

Philologie numérique et syntaxe historique

1. Principes généraux de la syntaxe historique et diachronique

Les nouveaux outils numériques, dont les linguistes peuvent bénéficier, mais aussi qu'ils peuvent faire évoluer en fonction de leurs besoins, ont-ils modifié nos comportements, et même plus fondamentalement notre démarche ? Ou bien ne s'agit-il que d'une couche de modernité qui, sans rien changer, a pour effet de faciliter un peu notre travail – « nouveaux outils, ancienne démarche » en quelque sorte ?

Je voudrais tenter de présenter ici de quelle façon, me semble-t-il, les possibilités ainsi ouvertes ont, en peu d'années, profondément affecté, modifié les pratiques, la démarche, et même la conception que le linguiste diachronicien se faisait de ses tâches et les théories que l'on pouvait formuler.

La linguistique *historique*, et la syntaxe *diachronique*¹ en particulier, ont pour but la recherche, non seulement des *régularités* (comme en synchronie), mais également et surtout des *irrégularités* de la langue d'une époque et d'un lieu (et en cela elle se distingue de la linguistique synchronique).

Les *régularités* concernent non seulement la grammaire en elle-même, mais également les variantes repérées et intégrées à la grammaire. Les *irrégularités* en revanche, ce sont les *variantes non attendues* ; il peut s'agir d'une *innovation ponctuelle* – un hapax, rencontré une seule fois et que l'on ne retrouvera plus, ou bien d'une *nouvelle forme* ou *construction* qui, de texte en texte, de témoin en témoin, perdure et s'installe – et dans ce cas on peut parler véritablement de *changement*, ou encore d'une *variation* dans l'emploi d'une forme ou construction – sa progression, sa régression -, ou enfin sa *disparition*.

Rappelons une autre spécificité de la linguistique diachronique ; elle travaille sur un objet linguistiquement singulier, la ou les langues de siècles passés, pour lesquels il n'existe plus de locuteur natif pour *valider*, ou invalider, l'analyse que peut en donner le linguiste. Dans ce domaine, l'introspection n'est d'aucun secours, non plus que la

¹ On distingue les deux noms (histoire, diachronie) et les adjectifs correspondants (historique, diachronique), par leur extension. Pour nous, la *linguistique historique* recouvre aussi bien l'histoire *interne* de la langue que son histoire *externe*, et l'approche diachronique ou même synchronique de la grammaire des époques anciennes. La *linguistique*, ou la *syntaxe, diachronique*, désigne de façon bien plus précise et restreinte uniquement l'étude de l'*évolution de la langue, ou de la syntaxe*.

connaissance que le spécialiste peut avoir acquise de sa fréquentation des textes de ces époques, ou des grammaires historiques qu'il a assimilées. Il faut rappeler d'emblée que *le linguiste diachronicien ne peut donc produire que des hypothèses, jamais de certitude*. Dès lors son seul recours est celui des textes et des documents écrits, jusqu'au moment où apparaît l'enregistrement qui conserve l'oral. Il s'ensuit que plus on a de témoins, plus les hypothèses sont solides, ce qui, à l'évidence, limite la portée et les acquis possibles du travail 'à la main'. Pour le linguiste diachronicien, on le voit, d'emblée, le recours aux *corpus numérisés* est une façon de multiplier ses capacités d'exploration et donc de trouvailles, et de se donner ainsi de plus fortes possibilités pour la formulation d'hypothèses.

La *syntaxe historique* est une discipline jeune : ce n'est guère que depuis un siècle et demi qu'elle existe de façon autonome. Pendant quelques décennies elle s'est développée sur le terreau théorique de la linguistique synchronique, utilisant, en les gauchissant plus ou moins, les concepts et méthodes qui se développaient, avec quelle ampleur et quelle richesse, pour la linguistique synchronique (structuralisme, grammaire générative, sémantique fonctionnelle). Mais peu à peu ont vu le jour des théories spécifiquement dédiées à la diachronie, c'est-à-dire portant sur l'évolution des langues et le changement linguistique (grammaticalisation, réanalyse, analogie, contact entre langues).

Le *traitement numérique des textes* est, lui, bien plus jeune encore, il n'a guère que cinq décennies, ou plutôt, si l'on ne considère que le développement des outils nécessaires au linguiste, deux décennies. C'est dans les années 60 qu'a commencé à se développer par exemple le corpus de français historique *Frantext* ; mais les outils utiles ne sont venus qu'après.

Or ces deux domaines ont justement commencé à se développer – est-ce un hasard ? – en pleine période de réflexions nouvelles et extrêmement fécondes sur la production et la réception des textes (Zumthor, Jauss, Iser...).

La démarche des diachroniciens a dès lors pu bénéficier de ce triple 'progrès' – technique, théorique et interprétatif. Pour utiliser les multiples possibilités ainsi générées, tant dans le champ théorique que dans le champ des techniques, nous avons développé *une conception intégrée pour l'étude du 'changement linguistique' en syntaxe* : seul un travail d'équipe, long, régulier, solidaire, novateur a permis ces avancées (UMR 'ICAR'-Equipe ICAR-3 à Lyon, ANR 'SRCMF' Paris-Stuttgart, 'Corptef' Lyon-Paris).

2. Philologie numérique

2.1. Qu'appelle-t-on philologie numérique ('digital philology') ?

On range sous cette dénomination des activités relativement diverses, qui se déclinent sous plusieurs formes différentes, et en particulier pour ce qui nous intéresse sous deux formes complémentaires : la numérisation de textes existant déjà sous une forme

médiée différente (éditions imprimées), qui peuvent ainsi entrer dans la confection de corpus (nous y reviendrons) ; et d'autre part ce que l'on nomme 'édition numérique' ou 'édition électronique', qui désigne spécifiquement une édition conçue en fonction des outils désormais à notre disposition, et dont le support est d'emblée informatique.

Ces deux activités ont en commun de traiter des données numériques (qu'elles soient primaires, produites comme numériques, ou secondaires, c'est-à-dire numérisées) suivant une *méthode d'annotation numérique* qui permet de leur intégrer des informations nécessaires à une exploitation automatique, quantitative en particulier. Ces annotations concernent des informations ayant trait soit à *la mise en forme du texte*, soit à *son enrichissement linguistique*. Pour distinguer toutes les unités utiles à la mise en forme et à la structuration du texte (mots, phrases, vers, lignes, paragraphes, pages, folios, colonnes d'un manuscrit, frontières des passages en 'discours direct', etc.), on opère un balisage suivant les normes recommandées par la TEI (et éventuellement adaptées) dans le langage XML. Concernant l'enrichissement linguistique (pour les linguistes, les littéraires, les historiens aussi), on 'pose' des étiquettes morphologiques sur chaque forme (selon une liste de catégories explicite) ; et grâce à des développements récents, on est à même parfois d'opérer un 'passage syntaxique' des textes, suivant des étiquettes fonctionnelles et syntaxiques définies. Enfin, il s'avère toujours utile de pratiquer une lemmatisation des formes du texte.

