Français médiéval en diachronie : du corpus à la langue (mémoire de synthèse)

Sophie Prévost

▶ **To cite this version:**

Sophie Prévost. Français médiéval en diachronie : du corpus à la langue (mémoire de synthèse). Linguistique. Ecole normale supérieure de lyon - ENS LYON, 2011. tel-00667107

HAL Id: tel-00667107
https://theses.hal.science/tel-00667107

Submitted on 6 Feb 2012

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

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

FRANÇAIS MÉDIÉVAL EN DIACHRONIE : DU CORPUS À LA LANGUE

Sophie Prévost

Chargée de recherche au Lattice (UMR 8094), ENS /Paris-3

Mémoire de synthèse
en vue de l'obtention de
l'Habilitation à diriger des recherches
sous la direction de
M. Benoît Habert, Professeur des universités

Jury :

M. Michel Charolles
M. Bernard Combettes
M. Benoît Habert
Mme Christiane Marchello-Nizia
Mme Marie-Paule Péry-Woodley
Mme Lene Schøsler

Ecole Normale Supérieure de Lyon
2011

Remerciements

Je tiens à remercier en premier lieu Benoît Habert, qui accepté d'encadrer ce travail. Les conseils qu'il m'a donnés et les pistes qu'il m'a suggérées ont fait de cette démarche rétrospective et exploratoire une étape importante dans mon parcours.

Je remercie vivement aussi l'ensemble des rapporteurs et des membres du jury : Michel Charolles, Bernard Combettes, Christiane Marchello-Nizia, Marie-Paule Péry-Woodley et Lene Schøsler.

Benjamin Fagard, Christiane Marchello-Nizia, Thierry Poibeau et Bernard Victorri ont généreusement accepté de relire l'un ou l'autre des mémoires présentés : qu'ils en soient remerciés.

J'ai trouvé depuis dix ans au Lattice des conditions de travail d'une grande qualité, des échanges fructueux et une chaleur humaine que j'apprécie au quotidien. Que tous les membres du Lattice soient remerciés, en particulier pour leurs encouragements au cours de ces derniers mois.

Les projets avec l'ENS de Lyon autour de la *Base de Français Médiéval* n'ont jamais cessé. J'ai noué depuis des années avec Céline Guillot, Serge Heiden, et Alexei Lavrentiev des liens scientifiques et amicaux qui me sont précieux.

L'amitié de Fernande Dupuis à Montréal l'est tout autant.

Enfin, Christiane Marchello-Nizia et Bernard Combettes m'accompagnent depuis de nombreuses années de leur intelligence de la langue et de leur générosité intellectuelle. Je leur suis profondément reconnaissante de tout ce qu'ils m'ont apporté.

Préambule

Voilà plus de quinze ans que l'essentiel de mes recherches porte sur des états anciens du français et se situe dans une perspective diachronique.
Il est toujours utile de prendre du recul vis-à-vis de son activité, c'est-à-dire de se demander si les raisons et les motivations qui ont poussé à s'engager dans cette voie sont toujours présentes (ou bien si d'autres ont pris le relais), si ce que l'on fait a toujours un sens (pour soi) et un intérêt pour la communauté (ce qui n'est certes pas aisé à évaluer de manière objective).
Prendre du recul, c'est donc interroger le « pour-quoi », mais c'est aussi mettre en cause le « comment ». Arrive un moment où il est nécessaire de faire la synthèse de ces questionnements. Ce mémoire est l'occasion de mettre en forme cette réflexion, tout en présentant le bilan de mes travaux et de mon engagement dans la communauté des linguistes.

à mon père

TABLE des MATIERES

Introduction 1

Chapitre 1. Travailler en langue ancienne : l'absence de locuteurs 5

1.1. Quand le possible de langue rejoint le possible attesté 6

1.2. Construction des savoirs et émergence des grammaires 15
1.2.1. Les grammaires : du latin au français, du français à la langue ancienne 15
1.2.2. Observer la langue à travers le filtre des grammaires 17

1.3. Des classifications héritées aux classifications construites 18
1.3.1. Classification spontanée et classifications volontaires..... 18
1.3.2. Les parties du discours 21
1.3.3. Les fonctions syntaxiques..... 23
1.3.4. Des classes en construction 26

1.4. L'adéquation des cadres d'analyse 27

Chapitre 2. Travailler en diachronie 29

2.1. De l'intérêt d'étudier l'évolution d'une langue 29

2.2. Diachronie et synchronie : de l'opposition à la complémentarité 31

2.3. Expliquer l'évolution 34

2.4. Le facteur temps : la périodisation de la langue et la chronologie des phénomènes 37

2.5. Le facteur temps : le rythme des changements 41
2.5.1. La nature des changements : changements catastrophiques/progressifs 41
2.5.2. La vitesse des changements 42

Chapitre 3. Travailler en corpus 45

3.1. Les textes comme seul accès à la langue..... 45
3.1.1. L'absence d'oral 46
3.1.2. Une vision à la fois partielle et totale de la langue 47

3.2. Une tradition de travail sur les textes renouvelée par la numérisation des textes et la linguistique de corpus 49
3.2.1. Un nouveau rapport aux textes et aux données 49
3.2.2. D'une approche individuelle à un fonctionnement collectif 59
3.2.3. Allier la lecture outillée à la connaissance des textes 61

Chapitre 4. Classifier **65**

4.1. L'enrichissement des corpus 65
4.1.1. Fournir un outil de travail à la communauté 65
4.1.2. L'évolution des classes 68
4.1.2.1. La polycatégorisation des formes 70
4.1.2.2. Unités graphiques et unités linguistiques 73

4.2. Evolution des constructions et changement de classes 74
4.2.1. Evolution des caractéristiques des éléments d'une classe 76
4.2.1.1. Les sujets : d'une valeur sémantico-pragmatique à un statut d'indice grammatical 76
4.2.1.2. Les constructions détachées 79
4.2.1.3. Les éléments initiaux 81

4.2.2. Changements de classes et grammaticalisation 82
4.2.2.1. Migrations entre classes prédéfinies 82
4.2.2.2. Emergence de classes inédites : marqueurs de topicalisation et marqueurs discursifs 85

Conclusion **91**

Références bibliographiques **93**

Introduction[*]

Pendant des siècles, les études sur la langue ont porté sur les langues anciennes, le latin et le grec en l'occurrence. Objet d'attention exclusif tant que le vernaculaire sera considéré avec mépris, le latin gardera encore longtemps son statut privilégié quand le français aura acquis ses lettres de noblesse. A partir du 19ème siècle, les états anciens du français constitueront un vaste champ d'exploration, inaugurant une tradition philologique qui ne s'est pas démentie depuis, et qui a largement contribué à la description de la langue ancienne. Tradition philologique complétée, parfois concurrencée, par une linguistique qui sera d'abord « historique »[1], et prioritairement centrée sur la reconstitution de l'indo-européen. Philologie et linguistique : deux champs de recherche qui resteront longtemps en partie disjoints, représentés au tournant des 19ème et 20ème siècles par deux figures emblématiques : F. Brunot et A. Meillet[2].

Cent ans plus tard, il nous reste l'immense richesse complémentaire des apports des uns et des autres. Mais cent ans plus tard il apparaît aussi qu'il ne va plus autant de soi de travailler sur la langue ancienne, et cela pour au moins deux raisons, d'ordres différents.

La première tient à ce que l'on peut s'interroger sur l' « intérêt » qu'il y a à étudier des états anciens du français. Certains pourront en effet arguer que l'on connaît désormais bien la langue ancienne. C'est à la fois vrai et faux : l'accumulation des recherches depuis des dizaines d'années a certes permis d'acquérir une bonne connaissance de cette langue, mais il reste toujours des zones à explorer, et l'étude de nouveaux textes, ainsi que le renouveau des modalités de cette étude grâce à leur numérisation et aux technologies qui se sont développées conjointement[3] conduit parfois à réviser certaines descriptions et/ou analyses antérieures. A ceux qui sont sceptiques sur

[*]Je tiens à remercier Christiane Marchello-Nizia, Thierry Poibeau et Bernard Victorri pour leur lecture de ce mémoire et pour les remarques qu'ils m'ont faites.

[1] L'introduction du couple « diachronie » / « synchronie » est due à Saussure. La linguistique historique s'oppose à la linguistique diachronique en ce qu'elle tient compte de l'histoire externe (décisions institutionnelles, événements historiques, emprunts, contacts…), alors que la linguistique diachronique ne s'occupe que d'histoire interne. Les deux se distinguent de la linguistique synchronique dans la mesure où elles envisagent le changement. Dans les faits, l'expression « linguistique historique » sert parfois à désigner toute étude portant sur la langue ancienne, qu'elle soit ou non diachronique.

[2] Voir à cet égard l'introduction du livre de J.-C. Chevalier et P. Encrevé (2006) : *Combats pour la linguistique : de Martinet à Kristeva. Essai de dramaturgie épistémologique*.

[3] Nous reviendrons sur cette question en 3.2., mais disons d'ores et déjà que la massification des données traitées, grâce à leur passage (même partiel) sur support numérique, ainsi que le possible recours à des outils de recherche et de calcul a largement renouvelé l'approche des textes.

l'intérêt, en soi, de travailler sur une langue ancienne, c'est-à-dire sur une réalité déconnectée du présent, on peut répondre par le plaisir de la connaissance, pour soi, et qu'il est de notre devoir de transmettre : plaisir à se plonger dans une langue à la fois étrangère et familière[4], dans les textes, et donc dans une culture, puisqu'à travers les textes on s'y plonge toujours forcément un peu.

Ayant justifié l'intérêt d'étudier la langue ancienne, il convient de dire que cette exploration ne constitue cependant pas, dans ma recherche, un but en soi. Elle est liée au fait que la période pertinente pour étudier les phénomènes linguistiques auxquels je me suis intéressée a pour borne initiale l'ancien ou le moyen français, et pour borne finale, selon les cas, le 16ème ou le 17ème siècle, époque à laquelle les constructions impliquées ont acquis les caractéristiques majeures qu'elles ont en français moderne[5]. C'est donc d'emblée dans une perspective diachronique que je me suis située, approche dont l'« intérêt » sera exposé plus loin.

Il reste que mes recherches portent sur le « français ancien[6] », et que, seconde raison du nécessaire questionnement de cette activité, cela suppose des modes de raisonnement et une méthodologie spécifiques, qui peuvent sembler quelque peu « exotiques » à qui travaille aujourd'hui sur un état de langue moderne. Il serait d'ailleurs plus juste de considérer que l'on travaille *dans* un état de langue moderne, du fait même que celui-ci n'est pas révolu. Cette *localisation* différente du linguiste vis-à-vis de son objet distingue notablement l'étude d'une langue ancienne de celle d'une langue moderne.

Envisager la langue dans une perspective diachronique suppose aussi une démarche particulière, mais les deux aspects – langue ancienne et diachronie – doivent être disjoints. On opère parfois un amalgame entre les deux : travailler en diachronie

[4] Familière par sa fréquentation, mais aussi par les liens qu'elle entretient avec le français moderne.

[5] Autant le choix de la borne initiale s'impose lorsque celle-ci correspond à l'émergence d'une construction (c'est le cas pour les marqueurs de topicalisation : *à propos de*, *en ce qui concerne*, etc.), autant il ne va pas de soi pour ce qui concerne des constructions qui se rencontrent dès les premiers siècles du « français » : pourquoi ne pas remonter plus haut ? Le choix de considérer l'ancien français « classique » des 12ème-13ème siècles est doublement motivé : d'une part c'est une période qui connaît une relative stabilité linguistique, et d'autre part (et cela légitime d'ailleurs l'affirmation précédente), c'est une période bien documentée au sens où elle nous offre une abondance de textes, la période précédente (9ème-11ème) se caractérisant au contraire par la rareté des textes qui nous sont connus (d'emblée, on le voit, il est difficile de dissocier l'accès à la langue et le recours aux textes). Il reste que c'est un choix qui n'est pas complètement évident, et qui mérite d'être explicité : ne considérer la syntaxe du sujet qu'à partir du 12ème siècle revient à occulter la préhistoire du phénomène, aussi partielle que soit la connaissance que nous pouvons en avoir.

[6] Terme que j'emploie ici pour couvrir la période du 12ème au 17ème siècle, qui peut correspondre à la période dans son intégralité, ou bien seulement à l'ancien français, ou au moyen français : il s'agit d'une commodité langagière pour désigner un ou plusieurs état(s) ancien(s) du français.

impliquerait que l'on travaille sur une langue ancienne. Or on peut adopter une perspective diachronique pour traiter de l'évolution d'un phénomène linguistique durant ces 50 dernières années ; à l'inverse, l'approche synchronique sur un état ancien de la langue est fréquente. Il est vrai que la conjonction de la visée diachronique et du caractère ancien de la langue renforce certaines difficultés propres à chacun des deux points de vue, mais il est cependant préférable, par principe, de considérer les deux aspects de manière disjointe. De la même façon, je traiterai séparément, par la suite, la question des corpus : même si langue et corpus sont intrinsèquement liés quand il s'agit de langue ancienne, chacun d'eux a cependant une problématique autonome, et il est souhaitable de ne pas les assimiler l'un à l'autre.

Logiquement, c'est par les spécificités liées à l'objet « langue ancienne » que je commencerai, dans la mesure où elles sont présentes aussi dans la démarche diachronique dès lors que celle-ci est appliquée au français ancien.

Chapitre 1. Travailler en langue ancienne : l'absence de locuteurs

Les différents états de langue ancienne ont un point commun et une caractéristique majeure : ils sont dépourvus de locuteurs vivants. Non seulement nous n'avons pas de compétence pour ces états langagiers (mais c'est après tout aussi le cas pour qui travaille sur une langue vivante qu'il ne parle pas), mais nous ne pouvons faire appel à des informateurs. Si la langue du 16ème siècle nous semble plus proche de notre propre pratique langagière, l'on sait que ce sentiment de familiarité est trompeur. C'est encore vrai pour la langue du 17ème : certes, bon nombre de constructions syntaxiques sont acquises, avec leur valeur contemporaine, mais les faux-amis persistent, en particulier pour ce qui touche au lexique. Ainsi en est-il du mot *cœur*, dont la valeur sémantique de « courage », est encore bien vivante au 17ème siècle (à côté des autres valeurs : « organe » et « siège de sentiments »), comme en témoignent ces vers du *Cid* de Corneille :

(1) Rodrigue, as-tu du cœur ? (acte I, sc. 6)

(2) Non, non, en ce combat, quoy que vous veuilliez croire,
Rodrigue peut mourir sans hazarder sa gloire,
Sans qu'on l'ose accuser d'avoir manqué de coeur,
Sans passer pour vaincu, sans souffrir un vainqueur. (acte V, sc. 1),

le sens de « courage » ne s'étant maintenu en français moderne que dans quelques devises ou expressions figées et vieillies : *Haut les cœurs !* (titre d'un film de Sólveig Anspach (1999)), ou bien *le cœur me manque*.

Travailler sur une langue sans locuteurs a été la norme pendant des siècles, les textes constituant à la fois l'objet d'étude, et l' « outil » pour appréhender cet objet (sur ce point, voir 3.1. Les textes comme seul accès à la langue). Le fait de travailler sur des états de langue contemporains a permis d'autres approches. Je ne retracerai pas ici l'historique de ce développement ; je n'en retiendrai que le point d'aboutissement, acquis depuis quelques dizaines d'années : la linguistique s'est donné certains outils, au nom desquels elle prétend parfois au statut de science empirique. Mon propos n'est pas de discuter si cette prétention est légitime ou non[7]. Je m'en tiendrai au constat

[7] Voir Milner 1989 pour une réflexion sur la question.

qu'il existe aujourd'hui une méthodologie et des modes de raisonnement possibles lorsque l'on travaille sur une langue avec locuteurs, qui ne le sont pas lorsque l'on travaille sur une langue sans locuteurs. On ne cherchera pas à établir un jugement de valeur sur les différentes méthodes, ni à « justifier » celles utilisées dans le second cas, ce qui serait une forme de disqualification *a priori*. Or, à objet différent, méthode (sinon nécessairement en tout cas souvent) différente. Il ne s'agit donc pas ici de savoir si la démarche adoptée pour étudier une langue sans locuteurs peut prétendre au statut de « science », mais de savoir comment elle permet de connaître et décrire les propriétés de son objet d'étude de façon plausible[8].

Se positionner par rapport à d'autres « paradigmes » permet de questionner celui que l'on utilise, de ne pas le considérer d'emblée comme une évidence. C'est le sens de ma démarche.

1.1. Quand le possible de langue rejoint l'attesté

En tant que non-locuteur du français ancien, nous ne pouvons pas formuler sur une construction un jugement d'acceptabilité pour lequel nous nous appuierions sur notre « compétence » de la langue : précisément, nous n'en possédons pas. L'absence de compétence signifie aussi, et c'est lié, qu'il n'est pas possible de procéder à des opérations de manipulation des énoncés[9] (l'examen des variantes apparaissant dans un même contexte répond à un objectif analogue, mais elle n'opère, par essence, que sur l'existant). Tout cela implique que nous ne pouvons pas opérer sur les données langagières ce que Milner (1989) appelle un « jugement différentiel » :

> Le principe de ce jugement différentiel est que tout ne peut pas se dire. [...] Il suppose donc qu'il y a un impossible en langue. Toutefois cet impossible de langue n'est pas un impossible matériel. Autrement dit, une donnée de langue peut être possible matériellement, c'est-à-dire attestée, et impossible en langue, ou inversement (1989 : 57).

Or l'établissement de ce jugement différentiel est jugé constitutif de l'activité grammaticale :

[8] Le postulat étant que, en langue ancienne plus qu'ailleurs, il ne peut y avoir de certitude absolue quant aux propriétés énoncées.

[9] On sait d'ailleurs les limites de cette pratique : les variations d'appréciation inter- voire intra-locuteurs, tant en matière de grammaticalité que d'acceptabilité (voir Corbin 1980 : 123-138 pour cette distinction et son historique) sont grandes, et les critères qui président aux jugements pas toujours explicités.

> On en vient alors à suspendre la possibilité générale de toute grammaire à l'hypothèse suivante, qu'on peut bien nommer l'hypothèse grammaticale minimale : le possible de la langue et le possible matériel peuvent ne pas coïncider (1989 : 57).

Les limites de l'introspection vont au-delà de cette faiblesse. Comme le rappelle Corbin :

> …si l'introspection peut repérer certaines variations dans les pratiques langagières, elle est impuissante à décrire leur distribution dans la population : le social lui échappe par définition (1980 : 121).

Et plus loin, il insiste sur « l'espoir chimérique d'atteindre par introspection un « réel » linguistique ayant un ancrage social ». L'introspection a néanmoins ses vertus, et présente l'intérêt spécifique – et précieux – de pouvoir « envisager d'autres énoncés que ceux qui sont attestés. *L'introspection peut alors être conçue comme l'instrument privilégié d'une recherche sur les limites ultimes du possible prédictible à partir des observables* » (Corbin 1974 : 155).

Ces deux questions – établissement du jugement différentiel et possibilité d'une grammaire – se posent à vrai dire différemment pour la langue ancienne, puisque possible en langue et possible matériel tendent à être coextensifs, dans les faits et par principe.

En effet, ce qui nous permet d'établir le possible en langue d'une donnée langagière est précisément le fait qu'elle est attestée, observable, les textes existants étant notre seul outil d'appréhension de la langue (sur ce point, voir ci-dessous 3.1. Les textes comme seul accès à la langue).

On peut néanmoins considérer une donnée attestée comme impossible en langue, mais cette appréciation ne repose pas sur un jugement intuitif tel qu'on l'entend généralement, à savoir une introspection fondée sur notre compétence : il repose sur ce que l'on peut appeler une « intuition de reconnaissance »[10]. Il est un jugement permis par la construction progressive de notre connaissance de la langue ancienne, qui s'est faite par l'étude des grammaires (sur ce point, voir 1.2.) et par la fréquentation des textes. Notre aptitude à émettre un jugement sur le caractère possible ou impossible d'une donnée langagière provient donc de ce que nous avons *appris* des règles explicitement formulées (ou bien parfois en avons-nous inféré

[10] Marchello-Nizia (1995 : 22) suggère que le linguiste médiéviste peut parfois développer, à défaut d'une compétence de « production », une compétence de « reconnaissance », distinction qui rejoint celle entre compétence active et compétence passive.

certaines par une fréquentation assidue et une analyse des textes). Les modalités d'acquisition et de mise en œuvre de cette *grammaire* sont évidemment fort différentes de la grammaire intérieure que nous possédons pour notre propre langue. Contrairement à ce qui se produit pour celle-ci où « …nul sujet parlant n'est censé connaître les règles, parce que les ignorer n'empêche pas nécessairement de les appliquer » (Milner 1989 : 94), pour la langue ancienne, il faut connaître les règles pour juger de leur application.

