

REVUE DE
LINGUISTIQUE
FRANÇAISE
DIACHRONIQUE

2021

DIACHRONIQUES

REGARDS LINGUISTIQUES
SUR LES ÉDITIONS
DE TEXTES MÉDIÉVAUX

PDF complet – 979-10-231-2168-1

SORBONNE UNIVERSITÉ PRESSES

Regards linguistiques sur les éditions
de textes médiévaux

Regards linguistiques
sur les éditions
de textes médiévaux

Les SUP, sont un service général
de la faculté des Lettres de Sorbonne Université.
© Sorbonne Université Presses, 2021

Diachroniques n° 8
© Sorbonne Université Presses, 2021
ISBN papier : 979-10-231-0581-0

PDF complet – 979-10-231-2168-1

TIRÉS À PART EN PDF :

Glikman & Verjans – 979-10-231-2169-8
Bragantini-Maillard – 979-10-231-2170-4
Balon – 979-10-231-2171-1
Lavretiev, Guillot-Barbance & Heiden – 979-10-231-2172-8
Mazziotta – 979-10-231-2173-5
Bazin-Tacchella & Souvay – 979-10-231-2174-2

Maquette initiale : Compo-Méca (64990 Mouguerre)
Réalisation : Emmanuel Marc Dubois/3d2s

SUP

Maison de la Recherche
Sorbonne Université
28, rue Serpente
75006 Paris

Tél. (33) 01 53 10 57 60

sup@sorbonne-universite.fr

<https://sup.sorbonne-universite.fr>

Regards linguistiques sur les éditions de textes médiévaux¹

Julie Glikman

Université de Strasbourg, LiLPa

Thomas Verjans

Université Toulouse Jean Jaurès, CLLE

Pour celle-là, la langue est un moyen et le texte,
un but; pour celle-ci, la langue est un but et le
texte, un moyen².

Les débats concernant les possibles relations de la philologie et de la linguistique sont déjà fort anciens, puisqu'ils remontent *a minima* au milieu du XIX^e siècle³ et qu'ils sont régulièrement interrogés à l'aune de l'une ou l'autre des deux disciplines. Et longtemps, semble-t-il, les influences de ce débat sur les éditions de textes sont restées relativement faibles et les bilans

-
1. Le projet présenté ici fait suite à une série de séances menées au cours de l'année 2008 dans le cadre du groupe « Reverdie », associant de jeunes chercheurs en langue médiévale et fondé par Cécile Le Cornec-Rochelois, Anne Salamon et Anne Rochebouet à la faculté des Lettres de Sorbonne-Université. Qu'elles en soient ici remerciées, plusieurs documents demeurant sur le site internet <http://reverdie.free.fr> (rubrique « Séances »). Que soient encore remerciés l'ensemble des collègues qui, à un moment ou un autre du développement de ce projet, l'ont enrichi de leurs expertises et de leurs remarques, et en particulier Joëlle Ducos, Wendy Ayres-Bennet, Sophie Prévost et Lene Schøsler. Que soient enfin remerciés pour leurs relectures attentives et bienveillantes Annie Bertin, Daniéla Capin, Nicolas Mazziotta et Thierry Revol.
 2. Willy Bal, *Introduction aux études de linguistique romane, avec considération spéciale de la linguistique française*, Paris, Didier, 1966, p. 8.
 3. Voir Michel Bréal, *De la grammaire comparée à la sémantique. Textes de Michel Bréal publiés entre 1864 et 1898*, éd. Piet Desmet et Pierre Swiggers, Paris/Louvain, Peeters, 1995, p. 236; Sylvia Adamson et Wendy Ayres-Bennett, « Linguistics and philology in the twenty-first century: Introduction », *Transactions of the Philological Society*, « Linguistics and philology in the twenty-first century », dir. Sylvia Adamson et Wendy Ayres-Bennett, vol. 109, n° 3, 2011.

régulièrement réalisés sur les pratiques éditoriales, en France⁴ comme en Europe⁵, en ont régulièrement réitéré le constat. Ainsi, Gilles Roques remarquait encore que « [l]es éditions de type linguistique, plus rares, sont un genre qui mérite d'être développé⁶ » et esquissait, à la suite de cela, les directions possibles pour combler ce manque :

Mais il y aurait place aussi maintenant pour des travaux plus vastes : éditions de manuscrits dans leur entier et éditions synoptiques de manuscrits, œuvre par œuvre. L'intérêt de ces entreprises serait d'insister sur les variétés linguistiques et culturelles du français. On mettrait ainsi en lumière la part de liberté du scribe, confectionnant un recueil manuscrit de pièces variées, et la part de liberté qu'offrait le français médiéval avec ses variations régionales et stylistiques. En somme, on prendrait le contrepied de l'optique qui a présidé aux éditions antérieures qui cherchaient à offrir un seul texte, celui de l'original perdu ou celui du meilleur manuscrit. En montrant l'importance de la flexibilité qu'offrait la langue française, on progressera encore dans sa connaissance et dans l'appréciation fine de toutes ses nuances⁷.

Frédéric Duval, quelques années plus tard, dressait un constat similaire et appelait de ses vœux une prise en considération plus grande des faits linguistiques dans les introductions :

Dans l'immédiat, il semble important de revoir les introductions linguistiques. Doivent-elles ou non préparer le travail des phonéticiens, des spécialistes de morphologie et de syntaxe, sans compter celui des lexicographes ? Ce serait peut-être un juste retour des choses, car les linguistes permettent aux éditeurs de mieux comprendre leurs textes. Beaucoup d'introductions linguistiques sont aujourd'hui à la fois mal adaptées pour lever les difficultés de la langue médiévale que pourrait rencontrer un lecteur novice, ignorant par exemple la phonétique historique, et peu instructives pour un linguiste⁸.

4. Voir les travaux de Gilles Roques.

5. Voir les travaux de Frédéric Duval.

6. Gilles Roques, « Les éditions de textes », dans Bernard Cerquiglini et Gérard Antoine (dir.), *Histoire de la langue française (1945-2000)*, Paris, CNRS Éd., 2000, p. 877.

7. *Ibid.*, p. 878-879.

8. Frédéric Duval (dir.), *Pratiques philologiques en Europe*, Paris, École nationale des chartes, 2006. Depuis lors, une journée d'études a été consacrée à la question des

Cependant, et depuis plusieurs années, les relations des deux disciplines ont sensiblement évolué, notamment par une approche nourrie de linguistique de certaines éditions de texte récentes, et plus encore, sans doute, par les approches issues de la philologie numérique et par le développement des approches sur corpus.

Il ne saurait être question de présenter dans l'espace de cette introduction une histoire des pratiques philologiques, non plus d'ailleurs que des relations entre linguistique et philologie. Outre le fait qu'il existe des travaux historiographiques qui font autorité⁹, ce numéro s'inscrit résolument dans une approche linguistique, dont certaines des exigences sont sensiblement différentes de celles qui guident les philologues. En particulier il s'agit, pour ces derniers, de donner accès à des textes et, par là-même, à un savoir auquel peu auraient autrement accès, là où le linguiste s'attache à saisir la langue au plus près de sa réalité historique. C'est donc par le biais d'un regard proprement linguistique que doivent être lues les remarques qui suivent, lesquelles visent à esquisser quelques-uns des aspects de cette relation entre linguistique et philologie.

On notera en premier lieu que certaines éditions de textes médiévaux, relativement récentes, ont bien pris en compte l'appel relayé par Frédéric Duval, et que la part consacrée dans les introductions et, plus généralement, dans l'apparat critique à la dimension linguistique, avec une prise en compte des travaux les plus récents, y est sensiblement accrue¹⁰. De la même manière,

introductions linguistiques pour les éditions de texte (Frédéric Duval, Céline Guillot-Barbance et Fabio Zinelli [dir.], *Les Introductions linguistiques aux éditions de textes*, Paris, Classiques Garnier, coll. « Histoire et évolution du français », 2019).

9. Par exemple Frédéric Duval (dir.), *Pratiques philologiques en Europe, op. cit.* ; David Trotter (dir.), *Manuel de la philologie de l'édition*, Berlin/Boston, De Gruyter, 2015.

10. *Inter alia* Laurent Balon, *Transcription de Garin de Monglane à partir du manuscrit du XIV^e siècle Royal 20 D XI de la British Library. Description méthodique du manuscrit et analyses linguistiques (volume I) ; transcription des 12 590 vers de la copie de Londres (volume II) ; notes, glossaire, table des noms propres (volume III)* [thèse de doctorat en sciences du langage, Université Sorbonne Nouvelle, 2008] ; Jean Froissart, *Melyador. Roman en vers de la fin du XIV^e siècle*, éd. Nathalie Bragantini-Maillard, préface de Michel Zink, 2 vol., Genève, Droz, coll. « Textes littéraires français », 2012 ; Anne Rochebouet, *D'une pel toute entiere sans nulle cousture. La cinquième mise en prose du Roman*

et dans le sillage de l'ouvrage de Bernard Cerquiglini (1989) et des travaux menés en dialectologie par Anthonij Dees¹¹, la réflexion sur les variantes et leurs usages s'est développée et a notamment donné lieu à une journée d'études dont les actes ont été publiés¹².

Cette attention portée par les linguistes à la matérialité du texte ainsi qu'au statut du copiste et aux conditions de transmission réelles des manuscrits et de leurs copies s'inscrit pour partie dans le sillage de ce que l'on a nommé « la Nouvelle Philologie », et dont le texte, qui peut passer pour une forme de manifeste, a été accueilli dans le numéro de la revue *Speculum* dirigé en 1990 par Stephen G. Nichols¹³. Les critiques auxquelles celle-ci a pu donner lieu¹⁴, comme, par ailleurs, le caractère qui paraissait alors illusoire de certaines des préconisations de Bernard Cerquiglini concernant l'apport possible de « l'outil informatique »¹⁵, n'ont pas empêché que se développent certaines initiatives que l'on regroupe globalement aujourd'hui sous l'approche nommée « philologie numérique ».

Pour autant, l'essor somme toute encore relativement récent des travaux menés en linguistique diachronique a pu sembler s'éloigner un temps des préoccupations liées à la dimension éditoriale proprement dite, au profit des corpus informatisés. En effet, ceux-ci, s'ils ont le mérite de permettre de formuler des

de Troie, édition critique et commentaire [thèse de doctorat en études romanes, soutenue à l'université Paris Sorbonne sous la dir. de Gilles Roussineau, 2009].

11. Voir Anthonij Dees, « Dialectes et *scriptae* à l'époque de l'ancien français », *Revue de linguistique romane*, 49, 1985, p. 87-117 ; « Propositions for the study of Old French and its dialects », dans Jacek Fisiak (dir.), *Historical Dialectology. Regional and Social*, Berlin/New York/Amsterdam, Mouton De Gruyter, coll. « Trends in Linguistics. Studies and Monographs », 1988, p. 139-148.
12. Cécile Lecornec-Rochelois, Anne Salamon et Anne Rochebouet (dir.), *Le Texte médiéval. De la variante à la récréation*, Paris, PUPS, 2012 ; David Trotter (dir.), *Manuel de la philologie de l'édition*, *op. cit.*
13. *Speculum*, vol. 65, n° 1, « The New Philology », dir. Stephen G. Nichols, 1990.
14. Voir, en France, Philippe Ménard, « Réflexions sur la Nouvelle Philologie », dans Martin-Dietrich Glessgen et Franz Lebsanft (dir.), *Alte und Neue Philologie*, Tübingen, Niemeyer, 1997, p. 17-33 et, plus généralement, Frédéric Duval (dir.), *Pratiques philologiques en Europe*, *op. cit.*
15. Voir Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la philologie*, Paris, Éditions du Seuil, coll. « Des travaux », 1989, p. 112, et Philippe Ménard, « Réflexions sur la Nouvelle Philologie », art. cit., p. 27, pour les critiques de cette approche.

requêtes sur une quantité croissante de données et, partant, de suivre des évolutions avec une précision accrue, n'offrent que rarement accès à l'intégralité des informations historiques susceptibles d'être exploitées à partir des manuscrits, puisque ces corpus sont le plus souvent constitués à partir de la numérisation des éditions courantes. Et si des données complémentaires, telles que les datations, les traits dialectaux majeurs ou l'appartenance générique du texte édité sont naturellement présentes, ni les introductions, ni les variantes ne sont généralement accessibles. Proposant de suivre une voie médiane, qui puisse concilier ces deux exigences, Joëlle Ducos écrivait :

Sans doute faut-il inventer de nouvelles modalités d'éditions où le support informatique permettrait de rendre compte de la multiplicité des réalisations linguistiques d'un même support, où la variante pourrait être la base d'études diachroniques sur les systèmes du français¹⁶.

Or il est vrai que de tels projets apparaissent. L'on peut ainsi songer au projet *Questes*, initié par Christiane Marchello-Nizia, qui, dans son édition de la *Queste del saint Graal*¹⁷, donne à voir les différentes étapes éditoriales, depuis le manuscrit de base jusqu'à la version courante¹⁸. De la même manière, Lene Schøsler avait montré l'enjeu qu'il y avait, pour l'histoire de la langue, à revenir, lorsque cela est possible, à la réalité de la tradition manuscrite¹⁹.

Mais on rappellera tout de même que, dans le sillage de cette initiative aussi bien qu'en relation avec le développement

16. Joëlle Ducos, « Variantes, variance, variations », dans Cécile Lecornec-Rochelois, Anne Salamon et Anne Rochebouet (dir.), *Le Texte médiéval. De la variante à la création*, op. cit., p. 11.

17. Christiane Marchello-Nizia, Alexei Lavrentiev et Céline Guillot, « Édition électronique de la *Queste del saint Graal* », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, op. cit., p. 155-176.

18. Voir dans le présent volume les contributions d'Alexei Lavrentiev, Céline Guillot-Barbance et Serge Heiden (p. 77 sq.) et Sylvie Bazin-Tachella et Gilles Souvay (p. 121 sq.), pour davantage de références sur des projets de ce type actuellement en cours.

19. Lene Schøsler, « Historical corpora. Problems and methods », dans Andrea Bozzi, Laura Cignoni et Jean-Louis Lebrave (dir.), *Digital technology and philological disciplines*, Pisa/Roma, Istituti Editoriali e Poligrafici Internazionali, coll. « Linguistica computazionale », 2004, p. 455-472.

des corpus électroniques, des propositions commencent à être faites en vue de doter la « philologie numérique » des réflexions et des outils nécessaires à sa pratique et à leur harmonisation²⁰. Enfin, l'exploitation croissante des outils informatiques et des systèmes de balisage permet de coder des informations de plus en plus fines et diverses.

C'est dans le prolongement de ces travaux (trop) brièvement évoqués que souhaite s'inscrire le présent volume, en prolongeant un certain nombre de pistes, pour une partie d'entre elles liées à l'apport du numérique, mais sans que ne soit jamais remise en cause la pratique philologique elle-même, d'où le choix du titre de ce numéro : *Regards linguistiques sur les éditions de textes médiévaux*, et non sur la philologie.

En somme, avec le développement des études linguistiques du français médiéval sur de grands corpus, qui, à l'origine, étaient constitués de versions numérisées d'éditions papier, comme la Base de français médiéval (BFM)²¹ ou le corpus de textes du *Dictionnaire du moyen français*, Frantext médiéval, la question de l'édition de textes, et en particulier des informations qu'elle fournit, ou masque, pour le linguiste, se pose²². Les progrès techniques ont par ailleurs permis l'émergence d'une « philologie numérique » dont la souplesse permet d'allier informations utiles pour le linguiste et souci de lisibilité²³. La technologie ne résout cependant pas tout, et le

20. Voir en ce sens les recommandations du Consortium international pour les corpus de français médiéval (en ligne : <http://ccfm.ens-lyon.fr>). Par ailleurs on se souviendra du débat qui a opposé Lene Schøsler à Pieter van Reenen (voir Susan C. Herring, Pieter van Reenen et Lene Schøsler [dir.], *Textual parameters in older languages*, Amsterdam/Philadelphia, J. Benjamins, 2000) et Madeleine Tyssens (voir « Philologie “chevronnée”, Nouvelle Philologie », *Revue de linguistique romane*, 66, 2002, p. 403-420). Notons simplement que celui-ci concernait moins la philologie numérique en soi que certaines de ses applications, et en l'espèce la localisation des manuscrits par l'outil informatique sur la base d'un atlas des chartes. Voir le compte rendu qu'en a proposé Anne Salamon en 2008 sur le site internet Reverdie (<http://reverdie.free.fr/site/spip.php?article29>).

21. Ainsi que le rappellent Alexei Lavrentiev, Céline Guillot-Barbance et Serge Heiden dans leur contribution au présent ouvrage (p. 77).

22. Voir Jean Froissart, *Melyador*, éd. cit., et Laurent Balon, *Transcription de Garin de Monglane...*, thèse cit.

23. Voir par ex. Céline Guillot, Alexei Lavrentiev et al., « La “philologie numérique” : tentative de définition d'un nouvel objet éditorial », dans Richard Trachsler, Frédéric

rôle de l'éditeur moderne reste au centre du processus, comme le montrent dans leurs contributions Nicolas Mazziotta ainsi que Sylvie Bazin-Tacchella et Gilles Souvay. C'est sur ce dialogue entre exigences philologiques, parmi lesquelles le souci de lisibilité, et informations utiles au linguiste que se sont penchés les auteurs des contributions réunies dans ce volume.

Nathalie Bragantini-Maillard présente les principes d'édition retenus pour son édition du ms. Paris, BnF, fr. 99, qui contient une version du *Roman de Tristan* en prose, copiée en 1463 pour Jacques d'Armagnac et réalisée par le copiste Michel Gonnot. Elle insiste notamment sur l'importance de donner une édition du texte qui suit la version du manuscrit en limitant autant que possible les corrections, seul moyen selon elle de mettre en avant le rôle joué par les copistes (par exemple dans les procédés de rajeunissement, d'enrichissement, ou d'adaptation diaphasique ou diatopique de la langue), mais également de donner du texte une « version usagée » (c'est-à-dire telle qu'elle a circulé à une époque donnée, selon les termes de Christiane Marchello-Nizia). À l'ère des éditions numériques, loin d'abandonner la version papier traditionnelle, elle propose au contraire d'utiliser les deux versions et de créer des éditions multisupports communiquant entre elles, la version imprimée contenant l'édition critique et l'apparat critique, glossaire et index compris, tandis que la version numérique permet de donner accès aux données exploitables du manuscrit, comme les pratiques scripturales, à travers des transcriptions diplomatiques et fac-similaires associées. Ensemble, les deux supports permettent ainsi d'arriver à une fine compréhension de la langue médiévale, ainsi que de mesurer le degré de conscience grammaticale du copiste.

Pour Laurent Balon, le manuscrit doit aussi constituer le témoin privilégié pour l'édition, qui doit le restituer au mieux.

Duval et Lino Leonardi (dir.), *Actes du XXVII^e Congrès international de linguistique et de philologie romanes* (Nancy, juillet 2013). Section 13 : Philologie textuelle et éditoriale, Nancy, ATILF/SLR, p. 143-154, 2017 (http://www.atilf.fr/cilpr2013/actes/section-13/CILPR-2013-13-Guillot-Lavrentiev-Rainsford-Marchello_Nizia-Heiden.pdf).

Certains phénomènes dont le manuscrit est un témoin clé peuvent en outre particulièrement intéresser le linguiste, mais restent cependant encore trop masqués par les éditions traditionnelles, comme la question de la résolution des abréviations et celle de la segmentation ou de la séquenciation graphique de certaines unités. Son objectif est, tout en conservant les exigences de l'édition canonique, d'apporter aux textes littéraires le même souci de transcription que pour les documents non littéraires. Il propose ainsi d'adopter un marquage par soulignement, ou italique, pour la résolution d'abréviations. S'agissant de la segmentation, il préconise, à la suite de Clovis Brunel et Jacques Monfrin, l'utilisation du système des tirets (le trait d'union [-] indiquant une séquenciation, et le tiret long [—] une segmentation), développé au départ pour l'édition des documents d'archives, mais qu'il propose d'appliquer à l'édition des textes littéraires, en y ajoutant un troisième signe, le trait d'union redoublé [--] proposé par Nelly Andrieux-Reix pour indiquer un petit espacement – proposition qu'elle n'a cependant jamais appliquée. Il appuie sa démonstration sur de nombreux exemples tirés de son édition de *Garin de Monglane*.

La question de la segmentation, ou « agglutination graphique », est en effet l'un des enjeux de l'édition de textes médiévaux, pouvant permettre d'étudier notamment les figements en cours de certaines séquences. Après avoir montré l'impact possible des choix d'éditeurs sur l'étude d'évolutions de formes en cours de figement dans les textes numérisés de la Base de français médiéval, Alexei Lavrentiev, Céline Guillot-Barbance et Serge Heiden mettent en avant la manière dont les éditions numériques peuvent permettre d'encoder ces informations à différents niveaux, tout en conservant la possibilité de donner à lire une version du texte « normalisée » et sans surcharge, grâce au système de balisage. Chaque unité « mot » fait l'objet d'un balisage spécifique dans l'édition numérique, et le langage XML permet alors l'ajout d'un certain nombre d'informations. Celles concernant les segmentations, ou, à l'inverse, les soudures, telles qu'elles se présentent dans le manuscrit, peuvent faire

l'objet d'une annotation spécifique. C'est le choix qui a été opéré pour l'édition numérique de la *Queste del saint Graal* par Christiane Marchello-Nizia et Alexei Lavrentiev, disponible sur le portail de la BFM²⁴. Par le biais d'un moteur de recherche, ces informations peuvent ensuite faire l'objet d'une requête spécifique, permettant par exemple de sélectionner toutes les séquences agglutinées.

Nicolas Mazziotta s'attache à une question qui peut intéresser les linguistes, mais que l'édition moderne masque souvent : celle de la ponctuation du manuscrit. Le support numérique permet aujourd'hui une flexibilité dans le choix d'affichage donnant la possibilité d'alterner ponctuation ancienne et moderne. Nicolas Mazziotta rappelle cependant que le rôle de l'éditeur moderne reste important. En effet, il lui revient toujours de décider quelle distinction (entre différentes manifestations du *punctus* par exemple, placé plus ou moins haut sur la ligne) est pertinente ou non, et s'il est nécessaire de la transcrire, ce qu'il ne peut pas toujours déterminer *a priori*. Inversement, conserver toutes les distinctions, même minimes, peut se révéler inutile. Cependant la conservation de ces distinctions est capitale si l'on veut pouvoir rendre compte du fonctionnement de ces signes. En transcrivant la ponctuation ancienne, l'éditeur propose ainsi sa propre interprétation, reflet de sa compréhension du manuscrit, et dont il fait le pari qu'elle sera utile. La délimitation entre ce qui relève des variations de forme des occurrences, et des différences de classe, reste donc entièrement prise en charge par l'éditeur moderne, et ne peut être résolue par l'informatique.

Dans leur contribution, Sylvie Bazin-Tacchella et Gilles Souvay présentent les développements récents du lemmatiseur hors contexte LGeRM (pour « Lemmes, graphies lemmatisées et règles morphologiques »), développé à l'origine dans le cadre du *Dictionnaire du moyen français (DMF)* à l'ATILF, et qui devient aujourd'hui un véritable outil d'aide à l'édition de texte. En effet, LGeRM permet d'associer directement les formes du texte aux lemmes connus, mais aussi de faire ressortir une

24. Consultable à l'adresse http://txm.ish-lyon.cnrs.fr/bfm/catalog/qgraal_cm.

« Liste de mots inconnus ». Cette liste comporte des mots non identifiés par le programme, mais révèle également des erreurs de transcription ou de saisie, que les relectures n'avaient pas relevées. Cette étape permet ainsi une première correction du texte. Le lemmatiseur constitue également une aide à la construction de glossaires, et une aide à la lecture, et gagne en retour l'enrichissement de sa base. La lemmatisation finale des formes du texte reste cependant toujours de la responsabilité de l'éditeur, qui doit faire un choix entre les lemmes proposés dans le cas d'ambiguïtés, l'objectif de LGeRM n'étant en effet pas de proposer « une solution unique [mais] d'obtenir, parmi les choix multiples, la bonne solution²⁵ ».

25. Voir *infra*, p. 133.

Suivre la lettre du copiste : l'édition critique au service de la linguistique diachronique et diatopique. L'exemple du ms. Paris, BnF, fr. 99¹

Nathalie Bragantini-Maillard
Université Clermont Auvergne, CELIS

*Not just an ancillary discipline, philology is an attitude of respect for the datum, for the facts of the text and its contexts [...]*².

La connaissance des modalités d'évolution du français à la fin du Moyen Âge ne peut désormais s'affiner sans une reconnaissance véritable du rôle crucial que jouèrent les copistes au plan linguistique dans la diffusion et la survie des textes anciens. Leur action fut double : elle s'exerça sur un axe horizontal, d'un espace linguistique à l'autre, mais aussi sur un axe vertical, à travers les siècles. Dans la pratique scientifique, la reconnaissance de cette influence déterminante doit passer non seulement par une édition des textes respectueuse de la version procurée par un manuscrit donné, mais aussi par un examen documenté, exhaustif et précis des phénomènes linguistiques qui particularisent les témoins retenus. L'objectif serait d'établir une fiche d'identité linguistique de chaque copie étudiée et de tracer le profil linguistique du/des copiste(s) responsable(s) des modifications observées. À terme, l'information recensée

-
1. Une partie des réflexions menées dans le présent article a fait l'objet d'une communication lors de la journée d'étude organisée sous le titre « La fabrique du texte : aspects du travail d'édition (Moyen Âge-xvii^e siècle) » par Florence Bouchet et Jean-Luc Nardone à l'université Toulouse Jean Jaurès le 31 mars 2017.
 2. Siegfried Wenzel, « Reflections on (New) Philology », *Speculum*, vol. 65, n°1, « The New Philology », dir. Stephen G. Nichols, 1990, p. 11-18, ici p. 17-18.

devrait permettre de mieux appréhender les phénomènes de rajeunissement et d'enrichissement du français, ainsi que les procédés d'adaptation diatopique et diaphasique.

La tâche mobilise déjà des équipes de recherche³, qui, faute de disposer d'éditions exploitables, travaillent directement sur les manuscrits en constituant des corpus d'extraits, qu'il leur faut au préalable transcrire. Or, cette étape de préparation diffère une exploration approfondie du matériau linguistique, sans compter que la diffusion de ces corpus se limite à une communauté scientifique assez restreinte. Il conviendrait donc de prêter main-forte à ces projets en développant en amont non seulement des éditions d'intérêt nettement linguistique et stylistique, mais aussi des éditions qui donnent à lire un manuscrit strictement pour lui-même. Le vœu est exprimé par plusieurs linguistes⁴. Bien sûr, il ne s'agit pas de nier la dimension littéraire de tout projet d'édition. Mais, pour l'avoir expérimenté dans le cadre de nos propres travaux⁵, nous avons la conviction qu'il est tout

-
3. Nous pensons en particulier à l'équipe constituée autour d'Yvonne Cazal, Sylvie Lefèvre, Gabriella Parrussa, Cinzia Pignatelli et Richard Trachsler, dans le but d'étudier plus spécifiquement les pratiques graphiques de certains copistes à la fin du Moyen Âge ; voir, par exemple, Yvonne Cazal *et al.*, « L'orthographe : du manuscrit médiéval à la linguistique moderne », *Médiévales*, 45, « Grammaires du vulgaire », 2003, p. 99-118 ; Richard Trachsler, « Des chiffres et des lettres. Observations sur l'orthographe de quelques scribes de la fin du Moyen Âge », *L'Analisi linguistica e letteraria*, 2004, 1-2, p. 471-490 ; Gabriella Parussa et Richard Trachsler, « *Trasmisione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », dans Pierre Nobel (dir.), *La Transmission des savoirs au Moyen Âge et à la Renaissance*, Besançon, Presses universitaires de Franche-Comté, 2005, t. 1, *Du XII^e au XV^e siècle*, p. 348-362 ; Yvonne Cazal et Gabriella Parussa, « Orthographe pour l'œil, pour l'oreille ou pour l'esprit ? Quelques réflexions sur les choix graphiques à la rime dans deux manuscrits du XV^e siècle », dans Alexei Lavrentiev (dir.), *Systèmes graphiques de manuscrits médiévaux et incunables français. Ponctuation, segmentation, graphies*, Chambéry, Université de Savoie, 2007, p. 107-127.
 4. Voir en particulier Bernard Cerquiglini, « La parole étrange », *Langue française*, 40, « Grammaires du texte médiéval », dir. Bernard Cerquiglini *et al.*, 1978, p. 83-98, notamment p. 98 ; Gilles Roques, « Les éditions de textes », dans Bernard Cerquiglini et Gérard Antoine (dir.), *Histoire de la langue française (1945-2000)*, Paris, CNRS Éd., 2000, p. 867-882, plus spécifiquement p. 867-868 et p. 877-879 pour l'édition d'intérêt linguistique, jointe à l'édition de manuscrits pour eux-mêmes ; Frédéric Duval, « La philologie française, pragmatique avant tout ? L'édition des textes médiévaux français en France », dans *id.* (dir.), *Pratiques philologiques en Europe*, Paris, École nationale des chartes, 2006, p. 115-150.
 5. Voir Jean Froissart, *Melyador. Roman en vers de la fin du XIV^e siècle*, éd. Nathalie Bragantini-Maillard, préface de Michel Zink, 2 vol., Genève, Droz, coll. « Textes

à fait possible d'allier les deux champs au sein d'une même entreprise, sans toutefois perdre de vue les attentes de chacun⁶. Pour répondre notamment aux besoins immédiats des linguistes, il convient de rassembler un matériau exploitable sous diverses formes et à grande échelle. Pour cela, il peut être profitable de recourir aux technologies actuelles, qui peuvent donner plus d'ampleur à la dimension linguistique d'une édition.

Pour une alliance de l'édition imprimée et de l'édition numérique

Le travail de l'éditeur scientifique ne se borne pas à la transcription et à l'établissement d'un texte : dans l'idéal, il consiste également à rassembler et organiser une information aussi fiable et complète que possible, notamment au plan des particularités linguistiques et stylistiques. Ce matériau est censé se trouver en introduction à l'édition, mais, pour des raisons matérielles, tout un pan de la documentation historique décelable sur un manuscrit est nécessairement passé sous silence quand l'édition paraît en version imprimée. C'est précisément là que s'affirme l'intérêt du support électronique. Ces dernières années, celui-ci s'est révélé un allié précieux pour la mise en valeur et l'exploitation scientifique du manuscrit médiéval dans sa matérialité. Il ne s'agit pas de nier l'apport d'une version imprimée et de proclamer la supériorité du support électronique. Bien au contraire, le livre demeure irremplaçable pour la formation, mais aussi, au plan cognitif, pour une appréhension tridimensionnelle du texte, qui favorise une appropriation plus sûre et plus critique du contenu, ainsi qu'une circulation plus instantanée et analytique d'un lieu à l'autre du texte et de l'apparat critique. Cependant, nous aurions

littéraires français », 2012 ; Jean Renart, *L'Escoufle*, éd. Jean-Jacques Vincensini et Nathalie Bragantini-Maillard, Paris, Classiques Garnier, à paraître ; *La Queste del saint Graal. Édition numérique du ms. BnF, fr. 98 : Base du français médiéval*, dir. Christiane Marchello-Nizia, Alexei Lavrentiev et Céline Guillot-Barbance, ICAR-CNRS/ENS Lyon : <http://txm.bfm-corpus.org>.

6. Voir en particulier Raymund Wilhelm, « L'édition de texte, entreprise à la fois linguistique et littéraire », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, Berlin/Boston, De Gruyter, 2015, p. 131-152.

tout à gagner à utiliser de concert les deux supports, afin que leur complémentarité serve au mieux le travail éditorial et le développement des connaissances dans le domaine du français et de l'écriture littéraire au Moyen Âge. L'idéal serait de répartir les rôles sous la forme d'une édition multisupport. La version imprimée contiendrait l'édition critique, laquelle assortirait le texte établi d'un apparatus critique présentant la tradition manuscrite, l'intérêt littéraire et les particularités linguistiques et stylistiques du texte. C'est le modèle suivi par la plupart des éditions critiques, avec un espace plus ou moins généreux accordé à l'information linguistique. Quant à la version numérique, elle livrerait les données exploitables à partir du manuscrit, c'est-à-dire tout ce qui touche aux pratiques scripturales et à la variation au sein de la tradition manuscrite.

Des modèles de représentation numérique de textes médiévaux existent déjà. Conjointement à l'édition imprimée procurée par Albert Pauphilet en 1923⁷, *La Queste del saint Graal* contenue dans le manuscrit conservé à la bibliothèque municipale de Lyon (Palais des Arts 77 ; ms. K) et daté du XIII^e siècle bénéficie, depuis 2009, d'une édition numérique interactive, qui livre le texte sous un triple format⁸ : une transcription diplomatique permet de suivre l'usage scriptural du/des copiste(s) ; une transcription courante fournit une édition critique de type bédierien, à corrections limitées de sorte à rester fidèle au manuscrit ; une transcription fac-similaire reproduit quasi à l'identique la page manuscrite dans ses spécificités et ses variations scripturales (abréviations, ponctuation, regroupement de mots, etc.). Le tout est assorti d'une photographie numérisée du manuscrit, d'une traduction, et d'un apparatus critique complet, qui accorde une large place aux considérations linguistiques. Le texte intégrant un balisage XML de type TEI et un étiquetage morphologique, un moteur de recherche permet de faire des requêtes pointues

7. Voir *La Queste del Saint Graal. Roman du XIII^e siècle*, éd. Albert Pauphilet [1923], Paris, Champion, coll. « Classiques français du Moyen Âge », 1984.

8. Voir la Base de français médiéval, déjà citée ; Christiane Marchello-Nizia, Alexei Lavrentiev et Céline Guillot-Barbance, « Édition électronique de la *Queste del saint Graal* », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, op. cit., p. 155-176.

sur des lexèmes à forme variable, des catégories grammaticales, des constructions, des co-occurrences, etc. L'autre atout de ce travail est de livrer également d'autres versions de *La Queste del saint Graal*. Le texte transmis par le ms. Paris, BnF, fr. 98 (= *M*), copié au xv^e siècle et resté à ce jour inédit, profite ainsi également d'une édition multifacette, ce qui offre un premier corpus de comparaison substantiel notamment pour la linguistique diachronique. Le projet met en valeur chaque manuscrit à la fois pour son contenu littéraire et pour la pratique scripturale et le contexte de réception qui encadrent sa réalisation et son utilisation. Son objectif est bien de donner à lire, selon les termes de Christiane Marchello-Nizia, une « version usagée », c'est-à-dire une version telle qu'elle a circulé et été lue à une époque donnée, sans souci de retrouver une hypothétique version originale.

Un autre projet mérite d'être cité pour l'innovation qu'il propose dans l'intégration de l'hypertexte. Robert D. Sanderson a préparé une édition et une présentation du manuscrit M.804 de la Pierpont Morgan Library (New York), qui transmet une version des Livres I et II des *Chroniques* de Jean Froissart⁹. Le travail entre dans le cadre plus large du projet The Online Froissart, qui, mené par des chercheurs des universités de Liverpool et de Sheffield, offre une présentation holistique et dynamique de différentes versions manuscrites des *Chroniques* de Froissart¹⁰. Pour pallier l'inconvénient que constitue la consultation d'un appareil critique moderne, lequel contraint à quitter, un temps, l'espace

9. Voir Robert D. Sanderson, *Linking Past and Future: An Application of Dynamic HTML for Medieval Manuscript Editions* [thèse de doctorat en philosophie soutenue à l'université de Liverpool], 2003; voir aussi Peter F. Ainsworth, « *Ordinatores* et ordinateurs: à propos de quelques avatars récents de la glose », *Perspectives médiévales*, 26, « Actes du colloque "Translatio médiévale" (Mulhouse, mai 2000) », dir. Claudio Galderisi et Gilbert Salmon, 2000, p. 69-82. Pour un exemple de l'exploitation possible de ce type d'édition numérique, voir en particulier Godfried Croenen et Sofie Loomans, « Scribes or copy editors? Scribal behaviour and the production of manuscript versions of Jean Froissart's *Chronicles* in fifteenth-century Paris », *The Online Froissart*, dir. Peter F. Ainsworth et Godfried Croenen (<http://www.hrionline.ac.uk/onlinefroissart/apparatus.jsp?type=intros&intro=f.intros.GC-SLo, v. 1.5, Sheffield, HRIOOnline, 2013>).

10. Voir *The Online Froissart*, déjà cité.

du texte pour prendre un renseignement, l'outil développé par Robert D. Sanderson grâce à l'apport du DHTML offre une solution inspirée de la glose marginale pratiquée par les copistes médiévaux. Chaque feuillet du manuscrit des *Chroniques* est consultable sur une « page virtuelle *en feuilleté*¹¹ », qui garantit un espace unique pour la lecture de la copie numérisée du feuillet, de sa/ses transcription(s) et des commentaires linguistiques nécessaires à la compréhension. En outre, l'entreprise est exemplaire pour la question de la complémentarité essentielle entre support imprimé et support numérique, puisque cette présentation numérique vient compléter sans heurt l'édition partielle des Livres I et II procurée par Peter F. Ainsworth et George T. Diller, dans la collection « Lettres gothiques »¹².

Ce ne sont là que deux illustrations¹³, mais elles donnent une idée du remarquable potentiel que l'association des nouvelles technologies au livre offre au développement de l'édition critique dans une perspective linguistique¹⁴. Chaque projet laisse

-
11. Peter F. Ainsworth, « *Ordinatores* et ordinateurs : à propos de quelques avatars récents de la glose », art. cit., p. 76.
 12. Voir Jean Froissart, *Chroniques. Livre I (première partie, 1325-1350) et Livre II*, éd. Peter F. Ainsworth et George T. Diller, Paris, LGF, coll. « Le Livre de poche. Lettres gothiques », 2001.
 13. Plusieurs autres projets de présentation et édition numériques ont déjà vu le jour. Voir, par exemple, pour les romans de Chrétien de Troyes : <http://www.atilf.fr/dect>; pour *Le Roman de la Rose* : <https://dlmm.library.jhu.edu/fr/bibliotheque-numerique-du-roman-de-la-rose/>; pour le *Queen's Manuscript*, qui contient les œuvres poétiques de Christine de Pizan : <http://www.pizan.lib.ed.ac.uk/index.html>.
 14. Sur cette complémentarité qui existe entre support imprimé et technologies numériques et qui mérite d'être exploitée par les travaux éditoriaux, voir les réflexions méthodologiques menées dans Pierre Kunstmann *et al.* (dir.), *Ancien et moyen français sur le Web. Enjeux méthodologiques et analyse du discours*, Ottawa, Éd. David, 2003, en particulier Martin-Dietrich Glessgen, « La lemmatisation de textes d'ancien français : méthodes et recherches », p. 55-75; Yvan Lepage, « Transcriptions et éditions électroniques de textes en ancien français : bilan et perspectives », p. 105-121; et dans Andrea Bozzi *et al.* (dir.), *Digital Technology and Philological Disciplines*, Pisa/Roma, Istituti Editoriali e Poligrafici Internazionali, coll. « Linguistica computazionale », 2004, en particulier Andrea Bozzi et Maria Sofia Corradini, « Aspects and methods of computer-aided textual criticism », p. 49-66; Claus Huitfeldt, « Text technology and textual criticism », p. 259-275; Manfred Thaller, « Digital manuscripts as base line for dynamic editions », p. 489-511. Voir aussi Horst Fuhrmann, « Réflexions d'un éditeur », dans Jacqueline Hamesse (dir.), *Les Problèmes posés par l'édition critique des textes anciens et médiévaux*, Louvain-la-Neuve, Université catholique de Louvain, coll. « Publications de l'Institut d'études médiévales », 1992, p. 329-359.

une marge de manœuvre appréciable, qui permet à l'éditeur de moduler à sa convenance ses critères d'intervention sur le texte, sans pour autant priver les linguistes d'une transcription fidèle à la copie manuscrite, voire de documents complémentaires (autres versions manuscrites numérisées et retranscrites, relevés linguistiques et stylistiques, etc.) difficiles à intégrer à une édition imprimée, mais qu'un site internet peut tout à fait accueillir.