Par ailleurs, pour exploiter ces données numériques, il faut disposer d'un ou de plusieurs *outils de requêtes* permettant de travailler à partir de ces textes 'enrichis'.

Nous examinerons successivement les deux faces que nous avons évoquées, qui sont désormais à la base de nos recherches linguistiques : les corpus numérisés d'une part, et l'édition numérique de textes d'autre part.

2.2 La 'linguistique outillée' : constitution et exploitation de corpus

Un corpus est une entité 'raisonnée', dont la constitution et les possibilités sont adaptées au type de recherche visé. Dès les années 90, bien des publications ont traité de ces questions (voir par exemple Habert & al. 1997). Pour ce qui concerne la linguistique diachronique du français, plusieurs bases de données ont été constituées depuis quelques décennies, et FRANTEXT a été pionnier dans ce domaine. Généralement, ces bases de données sont constituées de textes déjà édités, que l'on numérise et balise. De ce point de vue, on choisit surtout des éditions fiables, peu interventionnistes, et même des éditions très fidèles, qu'on numérise en les balisant (mot, propositions, phrases, paragraphes, etc.), et éventuellement qu'on enrichit en y intégrant des étiquettes morphologiques ou syntaxiques. Une question de nature différente, mais non triviale car elle se pose encore régulièrement, concerne les droits juridiques de ces textes ; les éditions qui ne sont plus sous droits d'auteur ou d'éditeur échappent à cette difficulté, mais ce sont souvent des éditions anciennes, réalisées suivant des normes un peu rustiques ; dans le cas contraire, il est nécessaire de conclure des accords avec les éditeurs.

Pour étudier l'histoire du français, le linguiste bénéficie de plusieurs recueils de textes ainsi numérisés : FRANTEXT, la BFM (Base de français médiéval de l'ENS de Lyon : <http://bfm.ens-lyon.fr> ou <http://txm.bfm-corpus.org/>), le Nouveau corpus d'Amsterdam désormais hébergé à Stuttgart (<http://www.uni-stuttgart.de/lingrom/stein/corpus>), le corpus MCVF d'Ottawa (www.voies.uottawa.ca/corpus), etc.

2.3 Une édition fidèle à une « version usagée » d'un texte

L'autre voie pour constituer un corpus de français, en particulier pour la période la plus ancienne, est de réaliser directement des éditions numériques selon les normes les plus adéquates aux recherches en cours. Il faut alors choisir parmi des formes d'édition différentes: édition critique, avec un appareil critique éventuellement complexe ; ou édition d'un manuscrit choisi 'de base', dans la tradition 'bédiériste' ('continiste'? 'segriste'?). Cette deuxième option a l'avantage de fournir au linguiste un document dans l'état le plus proche du texte 'réel' tel qu'il a été copié et lu à l'époque de sa production. La transcription peut se faire – et ensuite s'afficher – par exemple sous trois formes, qui reflètent à des degrés divers la forme même du manuscrit :

- Édition 'courante' (ou 'normalisée'), où les abréviations sont résolues, les lettres ramistes insérées, les signes diacritiques (accents, point sur le *i*, cédille) employés selon les normes habituellement pratiquées pour l'édition de texte, les mots séparés comme dans la langue moderne.
- Édition 'diplomatique' n'introduisant aucune de ces modifications, mais n'imitant cependant pas les graphies spécifiques de l'époque (pour *s* et *r* en particulier).
- Édition 'facsimilaire', qui représente les variantes graphiques des lettres, et conserve les abréviations dans leur forme originelle.

Un exemple de ce type d'édition 'très fidèle', est celle que nous avons réalisée de la *Queste del saint Graal* (composée entre 1215 et 1225), à partir d'un manuscrit copié à la fin du XIII^e siècle sans doute, le Ms K dans la terminologie de Pauphilet 1921 (Cote : Ms P.A.77 de la BM de Lyon) (<http://txm.bfm-corpus.org/txm>: version librement téléchargeable selon la licence Creative Commons). Seule une édition de type 'bédiériste' et 'pragmatique' (pour reprendre le terme de F. Duval 2006, ce qui signifie : 'adéquat à la situation'), permet d'utiliser au maximum les possibilités du numérique, notre but étant de disposer d'un texte tel qu'il a été pratiqué, lu et peut-être recopié par les contemporains.

Parmi la cinquantaine de manuscrits qui nous sont parvenus de ce texte, dont une vingtaine copiés au XIII^e siècle, il n'était pas encore possible de réévaluer les stemmas et les regroupements proposés antérieurement (Pauphilet 1921 [1980], Bogdanow 2006). Aussi avons-nous choisi le même manuscrit de base que Pauphilet (1923), le Ms K. Cela avait l'avantage de pouvoir comparer les deux éditions : pour ce texte d'environ 110 000 mots, il existe entre l'édition Pauphilet et la nôtre, faite près d'un siècle plus tard avec des moyens techniques et des principes un peu différents, plus de 20 000 différences. Elles concernent la ponctuation (près de 10 000 différences, ce qui n'est guère étonnant, car même les éditions 'diplomatiques' ne respectent pas toujours la

ponctuation du manuscrit, alors que souvent – c’est le cas du ms. K – elle est parfaitement acceptable pour nous), mais aussi des modifications opérées par Pauphilet dans la graphie des mots et leur segmentation, ou par la suppression ou l’ajout de mots ou expressions (d’après les mss de contrôle), par des changements dans l’ordre des mots (11 000 cas), et par un usage différent des majuscules (près de 3000 différences).

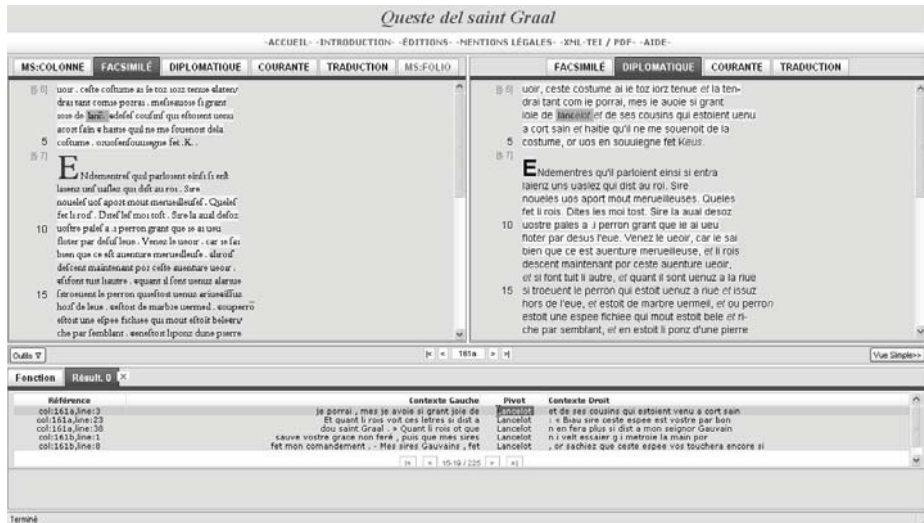
Ces trois versions ‘courante’, ‘diplomatique’ et ‘facsimilaire’, qui forment notre édition du même texte, peuvent être affichées, et lues en parallèle à l’écran grâce à une technique d’affichage multi-facettes en deux volets ; on peut jouer ainsi sur un co-affichage binaire entre 5 textes : les trois versions éditées, la photo du folio manuscrit correspondant, et également la traduction du passage édité.