Une donnée attestée mais néanmoins jugée impossible en langue correspond généralement à un hapax qui viole une règle établie (après vérification dans le manuscrit qu'il ne s'agit pas d'une coquille éditoriale). Si la donnée se révèle finalement non isolée, et qu'elle se rencontre dans d'autres textes, plus tardifs, on peut faire l'hypothèse qu'il s'agit de l'amorce d'un changement, et que la « règle » doit être révisée (se pose la question du seuil quantitatif au-delà duquel il faut repenser la règle) : un impossible en langue à un moment T0 peut devenir possible en langue en T1, c'est précisément la caractéristique du changement.

Ainsi jusqu'au début du 13ème siècle, les pronoms personnels régimes, atones, ne peuvent se trouver en première position dans l'énoncé (qu'il soit déclaratif, impératif ou interrogatif), celle-ci devant être occupée par un élément accentué (ce fait est décrit par la « loi » de Tobler-Mussafia). On opposera ainsi (3) :

(3) En tel manière **le me** di (*Jeu de Saint Nicolas*, fin 12ème)

avec un élément circonstanciel en tête, aux exemples (4)-(6) sans élément initial, et où c'est donc le verbe qui, exceptionnellement, apparaît en tête :

(4) Voit **le** Guillelmes, tot en est abosmez (*Aliscans*, 2nde moitié 12ème)

(5) Lessiez **moi** en votre manoir (*le vair palefroi*, 12ème)

(6) Rois, voudroies **le** faire issi? (*Tristan* de Beroul, fin 12ème)[11]

Le début du 13ème siècle marque un tournant, puisque la première position peut désormais être occupée par un pronom régime atone. Un énoncé tel que (7) ne devient « possible » qu'à partir de cette date :

(7) Ce dites, et **me conseilliez**
De present a moy marïer.
Me voulez vous dont martirer

[11] L'apostrophe est considérée comme « hors-phrase » : elle n'occupe pas une position dans la phrase.

En moy lïant en marïaige ? (*Histoire de Griseldis*, 1395)

Si l'on n'est pas face aux prémices d'un changement, et que l'hapax en est bien un, plusieurs interprétations sont possibles, entre lesquelles nous n'avons pas les moyens de trancher : soit il s'agit d'une faute (peut-être d'inattention), de l'auteur ou du copiste, en tout cas d'un impossible en langue ; soit il s'agit d'un possible en langue, que sa rareté n'a pas permis de mettre au jour jusqu'ici. Il peut aussi s'agir d'un possible en langue propre à celui qui l'a produit : ce ne serait alors pas une faute en tant que telle, mais l'indice que l'auteur de la donnée aurait eu une grammaire différente de celle observée le plus fréquemment[12]. On touche ici à la question des usages et de la variation, et à celle des basses fréquences.

Le jugement d'acceptabilité peut donc s'appliquer, dans une certaine mesure, aux données attestées, même si les « données attestées impossibles en langue » sont rares, et difficiles à établir.

On peut par ailleurs envisager de formuler un jugement d'acceptabilité sur une donnée non attestée. A cet égard, la question posée par Milner : « que signifie pour une donnée de langue qu'elle se rencontre ou pas ? » (1989 : 130) prend, pour la langue ancienne, un relief différent : s'agit-il d'un impossible de langue, ou d'un possible non attesté dans les textes explorés jusqu'ici mais susceptible de se rencontrer dans des textes nouvellement explorés, ou bien encore d'un possible n'ayant jamais été produit ? En effet, les productions langagières étant infinies, tous les possibles de langue ne sont pas attestés matériellement (et ce d'ailleurs pour tous les états langagiers[13]). La question qui se pose est celle de savoir si nous sommes capables, face à une absence, de choisir entre ces diverses explications, et donc de statuer sur le caractère du non-attesté : s'agit-il d'un possible en langue ou non ? Si l'on postule que toute donnée peut être attestée (seule notre exploration incomplète des textes ou notre impossibilité de connaître l'ensemble des données susceptibles d'avoir été produites nous la ferait ignorer) on en arrive à la conclusion que tout est possible dans la langue. Or ce n'est évidemment pas souhaitable. D'une part, ce n'est guère plausible, dans la mesure où nous connaissons désormais une proportion importante des productions

[12] Voir à ce sujet les remarques de Milner (1989 : 77 et suivantes) sur les différents usages, et l'analyse de la scène II, 6 des *Femmes savantes*.

[13] De même, pour le français moderne, il existe des données attestées non recensées : l'indexation des textes et de leur contenu n'est que partielle, et elle se heurte en outre à leur production continue.

langagières – nécessairement écrites –parvenues jusqu'à nous, ce qui réduit fortement la probabilité de faire des découvertes totalement surprenantes et qui bouleverseraient radicalement les propriétés mises au jour jusqu'ici. Une telle position serait d'autre part contre-productive d'un point de vue heuristique, dans la mesure où elle interdirait la possibilité de toute grammaire (entendue comme ensemble des propriétés de la langue).

C. Marchello-Nizia (1995 : 23) suggère que l'absence d'une construction attendue et la présence en sa place d'une paraphrase peut faire conclure au caractère agrammatical de la première. Ainsi face à *'je si ferai'*, on ne rencontre jamais (avant le 15ème siècle) *'Je ne si ferai'*, mais une construction paraphrastique : *'je non ferai'*. On peut conclure à l'agrammaticalité du tour négatif attendu.

Arrive donc un moment où il faut décider que les données langagières connues sont suffisamment importantes et représentatives, en quantité et en variété, pour que l'on puisse prétendre en dégager les propriétés (quitte à réviser ces dernières face à des données « contradictoires », le cas échéant).

On peut donc admettre que le recours à une intuition de reconnaissance (qui consiste *a minima* en une excellente connaissance des règles de la langue mises au jour au fil des siècles) permet d'appliquer un jugement différentiel aux données non attestées : à propos des énoncés non rencontrés dans les textes, on postulera que certains sont impossibles alors que d'autres sont possibles et plausibles. Par exemple, dans la mesure où l'on a établi que, dès l'ancien français, le pronom personnel sujet postposé au verbe lui est conjoint (alors qu'il ne l'est pas en position préverbale) et qu'il ne peut en être séparé que par un autre pronom conjoint ou par la négation *ne*, on peut prédire que l'on ne rencontrera pas la séquence XVOnSp : < lors apele le chevalier il > n'est pas possible en langue, ni tout autre énoncé instanciant la même construction.

En revanche, il est des énoncés non attestés qui sont possibles, même pour une langue morte. L'énoncé <le chevalier apele li rois> (« le roi appelle le chevalier ») n'est pas attesté à ma connaissance, mais il est tellement plausible qu'il sert d'exemple prototypique dans les cours de grammaire pour illustrer l'existence conjointe d'une déclinaison et de la relative souplesse de l'ordre des mots en ancien français. Il est en effet la possible réalisation de la construction 'objet nominal-verbe-sujet nominal', qui, elle, est belle et bien attestée, et instanciée par de nombreux énoncés (même si cet ordre n'est pas majoritaire) :

(8) Dis blanches mules fist amener Marsilie (*Chanson de Roland*, vers 1100)
OBJET VERBE SUJET

(9) Iteus paroles distrent li frere de Lancelot (*La mort Artu*, déb.13ème)
OBJET VERBE SUJET

Il convient donc de distinguer l'attestation d'une construction, et celle d'une de ses réalisations spécifiques. C'est l'existence d'une construction qui nous importe véritablement, plus que celle de ses diverses instanciations, potentiellement infinies[14], mais ce sont évidemment les secondes qui permettent de formuler la première. Ce qui permet de dire que, en ancien français, le pronom personnel sujet préverbal est disjoint (il peut être séparé du verbe par un élément tonique), et que la construction 'sujet pronominal-X-verbe' est possible tient à l'existence d'énoncés tels que :

(10) Et **ele tant le conforta**
[….]
Que ele **en santé** le remist (P. de Rémi, *Jehan et Blonde*, vers 1230)

Aucune grammaire cependant, pendant longtemps, n'a donné d'exemple où l'élément séparateur est un objet nominal (On)[15], indice que la construction ne se rencontrait pas dans les textes. On aurait pu en inférer une règle stipulant que la construction 'Sp-On-V' est impossible. Or Skårup (1975) mentionne deux occurrences, dont l'une appartient à un texte très ancien et franco-provençal, et dont voici la seconde :

(11) Ne **il le roi ne desfïa**
Ne li rois droit ne li vea (*Roman de Thèbes*, mil. 12ème)

L'attestation, qui ne viole par ailleurs aucune règle établie (on rencontre des éléments autonomes entre le pronom sujet et le verbe), exclut de postuler une règle qui interdirait la présence de l'objet nominal, même si cette dernière est extrêmement rare.

Reste la question de savoir si l'absence d'instanciation d'une construction signifie que cette dernière n'était pas possible en langue, ou que le « hasard » ne l'a jamais produite, ou bien si rarement que l'on ne l'a pas (encore) rencontrée.

[14] Il n'est cependant pas certain que toute instanciation d'une construction attestée soit possible. Des paramètres indépendants du schéma de la construction peuvent « bloquer » la réalisation de cette dernière. Ainsi, en français moderne : <la souris mange l'éléphant> est syntaxiquement correct mais inapproprié du point de vue sémantico-pragmatique (même s'il est vrai que l'on pourra toujours arriver à construire un contexte approprié).
[15] « Objet nominal » pour nous, locuteurs modernes : voir 1.3.3. à propos de la projection des concepts modernes.

On touche ici aux limites du raisonnement : d'une part souvent il n'est guère possible de répondre[16], et d'autre part la question n'a guère d'intérêt. Il semble donc plus raisonnable *scientifiquement* de s'en tenir au constat que certaines constructions ne sont pas attestées.

On en arrive à la conclusion que, lorsqu'il s'exerce en langue ancienne, le jugement différentiel s'appuie, *in fine*, sur l'attesté.

Selon Milner, l'activité grammaticale repose essentiellement sur la disjonction entre le possible de la langue et le possible matériel : cette disjonction étant en partie mise en cause pour la langue ancienne, faut-il en conclure, que l'activité grammaticale se réduit au simple enregistrement des données, telle celle d'un « magnétophone » :

> […] si effectivement le possible et l'impossible de langue se confondent entièrement avec le possible et l'impossible matériels, alors on peut et on doit se borner à constater ce qui se dit et ne se dit pas matériellement : la raison grammaticale n'est rien de plus et rien de moins qu'un magnétophone (Milner, 1989 : 89).

L'affirmation est discutable, pour des raisons d'ordres différents. Elle l'est tout d'abord en raison de la productivité des règles de la langue : on ne peut être confronté à toutes leurs réalisations et à toutes leurs combinaisons, il n'est pas possible de toutes les enregistrer. C'est d'ailleurs là que réside, comme le souligne Corbin, l'intérêt spécifique de l'introspection (1980 : 155[17]).

L'affirmation est par ailleurs contestable dans la mesure où ce simple enregistrement des données, envisagé par Milner avec un certain dédain, constitue, pour la langue ancienne, un préalable indispensable (ce qu'il n'est pas pour un état de langue moderne), puisqu'il donne son existence même à l'objet. Cet enregistrement des données n'est cependant pas si « simple », et ne se réduit assurément pas à l'activité mécanique d'un magnétophone (image au demeurant incongrue lorsqu'il s'agit de langue ancienne…). Enregistrer les données suppose dans un premier temps de les recenser, et donc de répertorier les textes dans lesquels elles apparaissent, cela d'une manière aussi exhaustive que possible. Concrètement, de nos jours, enregistrer les données passe par leur « numérisation » en mode texte[18]. La mise sur support

[16] Hormis lorsqu'il existe une construction paraphrastique : on peut alors conclure au caractère agrammatical de la construction attendue. Cf remarque de Marchello-Nizia (1995 : 23).
[17] Voir ici p. 5-6.
[18] Le mode image est possible aussi, mais il ne permet pas la manipulation ultérieure des données ; il répond à d'autres objectifs.

numérique s'est d'abord effectuée manuellement, par la saisie des textes dans un traitement de textes, puis, de plus en plus fréquemment, par la technique d'*océrisation* (mise en œuvre d'un logiciel de reconnaissance optique de caractères). Cette forme moderne de mémorisation des textes permet l'accès ultérieur aux données (voir 3.2.), et leur possible manipulation, à condition que la numérisation ne soit pas « brute » : les données, les textes doivent être indexés (par période, par genre…), sous peine de n'être pas – ou mal – accessibles et exploitables. Aussi rudimentaire que soit cette indexation (nous aborderons en 4. la question de l'enrichissement des textes), elle met néanmoins en œuvre une forme de classification. Enregistrer les données n'est donc pas « simple » à deux égards : c'est à la fois une tâche lourde et une opération complexe.

Ces données doivent ensuite faire l'objet d'autres formes de classification (parties du discours, fonctions, catégories sémantiques, etc.) permettant de mettre au jour leurs propriétés. Celles-ci peuvent être énoncées sous forme de règles, qui permettent de prédire les énoncés possibles et les énoncés impossibles. Il reste que propriétés et règles sont susceptibles d'être falsifiées par la mise au jour de nouvelles données qui les invalident. C'est là un état de fait. Mais comme je l'ai souligné ci-dessus, la prise en compte sans cesse accrue de données permet de réduire le risque d'invalidation. En effet, les données de langue ancienne constituent un ensemble fermé : on peut concevoir que l'on aura un jour exploré la totalité de cet ensemble (même s'il est fort probable que des textes nous resteront inaccessibles, simplement parce que nous ignorons leur existence), et que l'on pourra donc mettre au jour des propriétés certaines et énoncer des règles définitives. Avec néanmoins cette double restriction : d'une part, ces propriétés et règles, aussi descriptives et objectives qu'elles puissent se vouloir, sont toujours le résultat d'une vision subjective de l'objet observé (mais cela est vrai pour tout état de langue), d'autre part, aussi exhaustif que soit notre recensement et notre étude de l'existant, cela n'empêche pas que cet existant ne nous permet qu'une vision partielle de la langue, dont la dimension orale ainsi que certains registres de l'écrit nous échappent dans une large mesure[19].

[19] Nous avons cependant, dès le Moyen Âge, quelques témoignages qui permettent un accès à une réalité autre que celle qui nous est parvenue à travers les textes que nous connaissons. Il s'agit en particulier des « manières de langage », petits manuels prescriptifs du bon usage.

Il faut bien admettre que la langue que nous étudions est celle et seulement celle des données textuelles qui nous sont parvenues (aussi précieux que soient les éventuels « témoignages » –voir note 19– les informations qu'ils livrent ne peuvent être traitées de la même façon que les données attestées). Un objet partiel dans l'absolu, mais qui constitue néanmoins une totalité finie pour nous. La situation est donc pour ainsi dire opposée à celle dans laquelle on se trouve pour la langue moderne, potentiellement accessible dans sa totalité, mais dont la totalité est infinie (de nouvelles données sont sans cesse produites). C'est là l'une des différences essentielles entre un état de langue révolu et un état de langue contemporain.

J'ai parlé jusqu'ici de recensement, d'établissement des données, de description classificatoire et de mise au jour de propriétés. Je n'ai pas parlé d'*explication*, activité qui pourtant, selon moi, tend à distinguer l'activité linguistique de l'activité grammaticale au sens traditionnel. Cela tient à ce que je considère que le champ d'exercice de l'activité explicative est davantage celui du changement, et donc des études diachroniques, abordées en 2. La visée explicative n'est cependant pas absente de l'étude synchronique d'un état de langue ancien : elle se manifeste dans toute activité classificatoire, et plus spécifiquement dans la mise en relation que l'on peut établir entre certains phénomènes : ainsi lorsque l'on dit que, en ancien français, l'existence d'une déclinaison et la souplesse de l'ordre des mots sont corrélées, on introduit une dimension explicative (en établissant, dans ce cas précis, une relation de causalité partielle entre la première et la seconde).

Je me suis placée jusqu'ici dans une position pour ainsi dire anhistorique, tant du point de vue des données disponibles que de l'établissement des règles, les deux étant liés. Or il est utile de quelque peu historiciser ces deux aspects.

D'où vient la connaissance que nous avons du français ancien ? Comment s'est-elle progressivement construite ? Quelle est notre position vis-à-vis de cette construction ? Il ne s'agit pas de proposer une histoire de la grammaire (ou plutôt des grammaires, comme on va le voir), ni des concepts qui l'accompagnent[20], mais simplement de rappeler quelques faits importants afin de resituer notre pratique actuelle.

[20] Ne serait-ce que parce que j'en serais incapable. Parmi l'abondante littérature sur le sujet, on citera en particulier : F. Brunot 1905-1938, Tome II, p. 125 et suivantes, J.-C. Chevalier 1994 et 2006, J. Julien 1988, et les recherches menées dans le cadre du laboratoire *Histoire des Théories Linguistiques* (NCRS/Paris-7), largement diffusées dans la revue *Histoire Epistémologie Langage* (voir en particulier : 1981, vol.3, fasc.1 ; 1984, vol.6, fasc.1 ; 1990, vol. 12, fasc.2).

1.2. Construction des savoirs et émergence des grammaires

1.2.1. Les grammaires : du latin au français, du français à la langue ancienne

L'activité grammaticale a pendant des siècles été consacrée au latin. Langue de prestige, mais aussi langue qu'il fallait apprendre, ce qui n'était pas le cas du français, langue maternelle qui n'avait pas besoin d'être rationnalisée pour être apprise. L'activité réflexive sur le français n'est cependant pas totalement absente au Moyen Âge, mais elle n'est pas mise en écrit, aucun impératif pratique n'incitant à sa systématisation. C'est donc de l'étranger, en l'occurrence d'Angleterre, que viendront, dès le 14ème siècle, les premiers « traités », destinés à ceux qui ont à apprendre le français en tant que langue étrangère[21]. Et c'est aussi d'Angleterre que viendra, au 16ème siècle, la première grammaire à proprement parler, celle de Palsgrave (*Leclarcissement de la langue francoyse*, 1530), pareillement rédigée à destination des Anglais[22]. Ce n'est qu'ensuite qu'apparaîtront en France des grammaires du français : la première est celle de Sylvius (*Grammatica latino-gallica*, 1531), mais elle est rédigée en latin et le français n'y est envisagé que dans sa relation au latin. Il faut attendre 1550 et Meigret pour que paraisse enfin la première grammaire du français rédigée en français (*Trettté de la grammere françoeze*).

Elle inaugure un mouvement qui ne cessera de prendre de l'ampleur. Les raisons qui ont motivé ces premiers ouvrages sont assez diverses, comme le souligne F. Brunot (1905-1938, tome 2 : 128-131) : il s'agit d'aider les étrangers, mais aussi de permettre aux générations futures de « se rendre compte de l'état de la langue sous Henri II ».

Ce mouvement résulte aussi, et peut-être surtout, du sentiment qu'il est nécessaire de fixer les règles de la langue, pour endiguer le changement[23] (et permettre au français de devenir une langue aussi prestigieuse que le latin). Beaucoup connaissent en effet les textes des siècles passés, et sont conscients que la langue a évolué : c'est non seulement Joinville et Villehardouin qu'il faut « rajeunir » pour les éditer, mais aussi des auteurs plus récents, comme Antoine de la Sale ou Villon. Connaissance de textes anciens donc, et conscience d'un écart linguistique, mais l'intérêt se porte

[21] Sur cette question, voir Lusignan 1986.

[22] Rappelons que certains anglais réclamaient la possession de fiefs en France ; pour cette raison, ils pouvaient souhaiter apprendre le français.

[23] Meigret n'avait cependant guère d'illusion sur l'efficacité de ce moyen pour fixer la langue, conscient que la grammaire repose sur l'usage, sans cesse innovant.

prioritairement vers la langue contemporaine, il ne se porte pas encore véritablement vers la langue, pour elle-même, des auteurs anciens.