Éditer en version usagée le ms. Paris, BnF, fr. 99

Le projet qui nous occupe autour du manuscrit français coté 99 à la Bibliothèque nationale de France appelle tout particulièrement une telle mise en œuvre éditoriale. Le témoin en question contient une version du *Roman de Tristan* en prose copiée en 1463 pour l'illustre bibliophile Jacques d'Armagnac, duc de Nemours. On sait que la copie fut réalisée par le copiste Michel Gonnot, qui signe l'ouvrage par le colophon suivant :

spécifiquement p. 336-338; Bernard Cerquiglini et Jean-Louis Lebrave, « Philectre: un projet de recherche pluridisciplinaire en philologie électronique », *Zeitschrift für Literaturwissenschaft und Linguistik*, vol. 27, n° 2, 1997, p. 83-93; Karl D. Uitti, « Informatique et textualité médiévale: l'exemple du "Projet Charrette" », *Le Médiéviste et l'Ordinateur*, 37, « Le texte médiéval sur Internet. I », 1998, p. 25-36; Gina Greco, « L'édition électronique de textes médiévaux. Théorie et pratique », dans Gabriel Bianciotto et Claudio Galderisi (dir.), *L'Épopée romane*, Poitiers, Université de Poitiers, Centre d'études supérieures de civilisation médiévale, 2002, t. II, p. 1045-1050; Cinzia Pignatelli, « Y a-t-il une vie pour les variantes après l'édition critique? Une recherche en cours sur les manuscrits du *Chevalier de la Charrette* », dans *Comme la lettre dit la vie. Mélanges offerts à Michèle Perret*, rééd. dans Dominique Lagorgette (dir.), *Littérature et linguistique: diachronie/synchronie. Autour des travaux de Michèle Perret*, CD-Rom, Chambéry, Université de Savoie, 2007, p. 439-452; Martin-Dietrich Glessgen, « L'élaboration philologique et l'étude lexicologique des *Plus anciens documents linguistiques de la France* à l'aide de l'informatique », dans Frédéric Duval (dir.), *Frédéric Godefroy*, actes du X^e Colloque international sur le moyen français (juin 2002, Metz), Paris, École nationale des chartes, 2003, p. 371-386; Pierre Kunstmann et Achim Stein (dir.), *Le Nouveau Corpus d'Amsterdam*, actes de l'atelier de Lauterbad (23-26 février 2006), Stuttgart, F. Steiner, 2007; Frankwalt Möhren, « Édition, lexicographie et l'esprit scientifique », dans David Trotter (dir.), *Present and future research in Anglo-Norman*, actes du colloque d'Aberystwyth (juillet 2011), Aberystwyth, Art & Humanities Research Council, 2012, p. 1-13, p. 4 pour notre propos; Pierre Kunstmann, « Le dictionnaire électronique au centre du travail d'édition des textes: quelques réflexions sur les rôles et rapports de l'éditeur et du lexicographe », dans David Trotter (dir.), *Present and future research in Anglo-Norman*, op. cit., p. 159-162.

« Au jourd'uy huitiesme jour du mois d'octobre mil cccc lxiij fu achevé ce ronmans par Micheau Gonnot de la Brouce, prestre demeurant a Crousant » (fol. 775v). Il est donc permis de penser que le manuscrit permettrait de suivre de manière privilégiée certains phénomènes de modernisation du français dans la seconde moitié du xv^e siècle, ainsi que l'influence possible du milieu de copie. Autrement dit, par l'examen de sa langue, on pourrait se faire une idée de la marque laissée sur le texte par le copiste. L'examen est susceptible de nous renseigner sur les interventions stylistiques qu'a pu s'autoriser Michel Gonnot en fonction de ses sensibilités et/ou pour satisfaire un goût du commanditaire, amateur éclairé de littérature arthurienne. C'est en particulier pour explorer ces questions, sur lesquelles nous reviendrons, que nous avons décidé d'éditer cette version du *Roman de Tristan* en prose selon une approche holistique (texte, image, paratexte, contextes social, culturel, etc.), qui vise à conjuguer des intérêts littéraires, linguistiques et stylistiques, mais aussi ceux de l'histoire de l'écriture et de l'histoire de l'art. Il s'agira de fournir un ensemble de documents exploitables par ces disciplines à l'aide des différents supports mis actuellement à notre disposition¹⁵.

Bien sûr, se pose la question de l'*exemplar* du fr. 99. On ignore encore en effet à partir de quel modèle travailla Gonnot. Dans ce cas, à quoi comparer le texte de fr. 99 pour en dégager les spécificités de tous ordres? Comment attribuer tel ou tel phénomène à Gonnot? Ces questions n'entament cependant

15. Sur cette approche globale qui reconnaît la matérialité du texte manuscrit médiéval et les conséquences de la mouvance textuelle au Moyen Âge, voir tout particulièrement les contributions au numéro spécial, déjà cité, de *Speculum*, ainsi que Andrea Bozzi et al. (dir.), *Digital Technology and Philological disciplines*, op. cit. Sur l'intérêt linguistique que représentent les manuscrits précisément documentés et sur la nécessité scientifique d'en procurer des éditions en version usagée, rassemblant des données aussi exhaustives et fiables que possible afin de faciliter la constitution de corpus, voir Alexei Lavrentiev, « Typologie textuelle pour l'étude linguistique de manuscrits français médiévaux », dans id. (dir.), *Systèmes graphiques de manuscrits médiévaux et incunables français. Ponctuation, segmentation, graphies*, op. cit., p. 49-66, plus spécifiquement p. 51-61; « Base de français médiéval et transcriptions de manuscrits : recherche de complémentarité », dans David Trotter (dir.), *Actes du XXIV^e Congrès international de linguistique et de philologie romanes, Aberystwyth, 1^{er}-6 août 2004*, Tübingen, Niemeyer, 2007, t. 1, p. 405-410.

en rien le bien-fondé du projet éditorial. Quant à l'examen linguistique, parmi les principes que nos prédécesseurs ont arrêtés à ce jour, il est établi que « le modèle transcrit par le scribe ne paraît pas entrer en ligne de compte de façon significative pour les phénomènes qui lui importent¹⁶ », étant entendu que, à ce stade de la recherche, lesdits phénomènes se situent essentiellement au plan graphématique. La comparaison avec les autres manuscrits de la famille permettra d'affiner, dans tous les cas, le profil scriptural de Gonnot, voire d'identifier un modèle.

En attendant l'heure de la comparaison, il convient de faire connaissance avec le texte du fr. 99, en tentant de cerner au mieux certaines de ses spécificités. Pour ce faire, nous prévoyons tout d'abord de consacrer la publication imprimée à une édition à intervention minimale du fr. 99. Une édition critique destinée à une publication sous forme de livre est apte à satisfaire tout à la fois intérêts littéraires, intérêts linguistiques et stylistiques, pourvu que des critères précis et cohérents soient mis en place et rigoureusement suivis lors de sa préparation. D'emblée, il faut décider du degré de fidélité.

Dans le cas précis du *Tristan* en prose, nous avons affaire à un texte dont différentes versions ont déjà été éditées. L'examen d'Emmanuèle Baumgartner a dégagé au moins quatre versions du *Tristan* en prose¹⁷. La version 1, dite V.1, correspond à une forme lacunaire, équivalant à la seconde partie de l'œuvre (version courte). Elle se lit dans une édition collective du ms. Paris, BnF, fr. 757 (fin du xiv^e siècle), dirigée par Philippe Ménard et parue chez Champion dans la collection « Classiques français

16. Richard Trachsler, « Des chiffres et des lettres. Observations sur l'orthographe de quelques scribes de la fin du Moyen Âge », art. cit., p. 473 ; voir aussi Yvonne Cazal *et al.*, « L'orthographe : du manuscrit médiéval à la linguistique moderne », art. cit. ; Gabriella Parussa et Richard Trachsler, « *Trasmissione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », art. cit., p. 356.

17. Voir Emmanuèle Baumgartner, *Le « Tristan en prose »*. *Essai d'interprétation d'un roman médiéval*, Genève, Droz, 1975, plus spécifiquement p. 29-87. Nous reprenons les abréviations qu'E. Baumgartner utilise pour désigner les quatre versions du *Tristan en prose*.

du Moyen Âge »¹⁸. La version II correspond à une version longue, cyclique, considérée comme la vulgate. Renée L. Curtis a réalisé l'édition de sa partie initiale (jusqu'à l'épisode de la folie de Tristan) d'après le manuscrit fragmentaire 404 de la bibliothèque municipale de Carpentras, produit au XIII^e siècle¹⁹, tandis qu'une autre équipe dirigée par Philippe Ménard a édité, chez Droz, dans la collection « Textes littéraires français », la suite des aventures tristaniennes, sur la base du ms. Vienne 2542 (ca 1300)²⁰. Les V.I et V.II résultent de remaniements composites d'une rédaction originale perdue, sans doute réalisée vers 1230-1240 : elles remontent donc, au plus tard, à la seconde moitié du XIII^e siècle.

Les deux autres versions, V.III et V.IV, respectivement datées de la seconde moitié du XIII^e siècle et des années 1330-1340, demeurent inédites. Pourtant, la version tardive IV présente un intérêt indéniable au plan littéraire, puisqu'elle constitue une compilation des trois autres versions et de textes étrangers au *Tristan* en prose. Elle mérite donc une édition qui ne dénature pas sa structure et son traitement de l'histoire. À cet égard, approche littéraire et approche linguistique s'accordent. Si, en outre, on retient le ms. BnF, fr. 99, seul témoin de sa famille à contenir l'intégralité de la V.IV et à présenter un contexte de copie nettement identifié, toutes les conditions sont réunies pour justifier une édition qui respecte au mieux la lettre du copiste²¹.

18. Voir *Le Roman de Tristan en prose. Version du ms. fr. 757 de la Bibliothèque nationale de Paris*, éd. dir. Philippe Ménard, Paris, Champion, coll. « Classiques français du Moyen Âge », 5 vol., 1997-2007.

19. Voir *Le Roman de Tristan en prose*, éd. Renée L. Curtis, Cambridge, D.S. Brewer, t. I [1963], 1986 ; t. II [1976], 1985 ; t. III, 1985.

20. Voir *Le Roman de Tristan en prose*, éd. dir. Philippe Ménard, Genève, Droz, coll. « Textes littéraires français », 9 vol., 1987-1997.

21. Par une heureuse coïncidence, notre choix du ms. BnF, fr. 99 combine à la fois des critères textuels de premier ordre (type de version, complétude du texte, etc.) et des critères matériels déterminants pour une meilleure connaissance linguistique (identité du copiste et du commanditaire, date et lieu de copie, etc.). Bien entendu, dans l'optique linguistique qui nous retient, ce sont avant tout les critères externes de la localisation de la copie et de l'identité du copiste et du commanditaire qui confèrent sa supériorité au fr. 99 sur les autres témoins de la famille. C'est une *scribal version* que nous éditons. Sur la validité à accorder désormais aux éléments de critique externe dans le choix d'un manuscrit à éditer, voir notamment Olivier Delsaux et Tania Van Hemelryck, « Réflexions sur le choix du manuscrit de base aujourd'hui ou la précaution inutile ? »,

En d'autres termes, il est possible de proposer une « version usagée » de ce *Tristan* en prose. L'édition de cette version se trouvera ainsi inscrite dans une approche pragmatique de la philologie, selon laquelle, affirme Suzanne Fleischman, « *the New Philologist must, insofar as possible, recontextualize the texts as acts of communication*²² ». Une telle approche impose, bien sûr, de renoncer au mythe d'une version d'auteur, authentique et immuable, et d'admettre que la transmission textuelle reposait avant tout au Moyen Âge sur « un concept de propriété collective », qui aboutissait inévitablement à des « versions vivantes »²³, fussent-elles même rédigées sous le contrôle de l'auteur²⁴.

Cela étant, pour mettre le texte en perspective, nous ménagerons une place de choix aux variantes pertinentes, susceptibles d'indiquer le degré d'intervention du scribe Gonnot. À ce jour, nous ne connaissons que trois autres manuscrits porteurs de la V.^{iv}²⁵. Ce petit nombre de témoins facilite un contrôle systématique des leçons et rend possible un relevé des variantes instructives aux plans littéraire et linguistique. De la sorte, le lecteur moderne pourra se plonger dans un texte

Perspectives médiévales, 34, « Les textes médiévaux face à l'édition scientifique contemporaine », 2012.

22. Suzanne Fleischman, « Philology, linguistics, and the discourse of the medieval Text », *Speculum*, vol. 65, n°1, dir. Stephen G. Nichols, 1990, p. 19-37, p. 37 pour la citation. Ce numéro de *Speculum* est entièrement consacré à la « New Philology », qui, ayant pris acte de la mouvance constitutive du texte médiéval, est convaincue de l'intérêt d'éditer un texte manuscrit pour lui-même et de suivre le travail d'appropriation du/des copiste(s), afin notamment de saisir la dynamique langagière mise en œuvre par le(s) copiste(s) et de donner ainsi un accès à un état de la langue des usagers de l'époque et à la variance du texte. En dehors de ce recueil, voir aussi Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la philologie*, Paris, Éditions du Seuil, 1989; Lene Schøsler, « Historical corpora. Problems and methods », dans Andrea Bozzi et al. (dir.), *Digital Technology and Philological Disciplines*, op. cit., p. 455-472.
23. Malachi Beit-Arié, « Transmission de textes par scribes et copistes. Interférences inconscientes et critiques », dans Jacqueline Hamesse (dir.), *Les Problèmes posés par l'édition critique des textes anciens et médiévaux*, op. cit., p. 173-196, p. 179 et 195 pour les citations respectives.
24. Les nombreuses réécritures que Jean Froissart réalisa des livres de ses *Chroniques* en les adaptant notamment aux sensibilités politiques des différents destinataires de l'œuvre en sont un parfait exemple.
25. Il s'agit des mss Chantilly, 645-646-647; New York, Pierpont Morgan Library 41; Leningrad, Fr. v. XII.2.

tel qu'il fut lu à la cour de Jacques d'Armagnac, tout en disposant d'un instrument de travail fiable pour un examen comparatif plus poussé. Ce relevé des variantes sera naturellement soumis aux limites imposées par l'édition imprimée, et toutes les variantes ne pourront être recensées. Pour autant, ce premier balayage ne manquera pas de donner un aperçu global et instructif des autres rédactions de la V.iv, tant au plan linguistique qu'au plan littéraire. Il sera relayé par un travail réalisé en parallèle sous format électronique, nous y reviendrons.

Toujours dans un souci de transparence, notre édition restreint les corrections. Une transcription respectueuse du texte manuscrit garantit une connaissance plus fine de la langue médiévale. Dans plusieurs éditions, anciennes ou plus récentes, trop de faits linguistiques marginaux sont passés à la trappe, sans avoir eu droit de cité dans les ouvrages linguistiques, à la suite d'une uniformisation systématique de la langue des textes. Il s'est pourtant avéré *a posteriori* que, en dépit de leur incongruité, nombre de particularités ont bel et bien existé et qu'en tant que telles, elles auraient mérité de passer intactes dans une édition critique afin de nourrir l'information des linguistes²⁶. Par conséquent, les interventions que nous nous autorisons sur le texte du *Tristan* visent essentiellement à l'intelligence du texte : elles portent, par exemple, sur des coquilles évidentes qui rendent un mot illisible, sur des lacunes de scribe qui entravent gravement la compréhension. Dans tous les cas, elles sont signalées par une mise entre crochets de la leçon exogène, que celle-ci s'inspire des manuscrits de contrôle ou relève de notre seule initiative. La leçon originale est reportée en bas de page, avec mention des variantes ayant servi à la correction, le cas échéant. Grâce au principe de respect du texte se trouve ainsi préservée toute leçon curieuse, mais qui pourrait

26. À ce sujet, voir en particulier Cinzia Pignatelli, « Y a-t-il une vie pour les variantes après l'édition critique ? Une recherche en cours sur les manuscrits du *Chevalier de la Charrette* », art. cit., p. 444-446 ; Lydie Louison, « Les dangers de l'harmonisation systématique : la transcription des termes en *-ié/-ie* dans *Cristal et Clarie* », dans Corinne Füg-Pierreville (dir.), *Éditer, traduire ou adapter les textes médiévaux*, Lyon, CEDIC, Centre Jean Prévost, Université Jean Moulin Lyon III, 2009, p. 55-62 ; Frankwalt Möhren, « Édition, lexicographie et l'esprit scientifique », art. cit., p. 5.

se révéler récurrente au moins à un autre endroit de la copie ou bien encore dépendre d'une influence diatopique.

Outre une présentation de la tradition manuscrite et de l'intérêt littéraire de la version transmise par BnF, fr. 99, l'apparat critique proposera, sous différentes rubriques, un premier recensement des particularités linguistiques et stylistiques du manuscrit. Tout d'abord, une étude synthétique rendra compte des principaux phénomènes touchant à l'état de la langue (graphiques, morphologiques, syntaxiques, lexicaux) afin d'esquisser le profil linguistique de la copie dans une triple perspective, diachronique, diatopique, diaphasique. Elle tentera ainsi de cerner l'état du moyen français utilisé et de le situer par rapport à sa date de copie (1463), dans l'espoir d'évaluer son degré de modernisation. Mais elle collectera aussi l'ensemble des indices susceptibles de nous renseigner sur l'adaptation linguistique que Gonnot aurait pu introduire dans sa *scripta*²⁷ sous double influence d'oïl et d'oc. Une première moisson des principaux traits stylistiques propres au ms. BnF, fr. 99 livrera aussi, à n'en pas douter, des éléments de connaissance sur les tics et les préférences du copiste. Bien sûr, les variantes seront utiles pour l'exploration de ces trois champs, ainsi que l'examen d'autres manuscrits rédigés par Gonnot²⁸.

Loin d'être un simple soutien à la lecture, ce type de panorama linguistique et stylistique serait censé ouvrir toute édition soucieuse de servir l'avancée des recherches linguistiques, puisque, à force de côtoyer un texte et sa langue, l'éditeur est le mieux placé pour en saisir les particularités en réduisant la marge d'erreur dans l'interprétation des faits de langue. Il est à même

27. La *scripta* se définit comme un système de langue vernaculaire écrite artificielle et hybride, composée de traits régionaux hétérogènes et variables. Elle est à distinguer de la langue vernaculaire parlée, laquelle reste pour ainsi dire insaisissable pour la période médiévale.

28. Outre le manuscrit ici considéré, de 1463 aux environs de 1476, Gonnot réalisa et signa trois copies pour Jacques d'Armagnac : *Marques de Rome* (Paris, BnF, fr. 93, en 1466), une compilation arthurienne qui contient une version abrégée du *Tristan* en prose (Paris, BnF, fr. 112, en 1470) et une compilation de textes religieux (Paris, BnF, fr. 916, en 1474). Il est probable que la compilation de textes religieux contenue dans le ms. Paris, Arsenal, 5121 (non daté) est aussi de sa main.

de contextualiser le plus finement possible la plupart des emplois lexicaux, des constructions syntaxiques, etc. Cette exigence de contextualisation a son importance au plan scientifique. Dans leur récente analyse des constructions clivées en ancien français, Achim Stein et Carola Trips ont souligné en particulier combien la prise en compte du contexte est déterminante dans la reconnaissance d'une mise en relief au sein d'une phrase et, partant, dans le recensement des constructions clivées employées dans un texte et à une époque donnée²⁹. L'éditeur est donc tout désigné pour résoudre des problèmes d'étiquetage liés à une interprétation contextuelle et/ou aux tendances stylistiques d'une copie. La synthèse linguistique qui accompagne une édition est le lieu par excellence de ce genre de relevé.

Glossaire et index seront également mis à contribution pour nourrir la réflexion et fournir une documentation substantielle à l'instruction des lexicologues et des lexicographes³⁰. Le glossaire offrira un relevé aussi complet que possible du vocabulaire employé dans le manuscrit, et ce, selon des principes scientifiques qui assureront un repérage rapide des phénomènes dignes d'intérêt pour l'enrichissement des dictionnaires par exemple (néologismes, régionalismes, hapax). Compte tenu de l'ampleur du texte, dont l'édition s'échelonnait sur plusieurs années et paraîtra en plusieurs tomes, il paraîtrait judicieux de rassembler les occurrences d'une même lexie sous un lemme (au cas régime pour les noms, les adjectifs et les déterminants) correspondant à la graphie moderne ou, pour les mots sortis d'usage, à la graphie adoptée par le *Dictionnaire*

29. Voir Achim Stein et Carola Trips, « Les phrases clivées en ancien français : un modèle pour l'anglais ? », *Revue de linguistique romane*, 78, 2014, p. 33-56, p. 41-42 pour notre propos.

30. Voir Albert Henry, « Lexicologie géographique et ancienne langue d'oïl », *Romanic Philology*, vol. 26, n° 2, 1972, p. 229-255, pour notre propos p. 229-236 et sa conclusion p. 254-255 ; Robert Martin, « Les lexiques d'auteurs et les lexiques de "genres" comme préalables à un *Dictionnaire du moyen français (DMF)* », dans *Le Moyen Français. Actes du 4^e colloque international sur le moyen français. Milan, 6-8 mai 1985*, Milano, Vita e Pensiero/Pubblicazioni della Università Cattolica del Sacro Cuore, t. II, *Le Moyen Français*, 1986, p. 131-139 ; Frankwalt Möhren, « Édition, lexicographie et l'esprit scientifique », art. cit.

*du moyen français*³¹. Mais cela pourrait aussi poser un problème de repérage à un lecteur non spécialiste. Il semblerait donc plus simple que le lemme corresponde à la forme la plus fréquente dans une portion de texte éditée. Si le texte ignore la forme du lemme (par ex. un infinitif), celle-ci sera encadrée de crochets, et les variantes graphiques seront relevées à la suite de la vedette. À la rigueur, pour être immédiatement localisables, les variantes rares figureront entre parenthèses après le renvoi textuel concerné. Chaque lemme sera catégorisé selon la terminologie en vigueur (*s. m.* = substantif masculin, *v. tr.* = verbe transitif, *adv.* = adverbe, etc.). Cette présentation ne manquera pas de dessiner progressivement une vue d'ensemble de la mouvance graphique du manuscrit et de révéler le degré de stabilité de la palette que déploie le scribe au fil de la copie.

Les locutions feront aussi l'objet d'un relevé systématique avec identification catégorielle (locution conjonctive, locution adverbiale, etc.) et glose. Les emplois relevant d'une phraséologie particulière et les associations synonymiques ou antonymiques seront mentionnés entre parenthèses au fil du relevé. Enfin, néologismes, hapax et régionalismes seront signalés par un système d'indications historiques ou diasystémiques. En outre, une partie de la synthèse linguistique consacrée à l'état du lexique employé leur sera réservée : elle classera les occurrences en fonction de leur(s) particularité(s).

S'agissant de l'interprétation sémantique des termes, certains linguistes reprochent aux glossographes de proposer des gloses traductives, qui plus est trop contextualisées, et ils regrettent que la plupart des glossaires n'aient pas recours à une définition sur le modèle des dictionnaires³². Le vœu est légitime,

31. Voir le *Dictionnaire du moyen français (DMF)*, Nancy, Université de Lorraine, ATILF/CNRS, v. 2015 (<http://www.atilf.fr/dmf>). L'abréviation *DMF 2015* renvoie spécifiquement à cette dernière version en date.

32. Voir, entre autres, Gilles Roques, « À propos d'éditions récentes de textes en moyen français. Problèmes et méthodes en lexicologie médiévale », dans Marc Wilmet (dir.), *Sémantique lexicale et sémantique grammaticale en moyen français*, actes du colloque organisé par le Centre d'études linguistiques et littéraires de la Vrije Universiteit Brussel de Bruxelles, 28-29 septembre 1978, Bruxelles, VUB Centrum voor Tal- en Literatuurwetenschap, 1979, p. 3-21; Kurt Baldinger, « Splendeurs et misères des

mais il occulte quelque peu la double vocation d'un glossaire d'édition, double vocation qui, à notre sens, est incontournable. Si un glossaire offre en effet une documentation exploitable à terme par la lexicographie, il n'en demeure pas moins un outil à usage immédiat lors de la lecture du texte. Le lecteur pouvant être aussi bien spécialiste que novice dans la connaissance du français médiéval, il appartient au glossariste de lever les difficultés sémantiques par une glose immédiate, qui évite le détour d'une réflexion supplémentaire pour trouver un équivalent moderne et une suspension trop longue de la lecture. La glose doit donc être aussi précise et contextualisée que possible. Cela assure au lecteur un accès rapide au contenu textuel et facilite une interprétation fine du fonctionnement littéraire, quel que soit son niveau de maîtrise linguistique. De surcroît, il peut s'avérer utile de dégager des gloses contextualisées dans la mesure où celles-ci apportent une connaissance plus fine de l'usage du vocabulaire médiéval. Il faut en effet se rappeler que les dictionnaires restent tributaires des corpus dépouillés, forcément limités au vu de la masse de manuscrits que renferment encore les réserves de bibliothèques. Aussi une lexie, à laquelle les lexicographes attribuent pour l'heure une gamme d'acceptions plus ou moins large, pourrait-elle tout à fait révéler des emplois plus subtils si l'on découvrait une nouvelle occurrence venant corroborer une glose qu'un éditeur aura contextualisée. Le *DMF*

glossaires (à propos de nouvelles recherches rabelaisiennes) », dans Sergio Cigada et Anna Slerca (dir.), *Le Moyen Français: recherches de lexicologie et de lexicographie*, actes du VI^e Colloque international sur le moyen français (Milan, 4-6 mai 1988), Milano, Vita e Pensiero, 1991, p. 265-288; Claude Buridant, « En passant par le *Glossaire des glossaires du moyen français* », *Revue de linguistique romane*, 55, 1991, p. 427-478; Claude Buridant, « Glossaire révisé du roman d'*Erec* en prose », *Travaux de linguistique et de philologie*, 31, 1993, p. 211-250; Frankwalt Möhren, « Édition et lexicographie », dans Martin-Dietrich Glessgen et Franz Lebsanft (dir.), *Alte und neue Philologie*, Tübingen, M. Niemeyer, 1997, p. 153-166; Jean-Pierre Chambon, « Lexicographie et philologie: réflexions sur les glossaires d'éditions de textes (français médiéval et préclassique, ancien occitan) », *Revue de linguistique romane*, 70, 2006, p. 123-141; *id.*, « De la glossographie à la glossaristique (ou "lexicologie philologique"): réflexions proposées aux occitanisants », dans Jean-François Courouau, François Pic et Claire Torreilles, « *Amb un fil d'amistat* ». *Mélanges offerts à Philippe Gardy*, Toulouse, Centre d'étude de littérature occitane, 2014, p. 327-337, p. 329-332 pour notre propos; Frankwalt Möhren, « L'art du glossaire d'édition », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, *op. cit.*, p. 397-437.

paraît avoir pris acte de cette donnée scientifique, puisque, loin d’aplanir le réseau des définitions recensées, il s’applique à enregistrer la plupart des gloses contextualisées fournies par les éditeurs, en les accompagnant de la mention (*Éd.*). Il en est ainsi, par exemple, de *soi ouvrir*, que Jean-Claude Delclos, éditeur de la *Chronique* de Georges Chastellain, comprend au sens de « se découvrir, cesser de se protéger » à propos d’un chevalier en combat singulier (*DMF 2015*, s. v. « ouvrir », II.A.4). Ailleurs, c’est le verbe *vider*, régi par un sujet désignant un coup, qui est enregistré avec deux propositions : la première « partir dans le vide, quitter sa trajectoire (?) », donnée, on le voit, sans certitude par le *DMF 2015* ; la seconde « frapper dans le vide, porter à faux (un coup de lance) », signalée comme reprise à Gilles Roussineau, éditeur de *Perceforest* (*DMF 2015*, s. v. « vider », II.A.2.b).

Dans notre édition de *Melyador*, c’est ce principe qui motive notamment les gloses nuancées du verbe *soi acoisier*, « rester inactif, inerte » (v. 24708) ou des expressions *sus fautre*, « sans délai, sur-le-champ » (v. 21796) et *en brief jour*, « peu de temps après » (v. 26479)³³. Dans le glossaire du *Tristan* en prose, nous poursuivons cette voie médiane, susceptible de concilier philologie et lexicographie, de faire œuvre de « lexicologie philologique³⁴ ». Soucieux de satisfaire les deux lectorats visés, notre glossaire proposera ainsi des gloses adossées au contexte tout en étant suffisamment proches des définitions données par le *DMF* pour faciliter leur classement par les lexicographes. En outre, dans cette même perspective, le relevé des occurrences sera exhaustif. Cependant, eu égard à l’ampleur du texte, les termes utilisés dans un sens moderne, encore vivant,

33. Voir Jean Froissart, *Melyador. Roman en vers de la fin du XIV^e siècle*, éd. cit., t. II, respectivement p. 1545, n. 24708 ; p. 1522, n. 21796 et p. 1558, n. 26479. Pour *sus fautre*, voir aussi Nathalie Bragantini-Maillard, « *Fautre* et *lance sor fautre* en ancien et en moyen français », *Revue de linguistique romane*, 73, 2009, p. 149-183, plus spécifiquement p. 164-167.

34. Jean-Pierre Chambon, « Lexicographie et philologie : réflexions sur les glossaires d’éditions de textes (français médiéval et préclassique, ancien occitan) », art. cit., p. 127 et « De la glossographie à la glossairistique (ou “lexicologie philologique”) : réflexions proposées aux occitanisants », art. cit., p. 330, qui emprunte l’expression à Helmut Stimm.

ne seront pas recensés, sauf si leur occurrence présente un intérêt historique. Le classement des acceptions s’alignera sur la structure dégagée par le *DMF* pour un mot donné, éventuellement selon des subdivisions alphanumériques.

Disposant encore de trop peu de matière à partir du *Tristan* en prose, nous prendrons un échantillon tiré du glossaire de notre édition de *Melyador*. Notre réflexion méthodologique ayant évolué depuis l’élaboration de ce glossaire, l’entrée citée ici est adaptée aux nouvelles exigences scientifiques que nous suivrons pour le *Tristan*. Une entrée du glossaire ressemblera au modèle suivant, en reproduisant le même code typographique :

fricement, frichement, *adv.* [région. picard-wallon; 1^{re} att.]: I. [contexte guerrier] *avec vivacité, gaillardement*, 8761, 12975; *vaillamment*, 19247; [position du chevalier] *fermement, avec assurance*, 5943, 8444, 12995, 15489, 18657; II. [équipement du chevalier, + *s’armer / s’appareillier*] *solidement, convenablement*, 10718, 11124 (– *couvers*), 14213, 15471, 29761; III. [habillement] *de manière élégante, avec goût*, 213*, 13466; [contexte de parade guerrière; sujet /chevalier/ + *passer outre frisquement + en faisant son moustre*] *de manière alerte, avec élégance*, 11162, 16147, 27548; [+ *amanagier / soi logier*] *agréablement, confortablement*, 6576, 15932; IV. [comportement courtois] *avec grâce, galamment*, 2721.

C’est la progression sémantique proposée par le *DMF* que suit le classement structuré à l’aide de chiffres romains, à l’exception de la quatrième subdivision, qui enregistre une application sans équivalent dans l’article du dictionnaire, à savoir celle qui développe le sème /agréable/ dans le contexte social de la courtoisie entre hommes et femmes, en insistant sur l’idée de comportement gracieux, galant au sujet de chevaliers qui servent les demoiselles. Enfin, l’intérêt historique et diatopique du terme est souligné en ouverture de l’entrée, par le commentaire « [région. picard-wallon; 1^{re} att.] », qui localise en régions picarde et wallonne l’emploi de l’adverbe *frichement* et signale le texte de *Melyador* comme premier témoin de son emploi³⁵.

35. L’astérisque (*) qui suit une référence de vers signale une note pour le terme ou le sens concerné.

Conjointement au glossaire, un index recensera avec précision les noms propres, un autre les énoncés de facture proverbiale ou sentencieuse, un autre enfin les formules imagées, chaque fois en enregistrant les éventuelles variantes. L'*index nominum* répondra aux principales exigences méthodiques formulées à ce jour³⁶. Distinction pourrait être faite entre anthroponymes et toponymes, les premiers apparaissant en petites capitales, les seconds en minuscules. Les entrées seront au cas régime, à l'exception des noms uniquement employés au cas sujet dans le texte. Elles apparaîtront sous la forme majoritairement utilisée par le manuscrit ou par chaque partie éditée. L'index relèvera, en minuscules romaines, toutes les variantes orthographiques et morphologiques, ainsi que les différentes appellations (surnom, titulature). L'identification, en italique, de chaque nom médiéval consistera en une traduction, éventuellement accompagnée d'un développement explicatif, qui distinguera, à l'aide de parenthèses, les niveaux de commentaires (information fournie par le texte, information externe au texte). Le relevé des occurrences sera aussi exhaustif que possible pour une œuvre d'une telle étendue. De ce point de vue, les outils informatiques pourront pallier les limites de l'édition imprimée, en procurant des relevés complets. Dans tous les cas, pour les personnages ou les lieux récurrents, dont les apparitions sont soumises au principe narratif de l'entrelacement, le relevé enregistrera les premières occurrences de chaque nouvelle apparition dans le récit. Les formes particulières, trop éloignées de la forme d'entrée, bénéficieront d'un système de renvoi, qui facilitera le repérage du lecteur. Enfin, les personnages anonymes recevront une entrée et, si un personnage entre en scène sous

36. Voir Paul Meyer, compte rendu d'Ernest Langlois, *Table des noms propres de toute nature compris dans les chansons de geste imprimées*, Paris, E. Bouillon, 1904, dans *Romania*, 35, 1906, p. 130-133; Paul Meyer, « Instructions pour la publication des anciens textes », *Bulletin de la Société des anciens textes français*, 35, 1909, p. 64-79, p. 77-78 pour notre propos; Françoise Viellard et al., *Conseils pour l'édition des textes médiévaux*, Paris, École nationale des chartes/CTHS, t. III, *Textes littéraires*, 2002, p. 173-179; Denis Collomp, « La table des noms propres : du *pensum* à une fin en soi », dans Corinne Füg-Pierreville (dir.), *Éditer, traduire ou adapter les textes médiévaux*, op. cit., p. 33-53.

l'anonymat avant d'être nommé, les occurrences de sa/ses première(s) apparition(s) s'ajouteront entre parenthèses à la liste des références.

Ce travail éditorial gagnera à être renforcé par une présentation qui se rapproche du manuscrit dans sa matérialité, afin de satisfaire d'autres pans de la recherche linguistique. L'avantage d'une présentation sur support numérique telle que l'ont mise en œuvre Christiane Marchello-Nizia et Robert D. Sanderson est de faciliter l'examen des pratiques scripturales. Un tel examen est difficile dans le cadre d'une édition imprimée, pour des questions matérielles évidentes, surtout lorsque le texte s'étend sur des centaines de feuillets. Ainsi, sur le modèle de l'édition de *La Queste del saint Graal*, nous envisageons d'élaborer une édition numérique au format XML, conforme au cadre du TEI (Text Encoding Initiative). L'encodage du texte assurera ainsi la pérennité des données informatiques. Ce système constitue l'un des outils les plus performants pour créer et intégrer des données textuelles informatisées, mais aussi pour gérer les grandes quantités d'informations que contient un manuscrit aussi volumineux. Le texte sera présenté sous la forme d'une transcription (semi-) diplomatique et d'une transcription fac-similaire. Ces différentes transcriptions du fr. 99 donneront à voir les spécificités des systèmes de ponctuation, d'abréviation, de graphie en usage chez Gonnot, sans oublier ses repentirs, ses hésitations, ainsi qu'une éventuelle évolution de chaque système au sein du manuscrit. On pourra ainsi mesurer le degré de conscience grammaticale du copiste.

À terme, il sera aussi possible d'enrichir la documentation avec des transcriptions, intégrales ou partielles, des autres témoins de la famille. Avoir accès à une transcription complète et fidèle de l'ensemble des témoins est non seulement indispensable à une description précise du système propre à Gonnot, mais rend aussi possible l'étude des invariants linguistiques et stylistiques. Ce que rappelait Stephen G. Nichols

en référence aux travaux stylistiques de Leo Spitzer vaut finalement aussi pour les recherches linguistiques :

literary language could only be adequately described by multiple examples from many texts, which then permitted him to identify the invariant signaling a universal or to demonstrate the normative deviation signaling stylistic originality [...]³⁷.

Avoir accès en outre à une copie numérisée d'un manuscrit constitue un apport documentaire considérable. Par chance, le fr. 99 a récemment bénéficié d'une numérisation et est consultable sur le site internet de la Bibliothèque nationale de France³⁸. En ce qui concerne les autres témoins de notre version du *Tristan* en prose, si les bibliothèques le permettent, il n'est pas exclu qu'une copie numérisée de tout ou partie de leurs feuillets vienne compléter le dossier. Ce format de consultation des manuscrits livrera des informations inédites. Il reste en effet encore beaucoup à découvrir sur la pratique graphique de Michel Gonnot. L'étude menée par Cedric Edward Pickford sur l'écriture des cinq manuscrits attribuables au copiste dégage une évolution dans la calligraphie, d'une écriture « d'abord pointue et serrée en 1463 » dans le fr. 99 « vers une écriture moins ferme, moins soignée » dans les mss BnF, fr. 916 et Arsenal 5121³⁹. De notre côté, à l'issue d'une première lecture, nous avons cru déceler une variation calligraphique au sein même du fr. 99. Il s'agira d'en déterminer les modalités.

En définitive, on mesure combien l'alliance de l'édition imprimée et de l'édition numérique représente une des voies à privilégier pour que l'édition serve au mieux les intérêts de la linguistique. Dans l'idéal, l'approche bédérienne guiderait l'édition imprimée, avec une intervention plus ou moins modérée en fonction du projet. Si l'objectif éditorial est avant

37. Stephen G. Nichols, « Introduction. Philology in a Manuscript Culture », *Speculum*, vol. 65, n°1, « The New Philology », dir. Stephen G. Nichols, 1990, p. 1-10, ici p. 3.

38. Numérisation accessible en ligne : <http://gallica.bnf.fr/ark:/12148/btv1b52503846r>.

39. Cedric Edward Pickford, *L'Évolution du roman arthurien en prose vers la fin du Moyen Âge, d'après le manuscrit 112 du fonds français de la Bibliothèque nationale*, Paris, Nizet, 1959, p. 21-22, ici p. 22 ; Gabriella Parussa et Richard Trachsler, « *Trasmissione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », art. cit., p. 359-360.

tout littéraire, avec une volonté de diffuser un texte au plus grand nombre, la méthode bédérienne est appliquée à plein : un manuscrit de base est retenu et corrigé par une sélection de témoins. En revanche, si le projet s'intéresse principalement aux aspects linguistiques du texte et aspire à faire connaître un manuscrit, un état du texte dans une perspective diachronique, diatopique, diaphasique, etc., les principes de la méthode bédérienne doivent être suivis avec pondération, de telle sorte que le texte proposé à la lecture soit le plus fidèle possible à la copie manuscrite.

Quel que soit le degré d'intervention, il est souhaitable que l'édition d'inspiration bédérienne se dote d'un appareil linguistique complet et précis, avec présentation des particularités linguistiques et stylistiques de la copie et de l'auteur en distinguant manuscrit de base et manuscrits de contrôle. Des notes linguistiques de qualité scientifique sont utiles pour expliciter dans le détail les faits de langue problématiques ou dignes d'intérêt, et enregistrer les références utiles à l'identification des innovations lexicales ou autres particularités lexicales remarquables. Ces notes sont soutenues par un glossaire exhaustif d'approche lexicographique, qui conjugue exhaustivité, précision et validité scientifique⁴⁰.

Parallèlement à la version imprimée, on voit aussi que le support numérique donne un accès privilégié au manuscrit, en s'inspirant de certains principes de la « nouvelle philologie » : il met à la disposition du lecteur et du chercheur différentes représentations d'une copie manuscrite, grâce auxquelles il est possible de juger sur pièces les états du texte ; d'autre part, les logiciels de recherche permettent de faire des requêtes sur une remarquable variété de critères (graphiques, lexicaux, syntaxiques, de co-occurrence, etc.).

40. C'est l'approche qui a guidé notre collaboration aux projets d'édition suivants, par ex. : Jean Renart, *L'Escoufle*, éd. cit. ; Wace, *Vie de sainte Marguerite. Conception Notre Dame. Vie de saint Nicolas*, éd. et trad. Françoise Le Saux, Françoise Laurent et Nathalie Bragantini-Maillard, Paris, Champion, coll. « Classiques français du Moyen Âge », 2019.

Pour une exploitation scientifique du texte édité du ms. fr. 99

Pour revenir au ms. BnF, fr. 99, les critères éditoriaux retenus jetteront une lumière nouvelle sur des questions aux enjeux tout autant (socio)linguistiques que littéraires, mais qui, à ce jour, demeurent en suspens.