Il faut souligner qu’un tel résultat n’est possible que grâce à un travail d’équipe très poussé et multi-spécialisé.

Voici ce que l’utilisateur peut voir sur son écran en affichant par ex. une colonne du manuscrit K et son édition courante :

Edition numérique de la Queste del saint Graal (ms K de Lyon) (C. Marchello-Nizia & A. Lavrentev) (ENS Lyon) online <http://txm.bfm-corpus.org/txm> (téléchargeable)

Ou bien, comme ci-dessous, deux versions différentes, facsimilaire et diplomatique, avec au bas la réalisation d’une requête sur les occurrences de ‘Lancelot’ – qui apparaît surligné de rouge dans le texte :



2.4 L'outil de requêtes et de calculs TXM

A partir de tous les textes dont nous disposons, soit numérisés secondairement, soit éditions numériques, nous utilisons le nouveau logiciel TXM (Serge Heiden, Lyon, 2011) pour effectuer des requêtes et des calculs. Cela nécessite d'avoir préalablement constitué un corpus balisé et étiqueté suivant un corps de normes flexibles et adaptables (XML-TEI), et de savoir utiliser le langage de requêtes adéquat.

Nous donnerons quelques exemples pour illustrer le fait que ce logiciel TXM permet certes les requêtes les plus simples, concernant par exemple un mot unique, ou encore la cooccurrence de deux ou plusieurs mots, mais aussi des requêtes bien plus complexes, combinant tout à la fois mots et catégories grammaticales.

Voici quelques exemples de requêtes (CQP) effectuées avec le logiciel TXM :

- Pour obtenir toutes les occurrences d'un mot précis (dans un texte précis, ou dans tout le corpus): soit en entrant simplement le mot dans la fenêtre dédiée, par ex. *chevalier*, ou en utilisant la formule canonique [word=>chevalier]; et si on veut également les formes avec -s final, la formule est : [word=>chevalier.*]
- Ce mot en 'discours direct' : [word=>chevalier & q=>[123]] (où q='quotation')
- Si l'on recherche les occurrences de l'adverbe de phrase *si*, on demandera ses emplois devant un verbe conjugué : [word=>si & q=>[123]] {0, 2} [pos= «VER-cjg»]
- A partir non plus de mots, mais de catégories (pos='part of speech'), on peut demander par exemple les adjectifs qualificatifs postposés à un nom, en 'discours direct', entre le X^e et le XIII^e siècle, ce qui s'exprime ainsi : [pos=>ADJqua & _.text_siecle=>10|11|12 & q=>[123]]

- Si l'on recherche les occurrences du pronom sujet indéfini ON entre le IX^e et le XVI^e siècle la formule sera : [word!=>m's'il|je|j'elle>%c] [word=>on|l'on|om|lom|un|l'un|um|l'um|en|l'en|nen|em|l'em|hom|l'hom|hon|l'hon|l'an|an> %c & ((pos!=>PRE|DET.* « & fropos!=>PRE|DET.*» & _text_siecle =>10|11|12|13|14|15») (frpos!=>PRP|NUM|DET:ART» & _text_siecle=>16»))]]

En combinant de façon encore plus complexe mots, ponctuations et catégories, et en jouant sur leurs exclusions, nous pouvons même aboutir à peu de choses près à la liste des propositions dans la *Chanson de Roland* (éd. Moignet) :

```
[word!=>AOI>] [|* (([word=>»][word=>p{Lu}.*» & word!=>AOI>] |
[word=>»tant>%c] [word=>cum|cume>] |[word=>»] [word=>il|jo|si|sim|sel|s
em|dist>)] [ (word=>[.!?:;->] | (word=>ki.*|qu.*|dunt|dun|sel|sem|sim>%c &
pos=>PROrel|CONsub.*|DETre|DETint>) | (word=>Si>) | (word=>sin>%c) |
(word=>Ne|N'|unches>))] within s
```

En un clic ou presque, en tout cas très rapidement, on obtient un tableau Excel de 4878 lignes, soit autant de propositions potentielles ; après corrections (près d'une dizaine d'heures tout de même), on parvient à un résultat correct de 4974 propositions pour l'ensemble de la *Chanson de Roland* – soit moins de 2% de bruit, ce qui est peu.

2.5 En quoi ces avancées techniques renouvellent-elles l'approche des faits syntaxiques ?

La rapidité, la quantité d'informations traitées (plusieurs millions de mots actuellement), le nombre des facteurs interagissant combinés, la complexité des possibilités de requêtes, atteignent désormais des niveaux jusque là inconnus, et qui étaient bien évidemment inatteignables par des recherches 'à la main'. Et pour l'étude de la syntaxe, qui est un champ complexe où priment les combinaisons de termes et même de construction, il est fondamental de disposer d'un outil qui puisse combiner et hiérarchiser plusieurs facteurs, et qui puisse effectuer non seulement des requêtes sur des termes ou des catégories, mais également des calculs complexes à travers des tris croisés de plusieurs facteurs. Dès lors qu'on dispose de ces possibilités, la pratique se modifie, et les hypothèses que l'on peut formuler en sont modifiées dans leur nature - une très grande complexité devenant possible, ainsi que leurs résultats, puisqu'on parvient à une granularité très grande. Grâce à la rapidité des calculs, on teste aussitôt une hypothèse (par exemple l'existence d'une construction), et les régularités ou les impossibilités apparaissent assez vite, puisqu'est offerte la possibilité de jouer sur les contextes, en modifiant l'un après l'autre chacun des facteurs de la requête.

Cette facilité pour fouiller de très gros corpus permet le repérage quasi immédiat des *innovations* (préluant à un possible changement), mais aussi des *raréfactions* (obsolescences), ou des *absences* (disparitions). Le diachronicien suit ainsi le *rythme* d'un changement, et dans tous ses contextes ; or on sait qu'une innovation apparaît toujours dans un contexte précis, et gagne ensuite successivement d'autres contextes.