Au siècle suivant, les Bénédictins de la Congrégation de Saint-Maur seront confrontés aux chroniques médiévales, religieuses et laïques : pour les comprendre, ils devront développer une connaissance spéciale[24]. Le mouvement s'intensifie au 18ème siècle, avec trois courants : le premier est juridique (et s'était d'ailleurs engagé dès le 17ème siècle, avec la « réfection des terriers » et l'établissement et la collation des *coutumes* par diocèse et province), le second est historique et incarné par La Curne de Sainte-Palaye, qui s'est consacré à l'étude des chroniqueurs du Moyen Âge, ainsi qu'à celle des manuscrits relatifs à l'histoire de la langue et des institutions françaises. Le troisième courant est littéraire : à partir de cette époque on édite de plus en plus de grands textes : *Roman de la Rose, Tristan…,* activité éditoriale nécessairement philologique. Le mouvement est donc bien lancé, mais il faudra cependant attendre le 19ème siècle pour que paraisse un ouvrage sur la langue ancienne : il s'agit, en 1816, de la *Grammaire romane ou grammaire de la langue des troubadours* de Raynouard, qui n'est d'ailleurs pas spécifiquement dédiée au français, comme son titre l'indique.

Le 19ème siècle verra se développer études philologiques et grammaticales, les secondes s'alimentant des premières. Progressivement les propriétés de la langue ancienne seront mises au jour, les règles énoncées, les unes et les autres complétées, amendées au fur et à mesure que de nouveaux textes seront étudiés. C'est là évidemment un résumé bien rapide et grossier de ce qui s'est passé mais il m'importe surtout d'insister sur deux points : la description de la langue s'est construite, progressivement, non seulement par un mécanisme d'accumulation mais aussi selon une méthodologie de type « corpus-driven » : en dégageant peu à peu des régularités[25], on a pu définir des propriétés et établir des règles. La première syntaxe de l'ancien français, celle de Foulet, fondée sur l'étude minutieuse de textes (13 dans la première édition en 1919, 19 dans la troisième édition en 1928), est une illustration de ce processus[26].

[24] Je remercie Gabriel Bergounioux et Christiane Marchello-Nizia pour les précieuses informations qu'ils m'ont fournies sur les premières études portant sur le français médiéval.

[25] Ce qui ne va pas de soi : les régularités ne s'imposent pas d'elles-mêmes, elles n'ont pas d'existence autonome mais sont le résultat d'une construction ; celle-ci implique une opération d'abstraction à partir des données observées, ce qui suppose une forme de classification de ces dernières, et donc une préconception de la langue.

[26] Certes elle illustre aussi le caractère partiel et aléatoire du corpus, mais l'on retiendra plutôt ici le caractère pionnier de ce travail.

1.2.2. Observer la langue à travers le filtre des grammaires

Nous bénéficions aujourd'hui de plusieurs grammaires relativement récentes (moins de trente ans) consacrées à l'ancien et au moyen français. Elles s'appuient de façon critique sur les connaissances accumulées depuis maintenant 200 ans, qui ont elles-mêmes bénéficié de la multiplication et de la diversification des données textuelles explorées[27] : on peut les juger fiables, même si la précision des phénomènes décrits est inégale (ce qui s'explique en partie par leur visée exhaustive qui, dans la pratique, ne permet pas de traiter avec la même finesse l'ensemble des phénomènes abordés, mais aussi par le fait que les différents domaines de la langue n'ont pas toujours été l'objet de la même attention : ainsi la syntaxe et la sémantique sont les parents pauvres des premières grammaires, concentrées sur les aspects phonétiques et morphologiques[28]).

C'est avec ces grammaires – et la lecture des textes – que le linguiste construit initialement sa connaissance de la langue. Elles sont donc un outil interprétatif pour observer et étudier la langue. Pourtant ces grammaires ne sont pas « neutres ». Elles ne l'ont d'ailleurs jamais été : le simple recours à une classification en parties du discours dénote déjà un parti-pris théorique, même si cette catégorisation est tellement intégrée à notre tradition grammaticale que nous ne la percevons plus comme telle, sauf lorsque les limites entre catégories se révèlent problématiques.

Qu'il s'agisse de considérer de nouvelles données textuelles pour enrichir ou rectifier l'étude d'un phénomène, ou qu'il s'agisse d'étudier un phénomène non encore (suffisamment) exploré, toujours le linguiste doit essayer de porter un regard neuf sur les données. Il ne s'agit pas pour autant d'adopter la posture du premier observateur et de réinventer la roue, mais de garder à l'esprit que l'on considère les données à travers différents filtres : filtre des acquis sur la langue considérée, dont certains sont peut-être à reconsidérer, filtre de notre compétence de locuteur moderne, filtre de notre position de linguiste, qui tend à projeter sur la langue ancienne des concepts et des théories qui n'ont pas été, pour la plupart, conçus pour elle. Cela ne rend pas pour autant ces derniers inopérants, mais il convient de ne pas ignorer cet état de fait, sous risque d'adopter une posture anachronique.

---

[27] Les modalités d'exploitation directe des textes par les grammaires sont cependant très variées de l'une à l'autre, tant en ce qui concerne le nombre de textes retenus que leur traitement (objets de dénombrements systématiques, ou simples réservoirs à exemples).

[28] La syntaxe de Foulet est à cet égard très innovante.

1.3. Des classifications héritées aux classifications construites

1.3.1. Classification spontanée et classifications volontaires

Quand on lit un texte, on opère à la fois une segmentation et une classification des données, (parfois même face à une langue que l'on ne comprend pas, mais dont on essaie d'accrocher les items successifs à une catégorie familière). L'opération est plus ou moins consciente, plus ou moins explicitée (ainsi le lecteur non linguiste, face à son journal ou son roman, n'aura pas la même activité réflexive que le linguiste face aux données langagières qu'il étudie), et elle ne recourt pas toujours aux mêmes catégories, ni même à des systèmes de classification identiques.

Au simple niveau du décryptage des données s'effectue une opération de segmentation-catégorisation, conditionnée par notre connaissance de la langue ancienne, mais aussi par notre compétence de locuteur moderne, ne serait-ce que parce que toute édition (ou presque) de texte ancien restructure le texte du manuscrit en le ponctuant à la moderne (et en re-segmentant des mots regroupés par les copistes), sans même le dire parfois : il s'agit là d'une manipulation implicite, et acceptée, qui d'emblée structure notre perception du texte et donc la grammaire que nous nous en faisons[29]. A cette opération peu contrôlée, succède l'ajout, volontaire, de différentes couches de classification, d'ordres et de niveaux différents, en fonction du type de données considérées, de l'approche adoptée (syntaxique, sémantique, pragmatique…) et de la théorie ou du cadre d'analyse choisi : dès lors il convient d'expliciter et de questionner les classifications adoptées.

La projection d'une classification peut par ailleurs être mise en œuvre dans le cadre de l'annotation d'un corpus, et la systématisation de la démarche ainsi que l'existence de contraintes « matérielles » (en particulier en ce qui concerne la segmentation des formes) entraînent des difficultés et une réflexion que nous aborderons plus loin (4.1. l'enrichissement des corpus). Enfin, cette démarche classificatoire, qu'elle soit ou non systématisée dans le cadre de l'annotation d'un corpus, peut être réalisée dans une perspective diachronique : aux problèmes inhérents à l'établissement des différents types de classes, à la délimitation des catégories en synchronie, s'ajoute alors la difficile appréhension de leur évolution dans le temps.

[29] Sur la ponctuation dans les manuscrits, voir Lavrentev 2009.

Nous reviendrons plus loin sur la question de l'annotation du corpus et sur celle de l'évolution des catégories. Avant cela il convient de dire quelques mots sur les classifications que nous utilisons en synchronie.

Les classifications sont toujours des constructions, elles ne sont pas naturelles, que la réalité préexiste à son observation (comme c'est le cas dans les sciences de la nature) ou non (comme c'est le cas dans les sciences sociales)[30]. On peut considérer que leur caractère conventionnel est plus ou moins marqué selon précisément l'existence autonome ou non des objets classifiés (la classification des catégories socio-professionnelles serait à ce titre plus conventionnelle que celle des espèces végétales), mais une question récurrente demeure, plus ou moins prégnante : pourquoi, et selon quels critères, décide-t-on de traiter des objets comme équivalents ?

Que le langage n'a pas préexisté à l'homme, c'est une évidence ; que l'on puisse considérer ou non que les langues ont préexisté à leur observation, c'est une question que je ne discuterai pas ; il est en revanche certain que les classifications des données langagières, de quelque ordre qu'elles soient, sont une construction conventionnelle, et que, en tant que telles, elles informent nécessairement l'objet auxquelles elles sont appliquées. Cette « information » est donc inévitable, mais lorsqu'il s'avère que les classifications mises en œuvre s'inscrivent dans un contexte historique différent, et surtout largement postérieur à celui des données auxquelles elles s'appliquent, il est nécessaire de les utiliser avec quelques précautions, et légitime de s'interroger sur leur pertinence.

A la lecture d'un texte ancien, nous opérons une double projection : celle de la représentation que nous avons de sa langue, construite par l'étude, et celle, indissociable, de la représentation que nous avons de notre propre langue. C'est précisément par le jeu des classifications que s'opère le pont entre les deux.

On peut néanmoins s'interroger sur la pertinence, et la légitimité, à projeter certains concepts, plus ou moins récents, sur la langue ancienne, tels que ceux, par exemple, de complément d'agent, ou de marqueur de topicalisation, ou bien encore de cadre de discours. Dans certains cas, la réalité que recouvrent certains concepts, pragmatiques,

[30] Sur la question de la classification dans les sciences sociales, voir Desrosières 1989 et 2001, et Desrosières et Thévenot 2000. Desrosières (2000 : 117) oppose ainsi l'étoile polaire qui a « existé bien avant tous les astronomes », et la population active et le chômage, dont la définition et la mesure « relèvent d'une autre épistémologie que celle de l'étoile polaire ».

n'a pas toujours existé. Ainsi les marqueurs de topicalisation (expressions qui introduisent un référent qui a un rôle de topique dans la suite de l'énoncé) n'apparaissent, progressivement, qu'à partir du moyen français :

(12) **Quant aux clercs, ilz** ont aussi charge de justice ; faictes savoir se ilz resident, et se il y a nulz abus. (Juvenal des Ursins, *VERBA MEA AURIBUS PERCIPE, DOMINE*, 1452)

(13) **A propos de nos cochetz**, ceste dame dont nous parlons **les** faisoit garder à part des poulles, pour servir à table en lieu de chappons, dont elle se trouvoit bien. (B. des Périers, *Les Nouvelles récréations et joyeux devis de feu*, 1558)

Avant cette période, le concept ne correspond à aucune réalité. Il en va différemment pour d'autres concepts, qui recouvrent une réalité existant déjà dans la langue ancienne, mais nommée et/ou décrite autrement, ou s'inscrivant dans un paradigme classificatoire différent, ou bien même dans aucun paradigme. La construction à laquelle correspond ce que nous appelons 'complément d'agent' est attestée dès l'ancien français :

(14) sachiez, et itant vous en di,
que, se [si] **je sui par vous trahi**,
vous en receverez la mort. (*Châtelaine du Vergy*, milieu 13ème) ,

mais elle ne sera discriminée et désignée comme telle qu'à partir de la seconde moitié du 19ème siècle.

Face à ce constat, plusieurs questions se posent : en quoi ces concepts avaient-ils une validité à l'époque, et est-il nécessaire qu'ils en aient eu une ?

Il ne s'agit nullement d'envisager l'analyse des textes étudiés dans le cadre des concepts et des classifications qui pouvaient prévaloir à l'époque de leur production : ce serait faire le grand écart avec notre compétence de locuteur-linguiste, et cela n'aurait pas grand sens. Il s'agit simplement de garder à l'esprit l'existence d'un décalage entre les concepts et les catégories que nous manions, les analyses que nous opérons, et l'état de langue que nous considérons. Or ce décalage n'a pas la même ampleur selon les types de classification envisagés.

1.3.2. Les parties du discours[31]

Des apprentissages de la grammaire scolaire, il reste généralement aux moins deux choses à ceux qui ont pris des chemins éloignés des considérations grammaticales et linguistiques : les parties du discours (au moins le concept, et la capacité à en mentionner certaines) et la tripartition de la phrase : sujet-verbe-complément. Sorte de fonds commun qui nous accompagne depuis l'école élémentaire, et qui accompagne, au moins pour les parties du discours, la langue française depuis les origines. C'est même bien au-delà qu'il faut remonter, jusqu'à Denys le Thrace[32], chez qui l'on trouve déjà huit parties du discours, les différences avec les neuf que nous utilisons aujourd'hui tenant à la présence du « participe », à la conception de l' « adjectif » comme sous-partie du « nom », et à l'absence de l' « interjection ». *Partes orationis* que l'on retrouvera identiques chez Priscien, si ce n'est, latin oblige, que l'article disparaît et que l'interjection fait son entrée. Il faudra attendre le 16ème siècle pour que réapparaisse l'article[33], et le 19ème pour que le participe s'éclipse et que l'adjectif devienne catégorie de premier niveau, le système que nous connaissons aujourd'hui étant désormais en place.

Certes la vie des parties du discours n'a pas été un long fleuve tranquille : largement critiquées depuis la fin du 19ème siècle, principalement en raison de l'hétérogénéité de leurs critères définitoires[34], elles ont connu bien avant cela de nombreuses turbulences. Elles n'ont cessé d'être questionnées, en particulier par les modistes, par les premières grammaires du français, par la Grammaire de Port Royal, et par la Grammaire générale du 18ème siècle, pour ne citer que quelques moments-clés de leur histoire. Sont entre autres mises en cause, au fil des siècles, leurs propriétés définitoires, plutôt sémantiques ou plutôt syntaxiques (constructionnelles), la nature de ces propriétés, substantielles ou accidentelles, mais aussi la bipartition entre parties signifiantes ou consignifiantes (que ne recouvre que partiellement la distinction entre parties déclinables et indéclinables, les secondes étant d'ailleurs largement sous-estimées jusqu'à la prise en compte du français), et enfin leur ordre d'énumération.

---

[31] Voir *Langages*, 1988, n°92, consacré aux parties du discours. En particulier l'introduction de B. Colombat et les articles de J. Lallot sur la tradition grecque, de I. Rosier sur le Moyen Âge, de B. Colombat sur le 16ème siècle et de S. Auroux sur le 18ème siècle. Voir aussi Rosier 1981.

[32] Voir Lallot 1988.

[33] « Article » parfois remplacé par le « déterminant » dans les actuelles classifications, le second ne recouvrant pas exactement les mêmes formes que le premier.

[34] Voir Lagarde 1988.

On constate donc que, très tôt, apparaissent des questions qui demeurent prégnantes dans la réflexion contemporaine. La première tient à l'ambiguïté des *partes orationis*, soit partie de phrase, « constructible », soit catégorie grammaticale, ambiguïté qui existe dès l'origine[35]. Comme le souligne Milner, la partie du discours est une notion double :

> D'une part, elle permet d'*analyser* quelque donnée de langue que ce soit, c'est-à-dire de justifier l'attribution à cette donnée d'une valeur de prédicat différentiel ; d'autre part, elle permet de *subdiviser* celle-ci en composants constitutifs.
> Il vaut la peine de souligner la singularité essentielle de la grammaire à cet égard. En elle–même, la notion de catégorie n'a rien à voir avec la notion de partie constitutive. Elle relève du jugement et en l'occurrence du jugement grammatical qui consiste dans la proposition « X est P »/ « X est non-P », où P est le prédicat différentiel : possible/impossible, correct/incorrect, acceptable/inacceptable, etc.
> Par ailleurs, toute grammaire suppose que l'attribution de P se fasse sur la base de la constitution de X en parties. On comprend alors que les parties soient confondues avec les catégories et réciproquement. Il n'en reste pas moins que d'un point de vue strictement logique, les deux notions sont distinctes (Milner 1989 : 60).

Les langues agglutinantes illustrent bien la non-coïncidence des deux notions. Plus près de nous, le français fournit des exemples de non recouvrement entre segment et partie du discours, en particulier lorsque l'on considère l'ancien français, riche en formes contractées (*je + le > jel* ; *ne + le > nel ; en + le > ou*…), ou au contraire en formes segmentées dont on ne sait comment analyser les différentes composantes. Ainsi en est-il de *quel que leu que*[36], construction diachroniquement intermédiaire entre *en quel leu que* et *en quelque leu que* (voir Foulet 1919), dans laquelle on peine à analyser le premier *que*, du point de vue syntaxique, mais aussi en termes de parties du discours. Nous reviendrons sur ces exemples en 4.1.2.1.

La seconde question précocement posée est celle de la possible polycatégorisation des unités (voir Auroux 1988), déjà formulée chez les Grecs (voir Lallot 1988), même si la distinction entre homonymie et polysémie n'y apparaît pas encore. Différentes « options » sont discutées : changement de catégorie (décatégorisation / recatégorisation), appartenance simultanée à plusieurs catégories (polycatégorisation), ou bien au contraire appartenance à une seule, l'ellipse ou l'homonymie venant alors au secours du principe de monocatégorisation (en dépit de cas récalcitrants).

[35] Voir Colombat 1988b.
[36] …come cil qu'il fesoit porter o soi en quel que leu que il alast (*Queste del Saint Graal*).

La possible polycatégorisation des unités est une question centrale en linguistique. Pour qui travaille sur la langue ancienne, elle prend une dimension spécifique. Ainsi, il est parfois impossible de trancher quant à l'appartenance d'une forme à telle ou telle catégorie : c'est le cas – rare il est vrai – entre les catégories « nom » et « verbe » (infinitif) lorsque les critères retenus pour trancher en faveur de l'une ou l'autre catégorie (respectivement la présence d'un déterminant ou d'un complément régi) sont absents, ou au contraire sont co-présents, comme dans les deux exemples ci-dessous :

(15) Ore dist li contes que a heure de **disner**, quant les tables...(*Merlin*, déb. 13ème)

(16) Ostes, ja fustes vous compaing,
Puisque che vient **au dire voir** (Bodel, *Jeu de Saint Nicolas*, fin 12ème)

Au-delà des choix qui peuvent avoir été faits en matière de substantivation – favoriser le fonctionnement en contexte (syntaxique) ou au contraire la catégorie d'origine (morphologique) – se pose ici le problème de la compréhension de la construction, et de notre absence de compétence. Il n'est cependant pas sûr que la volonté de choisir entre l'une ou l'autre catégorie soit pertinente, et le recours à un continuum catégoriel s'avère parfois la meilleure option, même si cela suscite d'autres difficultés, que nous envisagerons en 4.1.2. dans le cadre de l'enrichissement des corpus.
La question du changement de catégorie et/ou de l'appartenance à plusieurs d'entre elles est au cœur de la problématique du changement : elle sera largement développée quand seront présentés, d'une part les phénomènes étudiés, d'autre part la démarche d'enrichissement des corpus.

Les catégories du discours, conçues pour une langue synthétique et flexionnelle, ont parfois eu du mal à se couler dans le moule de la langue française, suscitant discussions et adaptations, et révélant leur caractère non universel (ainsi de l'article, et de la préposition). Il reste que, au-delà de la variabilité des approches, on ne peut qu'être frappé par la relative stabilité de la nomenclature au fil des siècles.