La première question concerne les rapports que le fr. 99 entretient avec deux autres témoins bien connus. La bibliothèque de Jacques d'Armagnac abritait quatre manuscrits réalisés par Gonnot, mais peut-être aussi le manuscrit de Vienne 2542. Celui-ci, daté du début du ^{xiv}^e siècle, contient une version Π du *Tristan* en prose et fut sans doute acquis par héritage. Dans cette hypothèse, on peut imaginer qu'il ait pu servir de modèle à la copie du fr. 99. L'édition bédérienne dont il a bénéficié⁴¹ facilitera la comparaison. Par ailleurs, tout porte à penser que le fr. 99 fut peut-être un des *exemplars* que la compilation arthurienne du BnF, fr. 112, copié par Gonnot en 1470, suivit pour une partie du *Tristan* en prose⁴² qui s'y trouve inséré. Par conséquent, éditer le fr. 99 en version usagée rétablira un chaînon manquant dans l'étude de cet ensemble, chaînon susceptible de nous renseigner sur le degré de parenté des trois témoins et sur la méthode de travail d'un copiste au sein d'une cour située en domaine occitanisant, mais familière d'une forme d'oïl. L'apport est tout à la fois littéraire, linguistique et culturel. Seule une comparaison minutieuse de la lettre des trois versions permettra de confirmer ou d'infirmer les hypothèses de filiation entre les mss Vienne 2542 et fr. 99 d'une part et entre le fr. 99 et le fr. 112 de l'autre, mais aussi de déterminer dans quelle mesure Gonnot modernisa la langue de son modèle et/ou en adapta le style, et si, d'autre part, sa *scripta* connut une influence occitanisante. En outre, l'examen croisé des trois manuscrits en dira long sur les modalités de personnalisation du code graphique et

41. Voir *Le Roman de Tristan en prose*, éd. dir. Philippe Ménard.

42. Voir Cédric Edward Pickford, *L'Évolution du roman arthurien en prose vers la fin du Moyen Âge, d'après le manuscrit 112 du fonds français de la Bibliothèque nationale*, op. cit., p. 42-53 et 57-59; Emmanuèle Baumgartner, *Le « Tristan en prose ». Essai d'interprétation d'un roman médiéval*, op. cit., p. 21.

de ponctuation et sur leur évolution entre 1463 et 1470, années de copie respectives du fr. 99 et du fr. 112.

Cedric Edward Pickford et Emmanuèle Baumgartner ont également détecté une parenté étroite du point de vue littéraire entre le fr. 99 et deux autres manuscrits renfermant la même version du *Tristan*, à savoir Chantilly 645-646-647 et Pierpont Morgan Library 41⁴³. Une comparaison attentive des faits linguistiques qui se manifestent dans ces trois témoins permettra donc de préciser leurs rapports, voire d'esquisser une chronologie relative et, pourquoi pas, de reconnaître un modèle.

Comme nous l'évoquions plus haut, une autre énigme porte sur l'origine de Michel Gonnot et sur la *scripta* qu'il utilise pour rédiger les manuscrits destinés à Jacques d'Armagnac. À l'époque de la réalisation des manuscrits signés par Gonnot (1463-*ca* 1476), Jacques d'Armagnac était un seigneur puissant, dont le domaine s'étendait de la Creuse au Languedoc. Fils de Bernard d'Armagnac, comte de Pardiac et vicomte de Carlat, et d'Éléonore de Bourbon, comtesse de la Marche et vicomtesse de Nemours, il hérita de leurs domaines et fut également nommé lieutenant du Roussillon en 1463. Ses résidences principales alternèrent sans aucun doute entre les châteaux de Carlat, de Castres et de la Marche, où le duc constitua ses trois bibliothèques à partir de 1459⁴⁴. Au vu de ces données, on peut donc supposer que sa langue maternelle était la langue d'oc. Toutefois, les relations privilégiées que le duc entretenait avec les rois Charles VII et Louis XI l'amènèrent à souvent quitter ses terres occitanes pour Paris, capitale dont, du reste, il brigua le gouvernement. D'autre part, plusieurs de ses terres se situaient dans la Marche, cette zone administrative et linguistique aussi dite « du Croissant », située à cheval entre domaine d'oïl et domaine d'oc et s'étirant

43. Voir Cedric Edward Pickford, *L'Évolution du roman arthurien en prose vers la fin du Moyen Âge, d'après le manuscrit 112 du fonds français de la Bibliothèque nationale*, *op. cit.*, p. 37, n. 9, où la cote du manuscrit de Chantilly est numérotée de 315 à 317 ; Emmanuèle Baumgartner, *Le « Tristan en prose ». Essai d'interprétation d'un roman médiéval*, *op. cit.*, p. 71-76.

44. Voir Léopold Delisle, *Le Cabinet des manuscrits de la Bibliothèque impériale*, Paris, Imprimerie impériale, 1868, t. I, p. 86-90 ; Antoine Thomas, « Jacques d'Armagnac bibliophile », *Journal des savants*, 1906, p. 633-644, notamment p. 643-644.

de la Saintonge à l'Auvergne. Dans cet espace, français et occitan cohabitent sous leurs formes dialectales spécifiques et enchevêtrent leurs particularismes de façon variable d'un secteur à l'autre⁴⁵. Ce sont là autant de circonstances qui laissent penser que Jacques d'Armagnac devait sans aucun doute avoir une maîtrise sinon bonne, tout du moins fonctionnelle du français. Comme en témoignent les manuscrits de textes didactiques, religieux ou romanesques présents dans ses bibliothèques et vraisemblablement lus avec intérêt par ce mécène éclairé dans une perspective morale, spirituelle, voire politique⁴⁶, la lecture d'œuvres littéraires en langue d'oïl ne devait donc pas lui poser problème, encore que les copistes aient peut-être dû en adapter la *scripta*, surtout lorsque leurs modèles étaient diatopiquement très marqués. Les questions soulevées par ce cas particulier sont à relier à la complexité générale des rapports entre langue d'oïl et langue d'oc dans certains textes qui auraient été composés ou copiés en domaine méridional⁴⁷.

45. Voir en particulier Guylaine Brun-Trigaud, *Le Croissant : le concept et le mot. Contribution à l'histoire de la dialectologie française au XIX^e siècle*, Lyon, Université Jean Moulin Lyon III, 1990; R. Anthony Lodge, *Le Français. Histoire d'un dialecte devenu langue*, Paris, Fayard, 1997, p. 108; Lilliane Jagueneau, « Les limites linguistiques dans le Centre-Ouest », dans Dominique Guillemet et Jacques Peret, *Espaces, limites, frontières. Centre-Ouest français, Acadie, Poitiers*, Université de Poitiers, coll. « Les Cahiers du GERHICO », 2001, p. 81-98; Serge Lusignan et Diane Gervais, « La division de la France en oïl et oc du Moyen Âge à l'aube du XX^e siècle », dans Tim Pooley et Dominique Lagorgette, *Le Changement linguistique en français : aspects socio-historiques. Études en hommage au professeur R. Anthony Lodge*, Chambéry, Université de Savoie, 2011, p. 113-128, p. 116 pour notre propos. Au-delà de la zone mixte du Croissant, R. Anthony Lodge en particulier a toutefois discuté la conscience même d'une séparation linguistique entre langue d'oïl et langue d'oc dans la France médiévale; voir *Le Français. Histoire d'un dialecte devenu langue*, op. cit., p. 101-110; « Le clivage oc-oïl au Moyen Âge : fiction méthodologique », *Mélanges de l'École française de Rome. Moyen Âge*, vol. 117, n° 2, 2005, p. 595-613; Georg Kremnitz, « Les étapes historiques du recul de l'occitan. Un chantier de recherche à (r)ouvrir », dans Jean-François Courouau, François Pic et Claire Torrelles, « *Amb un fil d'amistat* ». *Mélanges offerts à Philippe Gardy*, op. cit., p. 625-637, p. 628-29 pour notre propos.

46. Voir Chrystèle Blondeau, « Du plaisir des sens à la passion du sens : l'illustration d'un cycle du Lancelot Graal et ses remaniements », dans Françoise Joubert (dir.), *L'Artiste et le commanditaire aux derniers siècles du Moyen Âge (XIII^e-XV^e siècle)*, Paris, PUPS, 2001, p. 99-114; Cécile Ravier, « Le Miroir historial de Jacques d'Armagnac : un monument bibliophilique pour un prince ambitieux », *Médiévales*, 67, « Histoires de Bohême », 2014, p. 143-168.

47. Que l'on pense, par exemple, à *Girart de Roussillon*, à *Aigars et Maurin* ou encore au *Jeu d'Adam*; voir en particulier W. Mary Hackett, *La Langue de Girart de Roussillon*,

En ce qui concerne plus particulièrement Michel Gonnot, nous l'avons vu, le colophon du manuscrit fr. 99 laisse entendre que le scribe travaillait à Crozant, dans la Creuse. L'hypothèse est plausible dans la mesure où le duc de Nemours pouvait alors résider aussi bien à Carlat qu'en son château de la Marche⁴⁸, ce qui lui offrait l'occasion d'être en contact avec des ateliers creusois. La Creuse faisant partie du comté de la Marche, la copie du fr. 99, destinée à un lecteur aristocratique vraisemblablement diglossique (à la fois occitanisant et familiarisé avec le français) pourrait renfermer des informations sur une éventuelle adaptation régionale qu'aurait pu subir ce texte oïlique du *Roman de Tristan*. Or, des études⁴⁹ ont relevé des marques du Nord et du Nord-Est dans la *scripta* de certaines productions de Gonnot. Cette donnée autoriserait-elle à faire du scribe un natif non pas du Limousin, mais du Nord de la France? La cour de Jacques d'Armagnac aurait-elle accueilli des professionnels de tous horizons? Le colophon du fr. 99 oriente plutôt vers un copiste originaire de La Brousse, qui *a priori* correspondrait à un village situé dans la Creuse, à quelques kilomètres à l'est de Crozant. Cependant, précisons qu'une autre commune Labrousse existe en Auvergne, au sud de Carlat, l'un des châteaux de résidence de Jacques d'Armagnac. À ce stade de la recherche, on ne peut

Genève, Droz, 1970; Olivier Naudeau, « Observations sur la langue de *Aigâr et Maurin* », *Romania*, 115, 1997, p. 337-367; *Le Jeu d'Adam*, éd. et trad. Geneviève Hasenohr, Genève, Droz, coll. « Texte courant », 2017, p. xxviii-xcvii.

48. Sur Jacques d'Armagnac et les manuscrits de Michel Gonnot, nous renvoyons à la bibliographie dressée par Gabriella Parussa et Richard Trachsler, « *Trasmisione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », art. cit., p. 357-358, n. 10 et 11. Sur le ms. fr. 99, voir plus particulièrement Susan Amato Blackman, *The Manuscripts and Patronage of Jacques d'Armagnac, Duke of Nemours, 1433-1477* [thèse de troisième cycle, University of Pittsburgh, 1993], Ann Harbor [Mich.], UMI dissertation information service, 1994, t. I, p. 50-59; Cedric Edward Pickford, *L'Évolution du roman arthurien en prose vers la fin du Moyen Âge, d'après le manuscrit 112 du fonds français de la Bibliothèque nationale*, op. cit., p. 14-24; Emmanuèle Baumgartner, *Le « Tristan en prose ». Essai d'interprétation d'un roman médiéval*, op. cit., p. 71-76.
49. Études fondées sur des extraits des mss fr. 99 et 112; voir *La Suite du roman de Merlin*, éd. Gilles Roussineau, Genève, Droz, coll. « Textes littéraires français », 2^e éd., 2006, p. LXXXVIII-XCIX; Gabriella Parussa et Richard Trachsler, « *Trasmisione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », art. cit., p. 359, n. 15; Richard Trachsler, « Des chiffres et des lettres. Observations sur l'orthographe de quelques scribes de la fin du Moyen Âge », art. cit., p. 477.

écarter la possibilité que Michel Gonnot fût natif de ce village. Quoi qu'il en soit, c'est la conjecture d'un copiste recruté dans la région du Limousin qui, jusqu'à présent, est partagée par la plupart des spécialistes⁵⁰.

Par conséquent, que faire de l'empreinte picarde qui se décèle çà et là dans le travail de Gonnot ? Nous approfondirons ailleurs la question de la *scripta* de Gonnot, mais, au vu des premiers relevés réalisés par nos prédécesseurs sur des échantillons des mss fr. 112 et fr. 99, il serait permis de dire que la coloration picarde détectée semble, somme toute, légère. De prime abord, même si ce point demande encore à être confirmé à l'aide d'un corpus plus étendu, la plupart des régionalismes apparaissent de manière isolée ou épisodique d'un manuscrit à l'autre. Il se pourrait que les quelques traits relevés aient été importés de modèles picardisants, le copiste ayant pu les reporter tels quels soit par inadvertance, soit parce qu'il ne les identifiait pas comme régionaux, soit encore parce qu'ils n'entraient guère l'intelligibilité du texte. L'hypothèse selon laquelle les formes isolées que compte un manuscrit peuvent tout aussi bien être attribuées à un *exemplar* fut d'ailleurs émise à propos d'autres corpus par Gabriella Parussa et Richard Trachsler⁵¹. En ce qui concerne la *scripta* de Gonnot, la conjecture suivante est donc envisageable : le coloris régional de sa rédaction peut fort bien tenir à une influence de la pratique picarde dont aurai(en)t été issu(s) son/ses modèle(s) ou une rédaction primitive. De toute évidence, en effet, au cours du XIII^e siècle ou à partir du début du XIV^e siècle, pour des raisons politico-culturelles, le picard devint

50. Voir Léopold Delisle, *Le Cabinet des manuscrits de la Bibliothèque impériale*, t. I, *op. cit.*, p. 86 ; *Documents historiques relatifs à la vicomté de Carlat*, éd. Gustave Saige et Édouard de Dienne, Monaco, Impr. de Monaco, 1900, t. II, p. CCLXXVI ; Cedric Edward Pickford, *L'Évolution du roman arthurien en prose vers la fin du Moyen Âge, d'après le manuscrit 112 du fonds français de la Bibliothèque nationale*, *op. cit.*, p. 22-23 ; *id.*, « A Fifteenth-Century Copyist and his Patron », dans Frederick Whitehead, Armel Hugh Diverres et Frank Edmund Sutcliffe, *Medieval Miscellany presented to Eugene Vinaver*, Manchester, Manchester University Press, 1965, p. 245-262, p. 256 pour notre propos ; Gabriella Parussa et Richard Trachsler, « *Trasmisione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », *art. cit.*, p. 358-359.

51. Voir Gabriella Parussa et Richard Trachsler, « *Trasmisione-trasformazione* ou comment comprendre l'apport d'un copiste vernaculaire », *art. cit.*, p. 357.

une sorte de *lingua franca* en domaine d'oïl. Il aurait participé à l'avènement d'une *koinè* littéraire, qui recourait à une *scripta* mêlant plus particulièrement traits du Centre et traits du Nord⁵². Les ateliers picards furent alors reconnus pour leur productivité et la qualité de leur travail. Dès lors, on peut conjecturer que les copistes du domaine occitan ont pu être formés par des copistes picards ou à partir de rédactions picardisantes, si bien que leurs modèles appartenaient essentiellement à cette zone linguistique alors même qu'ils lui étaient étrangers. Le statut de la *scripta* picarde utilisée dans les manuscrits littéraires tardifs reste toutefois à préciser quand on sait que, au cours du xv^e siècle – ce qui inclut la période d'exercice de Gonnot –, les traits picards tendaient à s'effacer des documents administratifs et juridiques⁵³.

Il est une autre pratique scripturale propre à Gonnot que l'édition du fr. 99 devrait aider à préciser : on pressent chez ce scribe une tendance archaïsante. Gilles Roussineau observe ainsi que, en copiant *La Suite du roman de Merlin*, Gonnot ne modernise que discrètement graphie et vocabulaire⁵⁴. Aux plans graphique et morphologique, son conservatisme se trouve confirmé par l'étude comparative menée par Richard Trachsler sur l'usage graphique d'une quinzaine de manuscrits du xv^e siècle⁵⁵. Au sein du corpus, dont il est pourtant le témoin le plus tardif, le fr. 99 fait

52. Voir en particulier Maurice Delbouille, « La formation des langues littéraires et les premiers textes. 1. Les plus anciens textes et la formation des langues littéraires », dans Hans R. Jauss et Erich Köhler (dir.), *Grundriss der romanischen Literaturen des Mittelalters*, t. I, Heidelberg, C. Winter, 1972, p. 559-622, p. 581 pour notre propos ; Phillip E. Bennett, « Le normand, le picard et les *koinès* littéraires de l'épopée aux xii^e et xiii^e siècles », *Bien dire et bien apprendre*, 21, « Picard d'hier et d'aujourd'hui », dir. Jacques Landrecies et Aimé Petit, 2003, p. 43-56 ; Serge Lusignan, *Essai d'histoire sociolinguistique. Le français picard au Moyen Âge*, Paris, Classiques Garnier, 2012. Pour une remise en cause des notions de *koinè* et de *scripta* aux xii^e et xiii^e siècles, voir Anthonij Dees, « Dialectes et *scriptae* à l'époque de l'ancien français », *Revue de linguistique romane*, 49, 1985, p. 87-117.

53. Voir en dernière analyse Serge Lusignan, « L'emploi tardif des *scriptae* régionales : réflexion sur l'anglo-français et le picard », dans David Trotter (dir.), *Present and future research in Anglo-Norman*, op. cit., p. 79-84.

54. Voir *La Suite du roman de Merlin*, éd. cit., p. LIX, LXXXVIII et XCIX.

55. Voir Richard Trachsler, « Des chiffres et des lettres. Observations sur l'orthographe de quelques scribes de la fin du Moyen Âge », art. cit.

en effet souvent figure d'exception par un recours prédominant à des formes archaïques⁵⁶. Reste à présent à étendre l'enquête à l'ensemble de la production de Gonnot. Si la propension de celui-ci au conservatisme se confirmait, elle pourrait tout aussi bien correspondre à un penchant nostalgique ou puriste du copiste qu'à une tradition scripturale picarde ou à une pratique spécifique des scribes œuvrant en domaine occitan face aux textes oiliques. N'étant pas forcément experts en dialectes et/ou en *scriptae* d'oïl, ces scribes purent fort bien adopter une *scripta* oilique « classique », une *koinè* littéraire plus ou moins conservatrice, qui leur permettait d'avoir accès aux textes sans nécessairement en maîtriser toutes les subtilités régionales⁵⁷.

À la lumière des critères qui sous-tendent le projet d'édition du ms. BnF, fr. 99 et des quelques cas étudiés ici, on mesure quel peut être l'apport d'une édition en version usagée qui accorde une place de choix à l'examen de la langue d'un texte. L'entreprise vise à fournir un outil exploitable pour une contribution à trois champs de recherche au moins :

- à une « typologie linguistique des genres textuels⁵⁸ » (phraséologie, faits stylistiques, registres formulaires typiques d'un genre littéraire) ;
- à une typologie des variations diaphasiques et des pratiques scripturales (préférences graphiques, morphologiques, syntaxiques, lexicales et stylistiques du copiste, profil du copiste, voire identification de la sensibilité linguistique et stylistique du commanditaire) ;

56. *Ibid.*, p. 476 et 488-489.

57. Voir en particulier, pour les XIII^e-XIII^e siècles, Jacques Pignon, *L'Évolution phonétique des parlers du Poitou (Vienne et Deux-Sèvres)*, Paris, Bibliothèque du français moderne/Éd. d'Artrey, 1960, p. 527 ; Jacques Monfrin, compte rendu de Thomas, *Roman de Horn*, éd. Mildred Katharine Pope et T.B.W. Reid, Oxford, Blackwell, 1964, dans *Romance Philology*, vol. 26, n° 3, 1973, p. 602-612, plus spécifiquement p. 611.

58. Martin-Dietrich Glessgen, « L'élaboration philologique et l'étude lexicologique des *Plus anciens documents linguistiques de la France* à l'aide de l'informatique », art. cit., p. 372.

- et plus largement, à une typologie des variations diachroniques et diatopiques (faits de modernisation, de régionalisation, d'adaptation interlinguale).

Qui plus est, l'intérêt dépasse la sphère linguistique : les pistes ainsi débusquées intéressent autant la critique génétique que l'histoire du livre et l'histoire des rapports culturels entre pays d'oïl et pays d'oc. On voit aussi combien respecter un texte manuscrit dans un souci avant tout linguistique ne nuit nullement à son intérêt littéraire, dès lors que l'on suit la conception que les médiévaux se faisaient du texte littéraire, objet intrinsèquement mouvant, indissociable de son contexte de commande, de (re)production et de réception. Une compilation comme celle du ms. BnF, fr. 99 est une œuvre en soi, qui, de par son contexte de copie, promet en outre d'heureuses trouvailles linguistiques. Une approche linguistique de l'édition critique a cette vertu de nous contraindre à accepter la déroutante diversité de l'écriture médiévale, à l'intégrer réellement et pleinement à notre lecture de cette littérature plurielle, qui, à chaque étape, fait intervenir la foule de ses lecteurs dans le processus d'existence de tout texte.

Pour une « troisième voie » en matière d'édition de textes d'ancien et de moyen français¹

Laurent Balon
Chercheur indépendant

La pratique de l'édition de texte littéraire d'ancien et de moyen français se trouverait face à un dilemme : en partant des conseils et principes d'édition formulés depuis le début du xx^e siècle concernant la transcription des graphies médiévales, on observe que les critères de choix des variantes aboutissent, bien souvent, à l'exclusion du matériau intéressant le linguiste qui, de son côté, aurait besoin d'un exposé intégral de toutes les données, sans tri. Ce besoin d'exhaustivité est important, mais la présentation des données intégrales du manuscrit, sauf à enfreindre la lisibilité et l'intelligibilité des textes, s'avère peu facile à satisfaire, voire impossible sur support papier. Or l'évaluation de seuils d'illisible préside à toutes les consignes éditoriales énoncées. Dans le même temps, on observe le recours à des méthodologies d'édition différentes selon le type de texte concerné, document d'archive ou texte littéraire, celui-ci n'étant pas toujours traité avec le même soin linguistique que celui-là : on observe ainsi que la question de la transcription, par exemple, de la forme *lequel* en un seul mot ne se poserait que pour les documents d'archives². Mais l'évolution de *le quel* en *lequel* est un fait d'ordre grammatical et historique

-
1. Cette contribution est exemplifiée par la partie du ms. BL, Royal, 20 D XI contenant *Garin de Monglane* que nous avons édité : « *Garin de Monglane* », *chanson de geste du XIII^e siècle – British Library, Royal, 20 D XI*, Paris, Classiques Garnier, coll. « Textes littéraires du Moyen Âge », à paraître. Toutefois, afin de ne pas alourdir la lecture de cette présentation, les références aux vers, sauf fait linguistique exceptionnel, ne seront pas données.
 2. Voir Françoise Vielliard et Olivier Guyotjeannin (dir.), *Conseils pour l'édition des textes médiévaux*, Paris, École nationale des chartes/CTHS, t. I, *Conseils généraux* [2001], 2014, p. 42.

révéléateur d'un état de langue qui opposerait une pratique « traditionnelle », et plutôt ancienne, représentée par *le quel* à un usage tendant à s'éloigner des « canons » de l'ancien français, et qui serait plutôt représenté par la graphie *lequel*, phénomène pour lequel la nature du texte importe peu ; or cette négligence présente le danger de voir les linguistes se détourner des données littéraires comme support à l'étude de l'histoire et de l'évolution du français.

L'objet de cette contribution est de présenter une méthode d'édition des textes littéraires permettant de mieux satisfaire certains besoins des linguistes par la mise en relief, dans certains cas que nous délimiterons, des graphies restituées par l'éditeur et des séquences graphiques du manuscrit.

La méthode que nous proposons constitue un compromis entre une pratique de la transcription diplomatique que certains appellent de leurs vœux³, très informative des pratiques de scription mais souvent peu lisible, et la pratique de l'édition de texte traditionnelle qui, en privilégiant la problématique littéraire et esthétique par rapport à la question linguistique, réduit certaines particularités graphiques, ce qui oblige le linguiste au recours constant aux manuscrits pour en vérifier les données et s'assurer qu'il ne s'agit pas d'une interprétation non signalée de l'éditeur⁴. On observe ainsi, par exemple, que les méthodes d'édition traditionnelles sont quelquefois amenées à opérer des choix contradictoires et arbitraires : sur la base de la lemmatisation

3. Voir les récents travaux, par ex., d'Alexei Lavrentiev (pour un aperçu, voir « Typologie textuelle pour l'étude linguistique de manuscrits français médiévaux », dans *id.* [dir.], *Systèmes graphiques de manuscrits médiévaux et incunables français. Ponctuation, segmentation, graphies*, Chambéry, Presses universitaires de Savoie, p. 49-66 et « Base de français médiéval et transcriptions de manuscrits : recherche de complémentarité », dans David Trotter [dir.], *Actes du xxiv^e Congrès international de linguistique et de philologie romanes, Aberystwyth, 1^{er}-6 août 2004*, Tübingen, M. Niemeyer, 2007, t. I, p. 405-410).

4. La valeur en tant que sources de données linguistiques fiables de certaines éditions « critiques » est ainsi parfois « critiquée » : voir Lydia Stanovaïa, « La standardisation en ancien français », dans Michèle Goyens et Werner Verbeke (dir.), *The Dawn of the Written Vernacular in Western Europe*, Leuven, Leuven University Press, 2003, p. 242-272 ; May Plouzeau, « À propos de *La Mort Artu* de Jean Frappier », *Travaux de linguistique et de philologie*, 32, 1994, p. 207-221 ; Alexei Lavrentiev, « Base de français médiéval et transcriptions de manuscrits : recherche de complémentarité », art. cit.

lexicographique, elle-même parfois inconstante et indécise⁵, on aura tendance à transcrire, par exemple, *defors* en un mot, mais *la fors* en deux mots, ce qui ne correspond pas toujours à l'usage que l'on observe dans les manuscrits. De la même manière, on note une tendance des pratiques éditoriales à figer un usage des formes enclitiques du type *al*, *del*, *nel* devant voyelle qui seraient en réalité généralement minoritaires dans les manuscrits face aux formes proclitiques⁶.

Ainsi, afin de fournir au lecteur et au linguiste des informations immédiatement exploitables et utiles à l'avancée de la discipline, le principe méthodologique proposé consiste à signaler dans le texte édité certains faits de langue relevant de la « ponctuation du mot » par l'emploi d'un code graphique qui en conserve la trace, à savoir un système de « tirets » déjà utilisé et suggéré par Clovis Brunel⁷ et Jacques Monfrin⁸ pour la transcription des documents d'archive, mais complété et appliqué pour la première fois à un texte littéraire par Nelly Andrieux-Reix⁹. Notre démarche s'inscrit dans la poursuite de cette innovation que nous affinerons : le projet n'ayant pas pour finalité de réduire le texte

-
5. Par ex., la combinaison *sitres* n'est pas considérée comme une unité lexicale par les dictionnaires d'ancien et de moyen français contrairement à *sitost*, *sicom* voire *sique(s)* (voir le *Dictionnaire de l'ancienne langue française et de tous ses dialectes du IX^e au XVI^e siècle* de Godefroy, le *Dictionnaire du moyen français (DMF)*, le *Petit dictionnaire de l'ancien français* de Van Daele ; l'*Altfranzösisches Wörterbuch* de Tobler et Lommatzsch ne les mentionne pas).
 6. Voir Nelly Andrieux-Reix, « Consignes et pratiques scripturaires au Moyen Âge : confrontations, enseignements », dans Claudio Galderisi et Jean Maurice (dir.), « *Qui tant savoit d'engin et d'art* ». *Mélanges de philologie médiévale offerts à Gabriel Bianciotto*, Poitiers, Université de Poitiers, Centre d'études supérieures de civilisation médiévale, 2006, p. 244-248. – Dans la partie du manuscrit de Londres signifiant *Garin de Monglane*, on relève 108 occurrences des formes clairement proclitiques *de l'*, *a l'* contre 5 occurrences des formes clairement enclitiques *del* et *al* (*nel* et *ne l'* n'apparaissant pas dans le manuscrit et les occurrences à l'initiale des vers, où la première lettre est, sauf exception, détachée du reste du vers, n'ayant bien évidemment pas été prises en compte dans ce calcul), ce qui confirme les observations de N. Andrieux-Reix.
 7. *Les Plus Anciennes Chartes en langue provençale. Recueil des pièces originales antérieures au XIII^e siècle. Supplément*, éd. Clovis Brunel, Paris, A. et J. Picard, 1952.
 8. *Documents linguistiques de la France*, t. I, vol. 2, *Chartes en langue française antérieures à 1271 conservées dans le département de la Haute-Marne*, éd. dir. Jacques Monfrin, Paris, Éd. du CNRS, coll. « Documents, études et répertoires – IRHT », 1974, p. lxx-lxvi.
 9. *Le Moniage Guillaume, chanson de geste du XII^e siècle*, éd. de la rédaction longue par Nelly Andrieux-Reix, Paris, Champion, coll. « Classiques français du Moyen Âge », 2003.

à un pur objet d'observations linguistiques, seront conservés dans l'édition les éléments qui dans l'écriture du copiste font sens dans l'histoire du français au plan graphique, lexicologique, sémantique et morphosyntaxique. Par ailleurs, afin de pas attester au plan statistique des graphies qui apparaissent « pleines » dans l'édition mais qui sont en réalité abrégées dans le manuscrit¹⁰, le principe méthodologique proposé consiste encore à signaler par l'emploi de l'italique ou du soulignement la restitution des parties abrégées non assurée par les données du manuscrit ou demeurant problématique.

Dans un premier temps, nous présenterons les modalités théoriques et pratiques de la solution proposée; dans un second temps, nous exemplifierons quelques données du problème par le ms. BL, Royal, 20 D XI.

Principes méthodologiques de la solution proposée

La méthodologie éditoriale pour laquelle nous plaidons ici s'appuie certes sur des travaux antérieurs, mais également sur notre propre édition de *Garin de Monglane* et l'étude linguistique qui lui est corrélée dont les leçons ont, pour une part, partie liée avec le système adopté¹¹.

L'objectif de la méthode est simple et clair: ménager la lisibilité et l'intelligibilité du texte littéraire tout en préservant le contenu linguistique transmis par le copiste avec le souci constant de ne pas créer d'artefacts graphiques, et par là lexicaux ou grammaticaux¹². Pour ce faire, une attention particulière est apportée à deux faits de langue ordinairement assez mal considérés dans les éditions traditionnelles:

10. Voir à ce sujet la critique par Stéphanie Brazeau et Serge Lusignan (« Jalon pour une histoire de l'orthographe française au XIV^e siècle: l'usage des consonnes quiescentes à la chancellerie royale », *Romania*, 122, 2004, p. 448) des travaux de Thera de Jong, « Parasite consonants: a homographic clash », dans Jacek Fisiak (dir.), *Medieval dialectology*, Berlin/New York, Mouton de Gruyter, 1995, p. 7-42.

11. Voir Laurent Balon, *Étude linguistique de « Garin de Monglane »* – *British Library, Royal, 20 D XI*, Paris, Classiques Garnier, à paraître.

12. Cette nécessaire lisibilité du texte littéraire appelant quelque restriction ou allègement d'usage du système proposé (voir *infra*).

les graphies suppléées issues de la résolution d'abréviations et les séquences graphiques¹³.

Dans cet objectif, une attention particulière est ainsi d'abord portée aux formes abrégées du manuscrit et à leur résolution, dont le traitement dans nombre d'éditions ne laisse pas d'étonner¹⁴; tout se passe comme si jamais l'interprétation des abréviations ne posait problème, ce qui paraît impossible. Or cette négligence des éditeurs¹⁵ est problématique et dommageable tant le système abréviatif du français médiéval apparaît complexe et en réalité, à y regarder de près, beaucoup moins aléatoire et beaucoup plus réfléchi qu'on aurait pu le penser¹⁶. Les conséquences pour l'étude de la langue sont importantes, par exemple pour l'attestation au plan statistique de graphies dont la restitution n'est pas assurée et dont l'interprétation ne s'impose pas par rapport à une autre également possible, notamment en cas de parité d'emploi entre plusieurs graphies, d'occurrence développée unique, d'absence de développement en toutes lettres ou d'hésitation de lecture, dans les formes développées en toutes lettres, des deux jambages pouvant valoir pour *u* aussi bien que pour *n* dans la résolution de certains tildes et *g* linéaires (par ex. : *coumence* ou *conmence* ?

13. La notion de « séquences » graphiques impliquant de fait la notion de « segments » graphiques (voir *infra*).

14. Voir, pour un ex. de problématisation des abréviations dans les éditions traditionnelles sur papier, *Le Moniage Guillaume, chanson de geste du XII^e siècle*, éd. cit., p. 27, et, dans les éditions imitatives électroniques, Nicolas Mazziotta, « Traiter les abréviations du français médiéval. Théorie de l'écriture et pratiques d'encodage », *Corpus*, 7, 2008, p. 191-218.

15. Preuve en est, sans doute, le paragraphe, souvent bref, consacré à la résolution des abréviations dans nombre d'éditions.

16. Nous avons, dans notre thèse (*Transcription de Garin de Monglane à partir du manuscrit du XIV^e siècle Royal 20 D XI de la British Library: description méthodique et analyses linguistiques [vol. I]; transcription des 12590 vers de la copie de Londres [vol. II]; notes, glossaire, table des noms propres [vol. III]*), soutenue sous la dir. de Nelly Andrieux-Reix à la Sorbonne Nouvelle en 2008), remis en cause l'idée que la recherche d'une tentative d'organisation en système des abréviations serait vaine (défendue notamment par Geneviève Hasenohr, voir « Écrire en latin, écrire en roman: réflexions sur la pratique des abréviations dans les manuscrits français des XII^e et XIII^e siècles », dans Michel Banniard [dir.], *Langages et peuples d'Europe: cristallisation des identités romanes et germaniques [VI^e-XII^e siècle]*, Toulouse, CNRS/Université de Toulouse II Le Mirail, 2002, p. 80).

moustrer ou *monstrer*?)¹⁷. Aussi utiliserons-nous l'italique ou le soulignement¹⁸ chaque fois qu'un doute sur la valeur d'une abréviation est permis, quel que soit le type de texte édité, et non pas seulement en « cas de textes spécialement importants pour l'historien de la langue ou pour le paléographe », ainsi qu'y invitent Françoise Vielliard et Olivier Guyotjeannin¹⁹.

Une attention particulière est ensuite portée à la concurrence, à l'intérieur d'un même manuscrit, entre trois pratiques de scription : la séquenciation graphique, la segmentation graphique et un « entre-deux » graphique que nous appellerons « petit espacement ».

Si l'opposition entre « séquences » et « segments » graphiques a été largement problématisée, la question du « petit espacement », pratique pourtant fréquente, l'a beaucoup moins été ; et, à notre connaissance, seule Nelly Andrieux-Reix en fait mention dans un article préparatoire à l'édition du *Moniage Guillaume*²⁰, mais dont elle n'a finalement pas tenu compte.

17. Plus généralement, les hésitations de lecture entre lettres apparentées devront, si besoin, faire l'objet, dans l'introduction, d'un relevé et d'une étude chaque fois qu'un doute est permis afin d'avertir le lecteur des choix opérés et pour quelles raisons, notamment dans la lecture, outre de *u* voyelle et de *n* envisagée ci-dessus, de *u* voyelle et *u* consonne (par ex. : *aussier* ou *avressier*?) ; *c* et *t* (et donc de *ct* et *cc*, par ex. : *destruction* ou *destrucion*? *destruction* ou *destruccion*? ; *s* et *c* (par ex. : *ains* ou *ainc*?) ; *e* et *o* (par ex. : *ouvre* ou *euvre*?) ; *a* et *o* (par ex. : *atot* ou *otot*?) ; *vu*, *vo* et *w* (par ex. : *vuidier*, *widier* ou *voidier*?) ; *s* et *f* (et donc de *sf* et *ff*, par ex. : *consuit* [vb. *consivre*] ou *confuit* [vb. *confuir*] ? *desfermer* ou *deffermer*?). Les implications pour l'étude de la langue sont là aussi importantes : en se concentrant sur ce seul dernier exemple, la variation graphique *des/def* ne relève pas simplement d'un problème de morphologie lexicale ; la graphie *des-* pourrait en effet indiquer un emploi du préfixe privatif encore senti comme un élément de construction *des* + *fermer*, tandis que *def-* suggérerait une fusion totale des éléments *des* et *fermer* par laquelle la valeur privative de *des* ne serait plus perçue corrélativement à sa transformation graphique en *def-* au contact du *f* de *fermer*.

18. Le soulignement ayant toutefois notre préférence en raison de son aspect plus « vilisible », pour reprendre Jacques Anis (« Vilisibilité du texte poétique », *Langue française*, 59, « Le signifiant graphique », dir. Jacques Anis, 1983, p. 88-102).

19. Françoise Vielliard et Olivier Guyotjeannin (dir.), *Conseils pour l'édition des textes médiévaux*, t. 1, *Conseils généraux*, op. cit., p. 31.

20. Nelly Andrieux-Reix et Simone Monsonégo, « Transcription, lisibilité, transgression : quelques problèmes posés par les éditions de textes médiévaux », dans Claude Buridant (dir.), *Le Moyen Français : le traitement du texte (édition, appareil critique, glossaire, traitement électronique)*, Strasbourg, Presses universitaires de Strasbourg, 2000, p. 59-60.

Cette variation graphique à l'œuvre dans tout manuscrit médiéval apparaît d'autant plus caractéristique que le mot graphique, selon Nelly Andrieux-Reix et Simone Monsonégo²¹, paraît avoir toujours existé en français, qui semble ne pas avoir réellement connu de *scriptura continua*, même si la question suscite le débat. Ainsi, la pratique de la séquenciation, en français, ne constituerait pas la survivance d'un fait historique de scription s'inscrivant dans un conservatisme graphique. Cette pratique pose dès lors la question des motivations linguistiques et des fonctions dont elle relève, la séquenciation ayant pu, de prime abord, apparaître anarchique, aléatoire et linguistiquement non pertinente²².

Les premiers travaux d'une étude systématique des séquences graphiques menés par Peter Rickard ont dressé un premier bilan des mécanismes régissant ces phénomènes²³. Nelly Andrieux-Reix et Simone Monsonégo ont conforté, tout en les approfondissant, les travaux et les premières conclusions de Peter Rickard : en s'appuyant sur un grand nombre de manuscrits, elles ont dressé une typologie de ces séquences – véritable « génome » de ces pratiques – mettant en évidence des invariants dans l'écriture des copistes, et ont formé l'hypothèse que la récurrence et la permanence de ces faits dans l'espace et dans le temps pouvaient révéler la possible existence d'une forme de norme, non officielle, évoluant en dehors de tout cadre prescriptif²⁴.

21. Nelly Andrieux-Reix et Simone Monsonégo, « Écrire des phrases au Moyen Âge. Matériaux et premières réflexions pour une étude des segments graphiques observés dans des manuscrits français médiévaux », *Romania*, 115, 1997, p. 289.

22. Voir Halina Lewicka, « Réflexions théoriques sur la composition des mots en ancien et en moyen français », *Kwartalnik neofilologiczny*, vol. 10, n° 2, 1963, p. 142 ; Robert Léon Wagner, *L'Ancien Français. Points de vue. Programmes*, Paris, Larousse, 1974, p. 41, 74.

23. Peter Rickard, « Système ou arbitraire ? Quelques réflexions sur la soudure des mots dans les manuscrits français du Moyen Âge », *Romania*, 103, 1982, p. 470-512.

24. Nelly Andrieux-Reix et Simone Monsonégo, « Écrire des phrases au Moyen Âge. Matériaux et premières réflexions pour une étude des segments graphiques observés dans des manuscrits français médiévaux », art. cit., p. 289-336 ; *ead.*, « Les unités graphiques du français médiéval : mots et syntagmes, des représentations mouvantes et problématiques », *Langue française*, 119, « Segments graphiques du français. Pratiques et normalisations dans l'histoire », dir. Nelly Andrieux-Reix et Simone Monsonégo, 1998, p. 30-51.

Aussi, pour conserver la trace des séquenciations et des segmentations, et « pour ne pas consacrer de formes en réalité non attestées ou à l’attestation incertaine et afin de fournir au lecteur [...] les informations nécessaires à l’approche linguistique des codes graphiques médiévaux²⁵ », Nelly Andrieux-Reix conseille le recours à l’utilisation de deux signes :

- le « tiret simple » (ou trait d’union) pour indiquer « une séquenciation (écriture en un seul mot graphique) pratiquée dans le manuscrit de deux éléments²⁶ susceptibles d’être individués, par exemple *par-tot, si-fait, em-prist*²⁷ » ;
- le « tiret double » (ou tiret long) pour indiquer « la présence d’un blanc entre deux éléments²⁸ susceptibles d’être employés en séquence ([qui] sont souvent les mêmes que ceux qui peuvent être dotés d’un tiret simple), par exemple *par—tot, si—fait, en—fuient*²⁹ ».