On gagne également ainsi la possibilité de *comparer* la grammaire de textes, d'auteurs, d'époques, de dialectes ou sociolectes, ou de genres différents – à la seule condition que ces paramètres aient été renseignés. Ces possibilités ont permis, par exemple, de 'pister' l'apparition, puis le développement sémantique du pronom indéfini *on*, ou bien l'histoire croisée, liée, des adverbes quantifieurs / intensifieurs *moult / très / beaucoup*, à travers les stades d'apparition, de changements de catégorie, d'extension ou de disparition ; de telles explorations exigent le traitement de millions de mots sur quelques siècles.

Nous illustrerons ces développements en syntaxe par quelques exemples pris dans trois domaines de recherche en cours.

3. Syntaxe historique et utilisation des corpus

Nous évoquerons trois domaines dans lesquels les apports de la 'philologie numérique' se sont révélés déterminants : la *punctuation*, l'*ordre des mots*, le *discours direct* (ou '*oral représenté*'). Pour chacun des cas, nous ferons préalablement un bref 'état des lieux'.

3.1 Punctuation et syntaxe : deux cas liés au discours direct, le comma et l'incise

Pour un état de langue ancien, antérieur à l'existence de l'enregistrement, et en l'absence de locuteur natif, il paraît impossible de formuler des hypothèses sur l'intonation des locuteurs. Et des travaux, en nombre réduit cependant (Llamas-Pombo, Lavrentiev), ont montré qu'une étude systématique des éléments segmentants que sont les signes de punctuation aboutit à des enseignements parfois très riches au plan syntaxique.

La détection de certains phénomènes, rendue possible par l'existence d'éditions numérisées respectueuses de la punctuation scribale, a permis, dans les deux cas que l'on va voir, de formuler des hypothèses neuves concernant la *syntaxe* et l'*intonation*. Ainsi, dans certains manuscrits français médiévaux, l'emploi du *comma* semble porteur d'une indication à l'adresse du lecteur concernant le type d'intonation à adopter. Et, inversement, l'absence de punctuation dans des contextes où la langue moderne ponctue systématiquement, invite peut-être à réinterpréter ces éléments. Mais sans l'accès à une transcription numérique très fidèle d'une part, et à un traitement automatique de toutes les unités et de leurs contextes d'autre part, ce type d'analyse n'est guère possible.

Le *comma* est assez peu utilisé au Moyen-Âge – mais peut-être passe-t-il souvent inaperçu ? Et quand il est repéré, il est interprété comme une punctuation forte, et transcrit soit par « ; », soit par un point d'exclamation « ! » (dont il est l'ancêtre semble-t-il).

Notre édition du ms. K de la *Queste del saint Graal*, très respectueuse de la punctuation du copiste, permet de repérer trois signes : l'un très courant, suivi ou non de majuscule, le *punctum*, les deux autres bien plus rares : le *punctum interrogatiuum*, et le *comma* (*punctum eleuatum*, que nous transcrivons imitativement par '·').

Sur les 60 folios occupés par le texte de la *Queste*, on a relevé 195 occurrences du *comma* : quelle en est la valeur ? Il se rencontre principalement dans trois contextes assez différents :

- a) 67 commas suivent immédiatement *Ha/ ha* ou *Hé/hé* ; ils sont en discours direct et leur valeur est clairement exclamative ;
- b) 18 sont employés entre deux membres de phrase corrélés, tels *aussi com...aussi*, ou entre deux propositions parallèles : les exemples sont extraits des versions ‘courante’ ou ‘diplomatique’ de l’édition :

«... car ausi com vos l’avez par achaison de fame .’ ausi le perdreiz vos par damoisele.»
(*Queste* éd. cour. § 64, fol 171c : ‘De même que vous l’avez obtenu grâce à une femme, de même le perdrez-vous à cause d’une demoiselle’)

«... car li filz n’amoit le pere .’ ne li peres l’enfant.» (*Queste* éd. cour. § 50, fol 168d)

- c) 82 sont employés entre une subordonnée initiale en *Quant, Se, Por ce que, qui que*, et la principale :

« Et se uos uolez sauoir donc ces granz auentures sont par tantes foiz auenues .’ alons a lui moi et uos » (*Queste* éd. dipl. § 42, fol 167a : ‘Et si vous voulez savoir d’où ces grandes aventures sont venues tant de fois, allons trouver cet homme, vous et moi.’)

« Mais quant tu ne cuideras pas que tu en puisses eschaper .’ lors descoeuere la croiz et di... » (*Queste* éd. dipl. § 45, fol 167c : ‘Mais quand tu ne croiras pas pouvoir en réchapper, alors découvre la croix et prononce ces paroles :...’)

«... por ce que li lyons estoit meins mesfesanz que li serpenz .’ corui je sus au serpent et l’ocis. » (*Queste* éd. courr. § 132, fol 183a)

Nous constatons que 75% des emplois du *comma* dans ce manuscrit se trouvent dans un épisode au discours direct. Cela pourrait être interprété comme une indication en vue d’une lecture à haute voix d’un passage qui se donne justement pour de l’oral, la prise de parole d’un personnage. Le *comma* indique qu’il existe bien à cet endroit une pause, mais sur laquelle il ne faut pas laisser retomber la voix, car un lien syntaxique et sémantique fort lie les deux portions de l’énoncé (corrélations et parallélismes : *de même que...*, *de même* ; *parce que...*, *alors* ; *quand...*, *alors*).

Quant à l’incise, tout éditeur d’un texte médiéval sait qu’il ‘faut’ la ponctuer en l’entourant de virgules, de même qu’il est habituel d’introduire des guillemets pour indiquer le changement de locuteur. C’est ce que nous avons fait dans la version ‘courante’ de notre édition de la *Queste*. Mais l’usage du copiste du ms. K est tout autre : l’incise *fet X* n’est presque jamais balisée par un signe de ponctuation :

et li rendent son salu, biau seignor fait il quele auenture uos a ca amenez. Sire fet Mes Sires Gauuains la grant fain et le grant desirrier que nos auions de parler a uos (*Queste* éd. dipl. fol 197a, § 192)

Quelle explication fournir à cette étonnante absence de démarquage ? Nous proposerons de voir dans l’incise, justement, une simple marque de changement de niveau énonciatif : l’auteur n’est plus l’instance de discours, c’est un personnage qui l’est

temporairement devenu, et c'est ce seul fait que marque l'incise. Deux phénomènes nous conduisent à cette interprétation. D'une part le fait que, dans ce texte, l'incise a toujours cette même forme canonique (relativement nouvelle alors) : verbe *faire* employé au présent, toujours avec un sujet exprimé et postposé – ce qui est doublement marqué : *fet il, fet li rois* : l'expression est tout à fait formulaire. Et d'autre part, on constate que l'incise ne cooccure pas avec une annonce de discours direct (du type : *et dist au roi*) ; les deux marques sont exclusives l'une de l'autre (à quelques rares exceptions près). Il s'agit simplement de marquer le décalage énonciatif, et l'une des formules suffit, toutes deux étant parfaitement figées dans ce rôle. Dès lors, point besoin de ponctuer, *fet il* a presque une valeur adverbiale.