1.3.3. Les fonctions syntaxiques

Le terme de syntaxe est ancien, mais son contenu a longtemps été étroitement lié à la morphologie, fondée sur une langue régie par les cas, le latin. La syntaxe sera d'ailleurs le parent pauvre de toutes les grammaires pendant longtemps : elle n'occupe encore que 4 pages (certes essentielles pour son avenir) dans la *Grammaire générale*

de Port Royal. C'est *L'encyclopédie*, au 18ème siècle, qui marque un tournant important : le terme de « complément » y apparaît pour la première fois, avec sa formulation grammaticale, à l'article *gouverner* sous la plume de Du Marsais. Mais comme le souligne J.-Cl. Chevalier (2006 : 10), l'article décisif est l'article *Régime*, rédigé par Beauzée, qui succède à Du Marsais, et qui explique que ce dernier, père du terme « complément », n'a pas osé rédiger un article qui lui soit dédié. « Evolution qui entraîne la séparation de la Syntaxe d'avec la Construction, avec laquelle elle a longtemps été confondue ; par là cette syntaxe devient autonome et installe un type inédit de représentation » (Chevalier, 2006 : 10). Il faudra cependant attendre la fin du siècle suivant pour que la notion de syntaxe telle que nous l'entendons aujourd'hui soit acquise, changement qui aura supposé aussi la stabilisation de la notion de phrase. L'évolution se fait dans un contexte nouveau, celui de la démocratisation de l'enseignement et du développement de la grammaire scolaire à visée pédagogique, ce dont témoigne le sous-titre de l'*Histoire de la grammaire scolaire* de Chervel (1977) : *Et il fallut apprendre à écrire à tous les petits Français...* La syntaxe est donc devenue, à la fin du 19ème siècle : « [l']ensemble des règles qui régissent l'arrangement des mots et la construction des propositions » (*Dictionnaire Général*, cité par Chervel, 1977 : 13). Elle ne se limite plus à la simple combinatoire des mots, et étend sa compétence jusqu'à la phrase complexe. Autre développement essentiel, ancré dans la visée pédagogique : la mise en place progressive des fonctions syntaxiques (complément circonstanciel, objet…), celles-là même qui constituent encore aujourd'hui le fonds de l'enseignement scolaire (au-delà des multiples réformes) et de la grammaire traditionnelle, celles-là même que les nouvelles approches syntaxiques du 20ème siècle ont pu contester, déconstruire. Il reste que, *in fine*, c'est souvent, sinon toujours, vis-à-vis de ces notions que se situe le débat. Qu'on les accepte, qu'on les réfute, elles restent au cœur de la discussion : peut-être parce qu'elles sont à la base de notre culture grammaticale commune, et qu'à ce titre, au-delà des divergences possibles, elles constituent un langage partagé.

Parties du discours et fonctions syntaxiques forment donc un socle commun pour les locuteurs modernes, et pour la réflexion linguistique des dernières décennies. Mais alors que les premières ont des siècles d'existence derrière elles et sont donc

contemporaines des productions langagières du français ancien, les secondes lui sont largement postérieures.

De ce constat émerge une double question. La première, déjà évoquée, concerne la légitimité à projeter des concepts et des classifications inexistants à l'époque dont on étudie la langue. Cette projection est selon moi légitime pour plusieurs raisons. En premier lieu, nous ne pourrons jamais nous placer dans la tête d'un locuteur médiéval, adopter son mode de pensée, sa langue… Si les informations que nous pouvons avoir dans ce domaine sont précieuses, elles ne nous autorisent pas pour autant à simuler une telle projection : nous restons des locuteurs du 21ème siècle. Par ailleurs, que des concepts et des classifications n'aient pas été formulés ni pensés à une époque passée ne signifie pas que la « réalité » qu'ils recouvrent pour nous n'existait pas. Projeter la notion de « complément », et non pas celle de « régime » sur « le chevalier » dans *li rois apele le chevalier* est une interprétation possible, considérer « adont » comme un complément circonstanciel et/ou comme un connecteur dans :

(17) Adont vint a la reïne li esvesques de Rovecestre, qui li dist …(*la Mort Artu*),

en est une autre. Raisonner en termes de « phrase », alors qu'il s'agit d'un concept relativement récent[37], relève d'une même démarche.

On peut donc légitimement adopter un point de vue contemporain sur un texte ancien, c'est d'ailleurs la pratique la plus courante. Il est cependant nécessaire de garder à l'esprit l'existence d'un décalage entre notre représentation de la langue ancienne et celle que pouvaient en avoir les locuteurs qui la pratiquaient. Cela nous conduit à la seconde question, qui porte précisément sur la conscience que pouvaient avoir les locuteurs de leur pratique langagière, non pas tant, comme pour certains d'entre eux, dans le cadre d'une réflexion volontaire et organisée (voir 1.2.1), mais au quotidien, spontanément si l'on ose dire. On ne sait évidemment pas grand-chose sur ce point, et l'on peut par ailleurs s'interroger sur l'intérêt de la question : nombreux (probablement majoritaires) sont les locuteurs modernes dont la production d'énoncés oraux ou écrits ne s'accompagne pas de la moindre activité réflexive sur leur pratique langagière. Pour autant, leurs énoncés ne sont pas moins « corrects » ni ne diffèrent de ceux émanant de linguistes. La situation est cependant différente en français ancien : les seules productions langagières qui nous sont parvenues, écrites, émanent de locuteurs « lettrés », dont on peut penser que beaucoup avaient une connaissance

[37] Voir Seguin 1993.

grammaticale, du latin et du français. Il faudrait certes établir des distinctions en fonction des époques : la conscience grammaticale n'est certainement pas la même au 13ème et au 16ème siècle, période à laquelle se multiplient les écrits sur la langue française.

Derrière cette question s'en profile une autre, celle de la possible influence, sur les écrivants, des règles en train de s'énoncer et de se systématiser, de la norme naissante.

Il ne s'agit pas tant d'apporter des réponses que de témoigner de certains questionnements qui ne peuvent totalement échapper au linguiste qui s'occupe de langue ancienne.

1.3.4. Des classes en construction

A côté des classifications précédemment évoquées, contemporaines ou non des faits étudiés, mais néanmoins relativement stabilisées (même si contestées), on peut envisager d'autres types de classifications. Plus récentes, elles font l'objet d'un consensus moindre, tant en ce qui concerne les membres qu'il faut y inclure que leur définition même, qui repose sur des critères pragmatico-discursifs (ce qui explique largement leur caractère moins consensuel, ces critères et leur application laissant une place plus grande à la subjectivité du locuteur/linguiste qu'en morphologie ou en syntaxe). C'est le cas par exemple des marqueurs discursifs (*à propos*), ou bien des marqueurs de topicalisation (*quant à X, à propos de X*) bien que pour les seconds les critères syntaxiques viennent en renfort des critères sémantico-pragmatiques.

Outre la difficulté à définir une classe – qu'est-ce qui en fonde l'unité ? – et à identifier ses membres, se pose ici de façon aiguë la question de projeter sur un état antérieur de la langue une classe conçue pour des états langagiers plus récents : quelle est la pertinence de la notion de marqueur discursif au 15ème siècle ? Le problème ne se pose pas dans les mêmes termes que pour les parties du discours ou les fonctions syntaxiques : les classes qu'il s'agit d'identifier ici ne recouvrent pas une réalité langagière qui a toujours existé (du moins sous cette forme : peut-être a-t-elle existé selon des modalités qui nous échappent). Il s'agit donc, non seulement d'expliciter la classe construite, mais aussi d'appréhender comment, en diachronie, des expressions ont construit leur appartenance à cette classe, *via* un changement avant tout fonctionnel (au sens de fonction discursive). Comment, par exemple, *à ce propos* est-il passé d'une valeur de complément régi :

(18) A laquele chose vault moult ajouster exemples manifestes, selonc le dit de Varron : « Tres clere maniere de enseignier est ajouster exemples ». **A ce propos** dist Aristote ou [en le] premier livre de Methafisique : … (J. Daudin, *De la erudition*, chap. 3, p. 22, 1360)

à celle d'introducteur de remarque incidente :

(19) Ce n'est pas chose estrange que les moynes hantent voluntiers les nonnains. **A ce propos** il advint nagueres que ung maistre jacobin tant hanta, visita et frequenta en une bonne maison de dames de religion de ce royaume.... (*Les cent nouvelles nouvelles*, p. 305, 1456)

Les différents types de changements de classe (que ce soit la caractérisation de la classe qui évolue ou celle des membres susceptibles d'y entrer) seront abordés dans la partie 4 (4.2. Evolution des constructions et changement de classes).

1.4. L'adéquation des cadres d'analyse

Projeter des notions et des classifications construites postérieurement à l'objet sur lequel elles sont projetées ne va pas de soi. Il n'est *a fortiori* pas facile d'utiliser sur un état de langue ancien un cadre d'analyse, voire une théorie, conçus pour les langues modernes. Le problème s'est posé en sens « inverse » lorsque l'on a voulu appliquer au français des modèles d'analyse (parties du discours en particulier) conçus pour le latin, langue synthétique et flexionnelle[38]. Même si, en soi, l'application d'un cadre à un nouvel objet peut soulever des problèmes intéressants, tant pour l'objet que pour le cadre même, elle peut se révéler inadaptée. La difficulté peut tenir à des différences trop importantes entre langue d'origine et langue cible (langue synthétique/analytique, systèmes référentiels divergents[39]…), mais elle peut aussi être due à l'impossibilité de mettre en œuvre certains « outils » du cadre d'analyse. Ainsi les approches de type pragmatico-informationnel reposent en partie sur l'analyse de la prosodie ; or la prosodie de la langue ancienne ne nous est pas accessible (la prosodie ne l'est pas davantage lorsque l'on travaille sur de l'écrit moderne, mais nous produisons néanmoins, à la lecture, une prosodie intérieure). Il faut alors inventer d'autres outils

[38] Un problème analogue a pu se poser quand on a voulu appliquer des théories conçues pour la langues indo-européennes aux langues non indo-européennes.
[39] Ainsi l'ancien français utilise beaucoup moins que le français moderne, et différemment, l'article défini ou indéfini.

qui puissent suppléer l'absence de l'outil prosodique[40]. Ce second cas de figure se rapproche *in fine* du premier : si l'absence de prosodie n'est pas inhérente à la langue ancienne, elle est cependant indissociable de la représentation qui nous en est parvenue, et cette représentation s'avère donc très éloignée des langues pour lesquelles ont été conçus les cadres d'analyse qui se fondent en partie sur la matière prosodique .

Pour conclure sur les spécificités conceptuelles et méthodologiques qu'implique le fait de travailler en langue ancienne, je réinsisterai sur cette posture très particulière que doit adopter le linguiste : locuteur moderne maniant des classifications et des cadres d'analyse postérieurs pour la plupart aux objets qu'il étudie, il doit toujours veiller à ce que sa position, nécessairement anachronique, le soit « raisonnablement », en ne se coupant pas totalement des conditions socio-historiques et culturelles dans lesquelles les données qu'il étudie ont été produites. Il ne peut parler des siècles passés que depuis le 21ème siècle, mais il doit faire en sorte que la représentation qu'il en donne soit plausible, c'est-à-dire compatible avec ce que l'on sait de la représentation que les locuteurs pouvaient avoir de leurs pratiques langagières, mais aussi cohérente, et non radicalement mise en question par l'exploration de nouvelles données langagières.

J'ai à plusieurs reprises évoqué des questions qui touchaient à la démarche diachronique. Il est temps de l'envisager plus longuement à part entière.

[40] Voir Prévost 2001 pour la mise en œuvre d'un modèle de structuration informationnelle sur un corpus de moyen français.

Chapitre 2. Travailler en diachronie

2.1. De l'intérêt d'étudier l'évolution d'une langue

Toutes les langues naturelles changent, et elles changent sans cesse. Il n'existe pas de contre-exemple à ce phénomène : il s'agit donc d'une caractéristique universelle, et l'on peut faire l'hypothèse que c'est un trait de la faculté même du langage. Point n'est besoin, dans cette perspective, d'expliciter l'intérêt qu'il y a à étudier le changement : l'espoir de mieux connaître et comprendre la faculté de langage, et au-delà, le cerveau humain, est infiniment stimulant, quelque peu enivrant même. Heureusement l'ivresse s'arrête vite : les objectifs à portée de connaissance du modeste linguiste sont, au quotidien, un peu moins ambitieux ; ils n'en demeurent pas moins passionnants.

En soi, l'étude des états anciens d'une langue et de leur évolution est digne d'intérêt, comme l'est tout objet historique saisi dans son devenir. En particulier, et nous reviendrons sur ce point par la suite, cela permet d'appréhender des changements de formes, de constructions, mais aussi de catégories, de classes, et de formes et de constructions dans les classes. Les études diachroniques permettent en outre une meilleure compréhension de la langue que nous parlons et que nous écrivons aujourd'hui, ceci sur le plan lexical, mais aussi morphosyntaxique : le caractère opaque d'une construction, l'existence en apparence peu justifiée de variantes, etc, ce sont là des phénomènes qui trouvent une explication dans l'évolution de la langue. L'observation des changements passés de la langue permet aussi de formuler certaines hypothèses quant à son devenir. En particulier, la mise en relation de différents changements corrélés ou convergents, c'est-à-dire la mise au jour de macro-évolutions[41], permet de formuler quelques pronostics sur son devenir. Une extrême prudence est néanmoins de rigueur, et il est à vrai dire souvent plus facile de se prononcer sur ce qui a peu de chances de se produire que sur ce qui est véritablement susceptible d'advenir. Le français a ainsi opéré, entre le 12ème et 15ème siècle, une distinction croissante entre des éléments de nature et de rôle différents[42]. Il s'est ainsi développé, pour les démonstratifs, une opposition morpho-syntaxique entre déterminants et pronoms (tandis qu'elle était précédemment de nature sémantico-

[41] Macro-évolutions que l'on peut rapprocher de la notion de 'drift' proposée par E. Sapir.
[42] Voir Marchello-Nizia 2003 et Combettes 2006.

pragmatique[43]), opposition qui a touché d'autres formes, en particulier les indéfinis : ainsi émerge en moyen français la forme *chaque*, qui prend en charge les emplois de déterminant précédemment assumés par *chacun*, ce dernier se voyant à partir de ce moment réservé aux seuls emplois pronominaux. Si une évolution dans ce domaine devait se poursuivre, on fait l'hypothèse qu'elle irait dans cette même direction, plutôt que dans celle qui consisterait à confondre formes et fonctions. Des évolutions cycliques sont cependant possibles : il suffit pour s'en convaincre de penser au fabuleux destin du futur (que l'on retrouve d'ailleurs dans plusieurs langues romanes) : synthétique en latin, il se développe en bas latin sous une nouvelle forme, périphrase analytique, dont l'évolution phonétique et morphologique a conduit aux formes synthétiques du futur du français moderne. Mais, dès le 16ème siècle, apparaissent de nouvelles périphrases exprimant le futur, dont beaucoup ont eu une durée de vie assez courte, mais dont l'une s'est maintenue et imposée : « aller + infinitif ». Doit-on prédire la réduction morphologique de cette structure[44] ?

En dépit des réserves émises plus haut quant à notre aptitude à faire avancer la connaissance de la faculté cognitive de langage, il reste que la mise en relation des multiples « travaux de fourmis » effectués dans des langues et des familles de langues différentes peut faire émerger des évolutions convergentes, et nous donne à penser que ces macro-évolutions constituent une caractéristique cognitive qui subsume l'individualité des langues. On a ainsi mis au jour certaines « chaînes sémantiques » (voir (Heine et Kuteva 2000), qui synthétise de nombreuses études antérieures), en particulier celle qui conduit une expression spatiale à acquérir une valeur temporelle. La forte représentation de cette évolution dans les langues du monde a conduit à la considérer comme une routine cognitive, et même à y voir la preuve d'une primauté du spatial sur le temporel. Outre la régularité de certains changements sémantiques, on constate que le changement dans les langues se traduit régulièrement par des gains, des pertes et des modifications de formes et de constructions. Plus précisément, il apparaît que toutes les langues connaissent des phénomènes de réanalyse (modification d'une structure en profondeur sans modification de la structure en surface), des cas d'analogie (simplification des paradigmes : la forme de surface change, mais pas la structure profonde), des emprunts et des phénomènes de grammaticalisation (évolution d'une forme d'un statut lexical vers un statut

[43] Voir en particulier Marchello-Nizia 2003 et 2006.

[44] Il semble que cela soit en train de se produire en castillan, et en espagnol de la région de Mexico.

grammatical). Ainsi, si le changement est universel, les types de changements eux-mêmes le sont aussi.

2.2. Diachronie et synchronie : de l'opposition à la complémentarité

On doit à Saussure d'avoir introduit le terme de « diachronie », opposé à celui de « synchronie », pour penser la dimension évolutive de la langue, la mutabilité du signe linguistique[45]. Mais, la diachronie, dans le *Cours de Linguistique générale*, est davantage définie en creux, permettant de mieux appréhender la synchronie. Surtout, elle est pensée d'une façon qu'on peut qualifier de négative, au moins au regard des objectifs assignés par Saussure à la linguistique : appréhender le système de la langue, ce que ne saurait faire selon lui la linguistique diachronique :

> La *linguistique synchronique* s'occupera des rapports logiques et psychologiques reliant des termes existant et formant système, tels qu'ils sont aperçus par la même conscience collective. La *linguistique diachronique* étudiera au contraire les rapports reliant des termes successifs non aperçus par une même conscience collective, et qui se substituent les uns aux autres sans former système entre eux (Saussure 1916 : 140)[46].

L'idée d'absence de système est reprise, plus ou moins directement, par les héritiers de Saussure. Ainsi, selon Malmberg, la langue serait par définition synchronique, et l'envisager comme variable et en évolution revient à adopter « un point de vue qui, au fond, est incompatible avec l'idée de la langue » (1945 : 26). Une langue qui évolue serait une *contradictio in adiecto*, « bien entendu si nous comprenons par langue un système au sens strict de ce terme ». Coseriu, en particulier, s'est insurgé contre une telle position, critiquant la dichotomie établie par Saussure, et la conception statique de la langue. Il met en avant un contre-argument fort : « …L'antinomie synchronie/diachronie n'appartient pas au plan de l'objet mais à celui de

[45] « Est synchronique tout ce qui se rapporte à l'aspect statique de notre science, diachronique tout ce qui a trait aux évolutions. De même *synchronie* et *diachronie* désigneront respectivement un état de langue et une phase d'évolution » *Cours de linguistique*, 1ère partie, chap. 3 : 116-117.

[46] Dans *Du primat de l'histoire*, Coseriu critique l'idée selon laquelle « les termes existant et formant système [sont] aperçus par la même conscience collective » : « on ne saurait en réalité saisir un système linguistique dans son ensemble : on le saisit toujours seulement de façon partielle » ; et on ne saisit que la partie qui se réalise dans le discours : « Strictement parlant, personne n'a à sa disposition l'ensemble des contextes dans lesquels peuvent apparaître les énoncés linguistiquement corrects ». Par ailleurs, en raison de la variation diatopique, diastratique et diaphasique, aucun locuteur n'est confronté à un système homogène : seule la langue fonctionnelle est homogène, et elle est de ce point de vue « 'syntopique', synstratique' et 'symphasique' », c'est-à-dire étrangère aux variations de lieu, de registre et de classe sociale, ou propres à l'individu.

l'investigation : elle ne se rapporte pas au langage, mais à la linguistique »[47-48]. Rappelant que « les auteurs cités [Saussure en premier lieu] ne nient pas que, en réalité, la langue change », Coseriu en tire la conclusion que :

> l'incompatibilité n'apparaît pas entre le changement et *la réalité de la langue*, mais entre le changement et une *certaine* idée de 'la langue'. Mais comme le changement est réel, cela signifie seulement que cette idée est inadéquate. Les conflits apparents entre la raison et la réalité sont toujours des conflits de la raison avec elle-même, puisque ce n'est pas la réalité qui doit se conformer à l'intellect, mais le contraire.

Coseriu, qui sans nier le bien-fondé de l'approche synchronique déplore simplement que l'affirmation de l'importance de celle-ci se fasse au détriment de l'approche diachronique, fait (ironiquement ?) remarquer que « pour démontrer qu'un objet quelconque ne change pas, il faut l'observer en deux moments distincts. Par conséquent, même si la langue était, par sa nature même, synchronique, cela, il faudrait le démontrer en diachronie ». C'est une raison supplémentaire pour concilier système et changement, d'autant que :

> la typologie, conçue adéquatement, montre le plus clairement que les structures linguistiques ne sont pas seulement liées entre elles sur le plan 'synchronique', mais aussi (et même en premier lieu, ou originellement) sur le plan 'diachronique', étant donné qu'elles ont été créées par l'utilisation de techniques analogues ou selon des principes fonctionnels analogues (« Du primat de l'histoire », traduction en ligne, p.10).

Le système langagier est supposé remplir deux fonctions : décrire et communiquer. C'est au regard de ces deux fonctions que le changement est supposé contre-productif, facteur de trouble dans l'intercompréhension : « Les langues changent sans cesse et ne peuvent fonctionner qu'en ne changeant pas » (Bally 1932 : 18). Mais tout ne change pas en même temps, le « système » change point par point, et l'on observe une succession de déséquilibres et de rééquilibrages[49]. Ce constat répond partiellement à

[47] Traduction en ligne établie par T. Verjans : « L'apparente aporie du changement linguistique. Langue abstraite et projection synchronique », 1er chapitre de E. Coseriu, *Synchronie, Diachronie et Histoire* (*Sincronia, diacronia et historiade*, seconde édition revue par l'auteur, 1973, Gredos).