Ainsi que nous l’évoquions plus haut, c’est à Clovis Brunel que l’on doit d’avoir initié l’ utilisation du trait d’union dans cette fonction ; à Jacques Monfrin la problématisation de la question des séquenciations dans une édition en s’appropriant le système mis en place par Brunel³⁰. Nelly Andrieux-Reix introduit donc le « tiret double », absent des transcriptions des chartistes où les blancs graphiques ne sont pas signalés autrement que par l’emploi d’une espace, conforme au code graphique moderne, entre deux mots. Le système originel, même s’il a fait ses preuves, apparaît donc imparfait car il ne fait pas le départ, dans les emplois en segments, entre ce qui relève de l’anecdotique, par exemple *il a*, qui serait écrit en deux mots dans le manuscrit, et du linguistiquement

25. *Le Moniage Guillaume, chanson de geste du XII^e siècle*, éd. cit., p. 25-26.

26. Nous précisons : de deux éléments ou plus (jusqu’à quatre, par ex. : *des-or-en-avant*).

27. *Le Moniage Guillaume, chanson de geste du XII^e siècle*, éd. cit., p. 26.

28. Nous précisons : de deux éléments ou plus (jusqu’à quatre, par ex. : *des—or—en—avant*).

29. *Le Moniage Guillaume, chanson de geste du XII^e siècle*, éd. cit., p. 26.

30. *Documents linguistiques de la France*, t. I, vol. 2, *Chartes en langue française antérieures à 1271 conservées dans le département de la Haute-Marne*, éd. cit., p. lxx. Toutefois, la question esthétique reste une préoccupation qui fait renoncer l’éditeur à utiliser le tiret simple dans le cas, par ex., de *a-ssavoir* en raison de la physionomie « disgracieuse » qu’il lui donne (*ibid.*).

pertinent, par exemple *en mener* qui le serait en deux et qui apparaît ainsi « noyé » dans la masse des segmentations plus ou moins « sémantiquement vides ».

Mais, ainsi que nous l'évoquions aussi plus haut, Nelly Andrieux-Reix a finalement renoncé à utiliser dans son édition un troisième tiret dont elle avait quelques années plus tôt suggéré et problématisé l'emploi et que nous utiliserons ; il s'agit du « tiret redoublé³¹ » (ou double trait d'union) pour indiquer un « petit espacement³² », par exemple noté tel dans *de--l-eschequier*, qui soulève un problème similaire à celui d'un emploi en séquence (*de--l-eschequier* = *de l'eschequier* ou *del eschequier?* = *deleschequier*). Le tiret signale alors une indécision dans la lecture du manuscrit et dans la pratique même du copiste entre un emploi en séquence ou en segments.

Faisant nôtre l'idée qu'« étant donné l'altérité de notre disposition instituée de l'écrit et de ce qui s'essaya au Moyen Âge, il importe de retenir de la disposition médiévale ce qui fait sens et doit être sauvé, aux dépens éventuels du reste³³ », nous maintiendrons dans le texte édité les éléments les plus informatifs d'une pratique de l'écrit en séquences ou en segments, variable dans les écritures successives du français mais variant aussi en synchronie selon les habitudes scripturaires individuelles ou d'atelier, en mettant en évidence ce qui dans l'écriture du copiste présente un intérêt particulier faisant sens dans l'histoire du français, sur quatre plans essentiellement : graphique, lexicologique, sémantique et proprement grammatical.

31. Nelly Andrieux-Reix et Simone Monsonégo, « Transcription, lisibilité, transgression : quelques problèmes posés par les éditions de textes médiévaux », art. cit., p. 60.

32. Par opposition au « grand espacement », équivalent de la « segmentation », et à l'absence d'espacement (= séquenciation ; soit l'ensemble des signes employés : séquenciation [-], petit espacement [--], segmentation [—]). Déterminer ce qui relève d'un « petit espacement » plutôt que d'une séquenciation ou d'un espacement assez « grand » pour nous permettre d'en déduire une segmentation n'est pas toujours chose aisée, notamment en cas d'écriture (re)serrée. Le principe sera de considérer qu'il y a « petit » espacement dès l'instant où il apparaît que le copiste a sans doute levé la main et/ou que les unités, séparées par une espace ténue, ne sont pas reliées par une ligature.

33. Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la philologie*, Paris, Éditions du Seuil, coll. « Des travaux », 1989, p. 43.

Les principes généraux consisteront donc à appliquer le système des tirets :

- au plan graphique : pour mettre en évidence les phénomènes relevant de la « syntaxe graphique³⁴ », autrement dit de transformations graphiques corrélées à des séquenciations touchant aux graphies de consonnes (redoublements, substitutions, réductions, vocalisations) :
 - dans des unités lexicales en cours de formation ou pressenties comme telles par le copiste et entérinées ou non par l'histoire ultérieure de ces unités : par exemple *de—sus*, mais *de-ssuz* ; *en—prendre*, mais *em-prendre* ; *mont—tirant*, mais *mon-tirant* ; *mal—gré*, mais *mau-gré*, voire *mou-gré* ;
 - dans des unités sans lendemain au plan du lexique et constituant sans doute des vestiges d'une ancienne pratique linguistiquement moins pertinente de la séquenciation : par exemple *em-pés* (*en + pés* (FM : *en paix*) ; *tou-droit* (*tout + droit*) ;
- au plan lexical :
 - dans des unités lexicales en cours de formation sans modification graphique : par exemple *en—mener*, mais *en-mener* ;
 - dans des unités lexicales en cours de formation ou non dont l'emploi, en séquence ou en segments, est non seulement variable en français médiéval mais aussi différent du code graphique du français moderne : par exemple *par-tout*, mais *par—tout* ;
 - dans des unités lexicales lemmatisées dans les dictionnaires mais qui apparaissent parfois en segments dans le manuscrit, comme un indice de non-lexicalisation (complète) pour le copiste : par exemple *a—savoir*, *par—*

34. Selon la création lexicale de Simone Monsonégo, voir « Les graphies et les mots dans les textes anciens. Problèmes de transcription et de codage », dans *Le Texte, un objet d'étude interdisciplinaire. Mélanges offerts à Véronique Huynh-Armanet*, Saint-Denis, Centre de recherche de l'Université Paris VIII, 1993, p. 175-190.

- amer*; dans tout type de segmentations atypiques: par exemple *en—viron*, *en—armé*;
- dans des unités lexicales non lemmatisées dans les dictionnaires mais qui apparaissent systématiquement en séquence dans le manuscrit, comme un indice (de début) de lexicalisation pour le copiste: par exemple *or-mier* (« or fin »); dans les hapax et créations lexicales non lemmatisés dans les dictionnaires: par exemple *en-margonna*; le plan lexical recoupant donc aussi parfois le plan grammatical (question du statut des éléments *en*, *a*, *par* dans ces exemples);
 - au plan sémantique, par exemple: problème de la différence de dénotation des formes du type *ma—dame/ma-dame* ou de la distinction de deux verbes comme dans *en—prendre/em-prendre* voire *em--prendre* (= « en prendre » ou « entreprendre »?). Dans ce cas, le problème de l'interprétation sémantique, corrélé à la formation en cours de l'unité lexicale *emprendre*, peut être accru ou résolu par des phénomènes de variations graphiques (*en—p/em-p/em--p*), le plan sémantique recoupant donc aussi parfois le plan graphique et lexical, voire proprement grammatical (question du statut de l'élément *en/em* ici);
 - au plan morphosyntaxique et proprement grammatical, par exemple: incertitude concernant le statut d'enclitique ou de proclitique, comme dans *de-l-eschequier* (= *de l'eschequier* ou *del eschequier*?); question du statut par exemple de l'élément *a* dans *l-a-lieve* (= *la lieve* ou *l'alieve*?), le problème grammatical recoupant dans ce cas la question de l'interprétation sémantique.

Si l'on est bien obligé d'admettre que l'ajout de ces quelques signes graphiques peut rendre, de prime abord, la lecture des textes un peu moins aisée par la charge typographique qu'ils imposent, cela ne tient qu'à l'habitude que nous avons des signes graphiques modernes, par exemple du tréma lui-même introduit dans le code graphique pour nous faciliter l'accès et la

lecture des textes médiévaux³⁵. Mais on constate, de-ci, de-là, une utilisation plus ou moins libre du tréma³⁶ phonétique dont la place, lorsqu'il est porté, varie en outre selon les traditions³⁷; on observe encore une restitution de la ponctuation plus ou moins en prise avec les habitudes graphiques modernes, une résolution de certaines abréviations plus ou moins normée³⁸; on note une certaine variabilité dans les principes d'établissement du texte parce que ceux-ci apparaissent en réalité plus ou moins utiles³⁹. Toutefois, afin d'éviter la surcharge typographique, deux allègements d'usage du système paraissent possibles: concernant les clitiques, essentiellement, et d'une façon générale les problèmes plus proprement grammaticaux qui sont ceux qui font le plus violence à la lisibilité, les formes clairement enclitiques ou proclitiques (par exemple *a l'eschequier/al eschequier*, *ne l'osent/nel osent* ou encore *la vespree/l'avespree*, *la mène/l'amène*⁴⁰) pourront se passer du tiret, de même que toutes

-
35. Voir Susan Baddeley, « Théorie et pratique de la segmentation graphique dans les textes français du premier tiers du XVI^e siècle », *Langue française*, 119, « Segments graphiques du français. Pratiques et normalisations dans l'histoire », dir. Nelly Andrieux-Reix et Simone Monsonégo, p. 52.
36. Voir Mario Roques, « Établissement de règles pratiques pour l'édition des anciens textes français et provençaux », *Romania*, 52, Paris, 1926, p. 245. Les restrictions d'usage du tréma préconisées par ces instructions sont parfois corrélées à des considérations qui apparaissent aujourd'hui dépassées par les possibilités du traitement de texte informatique.
37. Si la tendance actuelle est de placer le tréma phonétique selon la nature de la voyelle du groupe contenant une diérèse, par ex. *eü* (cf. Françoise Viellard et Olivier Guyotjeannin [dir.], *Conseils pour l'édition des textes médiévaux*, t. I, *Conseils généraux*, op. cit., p. 52-53), une tradition plus ancienne, représentée par ex. par Tobler et Lommatzsch, le place sur la première voyelle du groupe quelle que soit la nature de cette voyelle, par ex. : *ëu*.
38. Voir les avis divergents, sur le maintien ou le développement par ex. de l'*x* finale, de Pio Rajna (« *x = us* », *Romania*, 55, Paris, 1929, p. 528) et Jean Acher (« Sur l'*x* finale des manuscrits », *Revue de linguistique romane*, 56, 1913, p. 148-158).
39. Voir Mario Roques, « Établissement de règles pratiques pour l'édition des anciens textes français et provençaux », art. cit., p. 245, au sujet de l'inutilité dans certains contextes de l'adjonction du tréma.
40. Et non pas, par ex., *nel—osent/he—l-osent* ou *la—vespree/l-avespree*; en revanche, l'incertitude de lecture/scription sera indiquée: *la--vespree*, *l--avespree*, le tiret redoublé permettant dans ce dernier cas de mettre en évidence un emploi peut-être autonome de l'article élié *l*, l'emploi de consonnes entre deux blancs graphiques non pourvues d'un signe abrégatif étant un fait exceptionnel dans le code graphique du français médiéval dont on relève quelques occurrences dans la partie du manuscrit

les formes de forte occurrence systématiquement employées clairement en séquence ou en segments quel que soit le type de mots concernés (lexicaux ou grammaticaux); le fait devra être dûment signalé dans l'introduction par un relevé exhaustif par type de description des formes concernées.

Les rares consignes éditoriales promulguées en rapport avec ces problématiques avant l'édition du *Moniage Guillaume* concernaient essentiellement les questions relatives à la syntaxe graphique et aux faits d'enclise et de proclise.

Ainsi, Clovis Brunel⁴¹, davantage que Paul Meyer⁴² et Mario Roques⁴³, évoquait, avant l'introduction du tiret simple dans le *Supplément*, la question de « la séparation des mots grammaticaux » dans ses recommandations; mais l'appréciation de la segmentation ou non des éléments était laissée au libre arbitrage des éditeurs et, au bout du compte, seule la question des clitiques y était traitée dans des termes qui rappellent les instructions de Meyer. Dans les derniers principes édictés à ce jour, Françoise Vielliard et Olivier Guyotjeannin remettaient en cause le système mis en place par Brunel et continué par Monfrin en proscrivant l'emploi du trait d'union: « En latin comme en langue vulgaire, on n'emploie jamais le trait d'union⁴⁴ »; aussi l'appréciation de la segmentation ou non des éléments redevient-elle fonction de la « dénotation », et donc du sens (par exemple *ma dame/madame*), et de « l'analyse grammaticale [qui] peut rendre perceptible l'individualité ou la solidarité de chaque composant: *desorenavant* mais *des ore en avant*; *toujours* mais

de Londres consignait *Garin de Monglane: l' avon* (v. 304), *m ennuia* (v. 10652), *l apercevroiz* (v. 12088) et *n i a* (en trois mots graphiques) à la rime v. 8527.

41. Clovis Brunel, « À propos de l'édition de nos textes français du Moyen Âge », *Bulletin de la Société de l'histoire de France, 1941-1942*, p. 67-74.

42. Paul Meyer, « Instruction pour la publication des textes en ancien français », *Bulletin de la Société des anciens textes français*, Paris, 1909, p. 64-69.

43. Mario Roques, « Établissement de règles pratiques pour l'édition des anciens textes français et provençaux », art. cit., p. 243-249.

44. Françoise Vielliard et Olivier Guyotjeannin (dir.), *Conseils pour l'édition des textes médiévaux*, t. 1, *Conseils généraux*, op. cit., p. 39; voir aussi p. 41.

tous jours et surtout *touz jours*⁴⁵ ». Par ailleurs, l'évaluation de seuils d'illisibilité continue de présider aux conseils énoncés; le recours à des méthodologies d'édition différentes selon le type de texte y est affirmé en accordant plus d'attention au « cas très particulier de documents linguistiques importants » duquel le texte littéraire est implicitement exclu⁴⁶, comme si les préoccupations linguistiques et esthétiques ne pouvaient rien avoir à faire l'une avec l'autre.

La transcription des consonnes redoublées fait quant à elle l'objet de pratiques variables et aléatoires selon les éditeurs et les instructions. Nelly Andrieux-Reix prend pour exemple⁴⁷ l'« unanimité [qui] semble faite pour *assavoir* sur le maintien tel quel de l'ensemble de la séquence (par opposition à *a savoir* à transcrire en deux [segments]: *a savoir*) ». Mais ce « maintien tel quel » pour les éditeurs qui le pratiquent se heurte, sans signalement graphique spécifique, à la considération, sans nuance, que *assavoir* constituerait une forme complètement lexicalisée, ce qui ne semble pas être le cas en ancien français: selon Monfrin⁴⁸, ce verbe pourrait même ne jamais avoir existé; d'ailleurs, Tobler et Lommatzsch ne le lemmatisent pas et Godefroy n'en répertorie que trois exemples d'emploi dans le *Complément* à son *Dictionnaire*⁴⁹. Joan Williamson de son côté proposait une transcription en séquence de tout redoublement consonantique du type *assavoir*⁵⁰. Alfred Foulet et Mary Speer⁵¹ ainsi que Philippe Ménard⁵² préconisaient une transcription des consonnes

45. *Ibid.*, p. 41.

46. *Ibid.*, p. 39-42.

47. Nelly Andrieux-Reix et Simone Monsonégo, « Transcription, lisibilité, transgression: quelques problèmes posés par les éditions de textes médiévaux », art. cit., p. 59.

48. *Documents linguistiques de la France*, t. I, vol. 2, *Chartes en langue française antérieures à 1271 conservées dans le département de la Haute-Marne*, éd. cit., p. lxvi.

49. Voir toutefois, pour la période du moyen français, le *DMF*.

50. Philippe de Mézières, *Le Livre de la vertu du sacrement de mariage*, éd. Joan B. Williamson à partir du ms. Paris, BnF, fr. 1175, Washington D.C., The Catholic University of America Press, 1992, p. 25-27.

51. Alfred Foulet et Mary Blakely Speer, *On Editing Old French Texts*, Lawrence, Regents Press of Kansas, 1979, p. 60-62.

52. « Problèmes de paléographie et de philologie dans l'édition des textes français du Moyen Âge », dans Philip E. Bennett et Graham A. Runnalls, *The Editor and the Text. Mélanges Anthony J. Holden*, Edinburgh, Edinburgh University Press, 1990, p. 1-9.

doubles en pratiquant une coupure conforme au code graphique du français moderne par l'instauration d'un blanc (par ex. *il nou* [« ne le »] *ssalue*). Plus récemment, Françoise Vielliard et Olivier Guyotjeannin formulaient le même conseil⁵³.

Mais ces consignes n'envisageaient pas clairement les cas de transformations du type *empié* (*en + pié*), et passaient outre le problème récurrent dans tout manuscrit du « petit espacement ». Or, en dépit des apparences, « petit » et « grand » espacement ne sont pas nécessairement réductibles l'un à l'autre, ainsi que Nelly Andrieux-Reix l'avait déjà observé⁵⁴ et ainsi qu'il s'observe de façon encore plus nette dans la partie du manuscrit de Londres consignant *Garin de Monglane*⁵⁵. Et les conséquences pour l'étude de la langue sont là aussi importantes.

Illustration du système proposé par le manuscrit de Londres

[fol. 1a]

Ci coumence[*occurrence unique*]l'estoyre de Guerin de Monglenne, et aprez de Girart de Vienne et de Renier de Gennes et de Milon de Puille et de Hernaut de Biaulande. Et aprez d'Aymeri coument il ot Nerbone et Ermengart a moullier et des enfans qui d'euls issirent, c'est a—savoir de Guillaume d'Orenges et de Bernart de Brubant et de Buevon de Comarchis et de Guerin d'Anseune et d'Ernault le rous et d'Aymer le chetif et de Guibert qui fu roys d'Andrenas et de leur .V. sereurs coument elles furent mariees et a quelz seigneurs. Et y sont li ver Fouque de Sandie et tout li fait Renoart au tinel et de son filz Maillefer et tout li fait Guillaume d'Orenges d--usques a sa mort.

53. Françoise Vielliard et Olivier Guyotjeannin (dir.), *Conseils pour l'édition des textes médiévaux*, t. 1, *Conseils généraux*, op. cit., p. 42-43.

54. Nelly Andrieux-Reix et Simone Monsonégo, « Transcription, lisibilité, transgression : quelques problèmes posés par les éditions de textes médiévaux », art. cit., p. 60, qui notent que « l'ambiguïté dont le petit espace se révèle doté est comparable à celle qui sera plus tard attachée au trait maintenant dit "d'union" et qui, en fait, est autant de désunion que d'union ».

55. Précision rendue utile et nécessaire par le fait qu'il s'observe des différences notables de scription entre notre partie du manuscrit et celle consignant *Le Moniage Guillaume*, pourtant de la même main et principal support à l'étude d'Andrieux-Reix (voir Nelly Andrieux-Reix et Simone Monsonégo, « Transcription, lisibilité, transgression : quelques problèmes posés par les éditions de textes médiévaux », art. cit., p. 56).

I

Seignour, or escoutez, por Dieu omnipotent,
 Que Dame—Diex vos doit honor et joie grant!
 Oÿ avez chanter de Bernart de Brubant
 Et d’Ernaut de Biaulande, d’Aymeri son enfant,
 De Girart de Viane a-l-orgueilleux semblant
 Et de Renier de Gennes que Diex par—ama tant,
 Qui fu pere Olivier le compaignon Rollant,
 De Guillaume et de Fouque et du preu Viviant
 Et de la fiere geste dont on parole tant
 Qui tant orent de paine suz la gent mescreant.
 Cil jogleour n’en sevent la montance d’un gant
 Car tuit en ont lessié le droit coumencement
 De Guerin de Monglenne le chevalier vaillant
 Dont celle geste issi dont chantent li auquant.
 Ja sarez qui il fu et dont et de quel gent
 Et coument il conquist Monglenne et Mont—Tirant
 Et la terre en-viron une jornee grant
 Qu’a celui tans tenoient Sarrazin et Perssant
 Et qui fu celle dame dont furent cil enfant,
 Que l’en apele « geste » des le coumencement
 El royaume de France.

[fol. 1b]

II

Seignor, vous savez bien, quant Pepin fu feniz,
 Que Charles li siens filz fu chaciez du païs,
 Par force l’en--chacierent et Rainfroiz et Heudriz.
 Puis ala a Galafre le roy des Arrabiz,
 Aida li de sa guerre contre ses anemis.
 Quant il l’ot tret a fin et ot Bremant occis
 Et il ot Durendal le riche branc conquis,
 Si s’en revint arriere en France son païs.
 Tant fist par son grant sens, par force et par amis
 Que il fu couronez el moustier Saint Denis.
 Les .II. sers fist destruire et lor meillors amis,
 Puis conquist il maint regne et maint riche païs
 Par son fier vasselage.

Exemplification des données du problème par le manuscrit de Londres

Les problèmes éditoriaux posés par la résolution des abréviations médiévales sont nombreux, et leurs conséquences pour la langue importantes; la mise en relief d'une graphie restituée dans tous les cas où la résolution d'une abréviation est problématique doit permettre au lecteur, dans le cadre d'une analyse linguistique de l'écrit, de pouvoir se référer au texte édité sans risquer d'attester au plan statistique des graphies qui apparaissent développées en toutes lettres dans l'édition mais qui sont abrégées dans le manuscrit.

Pour illustrer notre propos, nous prendrons pour exemple le cas d'une abréviation polysémique: *g* linéaire ou suscrit employé devant *u* consonne ou *m*, et dans l'abréviation de certaines graphies grammaticales (du type *eux*).

La valeur *cou* de *g* linéaire employé devant *m* est attestée dans le manuscrit de Londres par une occurrence unique développée en toutes lettres dans le titre rubriqué: *coumence*. S'il apparaît très vraisemblable que ce titre a été rédigé de la main même du copiste du manuscrit, et que celui-ci l'a aussi peut-être créé, l'attestation en toutes lettres par une occurrence unique constitue nécessairement un frein à la restitution au plus près d'un état de langue sans créer d'artefacts; à quoi s'ajoute le problème de la fiabilité de la distinction *u/n*.

Certes l'examen des graphies *u* et *n* du manuscrit met en évidence un emploi généralement différencié et relativement stable des deux graphies, malgré quelques ambiguïtés et même si on lit parfois presque systématiquement *n* là où seul *u* semble possible⁵⁶. Or, de cette étude circonstanciée de toutes les formes pouvant commencer par *con-*, *cou-* ou *co-*⁵⁷ où une ambiguïté est possible⁵⁸, il ressort le fait contradictoire suivant, susceptible

56. Par ex., *ducau(t)/ducaus* se lit *ducan-* dans 15 de ses 17 occurrences.

57. La séquence *com-* ne posant pas problème.

58. La liste des formes concernées est assez restreinte. Dans le manuscrit, cela intéresse les formes et les paradigmes de *couvoitier*, *couvenir*, *couvent*, *couvenant*, *coument*,

de créer une impasse éditoriale : on lit invariablement *cou-* dans tous les cas où la forme est développée, en même temps qu'on note l'absence d'une forme abrégée *cō-*, semblant exclure, de fait, la possibilité d'une lecture *con-*⁵⁹ et confirmer celle de *cou-*. Mais on note aussi, pour ainsi dire inversement, l'absence remarquable de formes développées *co-* pour ces mêmes formes pouvant alterner avec *cou-*, alors qu'il se trouve, par ailleurs, que les graphies *o* et *ou* alternent dans le manuscrit dans tous les cas où, précisément, le problème de lecture *ou/on* ne se pose pas, comme, par exemple, dans *cortois/courtois, plorer/plourer*. Que penser de cet invariant ? Cette double constatation crée une impasse éditoriale : d'un côté on se retrouve devant une (quasi) impossibilité d'une lecture *con-* par l'absence d'une abréviation qui pourtant existe dans le manuscrit pour *com-*, et de l'autre, devant un invariant *cou-* vs *co-* qui peut apparaître suspect dans l'économie générale du système graphique du copiste et pourrait faire pencher la balance dans le sens d'une lecture *con-*. Que penser de cette constatation ? La résolution des abréviations reposant sur les formes développées en toutes lettres correspondantes, on devrait ici considérer qu'il ne s'agit là que d'un pur hasard. Mais cette assurance pourrait être remise en cause par l'occurrence abrégée, elle aussi unique, de *mōstrer* (au v. 2230, vs *moustrer*, 28 occurrences en toutes lettres).

Il semble en effet envisageable de considérer, dans ce cas, que l'abréviation puisse permettre d'atteindre la « vérité » du système graphique du copiste et aider à l'éclaircissement des formes développées pour lesquelles sans aucune exception on lit pourtant invariablement *moustrer*, et donc nous réorienter vers une lecture *monstrer*, la possibilité que la barre dite

couman-, *coumenc-*, *escousser* (mais aussi, le problème se répercutant finalement sur elles, les formes des paradigmes de *groucier*, *demoustrer*, *moustrer* – mais, peut-être, *monstrer*, voir *infra*).

59. Le problème est ici créé par le fait que ces formes s'abrègent conventionnellement par *g*, plus économique que *cō* dont on ne relève qu'un exemple lorsqu'un tilde est mis pour *n*, dans *recōneues*, alors même que la barre horizontale suscrite est assez fréquemment employée pour *m*, par ex. dans *c'om*, *com*, *comme*.

« de nasalité » puisse ici valoir pour *u*⁶⁰ se heurtant en réalité au système graphique du copiste qui ignore, partout ailleurs où cela aurait été possible⁶¹, la valeur vocalique de ce signe. Soit à nouveau une occurrence unique. Cela induirait peut-être aussi à penser que *u* et *n* pourraient ne pas avoir, pour ce type de formes du moins⁶², de valeur distinctive⁶³ en soi dans le manuscrit, bien qu'on lise clairement *coumence*, comme on lit toujours *covenant*, *couvent*, *couvoitier*, *(de)moustre*, *groucier*. Mais pourrions-nous pour autant en déduire une lecture *conmence*, *convoitier*, *conment*, etc. sur la base d'une possible lecture *monstrer* de *mōstrer*? Rien n'est évidemment moins sûr car, si l'abréviation *g* peut masquer l'existence par ailleurs toujours possible de *cō-*, inversement, *mō* ne peut pas être masqué par un signe qui aurait pu à lui seul abrégé toute la suite graphique *mon-* ou *mou-*, abréviation inexistante dans le système abrégatif médiéval. Aussi ces constatations, basées sur un nombre très faible d'occurrences et donc sur des manques graphiques, ne sauraient à elles seules battre en brèche la confiance, presque nécessairement aveugle, en une forme développée telle qu'elle se présente à nos yeux. Aussi doit-on considérer que *mōstrer* vaut pour *moustrer*, conformément à la lecture des formes développées. Or, dès lors que *con-* et *co-* sont inexistants dans certaines formes, que *o* et *ou* alternent dans toutes les formes non concernées par le problème posé ici et que *u* et *n* pourraient ne pas se discriminer clairement, il paraît impossible de déterminer si ce qu'on lit *cou-* vaut véritablement pour *cou-* ou pour *con-*. Qui plus est, on observe que les formes pouvant être graphiées par la suite *coum-* sont, à l'exception de *coumence*, systématiquement abrégées *gm* (170 occurrences). Cette sorte d'euphémisation graphique pourrait tout à la fois

60. Voir, à ce sujet, Philippe Ménard, « Édition de texte et paléographie: le problème de la transcription de *u* et de *n* », dans Maria Colombo Timelli et Claudio Galderisi (dir.), « Pour acquérir honneur et pris ». *Mélanges de moyen français offerts à Giuseppe Di Stefano*, Montréal, Ceres, 2004, p. 285; Anne Rochebouet, « Une "confusion" graphique fonctionnelle? Sur la transcription du *u* et du *n* dans les textes en ancien et moyen français », art. cit., p. 206-219.

61. Et cela aurait été possible dans plusieurs centaines d'occurrences.

62. En revanche, *n* et *u* consonne ont une valeur discriminante dans tous les cas où leur confusion serait sémantiquement ambiguë : *nos/vos*, *nous/vous*, *nostre/vostre*, par ex.

63. Parce que non ambigu sémantiquement (voir la note précédente).

cache et révéler, de la part du copiste lui-même, un embarras graphique où l'abréviation, par sa polyvalence, se substituant à la graphie sans en supporter les inconvénients, notamment phoniques, jouerait le rôle d'élément « cache-misère », indécision que le cas de *moustrer/mōstrer* pourrait révéler.

Un problème à certains égards similaire est encore posé dans le manuscrit de Londres par la mise en place d'un (micro)système d'abréviations, que nous appellerons « dichotomiques », concernant *g* suscrit dans l'abréviation de certaines graphies grammaticales.

Selon ce système, les graphies « nouvelles », non ou moins traditionnelles, et qui sont toujours des graphies « longues », par exemple *nous*, *vous*, sont représentées par des abréviations *no^o*, *vo^o* qui s'opposent aux abréviations *n^o*, *v^o* représentant les graphies traditionnelles « courtes » *nos* et *vos*; la validité de ce système paraît démontrée par l'exemple de l'abréviation *to^o*: alors que seules les graphies « longues » *tous* et *touz* se rencontrent en toutes lettres dans le manuscrit, *t^o* ne se rencontre pas, non plus que les graphies *tos* et *toz* qui lui correspondraient selon le système graphique du copiste. Or la tentation peut être grande pour l'éditeur de réduire – sans la signaler et quel que soit le choix opéré – la problématique de ces abréviations « dichotomiques » à une absence d'opposition : dans le cas, observé dans le manuscrit de Londres, des couples d'abréviations *c^o/ce^o* (= *ceus/ceulz* selon le rapport de la fréquence occurrence) et *e^o/eu^o* (= *eus/eulz* selon le même rapport), la fréquence occurrence étant toujours favorable aux graphies « nouvelles », on développerait ainsi par exemple *e^o* en *eulz* sans respecter le système de « double orthographe » mis en place par le copiste ainsi que semblent l'attester les données du manuscrit ; de la même manière, dans les textes où les graphies « longues » seraient majoritaires, on considérerait alors que *g* suscrit vaudrait par exemple pour *us* dans *no^o*, pour *ous* dans *n^o*; dans les textes où les graphies « courtes » seraient majoritaires, on considérerait alors que *g* suscrit vaudrait par exemple pour *os* dans *n^o*, pour *s* dans *no^o* – que d'aucuns qualifieraient d'« erreur

d'abréviation » – cette dernière valeur étant, certes, possible, mais seulement par défaut de graphie développée en toutes lettres correspondante : dans le manuscrit de Londres, l'introduction de *l* « étymologiques » dans la graphie de certains mots, notamment grammaticaux ainsi que nous l'avons vu plus haut, a motivé un système par lequel l'abréviation de *au^o* (= *auls*, *aulz* ou *aulx* selon le système graphique du copiste) est, de la même manière que les autres abréviations dichotomiques, supposée s'opposer à *a^o* (= *aus*) dans la notation ici de l'enclise représentant *a + les*; or le copiste associe une abréviation à une graphie (*auls*, *aulz* ou *aulx*) attestée par ailleurs dans le manuscrit dans quelques mots de façon très minoritaire (par ex. *Hernauls*, *haulz*, *maulx*), mais que ne connaît pas (encore) la forme développée en toutes lettres correspondante (*as*, *aus*); dès lors sommes-nous bien obligé d'admettre, ici, une valeur *s* de *g* suscrit dans *au^o* = *a + les*⁶⁴.

Pour résumer, on peut proposer le tableau suivant des abréviations dichotomiques observées dans le manuscrit de Londres :

Abréviations	Formes développées correspondantes selon les données du manuscrit	Forme développée retenue selon le rapport de la fréquence occurrenceielle
<i>n^o</i>	<i>nos</i>	=
<i>no^o</i>	<i>nous</i>	=
<i>v^o</i>	<i>vos</i>	=
<i>vo^o</i>	<i>vous</i>	=
<i>t^o</i> non attestée	<i>tos</i> , <i>toz</i> non attestées	∅
<i>to^o</i>	<i>tous</i> , <i>touz</i>	<i>touz</i>
<i>c^o</i>	<i>ceus</i> , <i>ceuz</i>	<i>ceus</i>
<i>ce^o</i>	<i>ceuls</i> , <i>ceulz</i>	<i>ceulz</i>

64. Un phénomène similaire est représenté dans le manuscrit par un autre emploi de *g* suscrit en collocation avec *a*, par exemple dans *bia^o*, où l'abréviation, conformément à la seule graphie représentée pour cette forme, *biax* (33 occurrences), a été « développée » en *x*, la quasi-exclusivité de la graphie *x* dans les formes parfois abrégées par *a^o* dans le manuscrit (*bia^o*, *chastia^o*, *chalumia^o*, *cheva^o*, *martia^o*) conférant au signe le caractère de graphie à part entière, *x* n'apparaissant plus avec le seul statut d'abréviation de la suite *us*. Autrement dit, le fait témoignerait ici d'un système particulier et original dans lequel, par une sorte de mise en abîme abrégative, *g* suscrit constituerait l'abréviation d'une abréviation, la valeur originelle de *x* étant *us*.

Abréviations	Formes développées correspondantes selon les données du manuscrit	Forme développée retenue selon le rapport de la fréquence occurrence
<i>e^o</i>	<i>eus, euz, eux</i>	<i>eus</i>
<i>eu^o</i>	<i>euls, eulz</i>	<i>eulz</i>
<i>a^o</i>	<i>as, aus</i>	<i>aus</i>
<i>au^o</i>	<i>aus graphie unique</i>	<i>aus</i>

La résolution du système abrégatif pose donc problème chaque fois que la graphie correspondante n'apparaît pas ou que son occurrence est unique, ce qui doit inviter à la prudence; la complexité du système abrégatif appelant par ailleurs la mise en évidence des graphies restituées dans tous les cas où plusieurs interprétations semblent possibles.

Les problèmes éditoriaux posés par la question de la séparation ou non des mots sont tout aussi nombreux, et les conséquences pour la langue tout aussi importantes. Parce qu'ils apparaissent comme un lieu d'observation privilégié des phénomènes notamment de lexicalisation et de grammaticalisation du français⁶⁵, les faits particuliers de « ponctuation du mot » ou de « syntaxe graphique » devront être sauvegardés dans la mesure du possible aux différents niveaux de la description linguistique distinguée plus haut.

Nous nous attacherons essentiellement ici aux problèmes posés par le « petit espacement », qui n'ont été que très peu abordés jusqu'à présent. Le petit espacement pose notamment problème chaque fois qu'il apparaît dans une unité en cours de formation, comme par exemple dans *en--chacier* (= *enchacier* ou *en chacier?*), ou entre deux éléments dont le premier est susceptible de subir une transformation graphique à la finale, comme par exemple dans *en--pié* (= *en pié* ou *enpié?*), le problème s'accroissant dans le cas des graphies complexes du

65. Voir Laurent Balon, « Les unités graphiques de l'ancien et du moyen français: un lieu d'observation privilégié pour une meilleure approche des phénomènes de lexicalisation et de grammaticalisation du français? », *Dossiers d'HEL* (SHESL), 9, « Écriture(s) et représentations du langage et des langues », dir. Julie Lefebvre, Jacqueline Léon et Christian Puech, 2016, p. 304-316.

type *ē--pié* (= *en pié, enpié, em pié* ou *empié*?). Pour illustrer notre propos, nous prendrons dans un premier temps l'exemple de ces graphies. Nous terminerons par l'aperçu de quelques difficultés posées par le recours à l'analyse grammaticale ou à la dénotation dans le traitement des séquences graphiques pratiquées dans les manuscrits avec l'exemple des formules *madame/ma dame, mes sires/messires, mon seignor/monseignor*.

Un espacement net, qu'on appelle « blanc graphique », laisse parfois place à un plus petit espacement que seule, semble-t-il, Nelly Andrieux-Reix a jusqu'à ce jour problématisé⁶⁶. Si cet « entre-deux » graphique correspond tantôt à une séquence, tantôt à une segmentation, ambiguïté que seule une étude de la fréquence occurrenceielle des formes permet parfois de lever, notamment dans les cas où la pratique de la séquenciation entraîne des modifications graphiques systématiques, la pratique pourrait dans quelques cas, notamment dans les unités en cours de formation, manifester une hésitation scribale concernant le statut grammatical ou la lexicalisation de certaines formes pouvant être employées en un ou en deux mots.

Dans le manuscrit de Londres, on observe la règle combinatoire suivante: *en*, pronom personnel régime ou préposition, employé entre deux blancs graphiques n'apparaît jamais sous une forme *em*⁶⁷. En revanche, lorsqu'il est employé en séquence avec un élément suivant dont l'initiale est *p* ou *b*, *en* est remplacé par *em* dans tous les cas (58 occurrences en toutes lettres) sauf un: *enporte* (v. 6587).

Ainsi, par exemple, *en pié*, en deux segments, devient systématiquement *empié* lorsque la forme est écrite en un seul tenant. Ce fait permet d'établir qu'un petit espacement comme dans *em--pié* équivaut donc, par le fait de la graphie *m*, à une séquence *empié*, la graphie *en--pié* n'étant en outre pas attestée

66. Voir Nelly Andrieux-Reix et Simone Monsonégo, « Transcription, lisibilité, transgression: quelques problèmes posés par les éditions de textes médiévaux », art. cit., p. 59.

67. La graphie *em* employée entre deux blancs graphiques semblant par ailleurs attestée: voir Eduard Schwan et Dietrich Behrens, *Grammatik des Altfranzösischen*, t. I, *Lautlehre*, Leipzig, O.R. Reisland, 1914, § 183, p. 110.

dans le manuscrit aussi bien que toute graphie du type *en--p-/b-* ; en conséquence de quoi il ressort que la pratique du copiste oppose « petits » et « grands » espacements partout où cela est vérifiable par des modifications graphiques. Cette observation permet de lever les ambiguïtés de graphies complexes du type « e tilde + petit espacement + base lexicale à initiale *p* ou *b* », comme dans *ē--pié* : si *m*, dans la forme en toutes lettres, constitue un indice sérieux d'un emploi équivalant à une séquence, la tentation dans ce cas peut être là aussi grande pour l'éditeur de considérer les deux éléments comme segmentés, et donc de transcrire *en pié* en deux segments : la mise en relief de cette règle combinatoire dans le système graphique du copiste permet ainsi de poser cette forme comme sans doute équivalente à *em--pié* et donc à *empié*. Toutefois, en raison à la fois de la variabilité selon les manuscrits, voire au sein d'un même manuscrit⁶⁸, et des préventions qui subsistent encore, la pratique du petit espacement sera préservée à tous les niveaux de la description linguistique distinguée plus haut, soit, pour la forme étudiée ici : *em--pié* ou *em--pié*.

L'apparition de « petits espacements » pourrait en outre procéder d'un geste plus ou moins réflexif, et revêtir par là une pertinence linguistique, corrélée à une interrogation du copiste sur le choix de scription à opérer.

Dans l'exemple de la triple alternance par exemple de *enchacier/en chacier/en--chacier* (v. 24), l'emploi indéterminé du vers 24 tiendrait à une difficulté sémantique particulière due au statut complexe de l'élément *en* et à la formation d'une nouvelle unité lexicale en cours ; cette perspective justifierait une scription *en--chacier*, irréductible aussi bien à une segmentation qu'à une séquence, au plan du découpage de la chaîne graphique du moins.

Soit l'exemple complet des vers 23-24 :

[...] Charles li siens filz fu chaciez du país,
Par force l'en--chacierent et Rainfroiz et Heudriz.

68. L'opposition entre petits et grands espacements n'étant pas systématique par ex. dans la partie du manuscrit de Londres consignant *Le Moniage Guillaume*, pourtant rédigé de la même main (voir *supra*, note 9).

Il est difficile ici, si l'on s'en tient à la présence d'un petit espacement, de présager si la forme *en--chacierent* est entendue au sens d'une unité lexicale et sémantique ou non. En effet, une segmentation univoque *en chacierent* aurait pu signifier un emploi de *en* comme substitut clairement anaphorique de *du pais*, plutôt que comme un préfixe se rattachant à *chacierent*. Or le recours au sens ne permet pas de lever réellement ici l'ambiguïté syntaxique et lexicale dans la mesure où, quel que soit le choix opéré, on obtient deux nuances sémantiques très proches et toutes deux acceptables en contexte : si *en* se rattache à *chacierent* en une unité lexicale, on entend : « Rainfroiz et Heudriz l'expulsèrent par la force » ; et si *en* est substitut, on entend : « Rainfroiz et Heudriz l'en chassèrent par la force ».