3.2 De l'ordre des mots

C'est sans doute, depuis un siècle et demi, le domaine de la syntaxe qui a suscité les plus nombreuses études et où les acquis sont les plus conséquents. Dans un premier temps, l'analyse a d'abord été centrée sur le pôle verbal et ses arguments (S, V, O : cf J.-J. Le Coultre, 1875, et L. Foulet, 1967 [1919 / 1930]). Puis l'essor de la grammaire générative a placé l'accent sur la relation S-V, avec la question du sujet non exprimé (ou 'sujet-nul' dans cette problématique) et de la position du verbe (V2), et depuis les premières explorations générativistes de la diachronie dans les années 85, l'attention a été totalement occupée par la périphérie gauche du verbe et sa cartographie, pour laquelle des affinements sont en cours.

La possibilité du traitement rapide de gros corpus a renouvelé et infiniment élargi le corps des énoncés étudiés, et a permis à d'autres orientations de se dessiner. Entre autres points, le nombre des facteurs pris en compte s'est multiplié, et en conséquence la chronologie des changements et leur définition se sont affinées. Désormais, on est à même d'analyser l'ensemble des éléments de la proposition, non seulement S et V, mais aussi l'objet et l'attribut, et tous les éléments, qu'ils soient ou non dans la valence verbale. Les avancées dans l'étude de la typologie des langues du monde ont permis de reformuler la question de la position et de l'optionnalité du sujet et de l'objet dans l'histoire des langues et du français en particulier.

Un grand nombre d'acquis peut servir désormais de socle à des études nouvelles. On sait en effet que la syntaxe des constituants propositionnels du français à époque très ancienne, du IX^e au XIII^e siècle, se caractérise par huit traits :

- 1) Le verbe est majoritairement en seconde position dans tous les types de proposition, mais aucune des autres positions n'est exclue (V1, V3, V4), la position finale méritant une étude en soi, car jusqu'au XIII^e siècle, le verbe final est fréquent en subordonnée.
- 2) La non-expression du sujet domine, plus ou moins largement jusqu'au XIII^e s., sauf dans certaines propositions subordonnées et à certaines personnes de l'interlocution où son expression domine (Vance 1997, Prévoist 2001, Detges 2003).
- 3) Lorsqu'il est exprimé, le sujet est majoritairement antéposé au verbe, mais sa postposition n'est pas rare.

- 4) Dès le XII^e s., l'expression du sujet (nominal et pronominal) et son antéposition au verbe croissent parallèlement.
- 5) En revanche, dès l'origine du français, l'expression de l'objet est obligatoire – sa non-expression étant d'une part très limitée et d'autre part jamais obligatoire (Schoesler 2001, Marchello-Nizia 2009, Rouquier / Marchello-Nizia 2012, 2013).
- 6) Pour le sujet et le verbe, leur syntaxe est dissymétrique entre les déclaratives et les subordonnées.
- 7) Les non-régies (indépendantes et principales) déclaratives sont le type de proposition le plus fréquent dans tous les textes.
- 8) Parmi les non-régies, les types les moins fréquents (interrogatives, injonctives, incises) ont d'emblée ou acquièrent dès l'époque ancienne une syntaxe spécifique.

Concernant l'objet direct, on sait que dès l'origine (IX^e s.) l'objet pronominal est cliticisé au verbe qu'il précède immédiatement, sa postposition existe cependant mais est fort rare. Son expression était obligatoire, sa non-expression restant très limitée, ne concernant que l'objet pronom personnel de 3^e personne - et même alors l'objet peut être exprimé. Quant à l'objet nominal, il peut s'antéposer ou se postposer au verbe, mais il se postpose majoritairement.

Mais récemment, cette fonction 'objet' qui était restée assez peu explorée en syntaxe diachronique, a bénéficié de nombreux acquis nouveaux, tant pour sa chronologie, que pour l'hypothèse de mono-actancialité. On sait désormais que l'objet nominal (nom, pronoms non personnels) est majoritairement antéposé au verbe avant le milieu du XI^e s. (Rouquier / Marchello-Nizia 2012, 2013). Mais il se postpose majoritairement, en toute proposition, dès le milieu du XI^e siècle (*Vie de saint Alexis*). Et dès le début du XIII^e siècle, la postposition de l'objet nominal est devenue la règle en prose (97% des objets nominaux sont postposés au verbe dans la *Queste*). Il est par ailleurs devenu contigu au verbe (*Paieront une contravention ceux qui..*), alors que le sujet continue de pouvoir en être séparé.

Enfin, grâce au traitement automatique d'un grand nombre de textes, l'hypothèse d'une 'monoactancialité' du français a pu se formuler. En effet, avant le XIII^e s., le sujet est absent dans plus de 50% des propositions verbales ; mais il apparaît qu'il est encore plus fréquemment absent dans les propositions à verbes transitifs (Rouquier / Marchello-Nizia 2012, 2013). En effet, l'objet étant obligatoirement exprimé (Schoesler 2001), et le sujet étant optionnel, il en résulte que si dans une proposition à verbe transitif il n'y a qu'un seul argument, c'est nécessairement l'objet. Dès lors, peut-on dire que l'ancien français était une langue monoactancielle orientée vers l'objet ? Peut-être, puisque complémentaiement on constate que le français moderne présente une situation inverse, l'expression du sujet y étant obligatoire, mais non celle de l'objet (Lambrecht / Lemoine 1996, Larjvaara 2000). Resterait à expliquer le passage d'un type de mono-actancialité à un autre.

3.3 L' « oral représenté » et son importance pour décrire le changement linguistique

L'insertion d'un discours dans un récit, et spécialement dans un *récit écrit*, est un phénomène courant, bien repéré et maintes fois étudié depuis plus d'un siècle (Marchello-Nizia 2012) : il s'agit d'un épisode de 'parole', donné comme oral (fictionnel ou ayant réellement eu lieu) et prononcé par un personnage. Les études sur le 'style direct', puis sur le 'discours direct', ont montré que ce changement énonciatif fait l'objet d'un marquage externe par des formules consacrées, avant l'invention d'un signe diacritique spécial, les guillemets, aux 16e-17e siècles. Mais anciennement, seules des formules annoncent d'une part l'arrivée d'un passage en 'discours direct', d'autre part sa clôture avec un retour au récit (cet élément-là n'est généralement pas étudié par les linguistes), et parfois le changement de locuteur est rappelé dans le discours même par une incise qui y est insérée suivant une syntaxe très spécifique. Ces trois formules, nous les nommons *l'annonce*, *le rappel*, et *l'incise*. Un tel marquage externe du discours direct, c'est-à-dire du discours rapporté se donnant comme l'exacte reproduction de paroles ayant été prononcées oralement, semble présent dans les langues du monde de toutes époques. On désigne souvent désormais par l'expression '*oral représenté*' les épisodes de discours direct spécifiques de l'écrit, et spécialement le discours direct inséré dans du récit et annoncé comme du discours rapporté.