[48] Plus loin il ajoute : « Saussure n'a pas fait de l'ontologie, mais de la *méthodologie* ; il s'est occupé à distinguer la linguistique synchronique et la linguistique diachronique, ou, mieux, le point de vue synchronique et le point de vue diachronique au sein de la linguistique. En conséquence, la distinction entre synchronie et diachronie n'appartient pas à la *théorie du langage* (ou de la langue), mais à la *théorie de la linguistique*. Dans cet ordre d'idées, la conception saussurienne de la diachronie, et surtout de son inéluctable « asystématicité », est discutable et doit être corrigée (cf. VII, 1. 2.) ; en revanche, transporter cette distinction à l'objet n'est pas une simple erreur, mais une confusion, et il est urgent de l'éliminer puisque, comme disait Bacon, *citius emergit veritas ex errore quam confusione.* »

[49] Bally (1932 : 18) : « A chaque moment de leur existence, elles [les langues] sont le produit d'un équilibre transitoire ».

l'apparente *contradictio in adiecto* du changement, et de ce point de vue, comme le souligne Coseriu dans *Du primat de l'histoire*, le changement est indissociable d'une continuité :

> Car une caractéristique saillante des langues par rapport à d'autres traditions collectives est plutôt que ce sont des traditions tellement figées, c'est-à-dire qu'elles se transmettent en principe sans mutations profondes […], et qu'une accélération du changement ne se produit que dans des circonstances historiques particulières (« Du primat de l'histoire », traduction en ligne, p. 6).

A ce premier constat – le système ne change que par « touches » – , qui vient minimiser les effets supposés nocifs du changement, s'ajoute l'idée que le changement, plus qu'inoffensif, est même intrinsèquement « normal », qu'il fait partie de l'essence du langage. L'idée a été formulée dans des approches et en des termes variables[50].

Une approche, la socio-linguistique, a profondément renouvelé la question, en apportant un éclairage nouveau, et en réconciliant synchronie et diachronie à travers la notion de variabilité. A propos de cette dernière, on rappellera le propos de Labov :

> Il serait faux de concevoir la communauté linguistique comme un ensemble de locuteurs employant les mêmes formes. On la décrit mieux comme étant un groupe qui partage les mêmes normes quant à la langue (1976 : 228)
> … Il paraît justifié de définir une communauté linguistique comme étant un groupe de locuteurs qui ont en commun un ensemble d'attitudes sociales envers la langue (1976 : 338, note 40).

Tout individu maîtrise donc plusieurs grammaires, et l'hétérogénéité serait par conséquent constitutive des langues : tout sujet parlant est capable de produire et d'interpréter ces variations :

> Une maîtrise quasi innée des structures hétérogènes […] fait partie de la compétence linguistique de l'individu unilingue. L'un des corollaires de ce point de vue est que, pour une langue utilisée par une communauté complexe (c'est-à-dire réelle), c'est l'absence d'une hétérogénéité structurée qui se révèlerait dysfonctionnelle » (Weinreich, Labov & Herzog 1968, traduit dans (Labov 1976 : 100).

[50] Selon Coseriu, qui reprend cette distinction à la théorie de Humboldt, la langue n'est pas *ergon* ('produit'), mais *energia* ('activité', 'production'). Dans une telle perspective, la question des causes du changement est mal formulée : « les langues en elles-mêmes ne sont pas des entités stables qui changeraient pour telle ou telle raison encore à définir : elles ne sont rien d'autres que l'objectivation historique des actes de parler ».

Une telle approche permet de dépasser le paradoxe du changement, puisque ce dernier peut être envisagé comme le résultat d'une variation :

> Pendant un certain temps, le point de départ et le point d'aboutissement de la mutation se trouvent coexister sous la forme de deux couches stylistiques différentes […] Un changement est donc, à ses débuts, un fait synchronique » (Jakobson 1952/63 : 37)

Il n'en demeure pas moins que si toute variante est un candidat au changement, et que si tout changement comporte d'abord une étape de variation, toute variation ne débouche pas pour autant sur un changement : la variante peut disparaître, ou la variation peut se maintenir (ainsi du double système de la négation en français moderne : *ne/ne pas*). Retenons principalement ici le lien étroit entre variation en synchronie et changement diachronique, qu'a permis d'établir la socio-linguistique.

Outre le renouveau qu'a apporté cette approche dans le champ de l'évolution de la langue, il est deux facteurs qui ont entraîné un renouvellement notable de la pensée et de l'approche diachronique. Il s'agit d'une part de la constitution de gros corpus numérisés, qui, par la massification des données traitées, ont permis d'appréhender différemment les changements, qu'il s'agisse d'amender la description de certains, ou bien de préciser la chronologie d'autres (pour ne considérer que ces deux apports majeurs). Ce point sera abordé en 3. Il s'agit d'autre part, et à un autre niveau, de la mise en place de cadres d'analyse (sinon de théories) conçus pour penser et modéliser le changement, au-delà des seules lois phonétiques et de leurs conséquences morphologiques. Le cadre de la grammaticalisation, en particulier, a permis de faire une large place à la dimension sémantique du changement, tout en la reliant à sa dimension syntaxique. Nous aborderons cette question dans la partie 4.

2.3. Expliquer l'évolution

Adopter une perspective diachronique est complexe à plusieurs égards. Rappelons tout d'abord qu'une démarche véritablement diachronique ne se réduit pas à la prise en compte d'états synchroniques successifs. Elle doit expliquer comment l'on passe de l'un à l'autre. Il ne suffit pas d'enregistrer les changements, il faut rendre compte de l'évolution : « Le changement, c'est le résultat, ce que l'on constate. L'évolution, c'est le processus invisible, et largement inconscient, que le linguiste a pour tâche d'expliquer, c'est-à-dire de reconstruire » (Marchello-Nizia, 1995 : 28-29). La tâche

est évidemment bien plus ardue que celle d'empiler les résultats de changements successifs, mais elle est aussi infiniment plus stimulante. Souvent, quand on aborde l'évolution d'une construction, d'un phénomène linguistique, on possède certaines hypothèses, fondées sur des recherches antérieures, sur des bribes d'explication. Elles sont généralement jugées insuffisantes (sinon pourquoi entamer cette étude ?), voire erronées. Parfois aussi, on ne part d'aucun élément d'analyse ou d'explication, ne disposant que des données textuelles pour juger. Comment dans l'un et l'autre cas expliquer, par exemple, qu'une construction passe d'une valeur A à une valeur B ? Il n'est pas de méthode unique, la démarche dépendant en partie de l'angle d'approche privilégié (morphologique, sémantique…) et du type de phénomènes ou de constructions envisagés.

Au regard des constructions et des phénomènes auxquels je m'intéresse et de la perspective sémantico-pragmatique que je privilégie, une double démarche s'avère assez probante. D'une part – et cela n'est à vrai dire pas spécifiquement lié à l'appréhension sémantico-pragmatique des objets que j'étudie – il convient de toujours chercher à mettre en relation le phénomène ou la construction avec un(e) autre qui présente une analogie quelconque. Mon hypothèse (fondée sur la réalité des changements, au moins en français) est que les changements souvent ne se produisent pas de manière isolée. Il se peut néanmoins que l'on découvre que deux constructions ou phénomènes dont on pensait les évolutions apparentées d'une manière ou d'une autre n'ont en fait aucun point commun. Au pire on a perdu un peu de temps.

Il s'agit d'autre part d'accorder une extrême attention aux contextes dans lesquels évoluent les constructions. Je reviendrai plus longuement sur cette question dans la partie 4, mais il est un point essentiel qu'il convient de souligner dès maintenant : bon nombre de changements se produisent d'abord dans des contextes spécifiques et restreints, qui permettent l'émergence de la nouvelle valeur, de la nouvelle construction, laquelle se généralise ensuite à un plus grand nombre de contextes. On doit aux études des phénomènes de grammaticalisation d'avoir mis au jour ce processus. La difficulté consiste à repérer ces contextes précurseurs (cet aspect est en partie lié à la bonne constitution des corpus), ainsi que les traits en eux qui ont pu favoriser le changement.

Il est une autre démarche, de nature différente des deux précédentes, qui peut parfois fournir des éléments d'explication complémentaires. Il s'agit de la prise en compte de

l' « histoire externe », au sens large du terme. Une telle attitude scandalise certains linguistes « purs et durs », partisans de ne considérer que la langue, rien que la langue. Attitude voisine de celle qui a consisté en littérature, à une certaine époque, à juger sacrilège la prise en compte de la « vie de l'auteur » : le texte, rien que le texte. Il ne s'agit pas pour moi d'expliquer le changement par les facteurs externes (même si certains changements, que nous n'aborderons pas, relèvent de ce type d'explication : ceux consécutifs aux réformes de l'orthographe, par exemple), mais d'envisager s'ils ont pu jouer un rôle. C'est plus précisément à la possible influence des grammairiens, remarqueurs et autres prescripteurs que je pense : quelle a été leur position – si positionnement il y a eu – à l'égard des faits qui nous intéressent ? A-t-elle déclenché, favorisé, accéléré, un changement, ou l'a-t-elle simplement entériné, ou au contraire ralenti ? Que le « facteur externe » soit moteur, ou facteur adjuvant ou opposant, sa prise en compte ne se substitue nullement à l'analyse précise du processus langagier, qui permettra d'ailleurs, en retour, d'évaluer l'effectivité de la norme prescrite, dans le cas où elle aura été formulée de manière explicite.

Ma position demeure celle d'une diachronicienne qui s'attache à l'histoire interne des changements. Mais écarter par principe l'histoire externe, selon moi, non seulement correspond à une vision réductionniste de la langue, mais aussi – et surtout – fait courir le risque de passer à côté de facteurs décisifs, d'avancer avec des oeillères[51]. De même qu'une construction évolue dans un contexte linguistique, une langue évolue dans un contexte socio-historique.

Une approche diachronique qui se veut exhaustive dans sa démarche explicative ne doit pas se priver de la prise en compte de facteurs externes s'ils sont pertinents. *A minima* il faut connaître ce qui s'est dit – le cas échéant – pendant la période considérée sur l'objet d'étude. Les « observations des grammairiens et remarqueurs classiques » citées par N. Fournier dans sa *Grammaire du Français Classique* (1998), à la fin de chaque question traitée, sont à cet égard fort précieux.

[51] La démarche est analogue à celle prônée en conclusion de 1 : il s'agit ici aussi de ne pas complétement se couper des conditions de production des constructions sur lesquelles on travaille.

2.4. Le facteur temps : la périodisation de la langue et la chronologie des phénomènes

L'approche diachronique a pour vocation d'*expliquer* les changements (voire l'évolution, processus invisible qui sous-tend les changements, et qui est plus complexe à mettre au jour). Expliquer est une tâche complexe en soi, mais elle l'est d'autant plus que l'on se trouve face à une masse continue, le temps, qui, pour des raisons à la fois cognitives et méthodologiques, ne peut être appréhendé ainsi, mais doit être segmenté. Il faut, d'une manière ou d'une autre, discrétiser le continu, procédure qui concerne à la fois, dans l'absolu, l'objet langue, et, dans la pratique, sa représentation et sa perception à travers les textes, et donc la constitution du corpus (aspect qui sera plus spécifiquement développé en 3.), d'où, on le voit bien, le risque d'une certaine circularité.

La période pertinente pour les phénomènes ou les constructions auxquels je m'intéresse s'étend potentiellement du 12ème au 17ème siècle, époque à laquelle les changements sont achevés, ce que nous prouve l'observation des faits dans les siècles suivants. De ce point de vue, travailler avec le recul du temps est un avantage certain.
Cela constitue néanmoins aussi un possible danger : la granularité du temps a tendance à perdre de sa finesse, cela d'autant que la période considérée est large. L'on ne peut cependant, ne serait-ce que pour des raisons pratiques, adopter le même grain temporel lorsque l'on travaille sur 7 siècles ou sur 50 ans.
Représenter la continuité temporelle est complexe ; lorsque celle-ci s'accompagne de changements, il semble y avoir à la fois nécessité et possibilité d'opérer une discrétisation de l'espace temporel, de distinguer des états de langue successifs et différents[52]. Il faut à cet égard distinguer deux choses : le fait d'opérer une périodisation générale de la langue, et celui de déterminer la chronologie d'un phénomène particulier. C'est le second qui nous intéresse prioritairement, mais il convient néanmoins de dire quelques mots de la périodisation générale. L'établissement de périodes successives du français est artificiel au sens où il est effectué *a posteriori*, et selon des critères modernes ; cependant, seul le recul du

[52] Même si : « on ne sait pas déterminer de façon certaine et fine quand l'on peut dire que deux langues sont la même ou sont différentes : c'est la question disputée, et en fait insoluble, des dialectes ; c'est aussi celle de la diachronie : quand le latin a-t-il cessé d'être le latin ? quand le français cesse-t-il d'être le français ? etc . ; du même coup, on ne sait pas les dénombrer. » (Milner 1989 : 46).

temps, précisément, le rend possible. Il est rare en effet que les locuteurs aient le sentiment de parler une langue suffisamment modifiée pour qu'elle soit perçue comme différente de celle qu'ils parlaient précédemment[53]. Le découpage est artificiel aussi parce qu'il peut donner à penser qu'il délimite des périodes homogènes, qui peuvent l'être d'ailleurs pour des raisons contraires. Ainsi l'ancien français est considéré comme une période stable, sans changement important, tandis que le moyen français est caractérisé comme le moment de tous les bouleversements : réduction des hiatus et des diphtongues sur le plan phonétique, simplification des paradigmes verbaux et chute de la déclinaison sur le plan morphologique, disparition de la contrainte du verbe en seconde position et fixation de l'ordre des mots sur le plan syntaxique, pour ne citer que les changements majeurs. Certes, ceux-ci ont bien eu lieu, mais ils avaient, au moins pour certains d'entre eux, commencé plus tôt. Ainsi la déclinaison (qui n'a au demeurant jamais été respectée de façon absolue) s'effrite dès le 13ème siècle, et l'ordre des mots commence à se fixer à la même époque : comme l'a montré C. Marchello-Nizia (1995), la position de l'objet nominal est devenue majoritairement postverbale au 13ème siècle, ce qui signifie que, en deux siècles, sa syntaxe a largement évolué (le sujet, lui, ne commencera à se fixer devant le verbe qu'au siècle suivant). L'ancien français lui-même ne correspond donc pas à une période entièrement stable, il connaît des changements qui introduisent une nécessaire dimension diachronique. Comme le rappelle C. Marchello-Nizia :

> En fait, tout état de langue est à la fois, nécessairement, 'transition' et 'stabilité', dans la mesure où toute langue naturelle change, continûment, tout en maintenant un équilibre indispensable à l'intercompréhension. (1979/1997 : 7)

Imaginer que la langue puisse rester parfaitement stable pendant des années et changer brusquement et massivement avant de s'installer dans une nouvelle phase de stabilité, d'une part est contesté par ce que nous savons de la réalité historique, d'autre part va à l'encontre de notre bon sens linguistique, donnant à imaginer un scénario pour le moins catastrophique au regard de la fonction communicative du système linguistique : pauvre locuteur pris entre les deux états de langue…(cela peut néanmoins parfois se produire pour le lexique : durant la période révolutionnaire on a

[53] Cela arrive néanmoins : c'est une telle situation qu'a entérinée le Concile de Tours en 813 en demandant aux ecclésiastiques de désormais prononcer leurs prêches en *lingua romana rustica*, le latin s'étant suffisamment érodé au fil des siècles pour que sa forme classique ne puisse plus être comprise des fidèles, ni que la langue parlée ne puisse porter ce nom.

ainsi radicalement – et assez brutalement - modifié le système des poids et mesures, de même que celui du calendrier, ce dernier, au vu de son impopularité, ayant finalement été aboli par Napoléon le 11 nivôse de l'An XIX (1er janvier 1806)).

La variation est constante, le changement continu, mais les ruptures sont plus ou moins marquées :

> Il y aurait ainsi des durées sans ruptures nettes, ce qui ne signifie pas absence d'évolution, et des moments où le changement serait plus perceptible, ne serait-ce que parce qu'il concernerait un nombre plus grand de sous-familles d'unités. Dans une telle perspective, la notion de période ne va donc pas systématiquement de pair avec celle de stabilité, d'homogénéité ; elle correspond à une absence de rupture, de palier, cette caractéristique se trouvant généralisée dans les différentes composantes du système linguistique, mais également dans les groupes – diversement définis – de locuteurs. La période se définit alors par défaut, en quelque sorte, comme l'espace entre deux moments où le changement se trouve plus nettement marqué (Combettes, à paraître[54]).

Dans cette perspective, l'ancien français correspondrait à une durée « sans rupture nette » (mais avec néanmoins des changements) et le moyen français à une durée où les changements seraient plus perceptibles, la rupture progressivement opérée trouvant son aboutissement à l'issue des deux siècles.

Bien que la délimitation de l'ancien et du moyen français ne repose pas toujours sur des critères exclusivement linguistiques[55], ceux-ci y occupent néanmoins une place prépondérante, ce qui ne signifie pas qu'il y ait consensus, comme en témoignent en particulier les divergences pour clore la période du moyen français, entre la fin du 15ème siècle et la fin du 16ème siècle selon les cas[56]. On peut d'ailleurs se demander si la difficulté tient au bornage final du moyen français ou au bornage initial de la période suivante, ce qui n'est peut-être pas exactement la même chose, mais qui témoigne en tout cas d'une difficulté à caractériser nettement la période qui suit. Et de fait, jusqu'il y a peu, la périodisation postérieure au moyen français non seulement s'appuyait pour une large part sur des critères non linguistiques (historiques, culturels, littéraires…), mais tendait à procéder à un découpage par siècles : 16ème siècle et sa « langue de la Renaissance » (élégant label, mais peu linguistique), 17ème siècle et son « français classique » …, comme si les changements linguistiques, disciplinés et ordonnés, avaient l'extrême délicatesse de se regrouper aux tournants de siècles, afin

[54] Sur la périodisation, voir aussi (Combettes et Marchello-Nizia 2010).

[55] Ainsi, selon Brunot (1905-1938, tome 1 : 421), le moyen français est borné par l'avènement des Valois sur sa gauche et par celui des Bourbons sur sa droite.

[56] Voir sur cette question l'introduction de (Marchello-Nizia 1979/1997a), ainsi que (Smith 2002).

de faciliter la tâche des linguistes. Prenant appui sur des arguments plus linguistiques, des découpages plus fins ont été proposés, en particulier pour délimiter une période dite « préclassique », qui s'étendrait du milieu du 16ème siècle au milieu du 17ème siècle. Nous ne discuterons pas davantage la question de la périodisation, renvoyant à (Combettes et Marchello-Nizia 2010) pour une synthèse éclairante de ses modalités et de ses enjeux (en particulier : 131-137). Citons-les pour conclure :

> Même si certaines époques de l'histoire d'une langue apparaissent comme plus ou moins stables, le changement linguistique est généralement considéré comme un mouvement continu, qui ne se caractérise pas par des ruptures brutales, des séparations nettes entre des états de langue successifs. Cette continuité est indissociable de la variation, moteur essentiel de l'évolution; […] Dès lors, la reconnaissance de « périodes » pose forcément problème dans la mesure où il s'agit de déterminer des ruptures dans ce qui apparaît comme un continuum. Il y a toutefois, dans cette volonté de reconnaître des étapes successives dans une histoire, une tradition constante dans l'analyse linguistique, qui concerne les langues des familles les plus diverses, pour autant que leur histoire puisse être documentée (p. 131).

Ainsi, si le découpage de la langue en périodes est artificiel, grossier, en partie subjectif, et donc très imparfait, il apparaît néanmoins nécessaire. Dans notre perspective – décrire l'évolution de certaines constructions et phénomènes linguistiques – il constitue une toile de fond sur laquelle se déroulent lesdits changements.

Ce sont en effet les délimitations chronologiques, et leurs étapes internes, liées à l'évolution de certaines constructions ou phénomènes qui nous intéressent plus spécifiquement (ce qui n'empêche pas de les inscrire dans une chronologie qui les subsume). Une diachronie, comme une synchronie, d'ailleurs, n'est jamais donnée dans l'absolu (les synchronies générales évoquées plus haut constituent à cet égard une sorte de compromis résultant de la prise en compte de l'ensemble des phénomènes linguistiques) : il faut toujours la construire.