L'indécision ici concernant la valeur et le statut grammatical de *en* trouverait un écho graphique dans le fait que les éléments *en* et *chacierent* ont été séparés de manière équivoque par un blanc graphique tenu en prolongement d'une réflexion sémantique et grammaticale. Le problème pour ce type de formes paraît en outre accru par la stabilité graphique de *en* employé en séquence avec un élément commençant par une autre lettre que *p* et *b* et qui interdit toute transformation graphique à la jointure des deux éléments aussi bien selon le code graphique du copiste que d'après celui du français moderne. Cet « entre-deux » graphique pourrait peut-être dès lors témoigner de la difficulté d'appréhension de la part du copiste lui-même d'une unité lexicale *enchacier* en formation, et peut-être encore mal perçue, face à la possibilité de conserver *en* en tant qu'unité indépendante et pouvant donc soit se rattacher à *chacier*, soit conserver une de ses valeurs intrinsèques de substitut. Le petit espacement constituerait dès lors, de la même manière que la séquenciation et la segmentation, une trace écrite, matérielle, d'une opération cognitive à l'œuvre prise sur le vif de l'acte de scription ; le petit espacement procédant alors, comme la séquenciation et la segmentation, d'un geste plus ou moins réflexif.

Enfin, si le recours notamment à l'analyse grammaticale et à la dénotation, ainsi que le recommandent Françoise Vielliard et Olivier Guyotjeannin, permet dans de nombreux cas de trancher entre segmentation et séquenciation, comme par exemple dans « Que tout envers labat dessus le pavement » (v. 1631, où *labat* = « l'abat » vs « la bat »), il reste quelques cas où la systématisation de ces principes affecte ou efface certaines particularités linguistiques qu'il conviendrait au contraire de sauvegarder.

Soit l'exemple des formules *madame/ma dame, mon seignor/monseignor, mes sires/messires* : en suivant les recommandations de Françoise Vielliard et Olivier Guyotjeannin, ces expressions seront transcrites en deux segments lorsqu'elles « dénotent nettement un lien de dépendance amoureuse ou féodale⁶⁹ », en un seul tenant dans les autres cas.

Dans l'exemple du vers 352 : « Ma dame la roïne avant hier me manda », *Ma dame* dénote aussi bien un lien de dépendance féodale qu'une marque de politesse, sans qu'il soit vraiment possible de trancher entre les deux solutions : quel choix de transcription devra-t-on dès lors adopter ? *Madame* ou *Ma dame* ? L'usage du tiret double, défini comme une possibilité d'occurrence en séquence d'une forme employée en segments, permet dans *Ma—dame* de conserver la pratique du copiste tout en signifiant la double possibilité dénotative.

Dans l'exemple *mon seignor/monseignor*, on observe que la distribution des deux graphies est fonction d'une dichotomie « formes pleines »/« formes abrégées » par-delà toute considération dénotative qui favorise l'emploi de *mon*[^ˈ. forme abrégée en séquence (2 occurrences) et de *mon seignor* forme pleine en deux segments (5 occurrences) dans tous les cas, sauf un : *mon--*[^ˈ. dont l'emploi peut suggérer un début d'autonomie du signe abrégatif alors que la solidarité ou non des éléments *mon* et *seignor* apparaît liée à l'occurrence ou non de l'abréviation dans les autres emplois. Dans ce cas encore, la méthode traditionnelle efface cette particularité.

69. Françoise Vielliard et Olivier Guyotjeannin (dir.), *Conseils pour l'édition des textes médiévaux*, t. 1, *Conseils généraux*, op. cit., p. 41.

Dans l'exemple *me sires* (v. 10025, 11681), la forme *me* atteste sans doute un amuïssement de *s* final qu'il convient d'indiquer dans la transcription : l'ensemble de la formule sera donc transcrite *mé—sires* dans ses deux occurrences⁷⁰. Sur ces deux occurrences, il s'en trouve une (v. 11681) pour laquelle il ne saurait y avoir aucun lien de dépendance amoureuse ou féodale en contexte, et une autre, au vers 10025, pour laquelle ce lien de dépendance est évident. Le recours à la dénotation devrait ainsi, dans le contexte dans lequel apparaît l'occurrence du vers 11681, nous conduire à transcrire la forme en une seule séquence *mesires* pendant que nous transcrivions l'autre *mé* (ou *me*) *sires* en deux segments au vers 10025. Or la forme ainsi séquenciée n'apporte aucune information quant à la réalité de son occurrence en deux segments dans le manuscrit, et n'indique plus l'amuïssement de *s* final : au contraire, elle devient l'équivalent d'une simple occurrence de *messires* dont on pourra se servir, dans le cadre d'une étude linguistique, comme d'un exemple de réduction consonantique de *ss* en *s* à la jonction de deux éléments entrés en séquence (*mes + sires* > *mesires*). Or la recommandation d'une transcription en deux segments de l'occurrence du vers 10025 selon une considération dénotative se heurte ici à l'analyse grammaticale du premier élément *me* qui, suite à l'amuïssement de *s*, « rend perceptible la solidarité de chaque composant⁷¹ », ce qui pourrait impliquer *in fine* une transcription par *mesires* dans cette occurrence, y compris selon une méthode traditionnelle.

Ainsi, la méthode que nous proposons apparaît-elle bien comme un moyen de conserver le contenu linguistique et littéraire des manuscrits par-delà certaines difficultés philologiques posées par les textes d'ancien et de moyen français.

70. Si la bizarrerie d'une graphie *me—sires* permet à elle seule de signifier un emploi ouvert de *e*, l'usage de l'accent se justifie par la présence par ailleurs de graphies comme *de or* (v. 6979), qui représente sans doute aussi un amuïssement de *s* final (*des or* > *dé or*, à moins qu'elle ne constitue une variante lexicale [*de or*], non attestée dans les dictionnaires d'ancien et de moyen français, de *de maintenant*), mais dont la lecture par *e* ouvert n'est pas spontanément perçue.

71. Françoise Viellard et Olivier Guyotjeannin (dir.), *Conseils pour l'édition des textes médiévaux*, t. 1, *Conseils généraux*, op. cit., p. 41.

La technique éditoriale pour laquelle nous plaidons constitue une tentative d'unification des différentes pratiques éditoriales en ce qu'elles ont chacune de plus pertinent et de plus informatif pour le lecteur au plan de la tradition manuscrite et littéraire et au plan linguistique, et propose, tout en ménageant certaines exigences de l'édition canonique, d'accorder aux écrits littéraires la même minutie dans la transcription que celle souhaitée pour les documents d'archives.

L'approche que nous proposons consiste d'abord à considérer le manuscrit comme un témoin historique authentique d'un état de langue donné, la transcription la plus fidèle possible de celui-ci autorisant seule une description méthodique et raisonnée des procédures de « mise en texte » du manuscrit et une analyse linguistique comprise non pas comme une étude de langue telle qu'on la mène dans une édition de texte traditionnelle, mais comme une analyse de phénomènes relevant de la linguistique de l'écrit⁷².

Ainsi, dans l'objectif de ne pas créer d'artefacts lexicaux ou grammaticaux, et dans l'intérêt même de la recherche selon une approche plus fine de l'ancienne langue, une attention particulière sera apportée aux faits de syntaxe graphique par l'emploi d'un système de tirets (le trait d'union [-] pour indiquer une séquenciation, le trait d'union redoublé [--] pour indiquer un petit espacement, le tiret long [—] pour indiquer une segmentation) chaque fois qu'une forme présente un intérêt particulier dans la formation du français au plan sémantique (par ex. : problème relatif à la différence de dénotation des formes du type *mes sire/messire*), lexicologique (par ex. : phénomènes relevant de la formation du lexique et des unités en cours de lexicalisation du type *en mener/enmener*), graphique (par ex. : transformations du type *en prendre/emprendre*), et morphosyntaxique (par ex. :

72. Cette perspective se différenciant de l'étude de langue traditionnelle par une approche argumentative des faits de langue et la mise en relief de microsystèmes dans la pratique des copistes, l'interprétation de ces microsystèmes pouvant éventuellement suggérer une réflexion grammaticale. Voir Laurent Balon, *Étude linguistique de « Garin de Monglane »*, ouvrage à paraître, cité n. 11.

question des formes clitiques et distribution en emploi pro- ou enclitique du type *del ange* ou *de l'ange*).

Parce que les systèmes abrégatifs apparaissent d'une complexité telle qu'ils doivent faire entrer l'éditeur dans l'ère du doute et du soupçon, une attention particulière sera aussi portée aux abréviations par l'emploi d'un système de soulignement (ou d'italique) partout où la restitution d'une graphie pose problème au regard des données du manuscrit, l'ensemble de ces marques typographiques (soulignements/italiques et tirets) devant faciliter le travail du linguiste et notamment lui éviter le recours constant aux manuscrits pour en vérifier les données.

Ainsi que nous l'avons par quelques exemples illustrée, la question de la méthode d'établissement du texte d'ancien et de moyen français pose foncièrement le problème de la technique de transcription à adopter. Or la finesse et la fiabilité de l'analyse linguistique, qui dépendent directement du choix éditorial opéré, en seront les reflets. L'édition du texte littéraire admet parfois des libertés importantes. La lisibilité du texte notamment devient l'excuse de normalisations qui ne rendent pas toujours compte de la langue réelle, et leurs conséquences pour l'étude de la langue sont considérables. Faire des choix, c'est imposer une interprétation, qui empêche les réappréciations ultérieures. Or il est bien souvent impossible, à partir des textes édités selon la méthode traditionnelle, de faire la part entre les données du manuscrit et les interprétations de l'éditeur. C'est ce genre d'inconvénients que tente de parer la méthode que nous proposons, où la reproduction de la lettre et le refus de défigurer le matériau auquel elle s'attache permettront de réexaminer à loisir le cheminement de la langue à travers le temps.

Enjeux philologiques, linguistiques et informatiques de la philologie numérique : l'exemple de la segmentation des mots

Alexei Lavrentiev, Céline Guillot-Barbance & Serge Heiden
Laboratoire IHRIM, ENS de Lyon/CNRS

Depuis les années 1960, le développement des ressources numériques a été particulièrement important en linguistique historique et diachronique, tout spécialement pour la période la plus ancienne du français. Trois types de ressources ont ainsi progressivement vu le jour : des dictionnaires électroniques pour l'ancien et le moyen français (*Dictionnaire du moyen français* [DMF], *Anglo-Norman Dictionary*, *Dictionnaire étymologique de l'ancien français* et *Dictionnaire électronique de Chrétien de Troyes*, notamment), des corpus textuels médiévaux (Corpus d'Anthony Dees devenu depuis le « Nouveau corpus d'Amsterdam », base textuelle du DMF, base textuelle du Laboratoire de français ancien, Base de français médiéval) et des éditions électroniques de textes phares du Moyen Âge (comme par exemple le *Chevalier de la Charrette*¹, *The Online Froissart*², cinq romans de Chrétien de Troyes édités par Pierre Kuntzmann³, le roman de la *Queste del saint Graal* édité par Christiane Marchello-Nizia et Alexei Lavrentiev⁴).

L'essor de ces ressources textuelles et linguistiques s'est accompagné, dans le même temps, du développement d'outils de recherche et d'analyse de plus en plus élaborés et adaptés à la perspective diachronique. La Base de français médiéval

1. <http://www.princeton.edu/~lancelot/ss>.

2. <http://www.hrionline.ac.uk/onlinefroissart>.

3. Ces éditions ont été réalisées dans le cadre du projet de *Dictionnaire électronique de Chrétien de Troyes* (DÉCT, <http://www.atilf.fr/dect>).

4. http://catalog.bfm-corpus.org/qgraal_cm.

que notre équipe développe à l'ENS de Lyon⁵, par exemple, est passée des concordanciers imprimés des années 1990 au logiciel d'interrogation en ligne Weblex dans les années 2000⁶, puis, à partir de 2012, à un portail web reposant sur la plateforme TXM⁷. Chaque changement technologique a apporté aux utilisateurs de nombreuses nouvelles fonctionnalités. Le développement conjoint et simultané des ressources textuelles et logicielles a permis la réalisation d'un grand nombre d'études en linguistique diachronique portant aussi bien sur le lexique, que sur la syntaxe, la sémantique grammaticale, la pragmatique, etc.⁸.

L'essor de la linguistique diachronique de corpus a également permis d'enrichir l'encodage et l'annotation des textes numériques en fonction des programmes informatiques chargés de les analyser et des questions de recherche qui leur sont posées. Après une phase d'accroissement quantitatif, les corpus et éditions numériques se sont généralement dotés d'outils et de systèmes de catégorisation en vue de l'annotation linguistique des données primaires (lemmes et catégories morphosyntaxiques dans le cas du Nouveau corpus d'Amsterdam⁹, catégories morphologiques et structures syntaxiques dans le projet « Les Voies de français »¹⁰ comme dans le corpus Syntactic Reference Corpus of Medieval French¹¹, lemmes dans les éditions

5. <http://txm.bfm-corpus.org>.

6. Bénédicte Pincemin, Céline Guillot, Serge Heiden, Alexei Lavrentiev et Christiane Marchello-Nizia, « Usages linguistiques de la textométrie. Analyse qualitative de la consultation de la Base de français médiéval via le logiciel Weblex », *Syntaxe & Sémantique*, 9, « Textes, documents numériques, corpus. Pour une science des textes instrumentée », 2008, p. 87-110.

7. Serge Heiden, Jean-Philippe Magué et Bénédicte Pincemin, « TXM : une plateforme logicielle *open source* pour la textométrie – conception et développement », dans Sergio Bolasco, Isabella Chiari et Luca Giuliano (dir.), *Statistical Analysis of Textual Data*, actes de la 10^e Journée internationale d'analyse statistique des données textuelles (Rome, juin 2010), Milano, LED, 2010, t. II, p. 1021-1032.

8. Voir, entre autres, Christiane Marchello-Nizia, *L'Évolution du français. Ordre des mots, démonstratifs, accent tonique*, Paris, A. Colin, 1995 ; Sophie Prévost, *La Postposition du sujet en français aux xv^e et xv^e siècles. Analyse sémantico-pragmatique*, Paris, CNRS Éd., 2001 ; Céline Guillot-Barbance, *L'Évolution sémantique du démonstratif en français (ix^e-fin xv^e siècle)*. Deixis, personne et espace, Louvain, Peeters, 2017.

9. <http://www.uni-stuttgart.de/lingrom/stein/corpus>.

10. <http://www.voies.uottawa.ca>.

11. <http://srcmf.org>.

de Chrétien de Troyes par Pierre Kunstmann¹² déjà citées, chartes lemmatisées éditées sous la direction de Martin-Dietrich Glessgen¹³, corpus lemmatisé de PALM/Meditext¹⁴, etc.).

Cette double activité d’encodage de sources textuelles et d’annotation pour l’extraction et l’analyse d’informations linguistiques basées sur l’indexation systématique de tous les éléments de surface des textes (mots, ponctuations, etc.) amène à poser une nouvelle fois, de façon empirique et sous un jour partiellement nouveau, les questions traditionnellement abordées par l’édition de textes à l’échelle d’un corpus entier : comment, par exemple, traiter la variation morphologique et graphique du texte, sa ponctuation, les délimitations de ses unités linguistiques (mots, locutions, segments de discours comme les passages au discours direct) ? Qu’elles s’appliquent à des ressources créées à partir d’éditions papier numérisées ou à des éditions numériques natives (*born-digital*), les opérations réalisées sur support numérique conduisent toujours à des réflexions méthodologiques et à des choix techniques concernant tous les aspects du texte pris en charge par le travail philologique d’édition. Quelques publications de nature plus méthodologique traitent de ces choix, dont certains seulement sont spécifiques aux états de langue anciens, et de leur exploitation possible pour la recherche linguistique et diachronique¹⁵.

L’accroissement et l’enrichissement continu des ressources numériques a par ailleurs conduit à la mise en place et à la diffusion progressive de bonnes pratiques partagées pour les textes médiévaux (métadonnées attachées à l’unité textuelle, structure

12. Dans le cadre du projet DÉCT ; citées *supra*.

13. <http://www.rose.uzh.ch/docling>.

14. <http://palm.huma-num.fr/PALM>.

15. Bénédicte Pincemin, Céline Guillot, Serge Heiden, Alexei Lavrentiev et Christiane Marchello-Nizia, « Usages linguistiques de la textométrie. Analyse qualitative de la consultation de la Base de français médiéval via le logiciel Weblex », art. cit. ; Céline Guillot, Alexei Lavrentiev, Bénédicte Pincemin et Serge Heiden, « Le discours direct au Moyen Âge : vers une définition et une méthodologie d’analyse », dans Dominique Lagorgette et Pierre Larrivée (dir.), *Représentations du sens linguistique 5*, Chambéry, Université de Savoie, coll. « Langages », 2013, p. 17-41.

formelle et sémantique interne au texte, annotations diverses)¹⁶. Ces principes communs permettent le partage, l'échange et l'exploitation des données par un grand nombre d'outils et de chercheurs. Ils rendent également possible l'incrémentation collective, la capitalisation et la pérennisation des traitements et des enrichissements apportés aux données. Par exemple, les textes lemmatisés du projet de *Dictionnaire électronique de Chrétien de Troyes* ont été intégrés à la Base de français médiéval où ils ont bénéficié d'une vérification supplémentaire de l'étiquetage morphosyntaxique, ce qui a permis aux auteurs du dictionnaire, à leur tour, d'améliorer le traitement des mots grammaticaux dans leur base. Ces principes permettent, enfin, la confrontation de pratiques diverses et leur mise en perspective avec les techniques plus anciennes mais à certains égards plus éprouvées de la philologie « traditionnelle ».

Notre article sera centré sur une problématique très précise en rapport avec ces questions générales, celle liée à la délimitation des unités-mots à l'intérieur du texte numérique. Nous essaierons de montrer à partir de cet exemple concret les spécificités et les nouveautés liées à la perspective numérique et leurs implications pour la recherche linguistique, d'une part, et l'édition de textes, d'autre part. Dans un premier temps, nous aborderons les enjeux de la segmentation en mots pour la linguistique diachronique, du point de vue de ses objets et outils de recherche. Nous montrerons dans un second temps quelles sont les difficultés liées au traitement de cette question, puis nous proposerons des solutions adaptées au domaine numérique.

La segmentation du texte en mots : quels enjeux pour la linguistique diachronique de corpus ?

La délimitation des unités lexicales à l'intérieur du texte apparaît comme une question centrale pour la linguistique de corpus, quelles que soient les périodes considérées. Les métho-

16. Voir notamment les recommandations du Consortium international pour les corpus de français médiéval (<http://ccfm.ens-lyon.fr>).

dologies d'analyse qui caractérisent ce cadre de recherche reposent en effet sur (i) la possibilité de repérer et d'extraire les unités sur lesquelles on travaille, (ii) la possibilité de les dénombrer, (iii) la possibilité de gérer la variation intrinsèque de ces unités, y compris dans leur découpage formel et graphique.

*L'apport de l'analyse de l'hétérogénéité
des graphies médiévales*

Les enjeux de la segmentation en mots, qui peut sembler un problème très pratique, sont en réalité particulièrement importants pour la recherche diachronique. On sait que la délimitation des unités graphiques dans les manuscrits médiévaux diffère dans des proportions très variables de celle du français moderne. La restitution au chercheur d'aujourd'hui des pratiques médiévales authentiques est une donnée importante pour sa connaissance des manuscrits, de leurs conditions matérielles de réalisation, des pratiques et habitudes d'écriture qui se développent et évoluent tout au long de la période médiévale. Elle peut également nous donner de précieux renseignements quant à la conscience linguistique des copistes médiévaux.

Deux types de phénomènes méritent une attention particulière. Les locutions en cours de figement vers un lexème, comme la préposition *parmi*, l'adverbe négatif *jamais*, la conjonction *lorsque*, etc., sont le résultat du processus de grammaticalisation et/ou de lexicalisation de deux unités initialement autonomes. La disparition progressive de l'espace blanc entre les éléments de ces locutions dans les pratiques sribales peut servir d'indice précieux pour l'analyse du processus de grammaticalisation/lexicalisation. Le phénomène inverse, qui tend à dissocier linguistiquement et graphiquement deux unités antérieurement soudées, est beaucoup moins fréquent, mais tout aussi intéressant. L'étude diachronique de l'adverbe *très* du français moderne¹⁷, par exemple, a montré l'importance de cette étape graphique dans le passage du préverbe ou de la

17. Christiane Marchello-Nizia, *Grammaticalisation et changement linguistique*, Bruxelles, De Boeck/Duculot, 2006, p. 166-172.

particule *tres*, toujours accolée au mot qui la suit, au statut d'adverbe autonome. Les habitudes graphiques et leurs changements progressifs entrent dans ce cas en relation directe avec les évolutions linguistiques et montrent l'importance pour le linguiste de pouvoir prendre en compte la réalité matérielle du découpage lexical du texte médiéval. Les deux phénomènes que nous venons de citer sont à distinguer nettement d'autres habitudes graphiques médiévales, sans lien direct avec ces processus de formation d'unités lexicales, comme la tendance à souder les unités grammaticales atones (pronoms personnels, articles, etc.) au mot plein qui les précède ou qui les suit. Mais, même dans ces derniers cas apparemment moins importants pour l'approche diachronique, la segmentation médiévale n'est pas sans intérêt pour le linguiste (elle n'est pas sans liens avec les phénomènes d'élision ou d'enclise, ou d'absence d'autonomie de mots grammaticaux).

Nous proposons d'adopter une terminologie précise pour la désignation de ces différents phénomènes : nous utiliserons le terme *agglutination* pour désigner la pratique scribale qui consiste à souder graphiquement deux unités lexicales ou les composants d'une locution figée en cours de lexicalisation et le terme *déglutination* pour la pratique inverse qui consiste à séparer graphiquement les parties d'une seule unité lexicale ou d'une locution figée.

L'apport de l'usage de graphies normalisées

La nécessité de rendre compte de la matérialité du texte médiéval, dans son caractère instable et évolutif, se heurte à un principe concurrent et antagoniste : la nécessité de devoir repérer et compter les unités lexicales sur lesquelles on travaille oblige à dépasser ou à éliminer les variations formelles qui les opposent. Ces variations sont de deux ordres : soit elles résultent de traits caractéristiques qui distinguent les manuscrits de types, de régions ou d'époques différents, soit ces variations sont internes à un même document manuscrit et ne font qu'illustrer la tendance forte de la culture médiévale à toujours varier sur le même motif.

Par ailleurs, les annotations et les enrichissements linguistiques dont il a été question plus haut portent, pour un grand nombre d'entre eux (lemmatisation, étiquetage morphosyntaxique, annotation sémantique), sur l'unité linguistique fondamentale qu'est le mot. L'association de son lemme ou de sa catégorie morphosyntaxique à une unité lexicale présuppose que ses limites formelles aient été déterminées au préalable et qu'elles l'aient été de manière suffisamment cohérente et systématique pour permettre une bonne exploitation de cette information linguistique. La normalisation la plus systématique possible est par conséquent souhaitable pour faciliter l'annotation et l'interrogation de corpus. Elle l'est également pour améliorer la performance des outils de traitement automatique de la langue (TAL), tels que les étiqueteurs morphosyntaxiques et les lemmatiseurs.

Ces deux principes divergents, à savoir le respect du manuscrit dans sa variation interne et externe et la normalisation de cette variation en vue de recherches portant non pas sur des formes mais sur des catégories linguistiques, s'appliquent en réalité à différents niveaux du texte ou à différents usages de la représentation numérique : les outils de visualisation doivent permettre de lire le texte tel qu'il se présente dans ses réalisations médiévales effectives ; les outils de requête et d'annotation doivent permettre de regrouper, comparer et analyser le fonctionnement d'unités linguistiques en contexte¹⁸. Nous proposerons à la fin de cette contribution quelques solutions pratiques qui répondent à ces usages multiples.

18. Voir notamment Nicolas Maziotta, « Le texte dans tous ses états. Philosophie d'encodage du projet Khartès », dans Gérald Purnelle, Cédric Fairon et Anne Dister (dir.), *Le Poids des mots*, actes des 7^e Journées internationales d'analyse statistique des données textuelles, Louvain-la-Neuve, UCL/Presses universitaires de Louvain, 2004, t. II, p. 793-803 ; Alexei Lavrentiev, *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 soutenue sous la dir. de Christiane Marchello-Nizia, ENS Lyon, 2009], en ligne.

La segmentation du texte en mots : difficultés théoriques et pratiques

Les difficultés théoriques et pratiques engendrées par la segmentation du texte en unités-mots se manifestent du côté de la philologie traditionnelle tout autant que des ressources numériques.

Les pratiques de segmentation de la philologie traditionnelle

On constate que la question de la segmentation en mots est très peu traitée dans les manuels consacrés à l'édition de textes et qu'elle n'est souvent abordée que rapidement, voire pas du tout, dans les introductions linguistiques aux éditions papier¹⁹. Ainsi, le rapport de la commission des romanistes chargée de l'établissement des règles pratiques pour l'édition des anciens textes français et picards²⁰ n'en porte aucune mention. Tel est aussi le cas du *Guide de l'édition de textes en ancien français* paru trois quarts de siècle plus tard²¹. Alfred Foulet et Mary Blakely Speer²² consacrent quelques pages aux problèmes des mots composés et de la réduplication des consonnes finales ou initiales en cas de soudure graphique (*derrechief < de rechief, dessus < de sus*, etc.). Ils notent que la frontière entre une locution figée et un mot composé peut évoluer dans le temps, et prennent comme exemple l'expression *avoir a faire a quelqu'un*, devenue plus tard *avoir affaire à quelqu'un*. Ils ne donnent cependant pas de règles ou d'indices pour définir le moment où ce changement de segmentation a lieu. Ils insistent seulement sur la nécessité d'être attentif aux cas où le choix de la segmentation influe

19. La situation est différente dans le domaine de l'édition diplomatique des chartes où l'attention aux aspects graphiques est plus importante dès les années 1970 (voir *Documents linguistiques de la France*. éd. dir. Jean-Gabriel Gigot, Jacques Monfrin et Lucie Fossier, Série française I, *Chartes en langue française antérieures à 1271 conservées dans le département de la Haute-Marne*, Paris, Éd. du CNRS, 1974).

20. Mario Roques, « Établissement des règles pratiques pour l'édition des anciens textes français et provençaux », *Romania*, 52, 1926, p. 243-249.

21. Yvan G. Lepage, *Guide de l'édition de textes en ancien français*, Paris, Champion, 2001.

22. Alfred Foulet et Mary Blakely Speer, *On Editing Old French Texts*, Lawrence, Regents Press of Kansas, 1979, p. 60-62.

sur le sens de la locution. Ainsi, la préposition *devers* indique la direction d'un mouvement, tandis que la locution *de vers* indique la provenance. Dans les cas de « désambiguïsation par segmentation », ils n'hésitent pas à recommander de supprimer les consonnes rédupliquées, comme lorsque *dessus* a le sens de « de dessus » et doit être transcrit *de sus*. Il convient de noter que cette différence sémantique n'est pas toujours évidente, dans les cas où le verbe ne dénote pas un mouvement orienté. Les éditeurs semblent dans la pratique privilégier la graphie soudée (97 % et 99,5 % respectivement des occurrences de *avoir affaire* et *dessus* dans la Base de français médiéval). Dans l'un des rares textes où l'on trouve les deux graphies, les éditeurs appliquent probablement la règle de Foulet-Speer, même si on peut se demander si la distinction créée n'est pas artificielle, comme dans les deux exemples suivants :

(1) Vit Achelor a la fenestre, / jouste un pilier **de vers** senestre²³.

(2) Si com l'aventure ert a estre, / ambedui joustent **devers** destre²⁴;

Les articles de recherche sur les pratiques sribales de segmentation des mots²⁵ ne semblent pas avoir influencé considérablement les habitudes des philologues dans les années 1990 et 2000.

Les auteurs des *Conseils pour l'édition des textes médiévaux*²⁶ recommandent de souder les « expressions composées usuelles ou celles qui sont passées soudées dans la langue moderne », en tenant compte d'une part des graphies (par exemple, distinguer *desorenavant* de *des ore en avant*) et, d'autre part,

23. *Le Roman de Thèbes*, éd. Guy Raynaud de Lage, Paris, Champion, coll. « Classiques français du Moyen Âge », 1966, v. 3510.

24. *Ibid.*, v. 5716.

25. Peter Rickard, « Système ou arbitraire? Quelques réflexions sur la soudure des mots dans les manuscrits français du Moyen Âge », *Romania*, vol. 103, n°412, 1982, p. 470-512; Nelly Andrieux-Reix et Simone Monsonégo, « Écrire les phrases au Moyen Âge. Matériaux et premières réflexions pour une étude des segments graphiques observés dans des manuscrits français médiévaux », *Romania*, vol. 115, n°459-460, 1997, p. 289-336.

26. Françoise Viellard et Olivier Guyotjeannin, *Conseils pour l'édition des textes médiévaux* [2001-2002], 3 vol., nouv. éd. revue et mise à jour, Paris, École nationale des chartes/CTHS, 2014-2018.

de la « réalité institutionnelle » (comme dans *lieux tenens* pour « lieutenants du roi »). En cas de lettres redoublées, ils proposent de séparer les mots en conservant la double consonne (par exemple *a sses parenz*). Cette dernière recommandation s'appuie notamment sur la pratique de Philippe Ménard dans son édition du *Roman de Tristan en prose*²⁷, mais semble être très peu suivie dans l'ensemble.

On peut conclure que la règle plus ou moins implicite qui ressort de ces recommandations est d'adapter les pratiques de division aux normes du français moderne pour faciliter la lecture et la compréhension du texte par le lecteur contemporain. Mais sur ce point comme bien d'autres, une certaine latitude est laissée au philologue, dont les choix pourront varier en fonction du texte et du manuscrit qu'il édite, en fonction du public visé et, finalement, en fonction de sa libre appréciation personnelle. Les règles communes n'étant ni très stables ni très explicites, les éditeurs ne mentionnent que rarement les problèmes de segmentation dans leurs introductions linguistiques.

Parmi les rares exceptions, on peut citer Philippe Ménard, qui, dans l'introduction à son édition du *Roman de Tristan en prose*²⁸, explicite les principes qu'il a retenus pour la segmentation des mots et tient compte, dans certains cas, de l'usage du copiste. Plus récemment, Nelly Andrieux-Reix a consacré une section entière à la « syntaxe graphique » dans l'introduction à son édition du *Moniage Guillaume*²⁹ et a utilisé les traits d'union et les tirets longs pour représenter les usages du copiste (le trait d'union pour les agglutinations et le tiret ordinaire long pour les déglutinations). Cette pratique n'a pas été suivie par les autres philologues, à l'exception de Frédéric Duval, qui, dans son édition du *Dit de la fleur de lis* de Guillaume

27. *Le Roman de Tristan en prose*, éd. dir. Philippe Ménard, t. I, *Des aventures de Lancelot à la fin de la « Folie Tristan »*, Genève, Droz, coll. « Textes littéraires français », 1987.

28. *Ibid.*, p. 53-54.

29. *Le Moniage Guillaume, chanson de geste du XI^e siècle*, éd. de la rédaction longue par Nelly Andrieux-Reix, Paris, Champion, coll. « Classiques français du Moyen Âge », 2003.

de Digulleville³⁰, distingue l'édition critique de l'édition des témoins, dans laquelle il utilise le tiret bas pour les agglutinations et le tiret pour les déglutinations. Par ailleurs, dans son édition commentée d'un extrait du *Pèlerinage de l'âme* (toujours de Guillaume de Digulleville)³¹, le même Frédéric Duval consacre toute une section de son introduction aux problèmes liés à la séparation des mots. Il y explique les pratiques du copiste et ses propres choix d'édition. Enfin, les problèmes de segmentation lexicale occupent une part importante de l'introduction apportée à l'édition numérique de la *Queste del saint Graal*³². L'expérience de cette édition a largement inspiré la solution philologique présentée ci-dessous.

Ces quelques exceptions mises à part, il est assez difficile de déterminer quels sont les motifs qui ont conduit l'éditeur à faire ses choix dans le texte qu'il édite. Cette situation engendre une grande diversité de pratiques, en particulier pour les expressions en cours de figement ou en voie d'autonomisation linguistique dont il a été question plus haut. La soudure, non signifiante, des mots grammaticaux (pronoms, articles et prépositions, etc.) n'est à peu près jamais représentée dans les éditions de textes du Moyen Âge, et, pour le reste, les usages ne sont pas très stables (voir la section suivante). La seconde conséquence de cette situation est qu'elle ne permet pas d'apprécier si les segmentations choisies par l'éditeur résultent d'une volonté de normalisation du texte ou du respect des solutions graphiques du manuscrit de base. On note, toutefois, que l'une des tendances les plus répandues est de choisir une segmentation homogène dans un texte donné (l'éditeur écrira de manière systématique *par mi* ou *parmi*), ce qui élimine la variation graphique interne

30. Guillaume de Digulleville, *Le Dit de la fleur de lis*, éd. Frédéric Duval, Paris, École nationale des chartes, coll. « Mémoires et documents de l'École des chartes », 2014.

31. Frédéric Duval, *Descente aux enfers avec Guillaume de Digulleville. Édition et traduction commentées d'un extrait du Pèlerinage de l'âme* (Paris, BnF, fr. 12466), Saint-Lô, Archives départementales de la Manche, 2006.

32. Christiane Marchello-Nizia, Alexei Lavrentiev et Céline Guillot, « Édition électronique de la *Queste del saint Graal* », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, Berlin/Boston, De Gruyter, 2015, p. 155-176.

à chaque manuscrit, permet une certaine harmonisation du texte et en facilite l'appropriation par le lecteur d'aujourd'hui.

Les pratiques de segmentation des textes numériques

Les corpus textuels étant généralement composés d'éditions papier numérisées, ils héritent des pratiques divergentes et souvent peu documentées des éditeurs de textes.

À titre d'exemple, nous nous sommes intéressés à quatre locutions dont la segmentation a évolué dans l'histoire du français : la préposition *par* suivie du substantif *mi* qui a abouti à la préposition *parmi*, le groupe d'adverbes *ja mais* qui a formé *jamais*, le préverbe *en* suivi du verbe *porter* qui a produit, dans certaines constructions, le verbe *emporter*, et enfin le groupe *à + venir* qu'il faut relier au verbe *advenir* et au substantif *avenir* et qui continue d'exister en tant que locution adjectivale (au sens de « futur »).

Nous avons recensé les occurrences des différentes formes graphiques (séparées et soudées) de ces locutions dans le corpus BFM2016 de la Base de français médiéval. Le moteur de recherche CQP³³ intégré à la plateforme TXM donnant accès aux textes permet en effet de sélectionner plusieurs variantes graphiques dans une seule expression de requête CQL³⁴. L'outil Progression (actuellement disponible dans la version bureau de TXM seulement) sert à visualiser la densité d'apparition d'un motif recherché au sein du corpus, tandis que l'outil Concordance présente l'ensemble des occurrences dans leur contexte en vue d'une analyse plus fine³⁵.

En ce qui concerne la locution *par + mi*, le graphique de progression (fig. 1) montre que les deux graphies, soudée (correspondant à la courbe cumulative des occurrences de l'expression de requête "`parm[iy]"%c`³⁶) et séparée (corres-

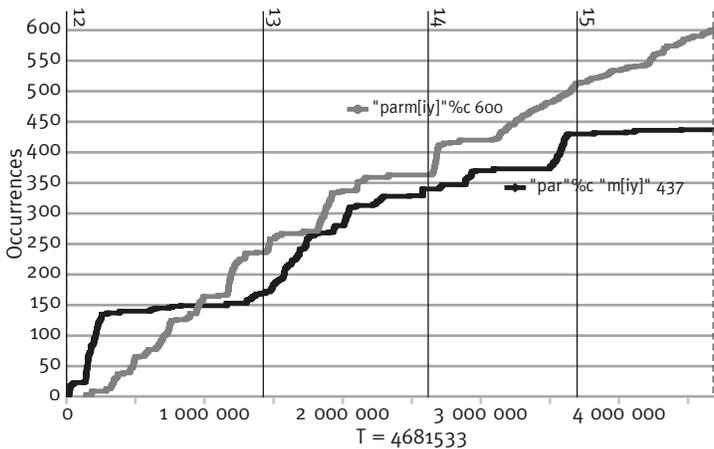
33. <http://cwb.sourceforge.net>.

34. http://cwb.sourceforge.net/files/CQP_Tutorial.

35. Nous remercions Bénédicte Pincemin (IHRIM), qui nous a apporté son aide pour l'interprétation des graphiques qui suivent.

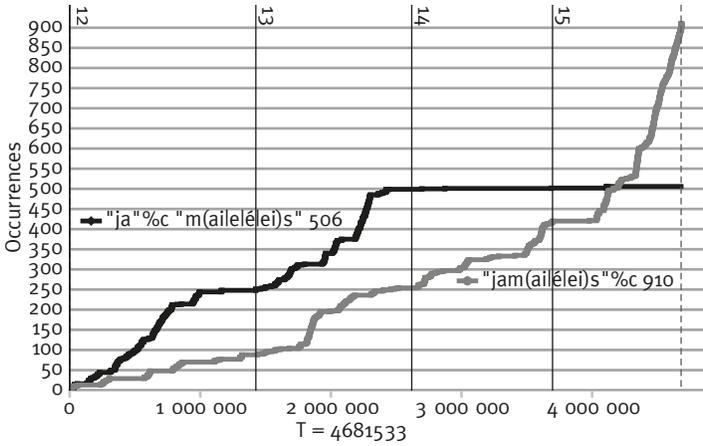
36. Cette requête permet de retrouver les formes qui commencent par *par* et finissent par *y*. L'opérateur « %c » permet d'ignorer la casse.

pondant à la courbe de l'expression de requête "par"%c "m[iy]"), sont utilisées avec une fréquence comparable dans les éditions de textes composés entre la fin du XII^e et la fin du XIV^e siècle (leurs pentes sont comparables, sachant que les apparitions des locutions au sein des textes sont représentées par des marches dans la courbe). La graphie soudée n'est pas utilisée dans les éditions des textes les plus anciens (avant 1150) et la graphie séparée disparaît presque entièrement dans les éditions des textes du XV^e siècle. Les courbes comportent cependant de nombreuses alternances de montées/plateaux, qui semblent manifester l'impact fort, sinon déterminant des choix éditoriaux (il y a les textes/éditions qui suivent telle segmentation, et les textes/éditions qui suivent telle autre).



1. Graphique de progression des occurrences des graphies soudées et séparées de *parmi/par mi* dans le corpus BFM2016.

Dans le cas de *ja + mais* (fig. 2), la graphie séparée (correspondant à l'expression de requête "ja"%c "m(ai|e|é|ei)s") est préférée par les éditeurs des textes du XII^e siècle. Pour les textes du XIII^e siècle, l'usage varie selon les éditeurs, et la graphie soudée (correspondant à l'expression de requête "jam(ai|é|ei)s"%c) devient pratiquement exclusive pour les éditions des textes composés à partir de la fin du siècle.

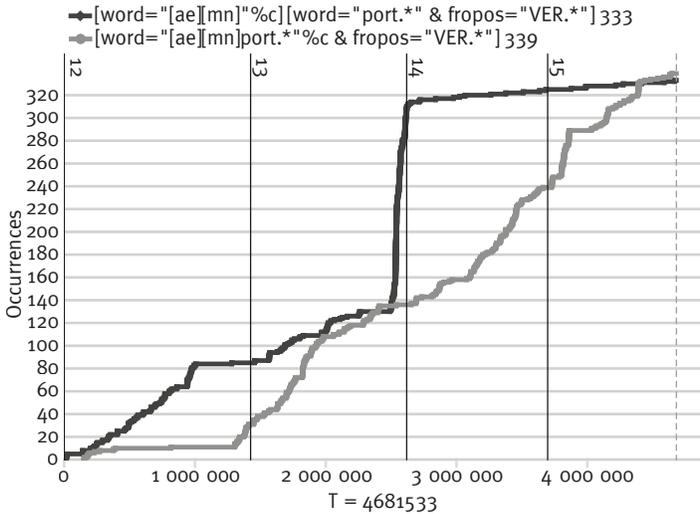


2. Graphique de progression des occurrences des graphies soudées et séparées de *jamais*/*ja mais* dans le corpus BFM2016.

En ce qui concerne le verbe *en porter/emporter*, le graphique de progression (fig. 3) est particulièrement intéressant, car on observe un « saut » énorme de la graphie séparée (correspondant à l'expression de requête [word="[ae][mn]"%c [word="port.*" & fropos="VER.*"]]) à la fin du XIII^e siècle, tandis que la graphie soudée (correspondant à l'expression de requête [word="[ae][mn]port.*"%c & fropos="VER.*"]) progresse d'une manière à peu près constante à partir de la fin du XII^e siècle. Le « saut » de la fin du XIII^e siècle s'explique par la fréquence exceptionnellement élevée de ce verbe dans les *Coutumes de Beauvaisis* de Philippe de Beaumanoir (181 occurrences sur 672 dans toute la BFM2016, toutes graphies confondues), toujours graphié séparément par Amédée Salmon³⁷. Notons qu'une grande partie des rares occurrences de la graphie séparée attestées dans les éditions de textes à partir du XIV^e siècle provient des constructions où *en* joue clairement le rôle d'un pronom et où la graphie soudée ne serait pas correcte du point de vue de l'orthographe moderne :

37. À l'exception d'une occurrence en bas de la p. 238, où il peut s'agir d'une coquille.