Plusieurs études se sont donné pour but de décrire comment l'auteur différencie les deux plans énonciatifs, celui du 'récit' et celui de l'oral (fictionnel ou non). Des particularités linguistiques de l' 'oral représenté' en ancien et moyen français ont déjà été repérées : parmi les premiers, P. Guiraud (1967) avait mis en évidence un statut différent et de la déclinaison, et des démonstratifs, dans les passages en discours direct. Par la suite, les études de B. Cerquiglini (1981), P. Zumthor (2000 [1972] , 1983), et plus tard et de façon plus approfondie D. Lagorgette (1998, 2006) et S. Marnette (2005, 2006), ont contribué à caractériser de façon de plus en plus précise cette grammaire de l'*oral représenté* : l'interjection, l'apostrophe dans ses diverses modalités, les réponses *oie*, *oïl*, *naje*, *nennil*, les exclamations comme *Ha*, *Hé*, la formule de serment et de véridiction *si m'aïst Diex* (Marchello-Nizia 1985), ne se trouvent guère que dans ce type d'épisodes.

Mais outre les morphèmes et expressions qui sont réservés à la représentation de l'oral, d'autres phénomènes grammaticaux distinguent les deux modes du récit et du discours direct, non par une exclusivité d'emploi, mais simplement par une fréquence plus ou moins forte dans chacun des deux types discursifs. Grâce à l'étiquetage morphologique de gros corpus, des requêtes automatiques contrastives menées sur le 'récit' d'une part et sur l' 'oral représenté' d'autre part ont conduit à mettre en évidence l'existence de deux grammaires partiellement différentes, chacune ayant ses « constructions préférées », pour reprendre l'expression de Glikman / Mazziotta (Guillot & al. à par., Glikman / Mazziotta à par., Donaldson à par.) ; un travail en cours sur l'emploi des temps verbaux a permis, pour l'instant à partir d'un corpus réduit, de formuler l'hypothèse que ce n'est pas le présent par exemple qui est le temps

« préféré » du discours direct, mais le futur. Déjà, Marnette (2006) avait mis en évidence six traits distinguant le ‘discours direct’ du récit. Plus récemment, sur un corpus bien plus vaste de 61 textes étiquetés et traités automatiquement (par ‘partitions’ sous TXT et analyses contrastives), Guillot / et al. ont mis en évidence 22 traits par lesquels on peut caractériser le discours direct (DD) et, ce qui est nouveau, 19 particularités du ‘non-DD’, ainsi qu’une série de traits non sensibles à cette distinction.

Il est certain que cette pratique d’une comparaison systématique est riche de découvertes. Ainsi par exemple, les deux paradigmes de démonstratifs CIL et CIST se rencontrent tous deux en récit et en discours direct (question qui avait troublé Guiraud...), alors qu’en très ancien français les formes en *ist* (qui disparaissent après le XII^e siècle) ne se trouvent qu’en discours direct. Mais la fréquence de CIL et de CIST dans chacune des deux modalités d’écrit est inversée : CIL est le démonstratif de loin le plus fréquent à l’écrit, CIST y étant peu fréquemment employé, essentiellement avec une valeur mémorielle et dans un contexte précis, celui du retour au récit après un discours direct (Guillot 2004) ; et inversement, à l’oral représenté c’est CIST qui domine, et dans des contextes qui ont permis de circonscrire plus exactement qu’on ne l’avait fait auparavant la valeur exacte, pragmatique, de ce paradigme : il désigne les éléments que le locuteur définit comme appartenant à sa sphère de locuteur (Marchello-Nizia 2004), tandis que CIL s’emploie, à l’inverse, pour caractériser un élément de la situation comme n’appartenant pas à la sphère du locuteur (Marchello-Nizia 2006). Cela, seule l’analyse automatique d’un très grand nombre de textes variés a permis de le décrire et surtout de l’interpréter.

Nous l’avons vu, le discours direct inséré en récit et balisé par les éléments que sont l’annonce, ou l’incise, et le rappel, peut lui-même, en tant que phénomène, faire l’objet d’une étude diachronique. Dans les plus anciens textes français, on rencontre surtout des monologues (*Vie de saint Léger*), et les dialogues restent relativement peu nombreux (*Sermon sur Jonas*, *Passion de Clermont*, *Vie de saint Alexis*, *Chanson de Roland*). Mais dès les premiers romans du milieu du XII^e siècle, ils deviennent courants et même fréquents. Et on rencontre même dès lors un discours direct enchâssé dans un autre (*Chanson de Roland* 377-391), et un siècle plus tard, dans la prose de la *Queste del saint Graal*, on atteint trois niveaux d’enchâssement énonciatif (éd. en ligne Marchello-Nizia / Lavrentiev, fol 167b, 168a, 190b).

Une ultime question se pose, cruciale pour la linguistique diachronique – et cette question a été bien des fois débattue aux débuts de la linguistique historique. Dès lors qu’on admet communément que c’est dans la pratique de la ‘parole’ (Saussure) et plus spécialement dans les réalisations orales d’une langue qu’apparaissent les innovations qui, par la suite, peuvent se pérenniser en ‘changements’, est-il légitime de considérer ce que nous avons nommé l’‘oral représenté’ comme un lieu où les changements seront mieux détectables qu’ailleurs ? Cette représentation hypothétique des pratiques orales n’est-elle pas un leurre pour le diachronicien ? Nous avons tenté (Marchello-Nizia à par.) de reprendre cette question et d’apporter quelques arguments empiriques en faveur d’une exploration ciblée des représentations de l’oral, en

l'absence de tout accès direct à celui-ci. Dans cette même perspective, Detges (2003) avait émis l'hypothèse qu'un changement syntaxique capital en français, le développement de l'expression du sujet, et d'abord du sujet pronominal, puis sa généralisation, aurait pris son essor en discours direct, dans l'emploi systématiquement 'auto-thématisé' du pronom sujet *je*. Nous avons montré que plusieurs changements dont on a pu récemment repérer les premières attestations montrent que celles-ci se situent très majoritairement dans des épisodes de discours direct, qu'il s'agisse d'interactions verbales en récit, ou bien d'interventions directes de l'auteur dans son texte. On évoquera simplement le cas de deux morphèmes quantifieurs / intensifieurs du français, *très* et *beaucoup*, qui ne sont pas hérités du latin, mais se sont développés longtemps après, et ont fini par remplacer le morphème *moult* (qui subsiste dans la majorité des langues romanes). Les plus anciens emplois de *biau cop* ou *grant cop* à valeur possiblement quantifiante (et non plus simplement métaphorique) sont attestés dès le XIII^e siècle et le début du XIV^e siècle, et sont liés à une énonciation en première personne :

Nos engins getoient au leur et les leurs aus nostres, mes onques n'oÿ dire que les nostres feissent *biau cop*.