Et pour cela, plusieurs questions se profilent, toutes en partie subordonnées à celle-ci : comment discrétiser le temps ? La question se pose à vrai dire à un double niveau : en amont, d'un point de vue méthodologique, il s'agit d'opérer la segmentation que l'on pense la plus pertinente, « stratégique », pour l'observation du phénomène à étudier. On peut procéder un peu au hasard, en effectuant des coups de sonde, ou bien s'appuyer sur des hypothèses. En aval, à l'issue de l'étude, il s'agit de mettre au jour la chronologie de l'évolution du phénomène observé, en dégageant les jalons décisifs.

Il s'agit là d'une discrétisation d'un autre ordre, non seulement parce qu'elle est déduite *a posteriori*, et non présupposée, mais aussi parce qu'elle doit servir à reconstituer une continuité, celle du changement étudié.

Mais c'est la segmentation « en amont », et sa dimension méthodologique, qui nous intéresse ici. Comment en effet opérer des coupes, c'est-à-dire discerner les frontières où les changements sont susceptibles de se produire ? Pour une approche diachronique, il faut d'abord déterminer la « taille » des synchronies (on ne peut faire autrement) : 20 ans, 50 ans …, puis leur fréquence : tous les 20 ans, 50 ans… Ces choix sont en partie conditionnés, ou corrigés, par des considérations pratiques : la disponibilité des textes, point que nous aborderons en 3. Le découpage prévu peut donc être parfois revu, par manque de textes pour l'une des tranches retenues. Ces considérations « matérielles » mises à part, il reste que les choix opérés en matière de taille et de fréquence des coupes mettent en cause, d'une part la granularité du découpage, d'autre part la conception de la nature même de l'évolution, c'est-à-dire le caractère brusque ou progressif des changements.

2.5. Le facteur temps : le rythme des changements

2.5.1. La nature des changements : changements catastrophiques ou progressifs

On observe deux positions majeures en ce qui concerne la nature des changements. L'une d'elle consiste à considérer que les changements surviennent brusquement : c'est la conception qui avait été développée dans un premier temps dans le cadre de la grammaire générative : selon Lightfoot (1979), un changement linguistique ne peut être que 'catastrophique', dans la mesure où il suppose une *réanalyse*, et que seuls les enfants sont capables d'accomplir une telle opération, en analysant différemment les énoncés entendus. Les changements linguistiques sont donc liés aux changements générationnels. Lightfoot a par la suite adopté une position moins radicale, tandis que les travaux de Romaine (1989) et de la socio-linguistique ont de leur côté conduit à réviser cette position. Il a en particulier été montré que la grammaire des adultes peut se modifier et qu'un adulte peut maîtriser plusieurs grammaires (cf. la notion de 'variante' et de 'grammaires en compétition'). En outre, l'idée que le changement doive être initié par un groupe influent est assez convaincante, or les enfants en bas âge ne constituent pas un tel groupe. Un changement peut être brutal chez un locuteur, mais néanmoins progressif à l'échelle de l'ensemble des locuteurs. De ce point de vue,

grammaire générative et socio-linguistique ont des visées fondamentalement différentes, qui expliquent partiellement leur différence de vue sur la nature du changement : la grammaire générative s'intéresse à la compétence du locuteur idéal alors que la socio-linguistique (et la grammaticalisation) s'occupe des usages.
La seconde position, développée dans le cadre de la grammaticalisation en particulier, correspond de fait largement aux vues de la socio-linguistique exposées ci-dessus. Elle envisage les changements comme inscrits dans un continuum, avec par exemple coexistence temporaire d'une forme nouvelle et d'une forme ancienne, la seconde, initialement dominante, devenant minoritaire avant, souvent, de disparaître ou de « se refaire une jeunesse » dans d'autres emplois. Les textes nous montrent que, au moins en syntaxe et en sémantique, c'est plutôt selon un mode continu qu'évoluent les constructions langagières. Cette réalité incite à réduire autant que possible les « ellipses » temporelles, qui pourraient occulter une phase importante du développement du phénomène étudié.

On constate que les changements se produisent d'abord dans certains « contextes » (qu'ils soient d'ordres linguistiques, liés aux genres, ou aux registres de langage…), puis s'étendent à un plus grand nombre (voire à l'ensemble) de(s) contextes. Il s'agit donc d'une diffusion à la fois « spatiale » et temporelle.
La difficulté majeure consiste à identifier les étapes, et donc les contextes où s'initient les changements. Pour les contextes textuels (*versus* liés aux genres ou aux registres), il s'agit souvent (toujours ?) de contextes ambigus, dans lesquels une autre interprétation est possible. Un contexte linguistiquement ambigu autorise et peut favoriser le développement d'une inférence, susceptible, ensuite, de se sémantiser progressivement.
Ces questions seront développées et exemplifiées dans la partie 4, dans laquelle nous aborderons la grammaticalisation de certaines constructions.

2.5.2. La vitesse des changements

Tous les changements ne se produisent pas à la même vitesse : on observe ainsi des différences selon les domaines. En particulier, comme le rappelle Combettes (à paraître), le lexique ne peut suivre exactement le mouvement de la morphosyntaxe, en

raison de sa dépendance de facteurs externes, tels que l'emprunt ou le rôle des traductions.

Par ailleurs, à l'intérieur même d'un domaine, tous les changements ne se produisent pas à la même vitesse : l'objet nominal semble s'être fixé derrière le verbe plus rapidement que le sujet devant le verbe, indépendamment du fait que le mouvement du premier a été plus précoce (dès l'ancien français).

Lorsque l'on aborde la question de la vitesse, il convient en fait, en toute rigueur, de distinguer deux aspects : d'une part la durée globale de chaque changement (une génération, un siècle, …), d'autre part la vitesse plus ou moins grande de la diffusion d'un changement : son rythme. A. Kroch (1989) a ainsi mis au jour, pour certains changements, un schéma d'évolution qu'il a appelé 'courbe en S' ('S-curve') : dans un premier temps les emplois augmentent lentement, gagnant progressivement de nouveaux contextes, puis dans un second temps leur fréquence s'accroît rapidement et pareillement en tous contextes, avant de ralentir, formant ainsi une sorte de palier.

La question de la vitesse des changements est infiniment compliquée à différents égards. Premier constat, un phénomène est souvent complexe. Ceux auxquels je m'intéresse comprennent au minimum une dimension syntaxique et sémantique (parfois aussi phonético-morphologique, dans le cas des pronoms personnels sujets par exemple). Mon hypothèse est qu'il y a toujours d'abord une réinterprétation sémantique – possiblement d'ordre analogique – qui peut impulser des changements syntaxiques (une réanalyse par exemple), laquelle à son tour accélère et/ou accroît le changement sémantique initial. Idéalement il faudrait distinguer différentes vitesses de changement pour les différents niveaux considérés (syntaxique, sémantique…), mais c'est difficile ; et vouloir disséquer le phénomène n'a peut-être finalement pas grand sens : tout est lié. Il reste que, établir les temporalités exactes – durée et rythme – n'est pas simple, et suppose évidemment de travailler sur un corpus suffisamment représentatif du ou des état(s) de langue dans le(s)quel(s) s'inscrit l'évolution du phénomène.

Il est un autre aspect intéressant à mettre au jour. Il est probable que l'évolution de certaines constructions connaît un effet d'entraînement, dû à l'évolution d'autres constructions appartenant à un même paradigme ou à un même micro-système. Mais il n'est pas toujours facile de distinguer ce phénomène du processus plus global que sont les macro-évolutions, c'est-à-dire le fait qu'une même évolution affecte différentes

constructions, pour des raisons qui nous échappent largement. Ainsi le français a opéré, dès l'ancien français mais surtout à partir du moyen français, une distinction des éléments de nature et de rôle différents, en particulier entre déterminants et pronoms (voir Marchello-Nizia 2003 et Combettes 2006) : le mouvement a affecté les démonstratifs, les possessifs, et, moins systématiquement et un peu plus tardivement, les indéfinis : l'évolution de ces derniers a-t-elle été « entraînée » (dans le double sens : causée et accélérée) par les deux autres, ou bien était-elle de toute façon *destinée* à se produire ? On touche ici à des questions auxquelles il est difficile de répondre.

Il est cependant fort probable que l'avancée dans l'établissement de la durée et de la vitesse des changements nous permettra de progresser dans la mise au jour de corrélations (voire de relations de causalité) entre certains changements.

Plus modestement, mais sans pour autant perdre de vue cet objectif, ma démarche consiste à essayer de repérer les étapes du changement à travers les textes. Il s'agit de dégager une chronologie, liée d'une part aux genres textuels (si cela s'avère pertinent : cela l'est par exemple pour les marqueurs de topicalisation, qui se développent d'abord dans les textes argumentatifs), liée d'autre part aux contextes linguistiques. Plus précisément, le but est d'identifier ceux où se produisent les premiers changements, et d'essayer de déterminer pourquoi, en particulier en dégageant les traits distinctifs, le cas échéant, desdits contextes. Dans une telle perspective, le modèle du changement sémantico-syntaxique en quatre étapes proposé par Heine (2002) s'avère fort intéressant, dans la mesure où il subdivise l'étape de transition (entre le stade originel et le stade final), et distingue plus précisément un contexte de 'transition' (*bridging context*), dans lequel la construction d'une inférence donne lieu à une nouvelle signification qui passe au premier plan, et un contexte de 'bascule' (*switching context*), incompatible avec la signification d'origine, laquelle passe au second plan.

Il reste que la mise en œuvre concrète de ce modèle est difficile. Notre point d'accroche principal réside dans l'établissement des fréquences des constructions : c'est en effet leur évolution qui permet de statuer sur les changements, d'où l'extrême attention à porter à la constitution du corpus (qui fournit les données).

Chapitre 3. Travailler en corpus[57]

3.1. Les textes comme seul accès à la langue

Le linguiste qui travaille sur une langue ancienne a un avantage sur ses collègues « modernistes » : il échappe au possible dilemme entre « linguistique de terrain » et « linguistique de bureau », puisqu'il ne peut recourir à l'introspection (voir 1.1.). Il est cependant vrai que les deux approches semblent depuis plusieurs années moins inconciliables, dans la pratique des linguistes, qu'elles ne l'ont été par le passé. Certes, comme le souligne Corbin :

> Elles [l'introspection, le corpus] découpent le champ des recherches linguistiques en deux domaines […], et dont aucun ne peut légitimement être présenté comme incarnant à lui seul LA linguistique : éventuellement complémentaires, voire présentant certaines intersections, ces deux linguistiques ne peuvent pas avoir globalement le même objet. (1980 : 121)

Mais ce partage du champ linguistique est désormais davantage perçu – comme le préconisait d'ailleurs Corbin en 1980 – comme une possible richesse complémentaire. Le revirement semble surtout s'opérer en faveur d'une utilisation accrue des corpus (qui vient légitimer les conclusions issues de l'introspection), le développement et l'enrichissement de ces derniers étant probablement l'une des raisons de cette évolution. Notons aussi que le développement et la vulgarisation (au sens positif du terme) des expérimentations psycho-linguistiques (et plus récemment des expériences neuro-psycho-linguistiques) ont encore reconfiguré la cartographie du champ des recherches linguistiques, dont on peut considérer qu'elles s'articulent désormais en trois domaines – corpus, introspection, expérimentations – qui nous fourniront peut-être un jour une vision complète des langues. Au moins des langues vivantes : si le recours à l'introspection est hautement problématique pour les états anciens de la langue (voir discussion 1.1.), le recours aux expérimentations psycho-linguistiques est évidemment exclu.

C'est donc uniquement à travers les textes qui nous sont parvenus que nous percevons la langue ancienne. Or les textes constituent par ailleurs l'objet d'étude. Le linguiste se trouve donc, à première vue, dans une situation quelque peu circulaire : étudier les textes avec une connaissance de leur langue fondée sur ces mêmes textes (même si

[57] Plusieurs des questions abordées ici ont été en partie développées dans Prévost 2005 et 2008b.

c'est indirectement, par le biais des grammaires qui se sont construites à partir de leur examen). Il est cependant possible d'échapper à la circularité, d'une part en se penchant sur des données non encore (pleinement) étudiées, d'autre part en prenant en compte des textes encore peu sollicités : ainsi peut-on confirmer, infirmer, corriger, compléter des propriétés précédemment mises au jour, ou en dégager de nouvelles. C'est donc grâce à l'accroissement des textes et des données pris en compte que peut progresser notre connaissance de la langue (et grâce aussi, nous le verrons, aux réflexions théoriques et méthodologiques sur la nature du corpus et son exploitation différenciée).

Le « corpus » est donc indispensable au linguiste travaillant sur une langue ancienne, mais il a néanmoins la possibilité de relativiser ce que révèlent les textes, par la maîtrise de leurs caractéristiques externes (même si cela n'est pas toujours une tâche aisée).

3.1.1. L'absence d'oral

Travailler sur des états anciens de la langue exclut toute prise en compte sérieuse de sa dimension orale. Mieux vaudrait parler d'ailleurs de **ses** dimensions orales. L'engouement, relativement récent, pour l'étude de l'oral[58] a ceci d'étonnant que celui-ci y est régulièrement présenté (au moins par ceux qui le découvrent comme objet possible d'investigation) comme un bloc monolithique : L'oral. Attitude surprenante dans la mesure où cet intérêt accru n'est pas complètement étranger à l'attention plus générale désormais portée aux « registres », elle-même consécutive et conséquente à la mise au jour par les travaux de socio-linguistique de l'importance de la variation langagière. L'explosion depuis une vingtaine d'années des linguistiques de corpus s'inscrit d'une certaine façon dans cette même lignée.

Que l'oral soit réduit par certains à un seul registre alors que c'est précisément l'attention portée aux registres qui l'a propulsé sur le devant de la scène linguistique a de quoi surprendre. L'oral, comme l'écrit, se décline sous de multiples formes (oral planifié, spontané, etc).

[58] J'entends dans le « grand public linguiste ». L'intérêt pour l'oral ne date évidemment pas d'hier, mais les travaux pionniers de C. Blanche-Benveniste (pour ne citer qu'elle), en dépit du grand intérêt qu'ils ont pu susciter, n'en ont pas moins été considérés par certains comme s'attachant à une dimension marginale de la langue. L'oral comme une sorte d'épiphénomène langagier.

La question de la diversité des registres de l'oral est néanmoins non pertinente pour ce qui regarde les états anciens du français. Aucun accès direct à l'oral n'est possible (ni ne le sera jamais) : nul n'entendra jamais le « son » de l'ancien ou du moyen français, les prononciations reconstruites gardant leur part d'inconnu et d'incertitude et ne nous disant par ailleurs que bien peu sur la prosodie de cette époque[59]. Outre la musique de la langue, ce sont aussi ses spécificités, en particulier syntaxiques, qui nous échappent. Or on peut supposer que, comme aujourd'hui, l'écart entre écrit et oral n'était pas négligeable. Nous possédons certes quelques indications, certaines provenant des grammaires, « manières de dire » et autres témoignages de l'époque, d'autres d'une mise en écrit de l'oral. La première source d'informations tend parfois à stigmatiser certains usages, dont on ne peut dire s'ils sont représentatifs de la majorité, et elle comporte une dimension prescriptive forte, dont on ignore si elle aura été suivie d'effets (problème de l'effectivité de la norme prescriptive). Pour ce qui est de la seconde source, qu'il s'agisse de dialogues dans un roman ou d'une pièce de théâtre, on sait que l'écart est souvent important avec l'oral réel, celui-ci étant passé au crible de la norme, qui peut conduire à une sorte d'hypercorrection stylistique, ou étant au contraire caricaturé vers le bas. Il faut donc considérer avec prudence ce témoignage : que l' « oral représenté » (Marchello-Nizia, à paraître) doive être pris en compte et ses spécificités mieux étudiées est un fait, mais qui ne doit pas faire oublier qu'il s'agit toujours d'un *oral* mis en écrit.

3.1.2. Une vision à la fois partielle et totale de la langue

Si l'accès aux registres de l'oral est fermé par principe, celui à certains registres de l'écrit l'est dans la pratique. Nous ne pouvons accéder qu'aux données langagières qui se trouvent dans les textes dont nous disposons. Or, d'une part nous savons que certains genres ne nous sont parvenus qu'à travers bien peu de textes, d'où la difficulté à caractériser leur « langue ». D'autre part il est fort probable que nous ignorons l'existence de certains. En outre, si les visées de certains textes nous sont connues, nous ne maîtrisons que bien imparfaitement leurs conditions de production et de réception. Comme nous l'avons souligné en 1.1., le linguiste qui travaille sur des états anciens de la langue a donc conscience qu'il ne peut accéder qu'à une vision

[59] Même si l'on peut supposer, comme le suggère C. Marchello-Nizia (1995 : chap. 6), que le français médiéval a connu un accent de micro-syntagme, entre l'accent de mot du latin et l'accent de macro-syntagme ou de phrase du français moderne.

parcellaire de l'état de la langue qu'il étudie. Mais dans la mesure où il ne peut prétendre à plus que cette réalité partielle (à moins de découvrir de nouveaux textes), on peut considérer qu'elle constitue la langue dans sa totalité. Il faut donc admettre que la langue appréhendée par le linguiste n'est qu'un sous-ensemble de la langue des locuteurs de l'époque. On peut faire un constat voisin pour qui travaille sur un état de langue contemporain : on n'appréhende jamais que des états langagiers, et non la langue comme ensemble de tous les usages possibles. Mais, dans ce dernier cas, le décalage est dû à une maîtrise nécessairement imparfaite de l'abondance des données langagières, dont la production est en outre continue (et qui ne peuvent donc être circonscrites dans leur totalité). Pour le linguiste qui travaille sur la langue ancienne, cela tient en en premier lieu à la nature de l'objet appréhendé : les textes dont peut disposer le linguiste ne correspondent qu'à un sous-ensemble de la langue des locuteurs de l'époque. Toutefois, ils constituent en même temps, pour lui, locuteur moderne, la totalité de cette langue. Il y aurait ainsi trois ensembles : tout d'abord celui de la totalité des textes produits dans un état de langue ancien, auquel nous n'aurons jamais accès ; ensuite celui de l'intégralité des textes qui nous sont restés, encore en partie inexplorés, et même inconnus pour certains, et enfin, sous-ensemble du précédent, les textes que nous connaissons et que nous exploitons. L'objectif étant de rendre le troisième ensemble coextensif avec le second, et d'accéder à la connaissance de tout l'existant. Evidemment il ne s'agit pas simplement de « connaître » l'existence des textes (opération de recensement, qui n'est cependant pas triviale), il faut les exploiter, ce qui suppose pour beaucoup d'en éditer le ou les manuscrit(s) (afin de les rendre accessibles à un grand nombre), de les numériser, d'indexer les données et de les analyser. Il s'agit certes d'une lourde tâche, de grande ampleur, mais si un tel projet voyait le jour (il n'existe pas à ma connaissance), peut-être serait-il susceptible de bouleverser notre vision de la langue, nous la rendant accessible dans ses multiples variations (qu'il conviendrait néanmoins d'objectiver en les mettant au jour et en les classant). Dans une telle perspective, un « corpus de référence » ne constituerait plus l'idéal inaccessible qu'il est actuellement, quel que soit l'état de langue considéré. Il l'est au regard de la variation qui sous-tend la langue, ensemble composite de différents « registres » dont il convient de dégager les caractéristiques, comme l'a souligné Biber[60]. Mais, comme le fait remarquer M.-P.

[60] Voir en particulier Biber (1988), (1990), (1993) et (1995).

Péry-Woodley (1995) en envisageant la constitution d'un corpus équilibré représentatif de la langue générale, la non-maîtrise de la variation langagière reste encore un obstacle majeur à cela :

> le corpus équilibré est sans doute celui qui a 'de tout un peu', mais encore faudrait-il savoir ce qu'est 'tout', c'est-à-dire quelles sont les classes à représenter – ce qui nécessite un modèle complet de la variation –, et avoir accès à des textes les représentant. (Péry-Woodley, 1995 : 218).

Si l'établissement d'un modèle de la variation reste assurément un enjeu de taille, la perspective d'accéder un jour à l'ensemble des textes constitue un premier pas vers la faisabilité d'un corpus de référence.

Un tel corpus n'existe pas à l'heure actuelle, et des considérations pratiques font que, même s'il existait, il ne serait probablement pas utilisé intégralement dans le cade d'une étude spécifique : se pose dans tous les cas, bien que selon des modalités différentes, la question de bien représenter l'état langagier que l'on souhaite observer, c'est-à-dire de bien « panacher » les textes. Il s'agit de préoccupations qui touchent à la représentativité de la langue étudiée et à la question de la constitution des corpus, que nous développerons dans le travail de recherche inédite.