(3) Vous estes chevalier? dist le Jouvenel, car vous **en portés** les enseignes, ce me semble³⁸.



3. Graphique de progression des occurrences des graphies soudées et séparées de *emporter/en porter* dans le corpus BFM2016.

Le cas de la locution à *venir/a(d)venir* est plus complexe (correspondant respectivement aux expressions de requête "ad?venir"%c | "[aà]d?"%c "venir"). Le graphique de progression (fig. 4) montre que les deux formes sont utilisées tout au long de la période médiévale, avec une prépondérance constante de la graphie soudée (sauf dans les plus anciens textes du XI^e siècle). Cette persistance de la double segmentation s'explique certainement par la diversité des fonctions de cette locution, qui a abouti à trois formes distinctes dans l'orthographe moderne : à *venir*, *avenir* et *advenir*. Dans les usages au sens de « survenir », la graphie soudée semble la seule possible dès les premières attestations :

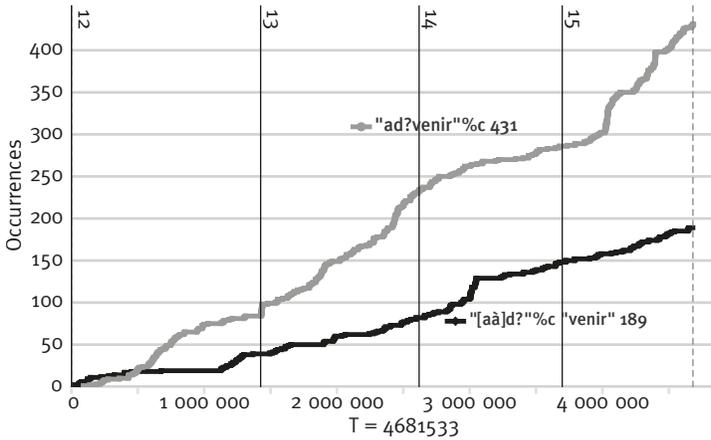
38. Jean de Bueil, *Le Jouvenel*, éd. Léon Lecestre, comm. Guillaume Tringant, Paris, H. Laurens, t. I, 1887, p. 99.

(4) Si m'ait Deus, qui ne menti, / jeo nel lerroie pur murir, / que
jeo ne l'auge ja ferir, / que ke m'en deie **avenir**³⁹.

Dans les usages où la préposition *à* est régime d'un verbe conjugué, la graphie séparée s'impose à son tour :

(5) Ci comencent angoisses dolentes **a venir**⁴⁰.

En revanche, dans les constructions *estre a venir* et *temps a venir*, on trouve aussi bien les graphies soudées que séparées.



4. Graphique de progression des occurrences des graphies soudées et séparées de *avenir/a venir* dans le corpus BFM2016.

Comme on l'a indiqué plus haut, les éditeurs sont généralement constants dans le choix de la graphie au sein d'un texte donné, même si des exceptions sont possibles.

Toutes ces observations nous permettent de conclure qu'il y a une tendance générale chez les éditeurs à préférer les graphies segmentées pour les textes les plus anciens (avant la fin du XII^e siècle) et que les graphies soudées dominent largement à partir du XIV^e ou du XV^e siècle, selon les cas. Pour les textes composés au XIII^e siècle, les pratiques des éditeurs sont divergentes.

39. Gormont et Isembart. *Fragment de chanson de geste du XI^e siècle* [ca 1130], éd. Alphonse Bayot, Paris, Champion, coll. « Classiques français du Moyen Âge », 3^e éd., 1931, v. 211.

40. *Li ver del juïse. Sermon en vers du XI^e siècle* [2^e quart du XII^e siècle], éd. Erik Rankka, Stockholm, Almqvist och Wiksell, 1982, v. 276.

L'hésitation dans les choix de segmentation se répercute également dans les dictionnaires. Dans le *Dictionnaire de moyen français* par exemple, on trouve des entrées pour les mots soudés, mais aussi des exemples sous les entrées du dictionnaire correspondant à l'une des parties de la locution. Ainsi des exemples de la forme *advenir* (au sens adjectival de « à venir ») peuvent être retrouvés sous les entrées « advenir » (III.A), « avenir » (I) et « venir » (II.A.4.a), et il n'y a pas de renvois systématiques d'un lemme à l'autre.

La segmentation du texte en mots : méthode de la philologie numérique

À ces variations du niveau graphique répond la tendance à l'harmonisation systématique des traitements numériques. La nécessité de normaliser les différences formelles à l'intérieur d'un même corpus textuel transpose le problème du niveau scribal au niveau éditorial : ce ne sont plus les variations constantes des copistes médiévaux qu'il s'agit de dépasser, mais les pratiques hétérogènes des éditeurs de textes, eux-mêmes confrontés à une documentation écrite aux règles instables et en constante évolution.

D'un autre côté, si les pratiques des éditeurs ne fournissent généralement pas de données fiables pour l'étude de la segmentation graphique dans les manuscrits médiévaux, elles constituent en même temps une donnée non négligeable de l'histoire de la pensée philologique. À ce titre, elles méritent d'être observées et conservées. Pour dépasser ces points de vue apparemment antagonistes, on peut souligner que la normalisation de la représentation numérique n'empêche pas l'encodage supplémentaire des pratiques des scribes et des éditeurs scientifiques. Le défi de la philologie numérique consiste ainsi à proposer des solutions d'encodage qui permettent de produire des éditions et des corpus numériques à la fois normalisés, faciles à utiliser et conservant les données primaires (pratiques des scribes) et secondaires (pratiques des éditeurs) potentiellement utiles.

Pour que la machine soit en mesure de relever ce défi, il faut modéliser informatiquement la segmentation des mots. Cette méthode repose sur trois opérations distinctes :

1. la représentation philologique numérique explicite de la séparation ou de la soudure lexicale ;
2. l'interprétation correcte de cette représentation par les moteurs de recherche ;
3. la possibilité de générer les visualisations de niveaux linguistiques nécessaires à partir de cette représentation.

Il est nécessaire de distinguer le versant philologique de la méthode qui consiste à définir les principes de normalisation et de restitution des données primaires ou secondaires non normalisées, et la solution numérique de la méthode qui permet d'implémenter les choix philologiques aux niveaux de l'encodage, de l'interrogation et de la visualisation des ressources.

Versant philologique de la méthode

Comme nous l'avons vu plus haut, les problèmes de segmentation lexicale ont commencé à attirer l'attention de certains éditeurs scientifiques à la fin des années 1980 et surtout au cours des années 2000. L'édition numérique de la *Queste del saint Graal*⁴¹, dont les débuts remontent à 1999, a été l'occasion de réviser la méthodologie traditionnelle d'établissement du texte afin de profiter des avantages offerts par le numérique et de créer une ressource d'un nouveau type, prototype pour de futurs projets éditoriaux.

Cette édition électronique présente la particularité d'offrir à l'utilisateur différents niveaux de transcription et de lecture du texte : (i) un « niveau normalisé », qui facilite la lecture en adoptant un mode de présentation plus conforme aux usages modernes ; (ii) un « niveau diplomatique », plus respectueux de la source primaire et signalant les interprétations de l'éditeur ; (iii) un « niveau fac-similaire », qui reproduit de manière

41. Éd. Christiane Marchello-Nizia et Alexis Lavrentiev, Lyon, ENS Lyon, en ligne : http://catalog.bfm-corpus/org/qgraal_cm.

aussi fidèle que possible le document médiéval dans ses particularités graphiques et matérielles⁴². Ces trois niveaux de lecture (ou « facettes ») sont bien distincts du niveau de base (ou « de référence ») correspondant aux unités linguistiques retenues pour les requêtes de recherches et de décomptes et l'étiquetage morphosyntaxique du texte. Et l'interface de la plateforme TXM permet de coupler l'affichage simultané d'un ou de plusieurs niveaux avec le moteur de recherche CQP⁴³.

Le traitement des locutions en cours de figement a fait l'objet d'une attention particulière dans cette édition pilote. L'introduction⁴⁴ explicite les règles de transcription retenues, mais elle présente aussi un tableau exhaustif des locutions en cours de figement avec leur traitement par le scribe (graphie soudée, séparée par un espace ou par un passage à la ligne) et les graphies retenues (indexées) pour les requêtes et l'étiquetage morphosyntaxique. Les graphies du scribe sont toujours préservées à l'affichage pour ces locutions, y compris dans la version normalisée du texte. L'expression de requête pour la forme est choisie au cas par cas et correspond le plus souvent à la graphie majoritaire du scribe. C'est par exemple le cas de la locution conjonctive *puis que* écrite avec une espace dans 74 occurrences sur 84, ou de l'adverbe *jamais* avec 78 graphies soudées sur 83. En revanche, la forme soudée a été retenue pour les requêtes portant sur la préposition *jusque*, bien que le manuscrit présente la graphie séparée *jus que* dans 72 occurrences sur 85. Dans le cas de l'adverbe ou locution adverbiale *a-tant*, la variation graphique du manuscrit (25 occurrences soudées contre 19 séparées) a été préservée dans les formes indexées pour les requêtes. Ces choix, parfois hésitants, témoignent de l'évolution de la réflexion méthodologique au cours du projet éditorial.

42. Christiane Marchello-Nizia, Alexei Lavrentiev et Céline Guillot, « Édition électronique de la *Queste del saint Graal* », art. cit.

43. Éd. cit. (voir *supra*, n. 40).

44. Base de français médiéval, introduction, p. 32-36.

Les douze premiers paragraphes (soit près de deux feuillets) du *Graal* ont été transcrits au niveau « fac-similaire » (ou « allographétique ») en plus des deux niveaux, « normalisé » et « diplomatique », disponibles pour l'ensemble du texte. Le niveau facsimilaire tient compte des variantes de lettres (ou « allographes », comme les *s* « rond » et « long ») et des marques d'abréviation, mais aussi des segmentations particulières (agglutinations et déglutinations), qu'elles soient ou non en cours de figement. On peut par exemple observer des graphies soudées *fuenu* (pour « fu venu ») et *figrant* (pour « si grant ») à la ligne 2 de la colonne 160a du manuscrit K. Le statut linguistique de ces agglutinations étant *a priori* différent de celui des locutions en cours de figement (elles ne sont pas liées à des phénomènes de grammaticalisation ou de lexicalisation), il est important de pouvoir distinguer les deux types d'agglutination tant au niveau de la visualisation qu'au niveau des requêtes. Les agglutinations sans figement ne sont jamais prises en compte dans l'indexation des mots du texte, et ne sont pas visibles aux niveaux normalisé et diplomatique de la transcription. En revanche, on peut rechercher les occurrences correspondantes grâce à une annotation spéciale (expression CQL [`rend="agg1"`]) et les visualiser dans la transcription fac-similaire.

Les sources XML-TEI de l'édition électronique de la *Queste del saint Graal* ont été distribuées et utilisées dans divers projets de recherche. Elles ont surtout été enrichies d'un niveau d'annotation supplémentaire dans le cadre du projet franco-allemand « Syntactic Reference Corpus for Medieval French »⁴⁵. L'annotation syntaxique basée sur les unités lexicales a permis de valider, et dans certains cas de réviser, la segmentation des unités d'après l'analyse syntaxique du texte. Et l'annotation parallèle de plusieurs autres textes médiévaux a permis de confronter nos choix à un corpus diachronique d'œuvres

45. Achim Stein et Sophie Prévost, « Syntactic annotation of medieval texts: the Syntactic Reference Corpus of Medieval French (SRCMF) », dans Paul Bennett, Martin Durrell, Silke Scheible et Richard J. Whitt (dir.), *New Methods in Historical Corpora*, Tübingen, Narr, coll. « Corpus Linguistics and International Perspectives on Language (CLIP) », 2013, t. III, p. 275-282; voir, en ligne : <http://srcmf.org>.

composées entre le ix^e et la fin du xiii^e siècle. Des principes de segmentation plus simples, constants et généraux ont été ainsi définis. Leur atout principal est d'être applicables à n'importe quel texte de français médiéval.

Ces principes reposent sur une règle fondamentale très simple : l'indexation des unités lexicales (ou *tokenisation*) se base systématiquement sur la segmentation maximale (la plus fine possible). Des solutions informatiques sont proposées pour (i) afficher la segmentation du manuscrit de base et/ou la segmentation décidée par l'éditeur scientifique, et (ii) faciliter les requêtes sur les locutions en cours de figement, qu'elles soient graphiées avec ou sans l'espace blanc.

Dans la pratique, l'application de ce principe amène parfois à des coupures contre-intuitives, comme dans le cas de *bon heur*, mais elle présente l'avantage d'offrir davantage de souplesse pour l'indexation et l'annotation du texte, car il est toujours plus facile de regrouper des éléments annotés que de créer des annotations sur des parties d'éléments indivisibles. La seule exception à cette « atomisation » de la segmentation concerne les cas où la fonction de l'un au moins des éléments séparables est très difficile à définir en contexte et à catégoriser à l'aide d'une étiquette morphosyntaxique. C'est par exemple le cas de *que final* dans *quelque* (même si des graphies séparées sont attestées dans les manuscrits).

Ce principe de segmentation est appliqué systématiquement dans les nouvelles éditions de la collection « Sources médiévales » mise en place à l'École normale supérieure de Lyon et liée à la Base de français médiéval. Progressivement, le *Graal* et l'ensemble des textes de la base seront ré-indexés en conformité avec ces nouvelles normes.

Versant numérique de la méthode

La méthode informatique utilisée pour traiter la segmentation lexicale comporte trois aspects :

1. l'encodage des données, qui établit une relation contractuelle avec les outils à travers un format associé à des conventions d'interprétation ;
2. l'ergonomie de la saisie, de la vérification et de l'enrichissement des données ;
3. l'ergonomie de l'exploitation des données (recherche, lecture et exportation des résultats et ré-annotation).

Les technologies d'encodage de données textuelles et paratextuelles sont extrêmement diverses et ont tendance à évoluer rapidement. Afin d'assurer une certaine pérennité à la solution adoptée, il est indispensable d'utiliser des standards internationaux soutenus par des communautés scientifiques importantes et ayant fait la preuve de leur capacité d'adaptation aux changements technologiques. Dans le domaine de l'édition numérique académique, les recommandations du consortium *Text Encoding Initiative*⁴⁶ répondent parfaitement à ces exigences. Mais dans le domaine de la linguistique de corpus et du traitement automatique du langage naturel (TAL), la TEI est loin de s'imposer. Néanmoins, le cadre proposé par la TEI (un jeu de près de 700 balises au format XML⁴⁷, bien documenté et extensible) constitue certainement l'une des meilleures options lorsqu'on veut combiner la richesse du balisage philologique avec la puissance des outils de TAL et d'analyse de corpus. C'est pourquoi la Base de français médiéval et la plateforme TXM ont choisi une représentation XML-TEI des textes numériques comme base commune d'interopérabilité.

L'élément `<w>` de la TEI⁴⁸ sert à encoder un mot ou une unité de segmentation lexicale (*token*). Cette unité peut porter de multiples annotations, dont peut faire partie la segmentation graphique primaire ou secondaire. La TEI ne spécifie pas quelle forme doivent prendre ces types particuliers d'annotation, mais une convention plus précise peut être adoptée dans le cadre d'un

46. TEI, <http://www.tei-c.org>.

47. <https://www.w3.org/TR/2006/REC-xml11-20060816>.

48. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-w.html>. Les références de toutes les autres balises de la TEI peuvent être consultées de manière analogue.

projet ou d'une communauté spécifique. Dans le cadre de la Base de français médiéval et de la collection d'éditions numériques associée, on a fait le choix d'indiquer l'absence « anormale » d'un espace après un *token*. La locution prépositionnelle *par + mi* soudée graphiquement dans la source sera alors encodée comme ceci :

```
<seg type="fgmt">
  <w rend="aggl">par</par>
  <w>mi</w>
</seg>
```

L'élément `<seg>` qui englobe l'ensemble est utilisé pour distinguer les agglutinations dans les locutions en cours de figement (signalées par la valeur « fgmt » de son attribut *type*) des autres cas (moins réguliers et *a priori* moins intéressants du point de vue linguistique). Dans les rares cas où il est nécessaire de noter la présence d'une espace à l'intérieur d'un seul *token*, nous avons fait le choix d'utiliser la balise `<space/>` :

```
<w>quel<space quantity="1" unit="chars"/>que</w>
```

Il s'agit en effet d'un cas exceptionnel, puisque le principe général consiste à segmenter le plus possible.

Bien entendu, les balises `<w>` et leurs annotations n'ont pas vocation à être encodées « à la main » par l'éditeur du texte. La « tokenisation » est réalisée par des outils de TAL plus ou moins sophistiqués et une méthode ergonomique peut être mise en place pour pré-annoter les cas particuliers, comme les agglutinations, et pour contrôler le résultat de l'application de ces outils⁴⁹. Dans la collection « Sources médiévales », par exemple, le caractère de pré-annotation « + » sera utilisé après le premier élément d'une agglutination simple (*des+ choses* : « des choses »), et un double plus « ++ » marque les agglutinations dans les locutions en cours de figement (*par++ mi*). Le tiret bas est utilisé dans les cas de déglutinations (*a_vons* : « avons »). Il est dupliqué dans les mots composés (*quel__que*). Dans certains

49. Voir Nicolas Maziotta, « Le texte dans tous ses états. Philosophie d'encodage du projet Khartès », art. cit. ; Alexei Lavrentiev, *Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIII^e au XV^e siècle*, thèse cit.

cas, la lecture du manuscrit ne permet pas de déterminer avec certitude si un espace blanc est présent entre deux mots ou entre deux parties d'un même mot. Il est alors possible d'encoder cette incertitude en ajoutant un point d'interrogation à la suite du caractère « + » ou « _ ».

Les caractères de pré-annotation sont interprétés par le logiciel de segmentation lexicale (ou « tokeniseur ») qui produit les balises TEI mentionnées plus haut. L'ensemble des règles de saisie de transcription des sources est présenté dans un document en ligne, constamment mis à jour⁵⁰.

Le balisage explicite des mots du corpus avec l'annotation des segmentations particulières permet d'obtenir tous les affichages souhaités pour la lecture du texte : on peut, au choix, imiter la segmentation graphique de la source ou, au contraire, « normaliser » tout ou partie des segmentations particulières. Dans le cas de la collection « Sources médiévales », le choix a été fait de conserver la segmentation graphique des manuscrits pour les locutions en cours de figement, mais il est relativement facile de revenir sur ce choix au besoin.

Lors de l'interrogation ou de l'analyse textométrique du corpus, la présence du balisage permet de retrouver assez facilement les occurrences concernées et d'ajuster éventuellement les calculs de fréquences⁵¹. À titre d'exemple, nous présentons ci-dessous quelques requêtes CQL exploitables sur un corpus balisé selon les exemples cités plus haut, après importation dans la plateforme TXM :

`[rend="aggl"]+ [rend!="aggl"]` : cette expression de requête permet de sélectionner toutes les séquences agglutinées (agglutinations simples ou locutions en cours de figement)

`[_ .seg_type="fgmt"] expand to seg` : cette expression cible toutes les locutions en cours de figement.

50. <http://goo.gl/EWZ7NC>.

51. Bénédicte Pincemin, « Lexicométrie sur corpus étiquetés », dans Gérald Purnelle, Cédric Fairon et Anne Dister (dir.), *Le Poids des mots, op. cit.*, t. II, p. 865-873.

Le cas précis de la segmentation en mots dans les éditions de textes, qu'elles se présentent sous forme papier ou électronique, a permis d'aborder de manière concrète quelques enjeux majeurs de l'essor du numérique dans le domaine de l'édition de textes et de ce qu'on peut désormais appeler la « philologie numérique ». Nous espérons avoir montré à partir de cet exemple particulier que le numérique offrait la possibilité d'inventer des manières de rendre compte, mieux que ne le faisaient les éditions traditionnelles ou de façon plus utile aux linguistes, de la réalité matérielle des manuscrits médiévaux. Ces technologies nouvelles offrent parallèlement des outils de recherche et d'analyse souples et puissants permettant de regrouper, de manier et d'analyser un volume de données très important tout en maîtrisant leur diversité.

La réflexion qui sous-tend les pratiques numériques émergentes vise à faire des choix de représentation du texte qui soient cohérents, consistants, documentés et aussi adaptés que possible aux objectifs scientifiques que l'on se donne. Ces choix incombent aux linguistes et aux philologues qui produisent et exploitent les ressources numériques. Ils permettent de renouer un dialogue actif entre spécialistes de disciplines récemment séparées. Ils sont également l'occasion de s'ouvrir aux sciences de l'information, et plus spécifiquement à l'informatique.

Les possibilités très riches offertes par le numérique ne doivent pas occulter le coût qui les accompagne. L'harmonisation et le partage des choix philologiques, linguistiques et techniques deviennent de ce fait une condition majeure des développements actuels et à venir. Les standards et les normes partagés, qui concernent les choix d'édition tout autant que les formats d'encodage, garantissent l'échange, l'enrichissement progressif, la diffusion et la pérennisation des ressources. Les licences libres qui sont utilisées de plus en plus largement, les licences *Creative Commons* en particulier⁵², contribuent à la même démarche participative et communautaire. Ce mouvement

52. <http://creativecommons.org>.

très vaste rend la collaboration interdisciplinaire d'autant plus nécessaire pour la survie des disciplines elles-mêmes. Nous pensons qu'il est aussi et surtout un facteur de progrès et d'enrichissement réciproque.

L'activité éditoriale comme démarche de représentation de la connaissance : l'exemple de la ponctuation médiévale¹

Nicolas Mazziotta

Université de Liège

Cette contribution s'inscrit dans une approche personnelle de la philologie en tant qu'activité de représentation de connaissances, que j'ai exposée dans un mémoire d'habilitation à diriger des recherches². Je parlerai principalement de mes propres travaux sur la ponctuation médiévale telle qu'elle se présente dans les chartes du XIII^e siècle et des questions que je me suis posées en réalisant ces études, puis en y jetant un regard rétrospectif.

Il faut constater que la ponctuation médiévale est un objet dont il est difficile de rendre compte au travers des signes modernes dans une édition de texte, en particulier sur un support traditionnel³. Elena Llamas-Pombo souligne les difficultés à « donner à lire » la ponctuation médiévale dans l'extrait suivant, qui introduit sa synthèse bibliographique des études sur cet objet :

-
1. Je tiens à remercier Alexei Lavrentiev et Elena Llamas-Pombo pour leurs remarques et suggestions stimulantes.
 2. Nicolas Mazziotta, *Représenter la connaissance en linguistique. Observations sur l'édition de matériaux et sur l'analyse syntaxique* [mémoire de synthèse, en vue de l'obtention de l'habilitation à diriger des recherches, en sciences du langage/linguistique générale et française, soutenu à l'Université Paris Ouest Nanterre-La Défense, 2016].
 3. On se reportera à Alexei Lavrentiev, « Traitement de la ponctuation dans les éditions de textes en français médiéval : normes, introductions, pratiques », dans Frédéric Duval, Céline Guillot-Barbance et Fabio Zinelli (dir.), *Les Introductions linguistiques aux éditions de textes*, Paris, Classiques Garnier, coll. « Histoire et évolution du français », 2019, pour une revue des pratiques.

[L]’édition ne peut que difficilement refléter la totalité des usages graphiques d’un manuscrit, à moins de devenir une pure transcription paléographique. Les règles instaurées pour l’édition des textes médiévaux français ne permettent pas d’en reconstituer la ponctuation originale, sauf pour ce qui est des paragraphes. De même, les rubriques et les divisions du texte ont beau être significatives, elles se prêtent parfois mal à l’édition si on les sait établies par une main autre que celle de l’auteur premier. C’est ainsi que demeurent exceptionnelles les éditions qui, cherchant à reproduire la ponctuation des manuscrits médiévaux, proposent un troisième type de présentation graphique, intermédiaire entre la transcription diplomatique et l’édition interprétative⁴.

Je parlerai en particulier des éditions plutôt imitatives (« diplomatiques »), en plaidant que la représentation des faits de ponctuation médiévale nécessite une démarche critique. J’accorderai donc beaucoup de valeur aux cas irrésolus et problématiques, qui sont révélateurs de la complexité de cette démarche.

Dans un premier temps, je présenterai les concepts de *connaissance* et d’*inscription* (1). Cette dernière notion sera utile par la suite. J’envisagerai ensuite trois aspects de ce rapport entre la connaissance et ses inscriptions en ecdotique au travers de quatre cas qui se répartissent en deux catégories : ceux qui ont trait à la compréhension de la ponctuation du manuscrit (2) et ceux qui ont trait à la question de la présentation de l’édition (3).

1. La connaissance et ses inscriptions

J’envisage ici les notions de *connaissance* et d’*inscription* de la connaissance en suivant les conceptions développées par Bruno Bachimont⁵, qui en a produit un cadre théorique général. Ces deux notions sont ici vues au travers de l’activité éditoriale. Cette dernière consiste à lire des textes, à les comprendre et à

4. Elena Llamas Pombo, « Ponctuer, éditer, lire. État des études sur la ponctuation dans le livre manuscrit », *Syntagma. Revista del Instituto de Historia del Libro y de la Lectura*, 2, 2008, p. 131-132.

5. Bruno Bachimont, *Ingénierie des connaissances et des contenus. Le numérique entre ontologies et documents*, Paris, Hermes science publications/Lavoisier, 2007.

produire un nouveau document, appelé *édition*, qui permet à d'autres de comprendre et d'exploiter à leur tour les matériaux. La question du statut épistémologique de ces éditions est le plus souvent posée en termes de pratique philologique, avec l'objectif de la rendre à la fois fiable (par rapport à un original) et lisible. Je voudrais aborder ici ce genre de question, mais sous un angle particulier.

Ce qui m'intéresse principalement, c'est que l'éditeur lit le texte transmis par les manuscrits, le comprend et choisit ce qu'il va sélectionner dans ce qu'il a compris pour ensuite le représenter dans son édition. Sa compréhension du texte est une *connaissance* personnelle qu'il a construite. Mais cette connaissance est intériorisée. Pour qu'elle soit accessible à autrui, il faut lui donner corps sur un support choisi (principalement le papier ou un fichier informatique), la transformer en un nouveau phénomène : une *inscription* de cette connaissance. Inversement, l'inscription est le seul moyen d'accéder à la connaissance au travers de l'interprétation que le lecteur en fait. Connaissance et inscription sont conceptuellement interdépendantes. En conséquence, la connaissance pure n'est pas accessible : toutes ses manifestations sont des inscriptions, qui se plient inmanquablement aux règles du support sur lequel elles se fondent (papier, fichier informatique, etc.). Je parlerai ainsi d'*inscriptions informatiques*, pour faire référence spécifiquement à l'encodage d'éditions électroniques dans des formats informatisés. J'oppose à ces inscriptions les *inscriptions traditionnelles*, c'est-à-dire les éditions « papier », ainsi que l'affichage, après transformation, d'inscriptions informatiques sous une forme statique destinée à la lecture. Les réflexions qui suivent concernent principalement les inscriptions traditionnelles, qui seront occasionnellement comparées aux inscriptions informatiques.

2. Comprendre

Lire des manuscrits médiévaux occidentaux demande un apprentissage qui vient se greffer à notre connaissance

générale de ce qu'est l'écriture occidentale (la démarche de tous les manuels de paléographie va dans ce sens). Étant enfants, nous avons appris à reconnaître le tracé des lettres et à interpréter l'agencement linéaire et les espaces entre ces dernières pour repérer les « mots ». Nous avons constitué ce que l'on peut nommer, d'un point de vue sémiotique et cognitif, une *encyclopédie visuelle*⁶. Cette encyclopédie est l'outil cognitif qui nous permet de reconnaître des occurrences particulières (*tokens*) de formes d'expression visuelles et de les rapporter à des classes abstraites (*types*). C'est parmi ces formes d'expression visuelles, que certains nomment *entités visuelles*⁷, qu'on retrouve les mots de la langue écrite.

Comprendre la lettre d'un manuscrit, c'est en reconnaître les entités constitutives. En somme, nous associons de nouvelles réalisations à des classes abstraites connues. Il est *a priori* assez simple reconnaître les lettres et les signes de ponctuation médiévaux. La plupart de ces entités ne sont pas très différentes de celles que nous avons appris à tracer sur les bancs de l'école : nous partageons avec les scribes médiévaux une partie commune de nos encyclopédies visuelles respectives. Par exemple, nous reconnaissons d'autres formes d'expression qui peuvent être rapportées à celles que nous connaissons déjà : ainsi tel tracé sera vu comme une réalisation de la classe <e>, tel autre comme celui d'un <A>. L'apprentissage de la paléographie n'est finalement rien d'autre qu'un enrichissement de notre encyclopédie ; par exemple il faut ajouter la classe <f> à notre encyclopédie, comme une alternative de <s> ; telle configuration est l'abréviation d'une autre configuration ; etc. La transcription telle que nous la pratiquons consiste ainsi à réduire les occurrences observées à une transcription qui rend compte de notre classification. Cela est inévitable, car l'encodage typographique impose une uniformisation qui correspond à l'identification de classes. C'est

6. Groupe μ , *Traité du signe visuel*, Paris, Éditions du Seuil, coll. « La couleur des idées », 1992, p. 147-148.

7. *Ibid.*

là un des impacts du support matériel sur la connaissance qui y est inscrite. Lorsque nous transcrivons une lettre ou un signe de ponctuation, nous choisissons une représentation qui exprime l'appartenance de cette occurrence à une classe. Lorsque nous mentionnons un élément graphique entre simples chevrons, comme <f>, nous identifions la classe qu'il représente.

Les choses se corsent toutefois lorsque nous ne pouvons facilement accorder de valeur à ce que nous lisons. Ainsi, quelle valeur accorder aux espaces entre les mots, qui peuvent être de taille variable et difficile à évaluer⁸ ? à quelles constructions syntaxiques ou énonciatives se rapportent les signes de ponctuation⁹ ? Malgré les études parfois très fouillées sur le sujet, la réponse à ces questions nous échappe encore partiellement et nous sommes amenés à transcrire des entités *que nous ne comprenons pas*. On touche là un des problèmes les plus fondamentaux de l'édition de texte. On a souvent reproché aux linguistes de recopier de manière un peu servile des phénomènes observés sur les manuscrits sans les comprendre et d'encombrer certaines transcriptions diplomatiques de

-
8. Voir notamment Peter Rickard, « Système ou arbitraire ? Quelques réflexions sur la soudure des mots dans les manuscrits français du Moyen Âge », *Romania*, 103, 1982, p. 470-512 ; Nelly Andrieux-Reix et Simone Monsonégo, « Écrire les phrases au Moyen Âge. Matériaux et premières réflexions pour une étude des segments graphiques observés dans des manuscrits français médiévaux », *Romania*, vol. 115, n° 459-460, 1997, p. 289-336 ; Alexei Lavrentiev *et al.*, « Enjeux philologiques, linguistiques et informatiques de la philologie numérique : l'exemple de la segmentation des mots », *supra*, p. 77 sq.
9. Voir notamment Christiane Marchello-Nizia, « Le *comma* dans un manuscrit du XIII^e siècle : grammaticalisation », dans Olivier Bertrand, Sylvie Prévost, Michel Charolles *et al.* (dir.), *Discours, diachronie, stylistique du français. Études en hommage à Bernard Combettes*, Bern, P. Lang, 2008, p. 293-305 ; Alexei Lavrentiev, « Pour une méthodologie d'étude de la ponctuation médiévale basée sur une approche typologique », dans Bernard Combettes et Christiane Marchello-Nizia (dir.), *Études sur le changement linguistique en français*, actes du colloque « Diachro 2 » (Paris, janvier 2004), Nancy, Presses universitaires de Nancy, 2007, p. 191-204 ; *id.*, *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 soutenue sous la dir. de Christiane Marchello-Nizia, ENS Lyon, 2009], en ligne ; Nicolas Mazziotta, « Inconstance ou consensus ? Ponctuation et protase dans les chartes écrites en français à Liège avant 1271 », dans David Trotter (dir.), *Actes du XXIV^e Congrès international de linguistique et de philologie romanes, Aberystwyth, 1^{er}-6 août 2004*, Tübingen, Niemeyer, 2007, t. IV, p. 225-238 ; *id.*, *Ponctuation et syntaxe dans la langue française médiévale. Étude d'un corpus de chartes originales écrites à Liège entre 1236 et 1291*, Tübingen, Niemeyer, 2009.

distinctions inutiles¹⁰. L'impulsion quasi positiviste qui pousse à conserver toutes les distinctions reconnues *a priori* pour les analyser ensuite ne se justifie qu'après coup.

2.1. Identification des classes

Le premier problème qui résulte de cette posture est une sorte de confusion entre les variantes de formes des occurrences (*tokens*) et les différences de classes (*types*) qu'elles actualisent. Ainsi, dans la **fig. 1**, les différentes manifestations du *punctus* ont l'air de se trouver à des altitudes différentes.



1. Altitude du *punctus* (Document 1236-07 : 1, détail)

À y regarder de près, on pourrait transcrire, en respectant ces différentes altitudes :

- [1] Jo Nicholes sire de Rumignj . et de florines . fai fauoir atos ceaz ki
or funt et auenir funt . que [...]
- « Moi, Nicolas, sire de Rumigny et de Florennes, je fais savoir à
tous ceux qui sont et qui seront que [...] »

Si l'on ne lit que le début du document, on peut avoir l'impression que la différence entre <.> et <·> revêt un caractère fonctionnel : le point serait plus élevé devant une conjonction (*et*, *que*). Or, il suffit de continuer la lecture du document pour constater qu'il comporte également des <.> devant les conjonctions. La distribution entre les deux altitudes ne paraît pas pertinente. En tout cas, nous ne la comprenons pas. Sans compter que dans de nombreux cas, il nous est impossible de déterminer si le point est au milieu de la ligne, en bas ou dans une position intermédiaire. Cette difficulté, les médiévaux eux-mêmes la ressentaient, et elle les a poussés à concevoir des systèmes qui ne sont pas fondés sur des distinctions

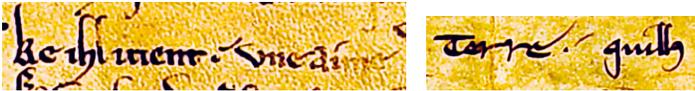
10. Raymund Wilhelm, « L'édition de texte – entreprise à la fois linguistique et littéraire », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, Berlin/Boston, De Gruyter, 2015, p. 131-151.

d'altitude¹¹. En tant qu'éditeur, nous n'avons qu'un contraste flou, anecdotique et non sémiotisé à présenter à nos lecteurs. À quoi bon le conserver ? L'éditeur représentera alors tous les points par <·>, symbolisant une classe qui accepte des variantes d'altitude dans ses réalisations.

Il arrive en outre que certaines distinctions soient pertinentes pour un document, mais moins évidentes pour d'autres. On part du principe que si différents documents sont écrits à la même époque dans le même espace variationnel et dans la même langue, il est raisonnable de postuler qu'ils partagent des conventions communes eu égard à la ponctuation. On pourrait donc penser que lorsqu'une distinction de classe est avérée pour un document, elle garde sa pertinence pour tous les autres documents examinés. Prenons l'exemple des signes « doubles », constitués d'un point et d'un autre tracé qui le surplombe (<./>, <.)>, <.'>, <.~>, etc.). On distingue au moins deux classes de signes originelles : le *comma* qui représente au départ une division de la période¹² et le *punctus elevatus*, qui représente une marque syntaxique et prosodique, initialement dans les textes liturgiques¹³ et dont on voit notamment se développer l'emploi « exclamatif » au XIII^e siècle¹⁴. Les deux classes sont réalisées par des signes très semblables que certains chercheurs ne distinguent pas. Il n'est pas déraisonnable de penser que certains médiévaux en sont venu à les confondre, en particulier avec le développement de la lecture silencieuse¹⁵. En conséquence, la question qui se pose à l'éditeur est la suivante : faut-il

-
11. Malcolm B. Parkes, « Punctuation, or pause and effect », dans James J. Murphy (dir.), *Medieval eloquence. Studies in the theory and practice of medieval rhetoric*, Berkeley/Los Angeles/London, University of California Press, 1978, p. 140.
 12. *Id.*, *Pause and Effect. An Introduction to the History of Punctuation in the West*, Cambridge, Scholar Press, 1992, p. 303.
 13. *Ibid.*, p. 306.
 14. Elena Llamas-Pombo, « Graphie et ponctuation du français médiéval. Système et variation », dans Gabriella Parussa, Maria Colombo Timelli et Elena Llamas-Pombo (dir.), *Enregistrer la parole et écrire la langue dans la diachronie du français*, Tübingen, Narr, 2017, p. 69-72.
 15. Malcolm B. Parkes, « Punctuation, or pause and effect », art. cit. ; Paul Saenger, *Space Between Words. The Origins of Silent Reading*, Stanford, Stanford University Press, 1997.

garder toutes les distinctions les plus subtiles en attendant de découvrir un document qui présente les oppositions qui les valideraient? On conserverait alors, comme représentants de classes supposées, des formes aussi variées que: <./>, <.>, <.'>, <~>, etc. Pour le corpus que j'ai exploré, j'ai confirmé qu'il fallait au moins distinguer deux classes, représentés ci-dessous par <.> et <.'>, car il est probable qu'ils correspondent à une opposition fonctionnelle d'ordre syntaxique dans un long testament liégeois de 1272¹⁶. Représenter cette différence de forme est capital si l'on veut préserver la possibilité de rendre compte du fonctionnement de ces signes.



2. Comparaison de <.> (gauche; Document 1236-05, l. 6).
<.'> (droite; Document 1236-12-15, l. 2)

Le corollaire apparemment évident de cette conclusion est qu'il est souhaitable de distinguer ces deux formes dans tous les documents (fig. 2). Toutefois, il apparaît que les tests statistiques sur l'essentiel des documents de la même époque ne montrent pas de différence significative d'emploi entre les deux formes d'expression¹⁷. La distinction conservée s'est ainsi révélée non pertinente.

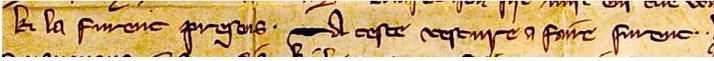
Les deux exemples ci-dessus montrent qu'il n'y a pas de choix proprement rationnel possible. Transcrire, c'est parier que la classification *a priori* que l'on propose sera utile.

2.2. Discretisation problématique

Un deuxième type de problème touche la discretisation des unités graphiques. La fig. 3 comporte un trait épais plus ou moins horizontal qui précède directement <A> (voir la transcription [2]).

16. Nicolas Mazziotta, « Étude linguistique de la ponctuation d'une charte originale en langue française (Liège, 1272) », *Orbis: Bulletin international de documentation linguistique*, 42, 2011, § 1.2.2.

17. *Id.*, *Ponctuation et syntaxe dans la langue française médiévale. Étude d'un corpus de chartes originales écrites à Liège entre 1236 et 1291, op. cit.*, p. 350-351.



3. Problème de discrétisation (Document 1277-03-23, l. 12)

[2] [...] · —A cefte vefture a faire furent ·· [...]

« À l'exécution de cette vêtue, furent [présentes les personnes suivantes] »

Cependant, selon mes observations sur un corpus de cent quarante-huit documents liégeois de la même période¹⁸, le signe <—> n'apparaît que trois fois (sans tenir compte de ses emplois pour clôturer une ligne, fréquents à la fin des documents, ou à la césure). Dans deux cas, le tracé est épais et il entre en contact direct avec la majuscule qui suit; dans le troisième cas, il est fin et en bout de ligne. Il est donc légitime de se poser la question suivante : le trait observé est-il bel et bien un signe indépendant (on aurait <—> + <A>) ? ou s'agit-il d'un trait adventice constituant <A> ? Il est impossible de trancher, dans la mesure où la distinction nous est complètement inaccessible.

L'escalade que je viens de présenter révèle des questionnements antérieurs à l'inscription de la connaissance à proprement parler. Nous en sommes à l'étape de la compréhension de la « lettre » du texte (qui, dans ces exemples, échoue au moins partiellement). Je voudrais à présent montrer que même si la compréhension aboutit, l'étape suivante, celle de la représentation dans l'inscription éditoriale, n'est pas toujours simple.