(Jean de Joinville, *Vie de saint Louis*, éd. Monfrin, § 193 : 'Nos engins tiraient contre les leurs et les leurs contre les nôtres, mais *je n'ai jamais entendu dire* que les nôtres aient réalisé *un beau coup*' / 'aient abouti à *beaucoup de résultats*')
 (Livre de Job, ms G, composé à la fin du XIII^e siècle, v. 1207-9 : 'Nous devons nous conduire beaucoup mieux à présent que ne le faisait Job, qui vivait avant l'époque chrétienne')

Mieux le devons faire *biau cop*
 De nos tamps que ne faisoit Job
 Qui fu devant le tamps de grasce.

(Livre de Job, ms G, composé à la fin du XIII^e siècle, v. 1207-9 : 'Nous devons nous conduire beaucoup mieux à présent que ne le faisait Job, qui vivait avant l'époque chrétienne')

Quant à l'autre marqueur, *très*, dont les premiers emplois comme intensifieur d'adjectif se rencontrent dans des textes du second tiers du XII^e siècle, on constate que dès cette époque un emploi lui est en quelque sorte réservé puisqu'on n'y rencontre jamais *moult* en cette position : c'est en début d'épisodes d' 'oral représenté', en apostrophe, comme intensif d'un adjectif 'affectif' épithète d'un nom de personne :

« *biax tres dous pere, et vos, bele tres douce mere...* »

(Floire et Blancheflor, court récit en vers écrit vers 1160, v. 3213)

Ce type d'apostrophe est fréquent ensuite, chez Chrétien de Troyes en particulier :

si li dist : " Ma tres chiere dame,
 vos qui..."

(Chrétien de Troyes, *Yvain*, éd. Roques, v. 2551 : 'Et il lui dit : « Ma très chère dame, vous qui... »)

Comme parfois dans le cas d'innovations, la forme introduite est destinée à occuper un emploi nouveau : avec le développement de la représentation des interactions orales insérées dans les récits, l'apostrophe, qui ouvre en général ces épisodes, s'est

peu à peu constituée en formule canonique, et *très* y a trouvé sa niche de développement, avant de se généraliser aux autres contextes.

L'exploitation ciblée des outils et techniques numériques dans le traitement des textes a ouvert la voie vers de nouvelles frontières, comme les exemples ci-dessus évoqués l'ont montré.

Elle permet une exploration approfondie des changements linguistiques, et rend possible la formulation d'hypothèses jusqu'alors hors d'atteinte concernant l'évolution des langues, et en particulier celle-ci, très importante pour la linguistique diachronique : après les discussions sur l'oral lieu privilégié de l'innovation, ne doit-on pas considérer l' 'oral représenté' comme un type discursif témoin privilégié du changement linguistique ?

UMR 5191 ICAR

Christiane MARCHELLO-NIZIA

Bibliographie

- Adams, Marianne P., 1987. *Old French, Null Subjects and Verb Second Phenomena*, Ph. D. : U. C. L.A.
- Bogdanow, Fanny (ed.), 2006. *La Quête du Saint-Graal, roman en prose du XIII^e siècle. Texte établi et présenté par F. B., traduction par Anne Berrie*. Paris, Librairie générale française (Livres de poche. Lettres gothiques).
- Cerquiglini, Bernard, 1981. *La parole médiévale : discours, syntaxe, texte*, Paris, Editions de Minuit.
- Detges, Ulrich, 2003. «Du sujet parlant au sujet grammatical. L'obligatorisation des pronoms sujets en ancien français dans une perspective pragmatique», *Verbum*, 25, 307-333.
- Detges, Ulrich, 2008. «How useful is case morphology? The loss of the Old French two-case system within a theory of Preferred Argument Structure», in: Barddal, J. / Chelliah, S. (ed.), *The Role of Semantic, Pragmatic, and Discourse Factors in the Development of Case, Studies in Language Companion Series*, Amsterdam, Benjamins, 93-120.
- Donaldson, Bryan, 2014. «Socio-Stylistic Reflexes of Syntactic Change in Old French», *Journal of French Language Studies* 24, 319-345.
- Duval, Frédéric, 2006. «La philologie française, pragmatique avant tout ?», in : Duval, F. (éd.), *Pratiques philologiques en Europe*, Paris, Ecole des Chartes, 115-150.
- Ecole nationale des Chartes, Coordination : Vielliard, Françoise / Guyotjeannin, Olivier / Bourgain, Pascale, 2001-2005. *Conseils pour l'édition des textes médiévaux* : Fascicules I, II et III, Paris, Ecole nationale des Chartes.
- Foulet, Lucien, 1967 [1930]. *Petite syntaxe de l'ancien français*, Paris, Champion.
- Glikman, Julie / Mazziotta, Nicolas (2013). «Représentation de l'oral et structures syntaxiques dans la prose de la Queste del saint Graal (1225-1230)», in : Lagorgette, Dominique / Pierre Larrivé (ed.), 43-64.
- Guillot, Céline / Lavrentiev, Alexei / Pincemin, Bénédicte / Heiden, Serge, 2013. «Le discours direct au Moyen Age : vers une définition et une méthodologie d'analyse», in : Lagorgette, Dominique / Pierre Larrivé (ed.), 17-41.