3.2. Une tradition de travail sur les textes renouvelée par la numérisation des textes et la linguistique de corpus

3.2.1. Un rapport nouveau aux textes et aux données

Les linguistes travaillant sur des langues anciennes ont toujours travaillé sur les données attestées dans les textes. Jusqu'à encore récemment, ils l'ont fait dans les livres et le crayon à la main, et donc, individuellement, sur un nombre de textes restreint (pour des raisons matérielles de temps). C'est toujours sur des données attestées qu'ils le font aujourd'hui, mais celles-ci sont désormais le plus souvent (aussi) disponibles sous forme numérisée, ce qui permet leur massification et un traitement facilité.

Il est à cet égard assez instructif de considérer l'évolution du nombre de textes (ainsi que de leurs modalités d'utilisation : analysés plus ou moins exhaustivement ou simples réservoirs à exemples) sur lesquels se fondent les grammaires, évolution largement liée aux possibilités nouvelles offertes par la numérisation (mais aussi

parfois à des choix personnels). Le tableau ci-dessous indique le nombre de textes utilisé par quelques grammaires de français médiéval :

	Date parution	**Nombre de textes utilisés**
Ancien français		
Petite syntaxe de l'ancien français, L. Foulet	1919 (1ère éd.) 1928 (3ème éd.)	13 19
Grammaire de l'ancien français G. Moignet	1984	102
Grammaire nouvelle de l'ancien français Cl. Buridant	2000	350 (+ index et concordances électroniques)
Moyen français		
Syntaxe du moyen français R. Martin et M. Wilmet	1980	24 (dont 3 numérisés et analysés exhaustivement)
La langue française aux 14e et 15e siècles C. Marchello-Nizia	1997 (1ère éd. 1979)	69

On constate que les écarts sont donc importants, même si l'on regrette de ne pouvoir les préciser en nombre d'occurrences, et que l'on aimerait savoir quelle utilisation exacte a été faite des textes. Ils peuvent en effet avoir été dépouillés exhaustivement, c'est-à-dire que la description des phénomènes linguistiques se fonde sur un relevé systématique de leurs occurrences ; celle-ci peut au contraire s'appuyer sur des sondages réguliers ; ou bien encore, les textes peuvent servir de simples réservoirs à exemples pour illustrer les phénomènes décrits. De plus, tous les textes utilisés pour un ouvrage ne sont pas nécessairement soumis au même traitement. Ces aspects méthodologiques ne sont généralement pas explicités (C. Marchello-Nizia distingue cependant les « textes dépouillés ou consultés » et ceux « auxquels il est fait occasionnellement référence »). Ces choix ne sont pourtant pas sans conséquence sur les descriptions proposées.

Au-delà de la constitution des grammaires, c'est plus généralement la méthodologie des études en langue ancienne et en diachronie qui a été bouleversée par l'utilisation des textes numérisés : cette dernière a en effet introduit des méthodes de travail inédites, et ce faisant un rapport différent aux textes et à la langue, avec des avantages, et des inconvénients.

Dans une histoire des études en langue ancienne, on pourrait ainsi distinguer différentes « générations », en fonction des types de support utilisés : les linguistes qui n'ont connu que les textes sur papier (jusqu'aux années 1960), ceux qui ont vu naître le support informatique et la représentation numérique, en continuant à travailler parallèlement sur le papier, enfin ceux qui, depuis une bonne dizaine d'années, ne travaillent (quasiment) plus que sur des textes numérisés. La tripartition suggérée est grossière, et la réalité évidemment plus nuancée, au moins pour ce qui est de la catégorie médiane. Ainsi, j'appartiens moi-même à cette génération intermédiaire, mais j'ai un pied dans la génération suivante ; à l'inverse, mes « aînés », ceux qui ont pris récemment leur retraite (d'enseignant), font eux aussi partie de cette génération, tout en appartenant à la génération précédente : ils ont d'abord travaillé en noircissant les marges des livres de commentaires et de comptages, avant de participer à l'émergence des textes numérisés et des outils permettant de les exploiter. De mon côté, j'ai débuté mon activité de recherche il y a une petite vingtaine d'années, et les textes numérisés faisaient déjà partie du paysage, bien que ce soit sans commune mesure avec ce qui existe aujourd'hui, tant au niveau du nombre de textes numérisés que des outils permettant de les manipuler. A vrai dire, les textes numérisés ne m'ont que peu servi dans mes premiers travaux, et cela pour deux raisons : je travaillais sur le moyen français, et les textes que j'avais sélectionnés pour ma recherche n'étaient pour la plupart pas encore numérisés. Surtout, je travaillais sur la position du sujet, tant nominal que pronominal. Or si le sujet pronominal est facile à repérer automatiquement, même avec des outils rudimentaires (il correspond à une liste de formes finies), il n'en va pas de même pour les sujets nominaux, pour lesquels il convient, idéalement de posséder des textes annotés syntaxiquement, ou bien de développer des règles de repérage complexes. Ces dernières supposent, d'une part, un outil capable de les mettre en œuvre[61], d'autre part une connaissance approfondie des

[61] Le logiciel *SATO* le permettait dès cette époque. Conçu par François Daoust (centre ATO, UQAM, Montréal), il s'agit d'un moteur de filtrage et d'annotation, qui permet des recherches fines en contexte s'appuyant sur l'utilisation d'expressions régulières. Voir <http://www.ling.uqam.ca/sato>.

conditions de positionnement des sujets : or c'est précisément ce que l'étude cherchait à établir. Elaborer des règles permettant de repérer automatiquement les sujets nominaux préverbaux et postverbaux aurait pu constituer une suite à mon travail de thèse, mais pouvait en revanche difficilement le précéder. C'est donc en quelque sorte « à l'ancienne » que j'ai travaillé pour cette recherche spécifique.

Mon insertion à partir de 1993 dans l'équipe d'accueil ELI (*Equipe Linguistique et Informatique*, ENS Fontenay/Saint-Cloud) dirigée par Christiane Marchello-Nizia et Benoît Habert, tout d'abord comme simple doctorante puis comme Attachée Temporaire d'Enseignement et de Recherche (ATER), m'a cependant permis très tôt de suivre l'émergence de la *Base de Français Médiéval* (*BFM*), et de contribuer à son développement. La *BFM* s'est constituée grâce à une politique de saisie et de scannage systématique, et grâce à des échanges continus. Ainsi, le premier texte saisi, *La Queste del Saint Graal*, a servi de « monnaie d'échange » avec une dizaine d'enseignants-chercheurs dans divers pays. D'autres textes ont été donnés par amitié, ou en retour de services rendus à des chercheurs.

Christiane Marchello-Nizia avait lancé le projet dès 1989 : constituer une base de textes intégraux d'ancien français, qui soit complémentaire de celle du *DMF* (*Dictionnaire du moyen français*) créée à l'initiative de Robert Martin à Nancy. Le but n'était pas de fournir à la communauté les textes eux-mêmes : sans doute à l'époque était-on déjà sensible à la question des droits d'édition, mais, surtout, fournir les textes tels quels avait peu d'intérêt à une époque où les outils pour les exploiter étaient encore peu développés, et en tout cas ignorés de la large majorité des linguistes médiévistes[62]. Ce sont donc des 'concordances' des textes que nous proposions aux étudiants et aux chercheurs, cela grâce au logiciel *Analyser*, développé par Pascal Bonnefois, qui permettait de générer des concordances et de les interroger. Nous réalisions les concordanciers des textes intégraux, ainsi que des concordances ciblées, souvent plus complexes (coocurrences de termes par exemple), que nous effectuions à la demande ; dans tous les cas ces concordances étaient ensuite consultables à l'aide d'un traitement de textes. Les plus débrouillards venaient jouer eux-mêmes avec *Analyser* (j'avais réalisé un petit manuel d'utilisation avec Pascal Bonnefois), formulant leurs requêtes et générant leurs propres concordances. Aujourd'hui, près de

[62] *A contrario*, l'équipe de Monique Lemieux au centre ATO à l'UQAM était déjà bien familiarisée avec le traitement informatique des textes, grâce au logiel *SATO* développé dans l'équipe (voir note précédente).

15 ans plus tard, les concordanciers nous semblent quelque peu désuets, pour ne pas dire archaïques, tant ont été considérables et rapides les développements dans les technologies et les outils permettant de manipuler les textes. Notons cependant que la fonction « concordances » n'a pas disparu des outils plus sophistiqués, mais elle voisine désormais avec de nombreuses autres fonctions.

Il n'en demeure pas moins qu'*Analyser* a constitué un précieux outil pour de nombreux linguistes, et qu'il a permis, par la collecte automatique et rapide de très nombreuses occurrences ou cooccurences de formes qui n'auraient pu être relevées manuellement, de mener des études qui sinon n'auraient pu l'être[63].

Le développement par Serge Heiden du logiciel *Weblex*[64] a ouvert une nouvelle ère dans les recherches. Accessible en ligne sur Internet, *Weblex*, dès ses débuts, mettait en œuvre l'ensemble de la méthodologie textométrique (méthode d'analyse quantitative s'appuyant sur des modélisations probabilistes et statistiques[65]), permettant la génération rapide de concordances, le calcul de cooccurrences et des analyses statistiques dans des environnements hypertextuels.

Parallèlement, la *BFM* continuait de se développer, tant sur la plan quantitatif que qualitatif, par la diversification des textes la composant.

L'aventure s'est poursuivie pour moi de 1998 à 2001 en tant que chercheure dans le laboratoire *Analyses de corpus* à Saint-Cloud, qui résultait de la fusion de l'équipe ELI et du laboratoire de textométrie créé par Maurice Tournier. Je ne m'occupais plus de la gestion de la *BFM*, mais continuais à participer activement aux projets lui étant liés, et à collaborer avec Serge Heiden, dont je partageais le bureau. A la suite de la délocalisation de l'ENS de Fontenay-Saint-Cloud à Lyon, et pour des raisons familiales, j'ai quitté le laboratoire en janvier 2002, pour intégrer le Lattice. Les liens

[63] Plusieurs articles et ou chapitres de C. Marchello-Nizia sur les démonstratifs ont été rendu possibles par la collecte massive d'occurrences à l'aide d'*Analyser*, en particulier les chapitres 4 et 5 de *l'Evolution du français* (1995). D'autres travaux de C. Marchello-Nizia se sont pareillement appuyés sur *Analyser* : la recherche sur l'accord des participes passés, pour le chapitre sur « Chevalerie et courtoisie », *Histoire des jeunes* (1994), G. Levi et J.-C. Schmitt (eds), Paris, Le Seuil, p. 147-197 ; puis sur les verbes supports : « Les verbes supports en diachronie : le cas du français », *Langages*, n° 121, p. 91-98.

[64] <http://weblex.ens-lsh.fr/wlx/>

[65] La méthodologie textométrique est issue de la méthodologie lexicométrique, mais elle vise à quantifier
tout niveau linguistique textuel, et pas seulement le seul niveau lexical. Cela est désormais permis d'une part par l'encodage XML-TEI quand il est disponible (codage manuel des unités) et d'autre part par le taggage automatique quand il est disponible (codage automatique des unités ou de leurs qualités). Pour une présentation détaillée de la méthodologie textométrique, voir le site du projet 'Textométrie' dirigé par Serge Heiden : <http://textometrie.ens-lyon.fr/>

de travail et d'amitié ne se sont pas défaits, et les projets communs autour de la *BFM* n'ont jamais cessé. Ils sont actuellement particulièrement actifs, comme on le verra plus loin.

J'ai eu la chance de me trouver au bon endroit au bon moment (ce qui ne signifie pas qu'il ne se passait rien ailleurs !), c'est-à-dire de participer activement à l'élaboration de ce qui est devenu depuis la plus grande base de textes intégraux d'ancien français (3 millions de mots), et d'être familiarisée assez tôt avec les textes numérisés ainsi qu'avec leur exploitation à l'aide de divers logiciels (*Analyser*, *SATO*, *Weblex*). Mais, comme je l'ai dit, certaines de mes recherches ne bénéficiaient que peu de la numérisation des textes et des développements technologiques. Il s'est néanmoins produit un changement notable dans mes pratiques : à un moment donné, j'ai pris l'habitude de lire à l'écran, et non plus sur papier, les textes sur lesquels je travaillais, qui étaient numérisés et auxquels je pouvais avoir accès, principalement parce que cela me permettait d'opérer des « copier-coller » des énoncés qui m'intéressaient (au lieu de devoir les saisir manuellement). Il reste que, à l'écran désormais, je continuais à parcourir les textes dans leur intégralité, en quête d'objets linguistiques qui ne pouvaient être captés automatiquement, mais aussi parce que ma formation universitaire (préparation de l'agrégation avec étude exhaustive d'une œuvre) m'y avait habituée. Cela restait possible parce que je travaillais à l'époque sur un nombre restreint de textes, que j'avais donc le temps de lire dans leur entier.

C'est un peu différent aujourd'hui. En 15 ans ma façon de travailler a changé : les corpus que je constitue pour mes recherches contiennent davantage de textes, non seulement parce que les textes numérisés sont plus nombreux mais aussi parce que j'ai travaillé ces dernières années sur des expressions facilement repérables (marqueurs de topicalisation et marqueurs discursifs : *quant à, à propos de, à ce propos…*), qu'il est possible de collecter automatiquement avec des outils de requêtes assez simples[66]. C'est parce que je n'avais pas à parcourir les textes que j'ai pu constituer de larges

[66] Il s'agit de constructions qui m'intéressent pour des raisons que j'exposerai plus loin, mais il est vrai aussi que la perspective de pouvoir collecter facilement un grand nombre d'occurrences a conforté l'envie de travailler sur ces expressions : on voit que l'interpénétration croissante entre sciences et techniques, dans notre discipline, peut agir très en amont, dans le choix même des constructions à étudier. Je reviendrai sur cette négociation récurrente entre « envie » scientifique et possibilités techniques.

corpus, mais c'est aussi parce que le corpus constitué comprenait de nombreux textes qu'il ne m'était plus permis de les lire, pour des raisons matérielles de temps.

Il reste que j'appartiens à une génération qui a toujours connu les textes numérisés et les technologies informatiques permettant de les exploiter : même si le développement de ces dernières modifie régulièrement ma manière de travailler et m'ouvre de nouvelles perspectives, il n'y a pas eu, dans mon parcours, de véritable rupture méthodologique ou intellectuelle.

Celle-ci s'est en revanche produite pour la génération qui m'a précédée. A la fois philologues et linguistes de formation, ceux qui ont débuté leur travaux dans les années 1960-1970 ont d'abord travaillé exclusivement le crayon à la main, parcourant systématiquement les œuvres en quête des objets linguistiques sur lesquels ils travaillaient, et en effectuant eux-mêmes les comptages.

J'ai sollicité certains d'entre eux, avec lesquels je travaille, pour la plupart, depuis de nombreuses années, pour qu'ils répondent à quelques questions et témoignent de la façon dont ils ont vécu le « passage au numérique ». Christiane Marchello-Nizia, Bernard Combettes, David Trotter et Fernande Dupuis ont eu la gentillesse de se prêter à cette petite interview, et de dire comment ils travaillaient « avant », ce qu'avait modifié pour eux l'apparition des textes numérisés et des technologies associées, tant dans leur méthodologie que dans leur rapport à la langue et aux textes. J'ai essayé de faire au mieux la synthèse de leurs réponses, qui se révèlent, sans grande surprise, largement convergentes.

Il ressort tout d'abord que, à certains égards, leurs méthodes n'ont pas été modifiées : ainsi le choix des textes composant un corpus était et reste conditionné par la construction ou le phénomène à étudier (on sait parfois que certains textes seront plus pertinents, révélateurs que d'autres, ce qui présuppose toutefois une relative connaissance préalable des textes…) ainsi que par la qualité de l'édition des textes. On continue à faire des comptages, à rechercher les premières occurrences, à élaborer des raisonnements, avant de retourner aux textes pour vérifier les hypothèses… La différence tient à ce que tout cela s'effectuait auparavant avec un crayon (et une gomme)[67] et des cahiers remplis de chiffres, et, surtout, à ce que les corpus étaient forcément limités par les possibilités humaines individuelles de comptage.

[67] Ce qui reste le cas pour les textes, encore nombreux, non-numérisés.

C'est un constat unanime qui ressort des témoignages : les textes numérisés ont permis un gain en temps et en souplesse. Gain de temps considérable dans les comptages (que des outils assez simples ont permis d'effectuer dès les premiers textes numérisés), et, du coup, fin de l'obsession liée à ces mêmes comptages, que l'on redoutait de devoir recommencer, soit que l'on ait commis une erreur, soit qu'une décision prise se révèle finalement fausse et qu'il faille tout reprendre depuis le début. Ce genre de mésaventures n'est jamais agréable, mais les outils informatiques rendent la tâche moins lourde. Cette relative facilité à (re)faire les calculs permet aussi de faire bouger le corpus (ajout/retrait de textes) : auparavant, étant donné que tous les calculs étaient faits à la main, on choisissait un corpus et on évitait de le modifier.

Le gain de temps permet aussi un enchaînement beaucoup plus rapide des raisonnements, et donc, potentiellement, un accroissement de la « productivité scientifique ». On peut aussi, grâce à l'augmentation du nombre de textes traités et grâce aux outils qui les exploitent, opérer des comparaisons assez rapides entre textes : dès lors devient possible la prise en compte plus systématique de la variation dialectale, dans ses différentes instanciations, de même que l'accroissement du nombre de textes étudiés permet l'établissement de diachronies à la fois plus larges et plus fines, les jalons textuels pouvant être multipliés.

Les avantages sont donc considérables… mais ils ont une contrepartie négative : on lit moins les textes, voire on ne les lit plus, à la fois parce que cela n'est plus indispensable et que ce n'est plus possible matériellement. Du coup, si l'activité est plus exploratrice au sens où elle permet de repérer plus facilement une première occurrence, un contexte inédit, elle est en revanche moins « curieuse », du fait qu'elle est davantage « pré-formatée » : en ne lisant plus au fil du texte, on ne repère plus, ou on repère moins, de phénomènes nouveaux, de constructions inédites. Comme le dit joliment Christiane Marchello-Nizia, on a une attitude moins *généreuse* envers les textes : on sait d'avance (ou l'on croit savoir…) ce qu'on attend d'eux, alors qu'auparavant c'était différent ; tout texte était une découverte exploratoire de sa langue aussi. Il est cependant vrai que, d'une part bon nombre des textes aujourd'hui explorés selon de nouvelles modalités ont déjà livré certains de leurs secrets dans le cadre d'études menées « à l'ancienne », et que d'autre part, et c'est en partie lié, il y a différentes façons d'être surpris par un texte : cela ne passe pas/plus nécessairement par la découverte, par exemple, d'une construction totalement inédite, mais davantage,

me semble-t-il, par des fréquences surprenantes, des relations ou corrélations inattendues entre constructions ou phénomènes…, que précisément les nouveaux modes d'étude des textes permettent de faire apparaître.

Il reste que ces linguistes-philologues ont gardé l'habitude d'aller voir « de près » dans les textes, et dans les manuscrits, les occurrences qu'ils collectent, et que leur formation en philologie les a en outre rendus familiers de nombreux textes. Ils sont en outre convaincus que la version numérique d'un texte ne remplace pas l'édition papier, dont l'apparat critique (introduction, variantes, notes,…) le plus souvent non numérisé, est aussi important que le texte lui-même. Que les éditeurs se rassurent néanmoins : nombreux sont les linguistes plus jeunes qui accordent une même importance à ces précieuses informations, qui témoignent en outre de la distance entre le texte publié et le/les manuscrit(s) d'origine.

La familiarité avec les textes me place aussi à un tournant : ma connaissance des textes est bien maigre comparée à la leur, et il est possible que la refonte partielle des épreuves du CAPES de Lettres entraîne un recul de l'étude des textes anciens, et par conséquent un appauvrissement de la connaissance que les futurs étudiants auront des textes[68]. Mais ne jetons pas la pierre à cette réforme, qui n'est, selon moi, qu'un signe (un symptôme ?) d'une certaine perte de l'intérêt porté à la langue ancienne[69] : la double casquette de philologue-linguiste risque de trouver moins de têtes à coiffer[70].

