nouvelles nouvelles*, éd. par F. Sweetser, Genève : Droz (TLF 127), 1966.
- Gombert et les deus clerics*, dans *Nouveau recueil complet des fabliaux (NRCF)*, éd. par Willem Noomen et Nico Van den Boogaard, Assen/Maastricht, Van Gorcum, 1998, t. IV, 296-301.
- Le Vilain de Bailluel*, dans *Nouveau recueil complet des fabliaux (NRCF)*, éd. par Willem Noomen et Nico Van den Boogaard, Assen/Maastricht, Van Gorcum, 1990, t. V, 246-249.
- Bonaventure des Périers, *Nouvelles recreations et joyeux devis de feu Bonaventure des Periers, valet de Chambre de la Roynne de Navarre*, Lyon, Rouillé Guillaume, 1561, publié en ligne par les Bibliothèques virtuelles humanistes.

Études :

- ARIEL M. (1990), *Accessing Noun-Phrase Antecedents*, London: Routledge.
- CARLIER A. (2001), « La genèse de l'article un », *Langue française* 130, 65-88.
- COLLET O., GINGRAS F. & TRACHSLER R. (éds.) (2012), *Études françaises* 48-3 : Lire en contexte : enquête sur les manuscrits de fabliaux, Montréal : Presses de l'Université de Montréal.
- GIVÓN, T. (1983), "Topic continuity in discourse: An introduction", in T. Givón (ed.), *Topic Continuity in Discourse. A quantitative cross-language study*, Amsterdam/Philadelphia: John Benjamins Publishing, 1-41.
- DUBUIS R. (1973), *Les Cent Nouvelles Nouvelles et la tradition de la nouvelle en France au Moyen Âge*, Grenoble : Presses de l'Université de Grenoble.
- GLIKMAN J., GUILLOT C. & OBRVY V. (2014), « Les chaînes de référence dans un corpus de textes narratifs médiévaux : traits généraux et facteurs de variation », *Langages* 195, 43-60.
- HEIDEN S., MAGUÉ J-P. & PINCEMIN B. (2010), « TXM : Une plateforme logicielle open-source pour la textométrie – conception et développement », in I. C. Sergio Bolasco (ed.), *Proc. of 10th International Conference on the Statistical Analysis of Textual Data - JADT 2010*, vol. 2, Roma: Edizioni Universitarie di Lettere Economia Diritto, 1021-1032.
- LABÈRE N. (2006), *Défricher le jeune plant. Étude du genre de la nouvelle au Moyen Âge*, Paris : Champion.
- LANDRAGIN F., POIBEAU T. & VICTORRI B. (2012), "ANALEC: a New Tool for the Dynamic Annotation of Textual Data", in *Eighth International Conference on Language Resources and Evaluation*, Istanbul, 357-362.
- LANDRAGIN F. & SCHNEDECKER C. (éds.) (2014), *Langages* 195 : Les chaînes de référence, Paris : Armand Colin.
- LANDRAGIN F. (2016), « Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT) », *Bulletin de l'Association Française pour l'Intelligence Artificielle (AFIA)* 92, 11-15.
- MÉLANIE-BECQUET F. & LANDRAGIN F. (2014), « Linguistique outillée pour l'étude des chaînes de référence: questions méthodologiques et solutions techniques », *Langages* 195, 117-137.
- PERRET, M. (2000), « Quelques remarques sur l'anaphore nominale aux 14^e et 15^e siècles », *L'Information grammaticale* 87, 17-23.
- PINCEMIN B. & MALRIEU D. (2014), « Caractérisation quantitative de textes -Application à l'oral représenté, en diachronie », in D. Ablali, S. Badir & D. Ducard (éds), *Documents, textes, œuvres -perspectives sémiotiques*, Rennes : Presses universitaires de Rennes, 43-56.
- SCHNEDECKER C. (1997), *Nom propre et chaînes de référence*, Paris : Klincksieck.
- SCHNEDECKER C. (2005), « Les chaînes de référence dans les portraits journalistiques : éléments de description », *Travaux de linguistique* 51/2, 85-133.
- SCHNEDECKER C. (2014), « Chaînes de référence et variations selon le genre », *Langages* 195, 23-42.
- SCHNEDECKER C. & LONGO L. (2012), « Impact des genres sur la composition des chaînes de référence : le cas des faits divers », *3^{ème} Congrès Mondial de Linguistique Française*, Lyon, France, 1957-1972.
- VELISSARIOU A. (2012), *Aspects dramatiques et écriture de l'oralité dans les Cent Nouvelles nouvelles*, Paris : Champion.
- ZUMTHOR P. (1983), « La brièveté comme forme », in M. PICONE, G. DI STEFANO & P. D. STEWART (éds.), *La Nouvelle : Formation, codification et rayonnement d'un genre médiéval*, Montréal : Plato Academic Press, 3-8.