3. Incrire

Incrire ce qui est compris ne pose la plupart du temps que des questions triviales. Les exemples de la section précédente le montrent. Une fois résolu le problème de la compréhension, il n'y a généralement pas de difficulté à choisir une forme apte à inscrire les signes reconnus, pour peu qu'on le désire. Ainsi, si l'on distingue deux hauteurs pour les points ou deux formes pour des signes qu'on souhaite distinguer (2.1), il est facile de choisir deux caractères pour inscrire ces différences. De même,

18. *Ibid.*, p. 353-354.

il est facile de rendre compte qu'on a choisi de discrétiser une ou deux unités (2.2). La philologie numérique a résolu une part des importants problèmes d'ergonomie qui empêchaient le développement d'éditions conservant la ponctuation médiévale (3.1), mais elle ne constitue pas une panacée. Je me concentrerai ici sur deux cas qui demeurent difficiles à représenter (3.2 et 3.3).

3.1. *Brouillage et ergonomie : la solution informatique*

Il faut se résoudre à constater que la préservation de la ponctuation médiévale dans une édition suffit à créer des problèmes d'ergonomie, c'est-à-dire, en l'occurrence, de compréhension et d'accès au texte. Ces derniers sont parfois bien trop importants pour le public habituel auquel ces éditions sont réservées.

On perd ainsi, pour certaines recherches, un très grand nombre d'informations, pour ne pas dire un objet même, que l'on ne peut donner à lire (sauf en retour à obscurcir, pour le bénéfice de quelques uns, le texte publié), que l'on ne peut guère verser en note (sauf à accroître l'inextricable maquis de l'édition hypersavante) et que l'on doit seulement pouvoir retrouver [...] ¹⁹.

On ne peut guère présenter une double ponctuation pour pallier le problème : cela ne fait que réduire l'ergonomie de l'édition. C'est pourquoi le philologue est souvent amené à se délester de cette ponctuation originale pour des raisons pratiques. Malheureusement, procéder de la sorte coupe de toute possibilité de se servir de l'édition pour résoudre des questions bien légitimes portant sur la mise en texte :

Or s'il nous faut admettre que les normes actuelles de ponctuation sont requises par les habitudes de lecture contemporaines et par notre image visuelle de la langue, il n'est pas moins juste de poser que la mise en texte des manuscrits répond à des habitudes différentes de lecture et de conservation des œuvres par l'écriture ²⁰.

19. Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la philologie*, Paris, Éditions du Seuil, coll. « Des travaux », 1989, p. 48.

20. Elena Llamas Pombo, « Ponctuer, éditer, lire. État des études sur la ponctuation dans le livre manuscrit », art. cit., p. 132.

Heureusement, les pratiques technologiques modernes, comme l'avait déjà mentionné Bernard Cerquiglini²¹, ont résolu le problème. Lorsqu'elles sont réalisées par des philologues conscients des enjeux de l'étude de la mise en texte, les éditions électroniques actuelles reposent sur des inscriptions informatiques dans lesquelles sont encodées à la fois la ponctuation moderne et la ponctuation ancienne. Pareilles inscriptions sont accompagnées de mécanismes de visualisation multiples qui permettent au lecteur de générer les inscriptions traditionnelles où il pourra ne voir que ce qui l'intéresse²². Cependant, comme on va le voir, l'encodage informatique n'est pas toujours suffisant pour résoudre complètement les problèmes ergonomiques.

3.2. *Variation entre manuscrits*

Le premier cas que je voudrais présenter concerne la tradition d'une charte dont nous possédons deux exemplaires originaux (Document 1268-08-02a et Document 1268-08-02b). En 1992, Céline Guillot-Barbance observait déjà une variation de la ponctuation dans cinq manuscrits du xv^e siècle transmettant le même texte²³. Il s'agissait là de copies indépendantes, regroupées en trois familles distinctes. Le cas qui nous occupe ici est différent, puisqu'il s'agit de deux expéditions originales d'un même acte. Leur ponctuation n'est cependant pas exactement identique : certains signes sont absents d'un des documents, ou bien ils sont de forme plus simple que dans l'autre. Si l'on

21. Bernard Cerquiglini, *Éloge de la variante. Histoire critique de la philologie*, op. cit., p. 115-116.

22. Sur les concepts, voir Nicolas Mazziotta, (2004): « Le texte dans tous ses états. Philosophie d'encodage du projet "Khartès" », dans Gérald Purnelle, Cédric Firon et Anne Dister (dir.), *Le Poids des mots*, actes des 7^e Journées internationales d'analyse statistique des données textuelles, Louvain-la-Neuve, UCL/Presses universitaires de Louvain, 2004, p. 793-803; pour des exemples d'éditions, voir Christiane Marchello-Nizia, Alexei Lavrentiev et Céline Guillot-Barbance, « Édition électronique de la *Queste del saint Graal* », dans David Trotter (dir.), *Manuel de la philologie de l'édition*, op. cit., p. 155-176 et *Documents linguistiques galloromans. Édition électronique*, dir. Martin-Dietrich Glessgen, en partenariat avec Hélène Carles, Frédéric Duval et Paul Videsott, 2016 (en ligne : <http://www.rose.uzh.ch/docling/>).

23. Céline Barbance, « La ponctuation médiévale : quelques remarques sur cinq manuscrits du début du xv^e siècle », *Romania*, vol. 113, n^o 451-452, 1992, p. 505-527.

observe les fac-similés de la fig. 4, on voit que celui du haut comporte </> devant <fe>, et l'autre non.



4. Variation de présence de *virgula* (Document 1268-08-02a, l. 11 et Document 1268-08-02b, l. 11)

Dans une étude parue en 2016²⁴, j'ai démontré que les deux documents étaient incontestablement véridiques et revêtaient la même valeur diplomatique au même moment. Il faut conclure que cette variation n'est pas accidentelle, mais que les signes de ponctuation sont des *spécifications facultatives* qui se superposent à un contexte « ponctuable »²⁵.

On peut bien entendu encoder la variation de la ponctuation dans un format qui permet de l'enregistrer et de générer des inscriptions traditionnelles *ad hoc*, comme indiqué ci-dessus (3.1). Une entreprise comme la Text Encoding Initiative fournit d'ailleurs des normes d'encodage satisfaisantes²⁶. Cependant, savoir comment encoder ne suffit pas : il est légitime de s'interroger sur la manière d'inscrire cette variation sous une forme traditionnelle. On peut ainsi concevoir trois alternatives²⁷ : une édition synoptique, comme celle des extraits proposés par Céline Barbance²⁸ ([3]) ; une édition où les variantes sont

24. Nicolas Mazziotta, « Variation de la ponctuation dans une charte originale en double expédition écrite en français à Liège au XIII^e siècle », dans Valérie Fasseur et Cécile Rochelois (dir.), *Ponctuer l'œuvre médiévale. Des signes au sens*, Genève, Droz, coll. « Publications romanes et françaises », 2016, p. 299-316.

25. Voir Alexei Lavrentiev, « Pour une méthodologie d'étude de la ponctuation médiévale basée sur une approche typologique », art. cit.

26. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, en ligne : <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/> (dernière mise à jour le 10 juillet 2017), § 12.

27. Nicolas Mazziotta, « Variation de la ponctuation dans une charte originale en double expédition écrite en français à Liège au XIII^e siècle », art. cit., § 3.

28. Céline Barbance, « La ponctuation médiévale : quelques remarques sur cinq manuscrits du début du XV^e siècle », art. cit.

reléguées à l'apparat critique ([4]); une édition intégrant les variantes dans la continuité du texte principal ([5]).

[3] **A**: lor done en almoine / fe nul droit i aj · **B**: lor done en almoine
fe nul droit i ai ·

« Je leur donne en aumône [mes droits sur un bien] si j'en ai. »

[4] lor done en almoine / fe nul droit i aj ·

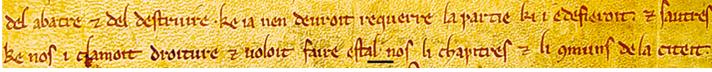
B: *almoine fe ; ai*

[5] lor done en almoine /^A fe nul droit i aj^A ai^B ·

Mais il y a lieu de parler de dilemme plutôt que d'alternative, car ces trois possibilités ont beau avoir leurs avantages respectifs, elles sont toutes imparfaites. [3] et [4] sont parfaitement lisibles s'il s'agit d'accéder simplement au texte et non de comparer les usages en matière de ponctuation. Dans ce cas, [3] oblige le lecteur à de constants aller-retours entre les deux versions en parallèle pour repérer les variantes. [4], qui simule ici un texte suivi d'un apparat critique en bas de la page, impose une démarche similaire en plus de hiérarchiser les deux documents. [5] obéit à des conventions particulières : il faut comprendre que chaque variante de présence ou de forme de signe de ponctuation ou de mot est étiquetée par la lettre du manuscrit où on la trouve, en exposant. Difficilement lisible en lecture linéaire, cette édition facilite en revanche la comparaison. On retrouve ici, exacerbés, les problèmes que Bernard Cerquiglini pointait déjà à propos de la simple représentation de la ponctuation médiévale (3.1).

3.3. *Édition génétique et accidents matériels*

Cette dernière sous-section sera consacrée à l'inscription de signes ajoutés par un scribe à la suite de la rédaction du document (en guise de correction). Les cas identifiables comme tels sont assez rares, mais extrêmement précieux : ils nous renseignent sur la conscience linguistique des écrivains et, partant, ils démontrent la nature conventionnelle des signes de ponctuation.



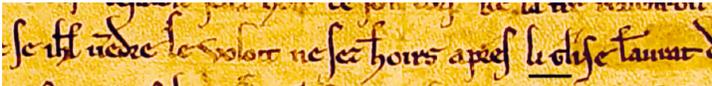
5. Ajout de ◌,◌ (Document 1237-12, l. 13 [soulignement ajouté ici])

Le premier exemple examiné ne pose pas de problème (fig. 5, signe indiqué par un soulignement). Dès lors que l'on a interprété la position de ◌,◌ comme un indice d'ajout *a posteriori*, il devient envisageable d'inscrire la dimension génétique dans l'édition. À nouveau, la Text Encoding Initiative fournit des normes d'encodage aptes à rendre compte de ces phénomènes génétiques²⁹ et, cette fois, leur inscription traditionnelle peut ne pas être trop encombrante. Ainsi, dans [6], le signe apparaît entouré de demi-crochets, conventionnellement interprétés ici comme une marque éditoriale signifiant l'ajout par le scribe.

[6] fautreil ke nof i clamoit droiture et uoloit faire estal ◌,◌ nof [...]

« s'il voulait la [= une terre] vendre, ou ses héritiers par la suite, l'église l'aura [...] »

À côté de ce genre de cas simple il arrive que l'éditeur ait à rendre compte de phénomènes plus complexes.



6. Ajout de ◌,◌ (Document 1236-05, l; 8 [soulignement ajouté ici])

Si l'on observe de près la fig. 6 (un retour au manuscrit est souhaitable, car le trait de plume est très fin), on voit que le mot ◌Glife◌ (dont l'initiale est soulignée) est précédé d'un signe de ponctuation qui lui est accolé : ◌,◌. En ignorant le problème de segmentation et en suivant la même convention que dans [6], la transcription diplomatique de cet extrait est donc la suivante :

[7] fe ihl uendre le uoloit ne sez hoirs apres li ◌,◌ Glife laurat [...]

« s'il voulait la [= une terre] vendre, ou ses héritiers par la suite, l'église l'aura [...] »

29. TEI P5: *Guidelines for Electronic Text Encoding and Interchange*, cit., § 11.7.

Le <·> est ici intéressant à plusieurs égards. Tout d'abord, le signe est difficile à discerner – il a d'ailleurs été ignoré dans l'édition de Wilmotte³⁰, qui rend pourtant compte de la ponctuation originale. Ensuite, sa position particulière, collé à la première lettre du mot qui le suit, est un indice que le signe a été ajouté après coup et qu'il n'est pas intégré au flux normal de l'écriture. Cette analyse est confortée par l'analyse morphosyntaxique et la connaissance que nous avons du système de ponctuation médiéval. En effet, <·> est principalement un marqueur de limite de complément circonstanciel³¹; en l'occurrence, il devrait marquer la fin d'une protase³². Toutefois, il est pratiquement certain que le signe est mal placé, car *li* est un article qui fonctionne avec *glise* et non un pronom disjoint introduit par *après*. Tout porte à croire que la personne qui a ajouté le signe a lu « se il vendre le voloit ne sez hoirs après li » et non « se il vendre le voloit ne sez hoirs après ».

Cela signifie que l'inscription [7] ne rend pas compte de la valeur du signe interprété par l'éditeur, mais uniquement de son expression matérielle. Pour inscrire la valeur reconstruite dans l'édition, il faudrait déplacer le signe et le positionner à l'endroit où il aurait dû apparaître.

[8] fe ilh uendre le voloit ne fez hoirs apref <·> li Glífe laurat [...]

[9] fe ilh uendre le voloit ne fez hoirs apref <·> li {··} Glífe laurat [...]

L'inscription [8] rend ainsi compte de l'interprétation de l'éditeur, mais pas de l'expression matérielle du signe. On pourrait combiner les deux informations dans [9], qui indique : d'une part que l'ajout du scribe, marqué par des crochets, doit être effacé (cela est indiqué par les accolades) ; d'autre part, que le même signe doit être ajouté à l'endroit où il est représenté encadré par des chevrons. Toutefois, bien que surchargée de marques éditoriales indigestes, l'inscription [9] n'est pas plus apte à rendre

30. Maurice Wilmotte, *Études de philologie wallonne réunies par ses amis et élèves à l'occasion de sa promotion à l'éméritat*, Paris, E. Droz, 1932, p. 71.

31. Nicolas Mazziotta, *Ponctuation et syntaxe dans la langue française médiévale. Étude d'un corpus de chartes originales écrites à Liège entre 1236 et 1291*, op. cit., chap. 7.

32. *Id.*, « Inconstance ou consensus? Ponctuation et protase dans les chartes écrites en français à Liège avant 1271 », art. cit.

compte du fait que le signe ajouté par l'éditeur *est* celui qui est ajouté par le scribe. Il faudrait, pour inscrire le déplacement, le représenter par une flèche ou un système d'indice, ce qui aurait pour effet de rendre l'édition moins lisible encore. Des solutions informatiques permettant l'encodage d'un tel déplacement sont envisageables, mais ne résoudront de toute manière pas le problème de l'ergonomie de l'inscription traditionnelle.

Les cas suivants ont été abordés au fil de l'article :

- i. reconnaître les lettres et les signes de ponctuation consiste à associer des occurrences à des classes de formes d'expression (2.1) ;
- ii. les conserver dans la transcription consiste à *parier* que les distinctions associées se révéleront pertinentes (2.1 également) ;
- iii. il est parfois difficile de segmenter les formes d'expression (2.2) ;
- iv. la représentation simultanée des variantes surcharge l'édition (3.2) ;
- v. la représentation des corrections surcharge l'édition (3.3).

Malgré les portes qu'elle ouvre (3.1), il apparaît qu'une part importante des problèmes fondamentaux qui se posent ne sont pas résolus par l'informatique, car ils se situent en amont ou en aval de la confection de l'inscription informatique : il faut toujours reconnaître les caractères avant de les encoder (i à iii) et il faut toujours donner à lire une inscription traditionnelle, qu'elle soit générée au départ d'un encodage complexe ou non (iv et v). L'éditeur ne saurait donc en aucun cas procéder mécaniquement. Premièrement, il est illusoire de penser que l'on va pouvoir tout transcrire *tel quel*, comme il est naïf de penser que transcrire n'implique pas de poser des choix. Deuxièmement, dès lors qu'il a une connaissance à transmettre, l'éditeur se doit de réfléchir à la manière dont il va l'inscrire pour la donner à lire. Les considérations ergonomiques lui incombent ainsi naturellement.

Sources d'archives

Tous les documents d'archives utilisés sont des chartes conservées aux Archives de l'État à Liège.

Document 1236-05, *Cathédrale Saint-Lambert* 134, mai 1236.

Document 1236-07, *Cathédrale Saint-Lambert* 136, juillet 1236.

Document 1236-12-15, *Cathédrale Saint-Lambert* 142, 15 décembre 1236.

Document 1237-12, *Cathédrale Saint-Lambert* 153, décembre 1237.

Document 1268-08-02a, *Abbaye du Val Saint-Lambert* 302, 2 août 1268.

Document 1268-08-02b, *Abbaye du Val Saint-Lambert* 302, 2 août 1268.

Document 1272-03, *Couvent des Dominicains* 29, mars 1272.

Document 1277-03-23, *Collégiale Saint-Martin* 98, 23 mars 1277.

Lemmatisation et construction automatique de ressources lexicographiques : les développements du lemmatiseur LGeRM

Sylvie Bazin-Tacchella & Gilles Souvay
Université de Lorraine, ATILF/CNRS

Le lemmatiseur hors contexte LGeRM¹ a d'abord été développé par l'ingénieur de recherche Gilles Souvay à l'ATILF dans le cadre du *Dictionnaire du moyen français* (DMF). Au départ, il s'agissait de faciliter la rédaction et la consultation du dictionnaire grâce à cet outil. Mais, depuis 2008, LGeRM a connu de nouveaux développements, en raison de son intégration dans d'autres projets. Il paraît intéressant d'en retracer l'histoire et de montrer son utilisation dans l'interrogation des bases textuelles et la constitution de lexiques ou glossaires informatisés.

La création du lemmatiseur LGeRM s'inscrit dans l'histoire du *Dictionnaire du moyen français*, aujourd'hui un dictionnaire électronique en ligne². Au moment de la rédaction de cette publication, sa version 2012 était encore en ligne, mais la nouvelle version en préparation, la version *DMF 2015*, l'a depuis remplacée³. C'est le propre de la lexicographie électronique que de fournir des matériaux évolutifs. Mais cette évolution

1. Acronyme de Lemmes, Graphies lemmatisées et Règles Morphologiques. Pour une présentation détaillée du lemmatiseur, se reporter à Gilles Souvay, « LgeRM: un outil d'aide à la lemmatisation du moyen français », dans David Trotter (dir.), *Actes du XXIV^e Congrès international de linguistique et de philologie romanes* (Aberystwyth, août 2004), Tübingen, Niemeyer, 2007, t. 1, p. 457-466 et Gilles Souvay et Jean-Marie Pierrel, « LGeRM. Lemmatisation des mots en moyen français », *Traitement automatique des langues*, ATALA, 50, 2009/2, p. 21 sq.

2. www.atilf.fr/dmf.

3. Les copies d'écran sont tirées de la version 2012. Même si l'utilisateur accède désormais à la version la plus récente sur le site, il lui est toujours possible de revenir à une version précédente grâce à un onglet « Versions antérieures » sur le bandeau de la page d'accueil.

passer par un développement cohérent et, de ce point de vue, il paraît intéressant de considérer avec attention l'origine et les évolutions de l'outil qui est désormais au cœur de l'ensemble du projet *DMF*.

Si la conception proprement dite d'un dictionnaire portant sur le moyen français remonte à une trentaine d'années⁴, un tournant méthodologique et conceptuel important a été pris en 2001 au moment où son promoteur, Robert Martin, a pris la décision d'en faire un dictionnaire totalement et uniquement électronique. En effet, le dessein initial avait été de construire méthodiquement le dictionnaire à partir de la synthèse de lexiques préalables augmentés de « dossiers de mots », réalisés notamment à partir de glossaires d'éditions critiques, lemmatisés dans le « glossaire des glossaires ». Cela correspond à la première période de rassemblement des matériaux (1984-2001). Pour des questions de coût et de durée, il a donc été décidé en 2001 d'abandonner le mode traditionnel d'avancée d'un dictionnaire, lettre par lettre, au profit d'une progression par étapes ; la première étape devait alors consister en une transformation des lexiques déjà disponibles à cette date et des lexiques ultérieurs en une base de données lexicales balisées en langage XML-TEI, de manière à construire progressivement le dictionnaire. Des lemmes avaient alors été créés pour regrouper les articles des différents lexiques portant sur le même mot au moment de la consultation, mais on était encore loin de la mise en place d'une procédure automatique de lemmatisation. Une deuxième période, entre 2007 et 2012, a vu la synthétisation des articles rédigés au cours des étapes précédentes. À partir de la version *DMF 2012*, une nouvelle étape d'enrichissement sélectif du dictionnaire et de développement de ses outils et de ses bases a été inaugurée.

4. L'histoire du *DMF* se greffe sur celle du *Trésor de la langue française (TLF)* développé à l'Institut national (aujourd'hui le laboratoire ATILF). Dans sa conception initiale, le *TLF* devait couvrir toute l'histoire de la langue, des plus anciens textes au *xx*^e siècle. Mais comme un tel projet était irréalisable dans les délais impartis, le *TLF* s'est cantonné à la langue des *xix*^e et *xx*^e siècles et Paul Imbs, fondateur et premier directeur du *TLF*, a souhaité que d'autres dictionnaires prennent le relai pour des périodes antérieures : c'est le cas du *DMF* pour le moyen français. Cette histoire commune explique aussi pourquoi le *DMF* s'appuie sur les mêmes principes lexicographiques que le *TLF*.

Le lemmatiseur LGeRM

La nécessaire gestion des variantes

LGeRM, conçu au départ pour apporter une solution aux difficultés de consultation rencontrées habituellement dans les dictionnaires qui portent sur des états anciens de la langue, devait rendre la consultation du *DMF* plus facile et par conséquent plus conviviale – ce qui est une attente forte pour un dictionnaire électronique. Ces difficultés sont liées au statut de la langue ancienne: en effet, le lecteur de textes médiévaux qui veut s'aider d'un dictionnaire traditionnel est souvent confronté à l'importance de la variation linguistique dans un état de langue non codifié. La langue médiévale est essentiellement variante, à la fois en raison de la transmission manuscrite et d'un système linguistique souple, sur les plans diachronique et diatopique, non normé, ce qui ne signifie pas pour autant aléatoire. Un dictionnaire construit selon les principes traditionnels ne permet pas d'identifier et de regrouper les formes variantes, ce qui rend son utilisation limitée pour le spécialiste et peu utile pour le néophyte.

Sous quelle entrée trouver les formes que l'on peut rencontrer dans un texte médiéval? La variation peut être seulement graphique, ainsi *tens/tans*; le copiste peut conserver des consonnes qui ne sont plus prononcées, comme *s* intérieure dans *teste*, ou choisir de les insérer pour rappeler l'étymologie, comme la labiale *p* dans *temps* ou *corps*, ou la palatale *c* dans *faict*; il existe même de véritables équivalences graphiques, telles *c/k*, *ss/s* ou *ai/ei/e*, *i/y*, souvent liées à l'évolution phonétique, caractéristiques d'une zone géographique et/ou d'une période temporelle. Quelle est alors l'entrée choisie par le dictionnaire? Celle qui se rapproche le plus du français moderne, celle qui est la plus fréquemment employée dans la période considérée? Lorsqu'un dictionnaire comme le Godefroy⁵ donne la liste

5. Frédéric Godefroy, *Dictionnaire de l'ancienne langue française et de tous les dialectes du IX^e siècle au XV^e siècle*, Paris, F. Vieweg, 10 vol., 1881-1902. Une version du Godefroy est accessible et interrogeable en ligne à l'adresse suivante: <http://www.micmap.org/dicfro>. On peut également y accéder à partir des liens présents dans les en-têtes

des formes rencontrées, il faut les retrouver sous une entrée qui n'est pas forcément celle qui pose problème à l'utilisateur. Il existe des renvois, mais ils ne sont pas systématiques.

Une consultation facilitée du dictionnaire

Grâce au lemmatiseur, le *DMF* s'est transformé en un dictionnaire totalement électronique, au-delà de la collecte et de la manipulation des données lexicographiques, notamment des corpus d'exemples. En effet, le lemmatiseur intervient lors de la consultation du dictionnaire. Il permet d'interroger à partir de la forme rencontrée dans un document, sans prérequis ou analyse particulière. Dans le cas de la forme adjectivale masculine *vis* ou de la forme verbale *menra*, la difficulté est de nature morphologique : la forme *vis*, qui peut être cas sujet singulier ou cas régime pluriel en ancien français, ou forme au pluriel lorsque la déclinaison disparaît, n'est pas une entrée du dictionnaire, il faut la chercher sous la forme non marquée⁶ *vif*; le dictionnaire ne permettra pas de retrouver *menra*, forme usuelle en ancien et moyen français du futur simple du verbe *mener*, avec disparition de *e* dans la séquence *-ner-*, puisqu'il faut chercher sous un infinitif dont le lien avec la forme considérée est loin d'être évident⁷. Parfois, cela

des articles du *DMF*, sous les sigles « GD » (pour « Godefroy ») et « GDC » (pour « Complément Godefroy »). Voici par exemple le bandeau correspondant dans l'article *CHERCHER* : [TL : *cerchier*; GD : *cerchier1*; GDC : *cherchier1*; AND : *chercher*; DÉCT : *cerchier*; FEW II-1, 695a : *circare*; TLF : *chercher*]. On remarque dans la liste que seul l'article du Tobler-Lommatzsch (TL) n'est pas consultable en ligne. Les articles accessibles sont ceux du *Trésor de la langue française (TLF)*, du *Französisches Etymologisches Wörterbuch (FEW)*, de l'*Anglo-Norman Dictionary (AND)* et du *Dictionnaire électronique de Chrétien de Troyes (DÉCT)*.

6. On appose en ancien français la forme marquée par un morphème flexionnel, en l'occurrence *-s*, au cas sujet singulier et au cas régime pluriel pour les masculins à une seule base, aux formes non marquées, cas régime singulier et cas sujet pluriel qui présentent la base « nue » ou suivie du morphème zéro. Par convention, c'est la forme non marquée qui sert d'entrée dans les dictionnaires ou les glossaires d'édition. Mais encore faut-il que le lecteur puisse retrouver à partir de la forme marquée qu'il rencontre dans un texte la forme non marquée du dictionnaire ! L'orthographe moderne, *viſs*, à travers l'insertion d'une consonne muette, souligne le lien entre le pluriel et le singulier.
7. On pourrait objecter qu'il en va de même des formes conjuguées modernes, qui présupposent la connaissance du système morphologique verbal. Cependant, en français moderne, s'il subsiste nombre d'exceptions ou d'irrégularités, les variantes se bornent à l'orthographe autorisée et ne touchent plus à la formation des paradigmes

peut sembler plus simple, ainsi pour une forme telle que *vendra*, que l'on aurait tendance à rattacher au verbe *vendre*, selon la morphologie moderne, alors qu'il peut tout aussi bien s'agir du futur du verbe *venir*, construit dans l'ancienne langue sur la base faible du verbe. Des variantes diatopiques se rencontrent également dans les textes en ancien et moyen français, selon la coloration dialectale des témoins, ainsi *loi/lei*, *bel/biel*, *ceval/cheval* ou encore *chacier/cachier* (latin **captiare*)⁸.

Le rédacteur du dictionnaire, confronté aux multiples graphies possibles d'un même terme⁹, a dû lui aussi choisir une entrée. Il peut s'agir d'une entrée moderne, lorsque le terme s'est maintenu – c'est l'option qui a été suivie dans le choix du lemme pour le DMF, ainsi *seignor* est rangé sous le lemme SEIGNEUR. Mais tous les termes n'ont pas subsisté. Lorsque le mot a disparu, le lemme peut être reconstruit comme une forme moderne acceptable, mais non attestée, comme PLENTÉ pour les formes du substantif *plenté/planté*¹⁰; ou alors la forme la plus fréquente dans les textes peut être choisie. Avec le lemmatiseur LGeRM, le choix des lemmes, même discutable, permet le regroupement des formes. L'enjeu de la lemmatisation ainsi comprise n'est plus tant d'offrir le seul accès pertinent aux informations contenues dans l'article que de permettre de naviguer d'une forme attestée à l'ensemble des formes auquel elle appartient. La lemmatisation ne se borne pas à une procédure de sélection d'entrées choisies une fois pour toutes, mais est convoquée à chaque interrogation sur une forme.

eux-mêmes, comme à l'époque médiévale où étaient attestés, par exemple, jusqu'à trois types de passés simples différents pour le verbe *voloir*.

8. La graphie *ei* pour *oi* est générale dans les textes en anglo-normand ; les autres variantes mentionnées se retrouvent dans les textes picards. Mais les textes peuvent être plus ou moins marqués et certains traits se retrouvent isolés dans certaines copies.
9. Ainsi, pour les mots cités précédemment, on relève des formes très variées dans la base textuelle : *lei*, *ley*, *loi*, *lois*, *loix*, *loiz*, *loy*, *loys*, *loyx* ; *ceval*, *cevaux*, *cevaus*, *chesval*, *cheval*, *chevalx*, *chevau*, *chevaux*, *chevalz*, *chevaus*, *chevaux*, *chevax*, *cheveaux*, *quevaux*. Cependant la fréquence permet de faire le départ entre les formes usuelles et des formes très marginales, voire douteuses.
10. Du latin *plenitate*. Le substantif, usuel en ancien et en moyen français, au sens de « grande quantité ou grand nombre, abondance » (DMF), n'a pas subsisté, alors qu'il est passé dans l'anglais *plenty*.

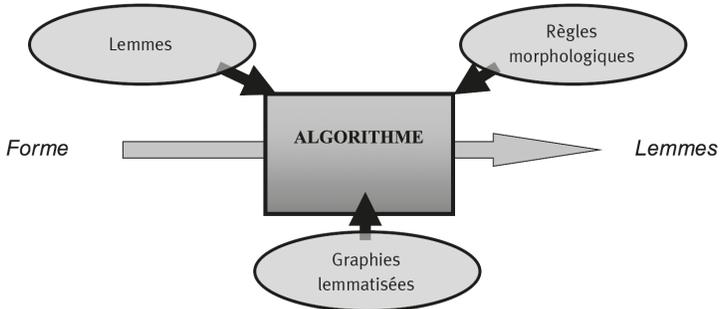
Le lemmatiseur est l'outil qui fait le lien entre les différentes composantes du *DMF*: dictionnaire, lexiques, base de textes et lemmatiseur. L'idée de départ était de permettre, par un simple clic sur un mot présent dans un exemple cité dans un article, d'accéder à l'article correspondant au lemme du mot. Mais il est apparu souvent difficile de savoir quelle entrée avait été retenue par les rédacteurs, si le mot avait subsisté ou non. Le lemmatiseur n'a donc pas permis d'interroger tous les mots des exemples rassemblés dans un article du *DMF*, parce qu'il aurait fallu lemmatiser tous les exemples. Mais il a facilité la consultation du dictionnaire: au lieu de chercher à deviner sous quelle entrée se cache telle ou telle forme, il offre la possibilité d'interroger par la forme rencontrée dans un manuscrit ou un texte édité. Le message qui s'affiche sur le site invite explicitement l'utilisateur à cette démarche: « Saisir un mot ou une forme sans se préoccuper de l'entrée du *DMF*; des propositions s'afficheront ». Ainsi des formes telles que *destroit*, *destreit* ou *destroict* sont-elles automatiquement renvoyées à deux lemmes de forme moderne, DÉTROIT 1 (adjectif) et DÉTROIT 2 (adjectif et substantif masculin). À l'utilisateur de faire son choix, en fonction du contexte, entre l'adjectif ou le substantif masculin, voire de consulter, en cas d'hésitation, les deux articles du *DMF*. Il aurait pu lui-même saisir directement la forme moderne *détroit*. En revanche, devant une forme comme *menra*, le dictionnaire traditionnel n'est d'aucune aide; le lecteur doit être familier des structures de la morphologie verbale médiévale ou, du moins, doit trouver des informations dans des ouvrages spécialisés à ce sujet. Dans un cas comme celui-ci, le lemmatiseur propose la réponse en ces termes: « La forme *menra* est connue du lemmatiseur avec l'analyse suivante: MENER, verbe ». Le lemmatiseur a reconnu la forme parmi l'ensemble des formes lemmatisées MENER et le lemme MENER sert alors de lien vers l'article du *DMF*. Lorsque la forme est ambiguë, le lemmatiseur propose plusieurs hypothèses: c'est le cas pour des formes telles que *destroit* ou *porte*¹¹. Il peut

11. Pour *destroit*, voir *supra*. Pour *porte*, le lemmatiseur propose trois réponses possibles: PORTE 1, subst. fém.; PORTE 2, subst. masc.; PORTER 1, verbe. Si la deuxième hypothèse est peu vraisemblable, car elle correspond à une seule attestation dans les bases avec le

arriver que la forme ne soit pas connue du lemmatiseur, dans ce cas le programme applique des règles, qui sont des procédures formelles, afin de retrouver une forme connue.

Conception de l'outil

Le schéma suivant résume le fonctionnement du lemmatiseur :



1. Architecture du système LGeRM. Gilles Souvay et Jean-Marie Pierrel, « Lemmatisation des mots en moyen français », art. cit., p. 154.

Ce programme a été conçu pour analyser les graphies du moyen français et éviter des manipulations fastidieuses en cas d'introduction de nouveaux mots par le biais des exemples. En effet, la simple projection du contenu de l'ensemble des lexiques sur les textes ayant servi à les constituer s'était avérée insuffisante, dans la mesure où tous les mots des exemples n'avaient pas été traités par les lexiques, qui avaient opéré des sélections sur des critères qui pouvaient d'ailleurs également varier d'un lexique à l'autre ; d'autre part, les mots grammaticaux n'avaient en général pas été traités dans les lexiques. Il y avait donc dans le corpus d'exemples des formes attestées sans lemme. L'outil d'analyse devait permettre de pallier ces manques : constitué d'une base de connaissances comportant les lemmes, les graphies lemmatisées et les règles morphologiques, qu'il

sens de « contribution », la première et la troisième renvoient à des lemmes usuels. Le lemmatiseur formule des hypothèses hors contexte, mais l'utilisateur cherche le sens de mots rencontrés dans un texte et, de ce fait, connaît leur catégorie grammaticale. Parfois cependant, même en contexte, subsiste une hésitation qui ne pourra être levée que par la prise en compte d'un contexte élargi, voire de connaissances extralinguistiques.

utilise pour fournir une liste d'hypothèses de lemmes associés à une graphie donnée, hors contexte. L'objectif est atteint si l'analyse juste est présente dans les hypothèses présentées; l'outil ne gère pas lui-même les ambiguïtés, c'est le relecteur humain qui valide telle ou telle hypothèse en fonction du contexte d'utilisation de la forme et plus généralement de l'ensemble de connaissances auquel participe le texte de référence. Si la graphie rencontrée n'appartient pas à la base de connaissances, le programme applique les règles morphologiques pour la ramener à une forme connue, si possible à une graphie lemmatisée. Le bruit généré est important, notamment en l'absence de balisage grammatical des formes étudiées. Le lemme retenu pour le *DMF* est une forme moderne lorsqu'elle existe; dans le cas contraire, la forme choisie est celle reçue par la tradition lexicographique. Parmi les règles morphologiques, un très grand nombre concerne la flexion verbale (environ 3 000): elles ramènent en général la graphie verbale à un infinitif ou elles modernisent une forme ancienne pour la ramener à une forme connue; d'autres règles portent sur la flexion nominale et adjectivale (environ 100): il s'agit notamment de trouver la forme correspondante du masculin singulier ou encore de moderniser une forme ancienne. Les autres règles sont beaucoup plus diverses (environ 400): modernisation ou vieillissement des graphies, règles de régionalisme, règles phonétiques... Le système permet enfin de prendre en compte des phénomènes de formation de mot, comme le lien entre l'adjectif et l'adverbe en *-ment*.

Élargissement des applications de LGeRM

Intégration dans de grands projets d'édition en ligne

Le lemmisateur LGeRM, conçu au départ pour le *DMF*, a vu ses applications s'élargir. En permettant la lemmatisation de textes de moyen français et, moyennant certaines adaptations, également la lemmatisation de textes des périodes antérieures et postérieures, il s'est transformé en outil au service de l'édition de texte. Ce développement a été favorisé par des collaborations scientifiques extérieures au laboratoire et des expérimentations

internes. La première d'entre elles a joué un rôle véritablement fondateur en rassemblant le *DMF* et des équipes britanniques travaillant sur de grands auteurs de moyen français : celles de l'université d'Édimbourg pour le projet consacré à Christine de Pizan, sous la direction de James Laidlaw, et les universités de Sheffield et Liverpool pour le projet dédié aux *Chroniques* de Froissart, sous la direction de Peter Ainsworth. Ce « *joint project* », intitulé « Medieval Vernacular Dictionaries: principles, methodology, practice, problems and solutions », s'est déroulé sur quatre années, entre 2008 et 2011. Les discussions sur les résultats de la lemmatisation des textes ont permis d'aller plus loin dans l'analyse des problèmes rencontrés et dans la recherche de solutions, notamment grâce à l'encodage des textes et au respect de la TEI. Chaque nouveau projet a offert la possibilité d'affiner le balisage, par exemple pour les noms propres. Ainsi l'utilisation de LGeRM dans d'ambitieux projets d'éditions électroniques a-t-elle permis de l'améliorer et d'en faire peu à peu un outil de construction de glossaire efficace et souple.

L'édition lemmatisée du Réceptaire de Jean Pitart

Parallèlement à notre participation à ces grands projets collaboratifs¹², nous avons lancé la lemmatisation de textes divers du moyen français, littéraires et non littéraires, notamment celle d'un recueil de recettes médicales, traditionnellement connu comme le « réceptaire de Jean Pitart ». Pour évaluer très concrètement les possibilités offertes par LGeRM sur un type de texte différent, d'ampleur limitée, une simple transcription du texte avait été lemmatisée avec l'outil¹³. Dans une première

12. Ces projets, toujours accessibles en ligne, renvoient au *DMF* par des liens sur chacun des mots du texte édité lemmatisé :

- a) Christine de Pizan, *The Making of the Queen's Manuscript*, édition électronique du ms. Harley MS 4431 de la British Library : <http://www.pizan.lib.ed.ac.uk/> ;
- b) The Online Froissart, *Édition électronique des Chroniques de Jean Froissart* : <http://www.hrionline.ac.uk/onlinefroissart> ;
- c) édition électronique en ligne du *Mystère des Actes des Apôtres* de Simon Gréban : <http://eserve.org.uk/anr/>.

13. Toute l'étude a été menée avec la version 2010 du *DMF*. Voir Sylvie Bazin-Tacchella, « Le "Réceptaire attribué à Jean Pitart" (xiv^e siècle) : projet d'une édition et d'un

étape, le texte ne présentait aucune espèce de balisage, sinon des éléments paratextuels habituels dans l'édition des manuscrits, indication de la foliotation et des colonnes, ponctuation et accentuation conformes aux normes en vigueur dans l'édition des textes médiévaux. Il avait été mis au format XML de façon à être lemmatisé avec LGeRM. Une première lemmatisation du texte avait mis à jour toute une série de termes pour lesquels LGeRM n'offrait pas de réponse, regroupés à part dans une « liste des mots inconnus ». Y apparaissaient quelques mots non identifiés par le programme, mais aussi des erreurs de transcription ou de saisie. Ainsi une première lemmatisation permet-elle de détecter et de corriger des erreurs souvent difficiles à déceler malgré les relectures, et de prendre conscience du balisage nécessaire pour que le traitement automatique isole de manière systématique un certain nombre d'éléments. Le texte au format XML TEI a donc été enrichi de nouvelles balises, avant qu'il ne soit procédé à une nouvelle lemmatisation :

```
<div><head><num>1.</num> La <rs type="remede">toile
maistre <name>Jehan Pitart</name></rs> contre toutes
bleceures de jambes et d'autres lieux et en ot la recepte du
<rs type="»personne»>roy de <name>France</name></rs>.
</head>
<p><c>P</c>renés oile d'olive <num>.iii.</num> livres; suif de
cerf <num>.i.</num> quarteron; ceruse poudree, <num>.ii.</
num> livres; serapin, opopanac, armoniac, litarge, mummie, de
chascun une once; stafizagre, <foreign>bdellium</foreign>, ase
fetide, orpiment, de chascun demie once; encens, mastic, mirre,
colofome, aloe, cicotrin, de chascun <num>.ii.</num> drames;
<foreign>ypoquistidos</foreign> une drame.</p>
<p><c>L</c>a maniere de confire est <pb n="3ra"/> tele:
mettés vostre oile et vostre suif de cerf sus lent feu tant que le
suif soit bien fondu et puis si mettés les poudres, c'est a savoir la
poudre de ceruse premierement et mouvés bien tousjours tant
qu'il se commence a espoissier et puis y mettés les gommés,
c'est a savoir serapin, armoniac, opopanac, qui aient esté par
```

glossaire électronique », dans Joëlle Ducos (dir.), *Sciences et langues au Moyen Âge / Wissenschaften und Sprachen im Mittelalter*, Heidelberg, Winter, 2012, p. 269-286. Voir également notre présentation détaillée du projet et de ce type de texte (<http://www.atilf.fr/dmf/JeanPitart>).

```

<num>.iii.</num> jours temprees en <num>.iii.</num> voirres
de bon vin aigre et quant ces gommess seront bien fondues sus
le feu, si y mettés toutes les autres choses, ce sont les autres
gommess l'une après l'autre a petit feu et laissiés boullir tant
que il soit espés et puis ostés du feu et y mettés les poudres
chascune par soy delieement poudrees, puis prenés vostre
toile et la moilliés dedens et puis la mettés sus une table qui
soit ointe d'uile d'olive et soit bien frotee d'une part et d'autre
d'une liche si que ele soit bien soueve et laissiés refroidier et la
gardés.</p>
</div>

```

Aux balises liées à la structuration du texte (divisions, titres, capitales, changement de feuillet), ont été ajoutées des balises permettant d'isoler :

- les quantités, indiquées par des chiffres romains encadrés de points :

```
<num>.iii.</num> livres ;
```

- les termes latins, notamment dans les énumérations d'ingrédients :

```
<foreign>bdellium</foreign> ;
```

- des syntagmes considérés

comme des noms composés :

```
La<rstype="remede">toilemaistre<name>JehanPitart</name>
</rs> ;
```

ou comme des noms propres :

```
maistre <name>Jehan Pitart</name>.
```

Analyse des résultats de la lemmatisation

Après repérage et correction des erreurs, il reste des mots inconnus, c'est-à-dire des termes que le lemmatiseur ne possède pas dans sa base de connaissances et pour lesquels il est incapable de remonter à une forme connue en appliquant des règles. Ces mots inconnus peuvent constituer de véritables apports lexicaux et doivent donc être étudiés à l'aide de la documentation disponible au laboratoire. Il peut s'agir de nouveaux lemmes ou de nouvelles formes correspondant à des lemmes existants. Certaines de ces formes ont pu être déformées

par le copiste ou résulter d'un enchaînement d'erreurs. Dans un certain nombre de cas rencontrés dans le *Réceptaire de Jean Pitart*, les formes, bien que déformées, ont pu être mises en relation avec des formes attestées et être correctement lemmatisées par le programme, ce qui est apparu encourageant pour une utilisation dans le cadre de l'édition électronique comme aide à la lecture des textes ; ainsi les formes *stafizagre*, *colofome* ont-elles été lemmatisées STAPHISAIGRE, COLOPHANE (DMF) ; pour la forme *tenoisiee*, le lemmatiseur a proposé deux lemmes : le verbe NOISER et le substantif TANAISIE (DMF). C'est le second qui était pertinent.