- Guiraud, Pierre, 1967. «L'assiette du nom dans la Chanson de Roland : II : le démonstratif», *Romania* 88, 59-83.
- Habert, Benoît / Nazarenko, Adeline / Salem, André, 1997. *Les linguistiques de corpus*, Paris, A. Colin.
- Heiden, Serge / Decorde, Matthieu / Bertrand, Lauranne, 2013. *Manuel de TXM. Version 0.7*, Lyon, ENS de Lyon. <<http://txm.sourceforge.net/doc/manual/manual.xhtml>>.
- Iser, Wolfgang, 1985 [1972]. *L'Acte de lecture : théorie de l'effet esthétique* (trad. de *Der Akt des Lesens*), Bruxelles, Mardaga.
- Jauss, Hans-Robert, 1978. *Pour une esthétique de la réception*, Paris, Gallimard.
- Lagorgette, Dominique, 1998. *Désignatifs et termes d'adresse dans quelques textes en moyen français*, Université-Paris-10, thèse de doctorat.
- Lagorgette, Dominique, 2006. «Etude des fonctions pragmatiques des termes d'adresse en ancien et moyen français», in : Taavitsainen, I. / Härmä J. / Korhonen, J. (ed.) *Dimensions du dialogisme*, Mémoires de la Société Néophilologique de Helsinki, LXVI, 315-338.
- Lagorgette, Dominique / Larrivée, Pierre (ed.), 2013. *Représentations du sens linguistique 5*, Chambéry, Ed. Université de Savoie.
- Lambrecht, Knud / Lemoine, K. 1996. «Vers une grammaire des compléments d'objet zéro en français parlé», *TCLC* 9, 270-310.
- Larjavaara, Meri, 2000. *Présence ou absence de l'objet : limites du possible en français contemporain*, Helsinki, Academia Scientiarum Fennica.
- Lavrentiev, Alexei, 2008. *Manuel d'encodage XML-TEI étendu des transcriptions de manuscrits dans le projet BFM-Manuscrits. Version 2.1*, Lyon, ENS de Lyon. <http://ccfm.ens-lyon.fr/IMG/pdf/BFM-Mss_Encodage-XML.pdf>.
- Lavrentiev, Alexei, 2009. *Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIII^e au XV^e siècle*. Thèse de doctorat en sciences du langage, Lyon, École normale supérieure Lettres et sciences humaines. <<http://tel.archives-ouvertes.fr/tel-00494914/>>.
- Lavrentiev, Alexei / Marchello-Nizia, Christiane, 2011. «The Development of the Demonstrative Determiner CE in the Manuscripts of Chrétien de Troyes' *Lancelot* or *Le Chevalier de la Charrette*», in : Greco, G. L. / Thorington, E. M. (ed.), *Dame Philology's Charrette, Approaching Medieval Textuality through Chrétien's Lancelot. Essays in Memory of Karl D. Uitti*, Tempe (Arizona), ACMRS, 127-140.
- Le Coultre, Jean-Jules, 1875. *De l'ordre des mots dans Crestien de Troyes*, Thèse, Dresde.
- Llamas Pombo, Elena, 2007. «Réflexions méthodologiques pour l'étude de la ponctuation médiévale», in : Lavrentiev, A. (ed.), *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies. Actes de la journée d'étude de Lyon, ENS LSH, 6 juin 2005*, Chambéry, Université de Savoie, 11-48.
- Marchello-Nizia, Christiane, 2004. «Deixis and subjectivity: the semantics of demonstratives in Old French (9th-12th century)», *Journal of Pragmatics* 37/1, 43-68.
- Marchello-Nizia, Christiane, 2006. «Du subjectif au spatial : l'évolution des formes et du sens des démonstratifs en français», *Langue française* 152, 114-126.
- Marchello-Nizia, Christiane, 2007. «Le comma dans un manuscrit du 13^e s. : grammaticalisation d'un marqueur de corrélation, ou marquage d'une intonation montante ?», in : Bertrand, O. / Prévost, S. / Charolles, M. / François, J. / Schnedecker, C. (ed.), *Discours, diachronie, stylistique du français, Etudes en hommage à Bernard Combettes*, Berne, Peter Lang, 293-305.

- Marchello-Nizia, Christiane, 2009. «L'ordre des constituants phrastiques en français : une évolution conduite par l'objet ?» in : Havu, E. / Helkkula, M. / Tuomarla, U. (ed.), *Du côté des langues romanes, Mélanges en l'honneur de Juhani Härmä*, Helsinki, Mémoires de la Société Néophilologique de Helsinki LXXVII, 33-54.
- Marchello-Nizia, Christiane, 2012. «L'oral représenté» en français médiéval, un accès construit à une face cachée des langues mortes», in : Guillot, C. / et al. (ed.), *Le Changement en français. Etudes de linguistique diachronique*, Berne, Peter Lang, 247-264.
- Marchello-Nizia, Christiane, 2014. «L'importance spécifique de l' "oral représenté" pour la linguistique diachronique», in : Ayres-Bennett, Wendy / Rainsford, Thomas M. (ed.), *L'histoire du français. État des lieux et perspectives*, Paris, Garnier, 161-174..
- Marchello-Nizia, Christiane / Lavrentiev, Alexei, 2013. *Queste del saint Graal. Édition numérique interactive du manuscrit de Lyon (Bibliothèque municipale, P.A. 77)*, Lyon, ENS de Lyon. <<http://txm.bfm-corpus.org/txm>>.
- Marnette, Sophie, 2005. *Speech and Thought Representation in French*, Amsterdam-Philadelphia, John Benjamins Publishing Company.
- Marnette, Sophie, 2006. «La signalisation du discours rapporté en français médiéval», *Langue française* 149, 131-147.
- Moignet, Gérard (ed.), 1969. *La Chanson de Roland*, Texte original et traduction par G. M. Paris, Bordas.
- Pauphilet, Alfred, 1921 [1980 Reprints]. *La tradition manuscrite et l'établissement du texte de la Queste del Saint Graal, attribuée à Gautier Map*, Paris, Champion.
- Pauphilet, Alfred (ed.), 1923. *La Queste del Saint Graal, roman du XIII^e siècle*, Paris, Champion (Classiques français du Moyen Âge).
- Prévost, Sophie, 2001. *La postposition du sujet en français aux XV^e et XVI^e siècles*, Paris, Editions du CNRS.
- Rouquier, Magali / Marchello-Nizia, Christiane, 2012. «De (S)OV à SVO en français : où et quand ? L'ordre des constituants propositionnels dans la *Passion de Clermont* et la *Vie de saint Alexis*», in : Dufresne, M. (ed.) *Constructions en changement. Hommage à Paul Hirschbühler*, Presses de l'Université de Laval, 111-155.
- Rouquier, Magali / Marchello-Nizia, Christiane, 2013. «La position de l'objet direct nominal et l'ordre des mots dans la *Passion de Clermont*, la *Vie de saint Léger* et la *Vie de saint Alexis*», *ACILPR* XXVI, 333-343.
- Schoesler, Lene, 2001. «The coding of the subject-object distinction from Latin to Modern French», in : Faarlund, J. T. (ed), *Grammatical Relations in Change*, Amsterdam-Philadelphia, John Benjamins, 273-302.
- Vance, Barbara, 1997. *Syntactic Change in Medieval French: Verb-Second and Null Subjects*, Dordrecht, Kluwer Academic Publishers.
- Zumthor, Paul, 1983. *Introduction à la poésie orale*, Paris, Le Seuil.
- Zumthor, Paul, 2000 [1972]. *Essai de poésie médiévale*, Paris, Le Seuil (Points).