Notre objectif en lemmatisant le *Réceptaire de Jean Pitart* était de préparer un texte électronique interrogeable en ligne, voire d'aider à la construction d'un glossaire systématique. Grâce aux interactions avec le DMF, les apports peuvent être réciproques :

- les éditeurs des textes médiévaux peuvent s'appuyer sur l'ensemble des ressources du DMF (le dictionnaire proprement dit, mais également toutes les bases qu'il exploite) ;
- le texte saisi peut offrir de nouvelles données lexicologiques qui enrichiront le DMF (compris au sens large).

Pour le premier type d'apport, il convient de distinguer ce qui relève du texte lemmatisé lui-même et ce qui relève du lemmatiseur ; en effet, il s'agit de deux procédures bien distinctes dans le formulaire de levée d'ambiguïté :

- la levée d'ambiguïté au sens strict, qui est opérée manuellement par la sélection d'un des lemmes proposés ou l'ajout d'un lemme nouveau ; elle ne concerne qu'une occurrence donnée dans le texte édité, mais elle peut être élargie aux autres occurrences ou à une partie de celles-ci ;
- l'enrichissement de la base du lemmatiseur par la nouvelle forme ou le nouveau lemme.

Confondre les deux procédures conduirait à limiter les possibilités du lemmatiseur. En d'autres termes, il s'agit de bien distinguer solution individuelle et programmation ; il faut laisser les choix ouverts, quitte à continuer à produire des ambiguïtés toujours en grand nombre – le but étant moins de parvenir toujours à une solution unique que d'obtenir, parmi les choix multiples, la bonne solution. Une solution ponctuelle, liée à un contexte particulier, qui permet la levée d'ambiguïté, ne saurait devenir une règle. En revanche, à travers telle ou telle solution que le lemmatiseur ne proposait pas, on peut formuler une règle qui faisait défaut au programme, notamment en matière d'équivalence graphique.

Les étapes du traitement

Le lemmatiseur applique un certain nombre de règles pour lemmatiser les formes inconnues : lorsque le nombre de règles appliquées ou le nombre de lemmes est élevé, on parle d'« analyses douteuses ». Le programme en donne la liste. L'utilisateur doit passer en revue tous les cas ambigus, dans certains cas valider un des lemmes proposés ou en proposer un autre. L'examen des analyses douteuses ne constitue qu'une première étape qui conduit à de nombreuses levées d'ambiguïtés. Les formes aberrantes sont détectées et corrigées. L'étape suivante sera l'examen des formes pour lesquelles le lemmatiseur propose un ou plusieurs lemmes ; dans le premier cas, il faudra valider ou non la proposition et s'interroger bien évidemment sur les « mauvaises solutions » proposées pour formuler des « parades » ; dans le cas de propositions multiples, l'examen des différentes attestations de la forme données en regard à l'écran devrait permettre de distinguer des homonymes, comme *face* (subst.) et *face* (forme verbale de FAIRE) ou *aux* (subst. pl. d'AUX) et *aux* (forme contractée de la préposition à et de l'article défini). Cet examen conduit souvent à établir des filtres pour éliminer des propositions non pertinentes pour le texte considéré. L'éditeur peut consulter en même temps le texte et le dictionnaire, mettre en regard les différentes attestations, ce qui est un gain de temps et

d'efficacité considérable au bénéfice de l'analyse proprement linguistique et sémantique du corpus.

Des matériaux pour des utilisations diverses

Ainsi, pour ce réceptaire, l'utilisation expérimentale de l'outil d'aide à l'édition appuyé sur le lemmatiseur du *DMF* a-t-elle permis d'offrir un texte sûr et interrogeable en ligne¹⁴. Bien évidemment, il ne s'agit pas d'une édition définitive et exhaustive, il manque un certain nombre d'apports attendus. Cependant, les matériaux qui sont accessibles donnent l'état de la recherche au moment de la mise en ligne ; ils seront complétés ou corrigés au gré de nouveaux apports ; une introduction a ainsi pu être ajoutée plusieurs mois après, offrant au lecteur toutes les informations disponibles sur la transmission du texte et sur ses enjeux. L'essentiel est cependant le fait de permettre d'interroger le lexique par forme, par lemme, en navigant dans le texte et dans le dictionnaire lui-même. Le projet n'a pas conduit à la réalisation d'un glossaire en bonne et due forme, mais tous les matériaux sont disponibles pour le faire et en l'état, il offre la possibilité de récupérer des éléments intéressants pour d'autres projets, par exemple des entrées et des exemples de termes scientifiques pour le projet Crealscience¹⁵, avec lequel le *DMF* a noué une convention permettant un échange de données :

Tableau 1. Extrait des exemples de termes de pharmacopée (lettre C) de Jean Pitart transmis au projet Crealscience

Lemme <i>DMF</i>	Forme attestée	Exemples sélectionnés dans le glossaire électronique de Jean Pitart
CALAMINE, subst. mineral de zinc, calamine	<i>calamine</i>	« Prenés couperose blanche, calamine cuite et en faites deliee poudre et qu'il ait de chascune egalment et mettés cele poudre temprer en belle yaue froide et clere. » 8vb « Et a faire ceste yaue faut a une chopine d'yaue le montant a une nois de noier de coperose et autant de calamine . » 9ra

14. <http://www.atilf.fr/dmf/JeanPitart>.

15. Il s'agit du projet de *Dictionnaire du français scientifique médiéval* mené dans le cadre de l'ANR Crealscience, codirigée par Joëlle Ducos (Sorbonne Université/Paris IV, EA 4509) et Xavier-Laurent Salvador (Paris I, I-DEF), voir <http://www.crealscience.fr>.

Lemme <i>DMF</i>	Forme attestée	Exemples sélectionnés dans le glossaire électronique de Jean Pitart
CAMOMILLE, subst. camomille	<i>camomille</i>	<p>« Item prenés jus de camomille et oile d'olive et mellés ensemble et en frotés au feu avant l'accés l'eschine du malade et les poux des bras, ce vaut moult. » 9rb</p> <p>« Prenés cire virge demie livre, raisine .i. quarteron, oile d'olive le tiers de demie livre, gabaron, cyroigne, mastic, encens, mirre, therebentine, dyauté, oile de camomille, de chascun une once. » 9vb</p> <p>« Baigniés vous en la decoction de foilles de fresne et de mauves et de camomille. » 11ra</p> <p>« Prenés camomille et triblés et boulés en vin aigre et en lavés souvent le chief, nulle chose n'est meilleur. » 16vb</p>
CAMPBRE, subst. substance aromatique extraite du <i>laurus camphra</i> ou d'autres plantes, huile ou substance brute	<i>campbre</i> <i>canfre</i>	<p>« Prenés .iv. onces d'oile rosat, cire blanche une once, campbre une drame, ceruse once et demie et puis les aubuns de .ii. oex et faites oignement. » 24rb</p> <p>« En teles fievres et en toutes autres eschivés a vostre pooir ire et toutes choses qui peuent le cuer esmouvoir ou eschauffer et tenés diete et mettés paine a dormir et a suer selonc les remedes qui sunt dis par devant pour dormir et pour suer et flairiés roses et violetes et semblables choses, especiaument canfre et yaue rose. » 14va</p>

Dans d'autres cas, l'utilisation de l'outil-glossaire a permis la construction du glossaire d'une édition parue sous un format papier. Chaque projet qui utilise l'outil a ses spécificités et les échanges avec nous sont déterminants. Des premiers essais jusqu'à la dernière version, les interfaces de l'outil ont constamment évolué pour s'adapter aux besoins, c'est pourquoi il est difficile d'en présenter une version figée. Nous nous bornerons à présenter ici les différentes étapes du travail avec l'outil d'aide à l'édition dans la version accessible au moment de la rédaction de cet article, et les différentes possibilités offertes à l'éditeur d'un texte en moyen français, ainsi que la richesse des matériaux à la disposition des utilisateurs.

Les possibilités de LGeRM

Mise en place

Des discussions préalables établissent les conditions de la collaboration entre un éditeur de texte médiéval et l'équipe chargée du projet *DMF*. Puis l'éditeur envoie le texte qu'il souhaite lemmatiser au format XML ; une fois la lemmatisation achevée, un mot de passe lui est fourni pour qu'il puisse accéder à son résultat. Sur la page d'accueil du *DMF*, il suffira alors de cliquer sur le menu « LGeRM : l'outil-glossaire » ou, mieux, d'utiliser un lien direct de type www.atilf.fr/dmf/pitart avec le nom du projet en collaboration¹⁶. Un formulaire s'affiche alors : à l'utilisateur d'indiquer son identifiant et son mot de passe, puisqu'il faut posséder les droits pour accéder au projet.

Toutes les données utiles sont rassemblées en ligne, sous la forme de synthèses accessibles sous le premier onglet du menu d'accueil « Présentation de l'outil-glossaire ». En voici la liste :

- Principes généraux
- Conditions d'accès
- Exemples de réalisation
 - Les étapes de la construction (1/3)
 1. Création du projet
 2. Lemmatisation du fichier
 3. Finalisation
 - Légende des couleurs
 - Gestion des mots grammaticaux
 - Présentation du lemmatiseur
 - Principe de fonctionnement
 - Bibliographie sélective
 - Prise en charge de la *TEI P5*
 - Généralités
 - Balises prises en charge
 - Segmentation des mots
 - Crédits
 - Projets en cours

16. <http://www.atilf.fr/LGeRM/glossaire>.

Procédures

À l'ouverture d'un projet, ce qui s'affiche par défaut est une synthèse brute, une sorte de tableau de bord du projet, avec toutes les données chiffrées après lemmatisation par LGeRM : nombre de lignes, de mots, de ponctuations, de mots connus, de mots non résolus, de formes inconnues, de noms propres, de noms étrangers, de lemmes proposés par le lemmatiseur. La date de la lemmatisation est indiquée, ainsi que les droits d'accès. À droite de chaque information chiffrée, un bouton « Voir » permet d'accéder aux données proprement dites. Il s'agit d'une première étape, comme le rappelle la mention figurant au bas du tableau indiquant que « [l]e glossaire n'a pas été finalisé », à droite de laquelle se trouve un bouton « Glossaire finalisé » qui permettra d'accéder à un nouvel état du glossaire, obtenu après validation des interventions de l'utilisateur. À gauche de la synthèse brute apparaissent les fonctionnalités offertes au rédacteur, qui peut ainsi accéder, par exemple, à différentes listes : « Liste des formes inconnues », « Liste des formes » ou « Liste des lemmes ». Précisons à nouveau qu'il ne s'agit ici que d'une première étape-test, dont les résultats ne peuvent être que provisoires, destinés à être améliorés par la relecture et la vérification du texte, son éventuel amendement, et parfois par un balisage renforcé. Le premier essai offre cependant la possibilité de « dégrossir » considérablement les difficultés rencontrées par le lemmatiseur et, partant, d'améliorer à la fois l'outil – dans ses fonctionnalités notamment – mais aussi le texte en cours d'examen lui-même.

Le texte étant débarrassé de ses erreurs de saisie – notamment grâce aux « mots inconnus » détectés par le lemmatiseur – et mieux préparé pour la lemmatisation, il est à nouveau lemmatisé et les résultats de la lemmatisation systématiquement contrôlés. L'éditeur dispose des différentes listes de formes ou de lemmes indiquées plus haut, qu'il peut parcourir en suivant l'ordre d'apparition dans le texte ou l'ordre alphabétique. Il peut également sélectionner une liste d'« analyses douteuses », c'est-à-dire des formes dont l'analyse implique l'utilisation

d'un nombre important de règles. Par défaut, le paramétrage porte sur l'utilisation de trois règles ou plus, mais l'utilisateur peut le modifier. Mais il peut d'emblée choisir de passer en revue toutes les formes ambiguës, c'est-à-dire pour lesquelles le lemmatiseur présente plusieurs lemmes possibles. Cette phase de « levée d'ambiguïtés », manuelle, demande beaucoup de temps et une très bonne connaissance de l'état de langue concerné. Pour chaque forme ambiguë, un formulaire permet de saisir le choix de l'utilisateur, en sélectionnant par un clic chaque attestation ou certaines seulement :

■ Travail sur la forme *vïelent*

<p>vïelent</p> <p>1 attestation</p> <p>Réinitialiser</p>	VIELLER , verbe 1 5 <input type="radio"/>	<input type="button" value="Choix de lemme avancé"/>
	VEILLER , verbe 2 20 <input type="radio"/>	<input type="checkbox"/> [1] mot 3446 <input type="checkbox"/> 16
	VÊLER , verbe 2 20 <input type="radio"/>	
	Lemme : <input type="text"/>	<input type="checkbox"/> Vïelors a dras vïelent par cez pavellons. d'ermine
	<input type="button" value="Code >"/>	
	<input type="checkbox"/> lemme absent du DMF	
	<input type="radio"/> mot étranger	
	<input type="radio"/> nom propre	
	<input type="radio"/> dénomination	
	<input type="radio"/> nombre	
<input type="radio"/> mot exclu		
<input type="checkbox"/> enrichir le lemmatiseur		
<input type="button" value="Lever l'ambiguïté"/>		

Cet exemple de formulaire présente une attestation unique dans le texte considéré. Le premier lemme proposé, qui est l'hypothèse pertinente, est doté d'une pondération faible, alors que les deux autres présentent une pondération importante liée à l'utilisation de règles par le lemmatiseur pour retrouver une forme connue. Il suffit à l'utilisateur de valider la première hypothèse proposée. Lors d'une nouvelle finalisation des données, l'ambiguïté disparaîtra et le seul lemme indiqué sera « VIELLER ». Lorsque la forme est attestée plusieurs fois dans le texte, le formulaire rassemble les différentes attestations et c'est l'utilisateur qui doit valider le lemme qu'il considère pertinent pour chaque exemple en cliquant dans la case placée à gauche du lemme retenu. En cas de doute, il peut toujours revenir

à un contexte élargi en cliquant sur le numéro de la page ou sur la référence indiquée. Dans certains cas, la levée d’ambiguïtés est la même pour l’ensemble des occurrences ; il est alors possible de sélectionner « Forcer l’analyse aux x attestations ». Mais il existe des formes véritablement ambiguës dans l’état de langue, pour lesquelles il faut examiner attentivement le contexte, en l’absence d’un « étiqueteur » morphosyntaxique précis et fiable. Il s’agit principalement des mots grammaticaux. Ainsi, seul le contexte permettra de dire si la forme *se* est un pronom réfléchi, un morphème introduisant une interrogative indirecte ou une hypothétique, voire une variante diatopique de l’adverbe *si*. Le lemmatiseur n’est d’aucune utilité dans ce cas ; il permet seulement de rassembler toutes les occurrences de la forme *se*. L’éditeur d’un texte en moyen français peut chercher surtout une aide pour construire son glossaire et rassembler toutes les données lexicales utiles à la compréhension du texte édité. Mais l’analyse des mots grammaticaux est utile dans les études linguistiques. Dans le formulaire de levée d’ambiguïtés, il est également possible de choisir d’autres options que les lemmes proposés, en saisissant un autre lemme de la base de connaissance ou en intégrant un nouveau lemme ; il est possible aussi d’abandonner la lemmatisation en sélectionnant par un clic l’un des cas suivants : nom étranger, nom propre, dénomination, nombre, mot exclu.

Ressources produites

Ajoutons que l’utilisateur de l’outil n’est nullement tenu de lever les ambiguïtés : il peut se contenter des ressources produites automatiquement pour générer de façon traditionnelle son glossaire ; il peut aussi lever les ambiguïtés de façon partielle et sélective ; le seul cas qui impose une levée systématique des ambiguïtés est celui de la diffusion du texte lemmatisé – ce que la plateforme en ligne du *DMF* nomme, par commodité, « édition électronique ». Depuis 2011, le texte au programme des agrégations de lettres est ainsi traité chaque année au laboratoire ATILF grâce à

l'outil-glossaire¹⁷. Le résultat du traitement automatique par LGeRM est vérifié complètement, les ambiguïtés sont levées et les lemmes contrôlés. C'est le projet finalisé qui est diffusé sur la plateforme du *DMF*: l'outil-glossaire offre un certain nombre de fonctionnalités liées à la lemmatisation du texte numérisé. Selon le statut du texte, le lecteur pourra ou non consulter le texte intégral: mais dans tous les cas, il aura la possibilité de connaître sa nomenclature, d'interroger par lemme ou par forme et d'exporter les résultats sous la forme d'un index lemmatisé ou d'une liste d'exemples. L'intérêt pour le linguiste comme pour le littéraire est évident. Le premier peut interroger les formes du texte, commenter leur distribution et leur fréquence. Le regroupement par lemme permet en outre d'avoir accès aux formes attestées des paradigmes morphologiques, notamment pour les verbes les plus usuels. Le second peut s'appuyer sur des données lexicales et sémantiques exhaustives et confronter différentes occurrences de mots-clés, des adjectifs tels que *courtois* ou *vilain* dans le *Roman d'Eneas*, ou un substantif comme *nonchaloir* dans les poèmes de Charles d'Orléans.

En deçà et au-delà du *Dictionnaire du moyen français*

Même si cette contribution, en insistant sur l'origine de l'outil, marque ses liens étroits avec le *DMF*, elle affirme aussi ses développements originaux. L'outil de lemmatisation, créé pour la gestion de la variation dans le dictionnaire, a élargi son champ d'intervention. Les textes médiévaux traités n'appartiennent pas tous à la période du moyen français, qui est la période de référence de notre dictionnaire; des textes d'ancien français

17. Charles d'Orléans, *Poésies* (2010-2011): <http://atilf.fr/dmf/CharlesOrleans>; Béroul, *Tristan* (2011-2012): <http://atilf.fr/dmf/Beroul>; Guillaume de Lorris, *Le Roman de la Rose*, éd. Armand Strubel (2012-2013): <http://atilf.fr/dmf/RomanRoseStrubel>; *Le Couronnement de Louis*, éd. Ernest Langlois (2013-2014): <http://atilf.fr/dmf/CouronnementLouis>; *Le Roman d'Eneas* (2014-2015): <http://atilf.fr/dmf/RomanEneas>; Jean Renart, *Le Roman de la Rose ou de Guillaume de Dole* (2015-2016): <http://atilf.fr/dmf/RomanRoseGuillaumeDole>; Christine de Pizan, *Le Livre du duc des vrais amants*, éd. Dominique Demartini et Didier Lechat (2016-2017): <http://atilf.fr/dmf/PizanVraisAmans>.

ont pu être lemmatisés avec des ajustements mineurs. L’outil-glossaire a également connu des développements au-delà de la période de référence : même si les vérifications et les interventions manuelles sont plus importantes que pour le moyen français, LGeRM peut s’adapter aux textes antérieurs (ancien français) comme aux textes postérieurs (xvi^e et xvii^e siècles).

Pour ces derniers, la participation de l’ATILF au projet européen Impact¹⁸ a été déterminante. En effet, pour chacun des pays partenaires, la collaboration entre bibliothèque nationale et laboratoire de linguistique historique avait pour but d’améliorer les logiciels d’océrisation utilisés sur les documents anciens. Dans ce cadre, Gilles Souvay a adapté son outil pour répondre à cette nouvelle utilisation. L’ajout de formes modernes Morphalou, autre ressource de l’ATILF¹⁹, lui a permis de construire un lexique capable de couvrir très largement la période choisie pour l’expérimentation (xvii^e siècle). Le lexique moderne « archaïsé » a été projeté sur un corpus textuel issu de Frantext et sur un corpus de seize textes ayant conservé leur graphie d’origine, numérisés dans le cadre du projet pour constituer la « vérité terrain ». L’utilisation combinée des ressources a montré son efficacité dans le processus d’adaptation à une période intermédiaire entre le moyen français et la langue couverte par le *TLF* et Morphalou. Un projet synthétique portant sur la variation dans les états anciens du français pourrait permettre de relier et d’approfondir les résultats déjà obtenus sur des périodes particulières.

Tous ces développements se greffent sur un projet initial, celui du *DMF*, qui a su prendre le tournant d’une lexicographie véritablement évolutive, ce qui implique de dépasser non seulement la construction lettre par lettre du dictionnaire, mais aussi d’en faire éclater les limites dans la consultation, grâce au balisage des données et aux liens hypertextuels. Au cœur du

18. « Improving Access to Text » : <http://www.impact-project.eu>.

19. Le lexique Morphalou est un lexique des formes fléchies du français, à large couverture (540 000 formes). Les données initiales proviennent du *TLFnome*, la nomenclature du *Trésor de la langue française*. Il est en accès libre à des fins de recherche et d’enseignement et sa mise à jour est assurée par l’ATILF.

dispositif, le dictionnaire proprement dit demeure la source et la référence, à travers des versions datées qui sont archivées et une dernière version, directement disponible en ligne, qui reprend, corrige et enrichit la précédente.

Deux thèses en lien avec ce projet indiquent des champs encore à défricher et à labourer : la première, achevée, s'est emparée d'un des corpus initiaux du *DMF* abandonné en cours de route, celui des premiers récits de voyage en français (xiv^e et xv^e siècles) pour en établir le lexique en s'appuyant sur l'outil-glossaire²⁰. Ce lexique des récits de voyage ne saurait ressembler aux lexiques initiaux qui devaient servir de ressources pour construire le futur *DMF*; il se définit plutôt comme une ressource pour un enrichissement raisonné du dictionnaire. La deuxième porte sur les prépositions en moyen français²¹, en lien avec le projet Presto auquel Gilles Souvay a également collaboré en fournissant des lexiques morphologiques²² – ce qui devrait permettre de réfléchir aux mots outils jusque-là un peu délaissés²³. À ces travaux étroitement liés au *DMF* s'ajoutent pour la lemmatisation de textes et le traitement des données lemmatisées des partenariats avec des projets extérieurs, l'un avec Michèle Goyens (Université catholique de Louvain),

20. Capucine Herbert, *Les Récits de voyage des xiv^e et xv^e siècles lemmatisés : apports lexicographiques au Dictionnaire du moyen français* [thèse de doctorat en sciences du langage sous la dir. de Sylvie Bazin-Tacchella, Université de Lorraine, 2016]. Voir également <http://www.atilf.fr/dmf/RecitsVoyage>.

21. Claire Schlienger, *Le Syntagme prépositionnel marquant l'inclusion : analyse diachronique de en, dans, dedans et à* [thèse de doctorat en sciences du langage sous la dir. de Sylvie Bazin-Tacchella, en préparation à l'Université de Lorraine].

22. Presto (*L'Évolution du système prépositionnel du français : approche diachronique et quantitative*) est un projet ANR/DFG (2013-2016), coordonné par Denis Vigier (Lyon II). Dans le cadre de ce projet, Gilles Souvay a participé au développement d'outils de lemmatisation adaptés à toutes les périodes du français en fournissant deux lexiques morphologiques, un lexique LGeRM médiéval et un lexique LGeRM xviii^e siècle : le premier, optimisé pour la période 1300-1500, comporte 66 976 lemmes et étiquettes *DMF*, soit 880 192 entrées dont 142 687 attestées dans la base Frantext; le second, optimisé pour la période 1550-1700, comporte 89 754 lemmes et étiquettes *TLF*, soit 2 959 371 entrées dont 116 161 attestées (3,9 %).

23. Il existe un autre projet collectif lié au *DMF*, sous l'égide de Bernard Combettes, sur les locutions temporelles. Il faut également signaler que la nomenclature des mots grammaticaux et un lexique des préfixes et suffixes (Robert Martin) sont déjà accessibles sur le site à la rubrique « Compléments au *DMF* 2012 », en attendant d'être publiés dans la version 2015.

sur la construction du vocabulaire médical au Moyen Âge et à la Renaissance²⁴ et l'autre avec les historiens médiévistes de l'Université de Lorraine²⁵.

Nous nous interrogeons en 2013 sur les nouveaux défis que devait relever le *DMF*²⁶. Cette contribution qui rappelle l'origine et les développements de l'outil-glossaire du *DMF* indique les chantiers actuels et à venir. L'adaptation d'un tel outil à la langue et aux textes du moyen français, puis l'élargissement de son domaine d'intervention s'inscrivent bien dans le dynamisme et le réalisme d'une lexicographie évolutive qui n'abandonne pas les projets ambitieux, mais choisit de les faire avancer par étape.

24. *Latin authority and constructional transparency at work: neologisms in the French medical vocabulary of the Middle Ages and the Renaissance and their fate.*

25. AMPLor: *Actes médiévaux des princes lorrains; vers un corpus numérisé*, projet coordonné par I. Guyot-Bachy.

26. Sylvie Bazin-Tacchella et Gilles Souvay, « Avec la version 2012, la fin d'un projet ou de nouveaux défis pour le *DMF*? », *XXVII^e Congrès international de linguistique et de philologie romanes, CLPR*, Nancy, 2013 [communication non publiée].

COMITÉ SCIENTIFIQUE

Hava BAT-ZEEV SHYLDKROT (Université de Tel Aviv)
Françoise BERLAN (Sorbonne Université)
Mireille HUCHON (Sorbonne Université)
Peter KOCH (Universität Tübingen)†
Anthony LODGE (Saint Andrews University)
Christiane MARCHELLO-NIZIA (École normale supérieure-LSH, Lyon)
Robert MARTIN (Sorbonne Université/Académie des inscriptions et belles-lettres)
Georges MOLINIÉ (Université Paris-Sorbonne)†
Claude MULLER (Université Bordeaux Montaigne)
Laurence ROSIER (Université Libre de Bruxelles)
Gilles ROUSSINEAU (Sorbonne Université)
Claude THOMASSET (Sorbonne Université)

COMITÉ DE RÉDACTION

Claire BADIOU-MONFERRAN (Université Sorbonne Nouvelle)
Michel BANNIARD (Université Toulouse 2-Le Mirail)
Annie BERTIN (Université Paris Ouest Nanterre La Défense)
Claude BURIDANT (Université Strasbourg 2)
Maria COLOMBO-TIMELLI (Università degli Studi di Milano Statale)
Bernard COMBETTES (Université de Lorraine)
Frédéric DUVAL (École nationale des chartes)
Pierre-Yves DUFEU (Université Aix-Marseille 3)
Amalia RODRIGUEZ-SOMOLINOS (Universidad Complutense de Madrid)
Philippe SELOSSE (Université Lyon 2)
Christine SILVI (Sorbonne Université)
André THIBAUT (Sorbonne Université)

COMITÉ ÉDITORIAL

Olivier SOUTET (Sorbonne Université),
Directeur de la publication
Joëlle DUCOS (Sorbonne Université-EPHE),
Trésorière
Stéphane MARCOTTE (Sorbonne Université),
Secrétaire de rédaction
Thierry PONCHON (Université de Reims Champagne-Ardenne),
Secrétaire de rédaction
Antoine GAUTIER (Sorbonne Université),
Diffusion de la revue

Résumés

Julie GLIKMAN et Thomas VERJANS,
Regards linguistiques sur les éditions
de textes médiévaux

Résumé

Cette contribution constitue l'introduction du volume. Elle présente le contexte dans lequel ce numéro a été préparé et la volonté des directeurs du volume d'interroger les rapports entre les pratiques philologiques et les études de linguistique diachronique. Ces rapports peuvent se mesurer dans la place accordée aux faits linguistiques dans les introductions d'édition, ou inversement la place accordée aux variantes et à l'apparat critique dans les corpus numérisés. Elle présente ensuite les différentes contributions du volume.

Abstract

This contribution is the introduction to the volume. It presents the context in which this issue was prepared and the willingness of the editors to question the relationship between philological practices and studies of diachronic linguistics. These relationships can be evaluated by considering the importance given to linguistic facts in the introductory sections of editions. Conversely, it can also be evaluated by according to the importance given to variants and critical apparatus in digitized corpora. The various contributions of the volume are also introduced.

Nathalie BRAGANTINI-MAILLARD,
 Suivre la lettre du copiste : l'édition critique
 au service de la linguistique diachronique et
 diatopique. L'exemple du ms. Paris, BnF, fr. 99

Résumé

La connaissance des modalités d'évolution du français à la fin du Moyen Âge ne peut désormais s'affiner sans une reconnaissance véritable du rôle crucial que jouèrent les copistes au plan linguistique dans la diffusion et la survie des textes anciens. L'action du copiste est en effet double, en s'exerçant à la fois sur le plan horizontal de la circulation des textes d'un espace linguistique à un autre et sur le plan vertical de la transmission des textes à travers les époques. Dans la pratique scientifique, la prise en compte de cet apport déterminant doit passer non seulement par une édition des textes plus respectueuse de la version procurée par un manuscrit donné, mais aussi par un examen documenté, exhaustif et précis des phénomènes linguistiques qui particularisent les témoins retenus et les modifications de scribe. À terme, l'information rassemblée par ces profils linguistiques devrait permettre de mieux appréhender les phénomènes d'adaptation, de rajeunissement et d'enrichissement du français au Moyen Âge. À titre illustratif, nous nous proposons de montrer l'intérêt que présente le ms. BnF, fr. 99 pour suivre de manière privilégiée certains phénomènes de modernisation du français dans la seconde moitié du xv^e siècle, ainsi que l'influence que put exercer le lieu de copie occitanisant sur l'adaptation linguistique du texte, autrement dit les conditions d'échanges entre oïl et oc.

Abstract

Knowledge of how French evolved in the late Middle Ages can no longer be refined without a genuine recognition of the crucial linguistic role played by copyists in the dissemination and survival of ancient texts. Copyists act both on the horizontal dimension of the circulation of texts from one linguistic space to another, and on the vertical dimension of the transmission of texts through

the ages. This decisive contribution must be taken into account, not only by providing edition of the texts that are faithful to the version of a given manuscript, but also by a comprehensive and precise examination of the linguistic phenomena that characterize the witnesses and scribal modifications. Ultimately, these linguistic profiles will provide information for a better understanding of the phenomena of adaptation, rejuvenation and enrichment of French in the Middle Ages. To illustrate this, we examine ms. BnF, fr. 99, which displays exceptionally well certain phenomena of the modernization of French in the second half of the 15th century. It also demonstrates the influence that the place of copying with an affinity for Occitan may have had on the linguistic adaptation of the text, i.e. the conditions of exchange between Oïl and Oc.

Laurent BALON,
**Pour une « troisième voie » en matière d'édition
de textes d'ancien et de moyen français**

Résumé

La pratique de l'édition de texte se trouve face à un dilemme : en partant des conseils trouvés dans les quelques articles sur la question et les manuels récents donnant des principes d'édition, on observe que les critères actuels de choix des variantes aboutissent à l'exclusion du matériau intéressant le linguiste qui, de son côté, aurait besoin d'un exposé intégral de toutes les données, sans tri. Ce besoin d'un non-choix est important, mais peu facile à satisfaire, voire impraticable à l'écrit, et la présentation des données intégrales du manuscrit se heurte à la lisibilité et à l'intelligibilité. L'objet de cette contribution est de présenter une méthode d'édition constituant un compromis entre l'édition critique traditionnelle et la transcription dite diplomatique, reposant sur un protocole de choix de variantes permettant de mieux satisfaire certains besoins des linguistes. Afin de fournir au linguiste des informations immédiatement exploitables et utiles à l'avancée de la discipline, le principe méthodologique proposé consiste à signaler dans l'édition

certains faits de langue relevant de la ponctuation du mot par l'emploi d'un code graphique qui en conserve la trace, à savoir un système de « tirets » déjà suggéré par Jacques Monfrin pour la transcription des documents d'archives, mais complété et appliqué pour la première fois à un texte littéraire par Nelly Andrieux-Reix. Le bien-fondé et l'intérêt de cette méthode seront illustrés par des études de cas en lien avec notre propre travail de recherche.

Abstract

Editors must cope with a dilemma: according to publishing principles in recent papers and textbooks, the current criteria for choosing variants excludes materials of great interest to linguists. They would need a comprehensive view of the data, without sorting. This is not easy to achieve, and even impossible on paper. The full presentation of the data of the manuscript hampers legibility and intelligibility. The purpose of this contribution is to present a compromise between traditional critical editing and diplomatic transcription, based on a protocol of choice of variants that better satisfies linguistic investigations. The proposed methodological principle aims at providing information that is immediately usable and useful for the advancement of the linguistics. This purpose is achieved by indicating facts relating to the punctuation of the word by using a graphic code that keeps track of them: a system of “dashes”, suggested by Jacques Monfrin for the transcription of archival documents. This system is expanded and applied for the first time to a literary text by Nelly Andrieux-Reix. The merits and interest of this method will be illustrated by case studies related to our own research work.

Alexei LAVRENTIEV, Céline GUILLOT-
BARBANCE et Serge HEIDEN,
Enjeux philologiques, linguistiques et informatiques
de la philologie numérique :
l'exemple de la segmentation des mots

Résumé

Les linguistes travaillant sur l'histoire de la langue ont toujours exploité et utilisé comme principale source d'exploration les éditions « classiques », bien que depuis longtemps on connaisse leurs limites pour la recherche linguistique. Le développement des technologies modernes a d'un autre côté rendu le recours à de nouveaux outils (concordances, index, calculs statistiques) peu à peu indispensable à la recherche en langue, et plus récemment, les progrès continus de la technologie ont également permis d'envisager la réalisation d'éditions d'un nouveau type. L'édition numérique, qui a déjà donné lieu à plusieurs réalisations concrètes, a ainsi permis aux linguistes auparavant bridés par le papier et les techniques traditionnelles d'exprimer plus librement leurs besoins et leurs exigences. Plusieurs recherches récentes déjà publiées montrent l'efficacité de ce mouvement et le caractère novateur des acquis ainsi obtenus. À partir d'un exemple concret d'édition numérique interactive, notre présentation détaillera les enjeux méthodologiques liés à ces nouveaux outils et à ces nouvelles pratiques, en proposant une réflexion sur le concept de « philologie numérique » et en montrant ses principaux apports pour la recherche diachronique. Cette question sera illustrée en particulier par la question de la segmentation des mots.

Abstract

Linguists working on the history of language have always exploited “classical” editions as their main source of exploration, although the limits of such resources for linguistic research have long been known. On the other hand, modern technology has gradually offered new tools (concordances, indices, statistical calculations), that now prove to be indispensable. More recently,

the continuous progress has also made it possible to produce new types of editions. Digital publishing, which has already produced several achievements, has thus enabled linguists to express their needs and requirements better than before, freed from the constraints of paper and traditional techniques. Several recent studies demonstrate the efficiency of digital publishing and the innovative nature of the results obtained. Based on an example of interactive edition, we survey the methodological issues related to these new tools and practices, by investigating the concept of “digital philology”, and by evaluating how it contributes to diachronic research. The specific issue of word segmentation will illustrate our point.

Nicolas MAZZIOTTA,
 L'activité éditoriale comme démarche
 de représentation de la connaissance :
 l'exemple de la ponctuation médiévale

Résumé

Cette contribution concerne le traitement éditorial de la ponctuation médiévale, selon une approche de la philologie comme activité de représentation des connaissances. Après une présentation des concepts de *connaissance* et d'*inscription* (des connaissances), le traitement de la ponctuation médiévale sert d'exemple aux questionnements que soulève toute activité éditoriale. Dans la démarche ecdotique, il s'agit d'identifier des classes de signes, pour distinguer ce qui est différent et rapprocher ce qui est similaire, mais également de segmenter correctement les unités observées. En outre, éditer consiste à « donner à lire », ce qui se manifeste par l'importance de choix ergonomiques importants pour garantir l'accessibilité de la connaissance inscrite. À bien des égards, l'inscription informatique de l'édition a beau ouvrir le champ des possibles, elle ne résout pas tout. Pour inscrire, il faut d'abord comprendre. L'édition ne pourra jamais se passer des *choix* foncièrement humains qui fondent le travail de construction de la connaissance.

Abstract

This contribution focuses on the editorial treatment of medieval punctuation, according to an approach of philology as an activity of *knowledge representation*. After a brief presentation of the concepts of *knowledge* and *inscription* (of knowledge), the treatment of medieval punctuation serves as an example for the questions raised by any editorial activity. Identifying classes of signs and distinguishing between what is different and what is similar are key parts of the ecdotic process. Moreover, by editing a text, one actually *makes it readable*. Consequently, ergonomic choices are prominent in this process, in order to guarantee the accessibility of the knowledge inscribed. In many respects, digital publishing opens up the field of possibilities, but it does not solve the fundamental problems. Understanding the text stands as the first step into building any valuable critical edition. Human *choices* will always remain the basis of any elaboration of knowledge.

Sylvie BAZIN-TACHELLA et Gilles SOUVAY,
Lemmatisation et construction automatique
de ressources lexicographiques :
les développements du lemmatiseur LGeRM

Résumé

Le lemmatiseur LGeRM, conçu à l'origine pour faciliter la consultation du *Dictionnaire du moyen français*, a connu depuis 2008 de nouveaux développements et est aujourd'hui utilisé dans de nombreux autres contextes, notamment dans l'interrogation de bases textuelles et la constitution de lexiques ou glossaires informatisés, autant d'outils qui peuvent servir d'aide à l'édition, le lemmatiseur ayant été intégré depuis à plusieurs grands projets d'édition en ligne. Cette contribution se propose de retracer l'histoire de la conception de LGeRM et de ses développements successifs, en montrant les différentes possibilités de l'outil illustrées à partir des projets récents.

Abstract

The LGeRM lemmatizer, originally designed to facilitate the consultation of the *Dictionnaire du moyen français*, has undergone new developments since 2008. It is now used in many other contexts. In particular, it helps the interrogation of textual bases and the constitution of computerized lexicons or glossaries. Additionally, the lemmatizer has also been integrated into several major online publishing projects in order to help the publishing process. This contribution retraces the history of the conception of LGeRM and its successive developments, by showing how recent projects make use of it.

Table des matières

Regards linguistiques sur les éditions de textes médiévaux Julie Glikman & Thomas Verjans	7
Suivre la lettre du copiste : l'édition critique au service de la linguistique diachronique et diatopique. L'exemple du ms. Paris, BnF, fr. 99 Nathalie Bragantini-Maillard	17
Pour une « troisième voie » en matière d'édition de textes d'ancien et de moyen français Laurent Balon	47
Enjeux philologiques, linguistiques et informatiques de la philologie numérique : l'exemple de la segmentation des mots Alexei Lavrentiev, Céline Guillot-Barbance & Serge Heiden	77
L'activité éditoriale comme démarche de représentation de la connaissance : l'exemple de la ponctuation médiévale Nicolas Mazziotta	103
Lemmatisation et construction automatique de ressources lexicographiques : les développements du lemmatiseur LGeRM Sylvie Bazin-Tacchella & Gilles Souvay	121
Résumés/Abstracts.....	147