[68] Il est à vrai dire trop tôt pour mesurer les conséquences de cette réforme. Après bien des péripéties – il a été question à un moment donné d'une suppression de l'épreuve de langue ancienne – la langue ancienne est maintenue au concours, mais elle ne donne plus lieu à une épreuve en soi : son étude constitue une sous-partie d'une épreuve générale de langue, et la dimension diachronique a été renforcée. Au regard de la mission professionnelle qui attend les candidats, la mise en avant de l'approche évolutive de la langue me semble une bonne chose (combien d'étudiants restaient perplexes devant cette épreuve d'ancien français trop déconnectée du français moderne). Il se peut en revanche que le niveau de la culture philologique des étudiants baisse. Il faut cependant bien admettre que l'épreuve dans son état précédent (de même que l'épreuve de l'agrégation qui, elle, reste inchangée) ne garantissait pas un vivier de philologues : les étudiants faisant « l'impasse » sur cette épreuve « technique » ont toujours existé.

[69] Même si la production d'éditions de textes se poursuit, l'université Paris-4 Sorbonne semblant fournir un important vivier de philologues-éditeurs, ainsi que les universités étrangères.

[70] Mon impression – mais je ne dispose pas de données chiffrées – est que le nombre de thèses portant sur la langue ancienne, dans une perspective synchronique ou diachronique, est en baisse. Plusieurs raisons peuvent être invoquées : un relatif désintérêt des étudiants, un recul des cours de langue ancienne (les deux étant partiellement corrélés), ainsi que le départ à la retraite de professeurs susceptibles de diriger des doctorants. Ces dernières années ont en effet quitté leur fonction de directeur/directrice de recherche d'éminents spécialistes de langue ancienne, qui ont formé de nombreux étudiants, devenus enseignants et/ou chercheurs : Michèle Perret, Christiane Marchello-Nizia et Bernard Combettes, en particulier. Or il existe actuellement un « vide générationnel » : rares sont les spécialistes de langue ancienne habilités à encadrer des doctorants, comme en témoigne depuis plusieurs années la difficulté pour les universités à trouver des candidats aux postes de professeurs. Le

Le vivier tend à se tarir, mais l'on observe conjointement un intérêt croissant, de la part de non spécialistes, pour ce qui touche à la diachronie, et donc souvent à la langue ancienne. J'y reviendrai plus bas (3.2.2.)

L'émergence des bases de textes numérisés et l'utilisation fréquente qui en est faite, à savoir l'extraction automatique de données langagières, a eu un effet majeur sur la « lecture » des textes. On lit moins les textes, ou, plutôt, on tend à les lire différemment : d'une lecture continue et linéaire, on est passé à une « lecture » par classes d'occurrences (quelles qu'elles soient) extraites plus ou moins automatiquement, dans un contexte plus ou moins large ; autrement dit, à une appréhension « syntagmatique » des textes se substitue désormais, souvent, une approche « paradigmatique ». Cette lecture à la fois verticale et transversale résulte du mode d'exploration et de filtrage qui consiste à générer les concordances d'une construction. Celles-ci peuvent être classées selon différents paramètres : respect de l'ordre d'apparition des constructions-pivots dans le texte, classement alphabétique en fonction de la construction pivot (à une même requête peuvent répondre des constructions différentes en surface), sur le contexte droit ou gauche… Dans tous les cas, la linéarité physique du texte est en partie brisée, mais c'est à un autre mode de représentation du texte que l'on accède, par classes. On peut mesurer l'homogénéité de celles-ci, leur variation interne… bien plus facilement que si les différences occurrences constitutives de ces classes étaient appréhendées au fil du texte. A cet égard, j'ai eu l'occasion à plusieurs reprises de vérifier l'étiquetage morpho-syntaxique de textes d'ancien et de moyen français. Je n'ai jamais procédé à une vérification linéaire des étiquettes : d'une part parce que jongler avec la bonne cinquantaine d'étiquettes composant le jeu me semblait un exercice assez vite épuisant, d'autre part parce que procéder par étiquette, et même par étiquette et par forme, me permettait de repérer bien plus facilement les erreurs (parfois de simples incohérences) ainsi que des régularités ou au contraire des variations, rendues saillantes par cette mise en paradigme.

Appréhender un texte par classes d'occurrences permet de découvrir des caractéristiques qu'une lecture linéaire n'aurait pas permis de mettre au jour (ou bien

désir de pouvoir encadrer des étudiants a d'ailleurs été le moteur principal à la réalisation du présent mémoire.

difficilement) ; c'est une structuration inédite du texte que l'on construit ainsi, non perceptible à l'œil nu, sans outils.

D'une manière générale, les outils rendent possible une massification des données traitées et ils permettent (et c'est en partie lié) de percevoir différemment les constructions langagières et les textes. Je ne développerai pas ici la question des différences de perception et d'appréhension des textes selon que l'on travaille à l'œil nu, ou bien de manière outillée (la complexité de l'outillage pouvant être très variable) : c'est une question qui sera développée dans le cadre de la recherche inédite présentée conjointement à ce mémoire de synthèse.

3.2.2. D'une approche individuelle à un fonctionnement collectif.

Le rapport individuel au texte a changé, et le fonctionnement de la communauté s'en est trouvé modifié.

L'émergence des bases de textes numérisés a bouleversé à différents égards notre appréhension des textes, et elle a aussi ouvert des perspectives à des linguistes non spécialistes de langue ancienne, qui ne se seraient pas aventurés dans la lecture de textes médiévaux. En effet, ce nouveau mode d'accès aux données, la possibilité de collecter facilement et massivement des occurrences, tout en sollicitant des spécialistes de langue ancienne (qui jouent un rôle d' « informateurs », au sens où ils connaissent la langue, et donc de garants), leur permet d'explorer de nouvelles terres linguistiques. Cet intérêt nouveau est corrélé, je le répète, à une attention croissante portée aux faits d'évolution de la langue : à ma connaissance, les linguistes modernistes qui se lancent dans l'aventure médiévale le font généralement parce qu'ils se penchent sur les états antérieurs d'un phénomène de français moderne dont ils souhaitent appréhender les changements passés, pour mieux le comprendre dans son fonctionnement actuel. L'une des conséquences majeures de cette situation est un décloisonnement des communautés : même si « modernistes » et « médiévistes » ont toujours communiqué, il me semble que se développe un travail en équipe qui restait beaucoup plus marginal il y a encore quelques années. Il n'est pas improbable que la réflexion autour des corpus, de leur constitution, de leur exploitation, a contribué à rapprocher les communautés.

On ne peut que se réjouir de l'intérêt croissant pour l'approche diachronique[71], mais se pose néanmoins la question de savoir jusqu'à quel point on peut travailler sur une langue que l'on connaît peu, ou mal, que l'on ne s'est pas appropriée. Certes, le français médiéval n'est pas sans affinités avec le français moderne, comparé à d'autres langues radicalement différentes, mais il faut se méfier de cette apparente familiarité : d'une part les « faux-amis » ne sont pas rares (qu'ils soient morphologiques, syntaxiques, sémantiques…), d'autre part notre regard de locuteur moderne tend vite à l'anachronisme (voir 1.2.2. et 1.4.) : par exemple la variation graphique est inhérente à la langue et l'adéquation entre unités graphiques et unités linguistiques bien moins systématique qu'en français moderne (voir 4.1.2.1. infra) ; il s'agit d'un état de faits sur lequel il n'y a pas lieu de porter un jugement ; gardons aussi à l'esprit que la ponctuation, qui conditionne largement notre lecture des textes, est l'œuvre, dans une très large mesure, des éditeurs modernes.

La coopération accrue entre linguistes modernistes et médiévistes s'inscrit dans un mouvement plus général : de plus en plus les projets sont collectifs. Cette évolution me semble avoir deux causes majeures, en partie liées. D'une part, l'émergence des bases textuelles de taille importante suppose une gestion collective (un chercheur seul ne peut s'en charger), de même que la massification des données accessibles et exploitables pour une étude donnée incite souvent à s'associer pour les traiter. D'autre part, et cela dépasse largement le microcosme des linguistes spécialistes de langue ancienne, la conception de la recherche s'est notablement modifiée depuis quelques années : difficile aujourd'hui de mener un projet seul, *a fortiori* sur le long cours (cela reste possible, mais comme une activité quelque peu marginale). Toute recherche, ou presque, pour bénéficier d'une reconnaissance et donc d'un financement, doit s'inscrire dans un projet collectif. Je ne ferai que quelques remarques à ce propos. Que les projets collectifs soient encouragés me semble une bonne chose (souvent plusieurs têtes pensantes valent mieux qu'une) ; qu'un relatif « contrôle » soit exercé sur les thèmes de recherche (par le biais des appels d'offre, par le financement ou non de projets soumis) me paraît justifié, même si l'on peut s'interroger sur la légitimité

[71] Et la vitalité des colloques qui lui sont dédiés en témoigne, qu'il s'agisse, depuis 2000 des éditions bi-annuelles du colloque Diachro que je co-organise, ou de la session Diachronie, que je coordonne, des deux éditions (2008 et 2010) du Congrès Mondial de Linguistique Française http://www.ilf.cnrs.fr/spip.php?rubrique4 et http://www.ilf.cnrs.fr/spip.php?rubrique54.

des décideurs et sur les critères qui président à la sélection[72]. Il reste que ces modalités de fonctionnement tendent à exclure des projets sur le long cours (or certains projets auraient besoin de plus de 2, 3, ou 4 ans pour pleinement se réaliser), ainsi que des projets exigeant un financement « moyen ». Tout projet ne nécessite pas le recrutement de personnes, et n'a donc pas besoin de centaines de milliers d'euros pour se concrétiser : le moule de l'ANR n'est donc pas adapté à tous les types de projet. Or il existe actuellement une sorte de « no man's land » en termes de financement, entre des projets ne requérant que quelques milliers d'euros et ceux dépassant la centaine de milliers d'euros[73].

Ainsi, l'actuel projet de *Grande Grammaire Historique du Français*, que je co-dirige avec C. Marchello-Nizia, B. Combettes et T. Scheer ne nécessite pas de recrutement à plein temps (les chapitres sont rédigés, bénévolement, par des linguistes spécialistes des différentes questions, qui sont en poste), mais simplement quelques dizaines d'heures de vacation (en particulier pour assurer la veille bibliographique), et le financement de réunions. Actuellement ce n'est pas un projet qui peut donner lieu à une demande de financement par l'ANR. C'est en outre un projet qui prendra plusieurs années (puisque, précisément, les rédacteurs n'y travaillent pas dessus à temps plein). Typiquement, en dépit du besoin auquel il répond – fournir, un siècle après le début de la parution l'*Histoire de la Langue Française* de F. Brunot, un vaste ouvrage de synthèse qui prenne en compte les nombreuses avancées dans le domaine de l'évolution de la langue – et de l'intérêt qu'il suscite, ce projet peine à trouver une structure qui le soutienne de manière pérenne.

3.2.3. Allier la lecture outillée à la connaissance des textes

A bien des égards, la méthodologie des études en langue ancienne et en diachronie a évolué, en grande partie en raison de l'émergence des bases de textes numérisés, qui ont modifié le rapport aux textes, aux données langagières, et à la langue.

---

[72] Sans pour autant se placer dans une vison purement utilitariste de la recherche (qui d'emblée discréditerait bon nombre de recherches en linguistique et en sciences humaines d'une manière générale), il n'est pas scandaleux de s'interroger sur le caractère plus ou moins nécessaire et/ou intéressant des thèmes de recherche. L'institution ne peut pas et ne doit pas tout financer, mais la question cruciale est de savoir qui, et selon quels critères, est à même de décider ce qui mérite ou non d'être soutenu : cela suppose une posture à la fois désintéressée, neutre, mais néanmoins impliquée, et cela exige aussi d'avoir une vision large et à long terme. Récurrente question de savoir qui, des scientifiques ou des institutionnels, est le mieux placé pour évaluer et décider…

[73] Effet pervers de cela : on en arrive parfois à gonfler le budget d'un projet pour le rendre éligible à un financement par l'ANR.

On a quitté un mode d'accumulation des savoirs individuels (chacun ne traitait qu'une petite part des textes et/ou des données, mais en avait une maîtrise et une connaissance « parfaites »), pour un mode ou chacun peut disposer de tout ce qui est accessible, sorte de mémoire commune et collective, mais que nous ne nous approprions plus de la même façon : on ne connaît pas véritablement les composantes de cette mémoire. On interroge des données brutes à l'aide d'outils de type concordancier, et des données enrichies à l'aide d'outils plus sophistiqués, mais nous sommes en partie « coupés » de ces données, des textes dont elles proviennent.

La massification des données traitées et les modes de représentation inédits que les outils permettent d'en donner (voir 3.2.1.) s'accompagnent donc d'une fréquente déconnexion de leur réalité textuelle. Ils n'en constituent pas moins un précieux apport, auquel il n'est pas question de renoncer, mais ils ne devraient pas (au sens déontologique du terme) s'accompagner d'une perte de contact avec lesdites données, sous peine de voir celles-ci se dématérialiser : une construction a un contexte linguistique immédiat, mais elle a aussi un contexte textuel plus large, et ignorer celui-ci peut faire passer à côté d'une dimension explicative de cette construction. Pour cette raison mais aussi parce que, plus généralement, travailler sur une langue que l'on ne fréquente pas un tant soit peu n'a pas grand sens selon moi, il faut, d'une part continuer à lire les textes, au moins en partie[74], et d'autre part aller visiter, autant que possible, les contextes des occurrences que l'on collecte automatiquement. Faire émerger des caractéristiques n'est véritablement pertinent que si l'on a une bonne connaissance de base des textes et de la langue : ce n'est que grâce à elle que l'on peut mesurer le caractère inédit, étonnant… de ce que l'on met au jour.

Il me semble que le défi majeur des années à venir consiste à trouver un équilibre entre un usage raisonnable et raisonné des ressources informatiques, et une fréquentation étroite des textes. Cela semble tomber sous le sens, mais l'on sait que l'obtention concrète de cet équilibre n'est pas une tâche aisée. Il n'y a pas de solution

[74] « en partie » : la recommandation est certes bien vague. Mais il n'est pas possible de faire une proposition précise dans l'absolu : cela dépend du nombre de textes explorés, de leur longueur, et du temps dont on dispose. Que dire sinon « autant qu'il est raisonnablement possible ». Idéalement la lecture des textes ne devrait pas se limiter au cadre temporel d'une étude, l'imprégnation de leur langue devrait se faire régulièrement. C'est de plus en plus difficile, tant la masse des publications à lire, sur le plan théorique, est conséquente, et sa croissance exponentielle, ou nous semble telle, du fait d'un accès facilité par la mise en ligne accrue des documents. A cet égard, être aujourd'hui à la fois un philologue compétent et un bon linguiste devient un vrai défi : il faut lire non seulement ce qui paraît, sans cesse, mais aussi ce qui a paru depuis une cinquantaine d'années. En effet, dans notre discipline, comme dans la plupart des sciences humaines et sociales, les publications récentes n'invalident pas, pour beaucoup, celles qui sont plus anciennes. De ce point de vue, notre tâche est plus lourde que celle de nos aînés.

toute faite, chaque étude est différente, il n'existe pas de règle absolue. Il faut néanmoins se doter d'une connaissance minimale des textes, qui, à l'instar de notre compétence de locuteur moderne lorsqu'il s'agit d'évaluer des énoncés de français moderne, nous permette, en quelque sorte, de tirer la sonnette d'alarme, devant des constructions ou des phénomènes inattendus, ou bien face à des fréquences suspectes (il est toujours possible de mal formuler une requête).
C'est la démarche que j'ai essayé d'adopter pour l'étude de l'évolution de la syntaxe du pronom sujet, présentée conjointement à ce mémoire. En préalable je reviendrai sur la question des corpus et des outils, en particulier sur certains aspects que je laisse volontairement de côté ici : la constitution du corpus, l'extraction et le traitement des données, les apports spécifiques des données numérisées outillées et les rapports entre outils informatiques et œil humain.

Parce que les textes numérisés étaient encore insuffisants, ou qu'ils ne bénéficiaient pas de l'enrichissement dont j'aurais eu besoin, parce que des contraintes matérielles de temps m'ont parfois conduite à réviser à la baisse mes ambitions initiales en matière de constitution d'un corpus de travail : pour ces différentes raisons je n'ai pas toujours mis en œuvre la « linguistique de corpus » que j'aurais aimé appliquer, mais faute aussi, parfois, de compétences méthodologiques (dans le domaine des calculs statistiques par exemple).
La recherche inédite qui est présentée avec ce mémoire de synthèse essaie de remédier à ces lacunes, et de mettre en œuvre une linguistique aussi raisonnée et cohérente que possible.

Dans la dernière partie de ce mémoire j'aimerais évoquer, brièvement, certains des travaux que j'ai réalisés ou qui sont en cours, en les envisageant selon un axe spécifique, celui de la classification.

Chapitre 4. Classifier

Etudier une langue qui évolue, c'est observer des classifications qui changent ou que le linguiste fait bouger. C'est selon ce fil directeur que j'aborderai deux aspects de mon activité : l'enrichissement des corpus et les études linguistiques. Même si les deux sont étroitement liés à bien des égards, ils diffèrent cependant sur un certain nombre de points. Tout d'abord, l'enrichissement des corpus s'inscrit dans un double projet collectif alors que j'ai mené seule la plupart des études linguistiques ; il vise par ailleurs à fournir à la communauté des ressources et des outils de travail directement exploitables par le plus grand nombre, alors que les études que j'ai menées (aussi intéressantes et stimulantes soient-elles…) ne constituent pas un outil aussi « prêt à l'emploi », elles supposent une réappropriation plus importante par leurs lecteurs. Enfin, enrichir un corpus suppose, dans les projets auxquels je contribue, d'appréhender la langue comme un système global, puisque c'est l'intégralité des formes des textes (ou des extraits de textes) qu'il faut étiqueter ou annoter : il faut donc concevoir un modèle cohérent qui vaille pour chacune des formes, des constructions, des relations… Les travaux linguistiques que j'ai réalisés étaient en revanche centrés sur un type de constructions (marqueurs discursifs, constructions détachées, relation sujet-verbe…), dont on verra en 4.2. qu'elles sont moins déconnectées qu'il n'y paraît les unes des autres. Même si l'on ne peut jamais isoler une construction du reste du système de la langue, la perspective est cependant différente.

4.1. L'enrichissement des corpus

4.1.1. Fournir un outil de travail à la communauté

Enrichir un corpus présente un double enjeu. Il s'agit en premier lieu de fournir un outil de travail utile à la communauté. Il est actuellement certaines études qui ne peuvent être menées à grande échelle parce que l'on ne dispose pas dans les corpus existants des informations permettant de collecter automatiquement les constructions : il faut les chercher à l'œil nu, ce qui limite forcément le nombre de textes qui peuvent être traités. Comme je l'ai déjà mentionné, je me suis heurtée à ce problème lorsque j'ai travaillé sur la position du sujet nominal. Plus récemment (Prévost 2003b et

2009), j'y ai à nouveau été confrontée pour les constructions détachées, non repérables si l'information n'a pas été encodée spécifiquement (*a priori* dans le cadre d'une annotation syntaxique) : impossible en effet, même avec un corpus étiqueté morpho-syntaxiquement, de distinguer automatiquement les différentes fonctions des syntagmes nominaux initiaux des énoncés suivants, et donc de repérer les constructions détachées (21a) :

(21a) **Cest chevalier**, je ne l'aime pas. (*Erec et Enide*, milieu 12ème)
CONSTRUCTION DETACHEE

(21b) **Le lendemain** commencèrent les escarmouches jusques aux portes de Paris (Commynes, *Mémoires*, fin 15ème)
CIRCONSTANT de TEMPS

(21c) **Iteus paroles** distrent li frere de Lancelot (*La mort Artu*, début 13ème)
OBJET

(21d) **Vostre mesage** fesime a Charlun (*Chanson de Roland*, fin 11ème)
OBJET

(21e) **Li rois** apele un escuier *(Tristan* de Beroul, fin 12ème)
SUJET

Le fait que l'ordre des mots soit relativement souple et que le sujet puisse ne pas être exprimé rend en effet inopérante toute stratégie s'appuyant sur la position respective des différentes catégories morpho-syntaxiques. C'est donc « manuellement » que j'ai collecté les constructions détachées dans différents textes, profitant en outre des précieux relevés qu'avait effectués Christiane Marchello-Nizia pour les deux études qu'elle avait réalisées quelques années plus tôt (Marchello-Nizia 1998a et 1998b). Dans la mesure où il s'agit de constructions qui restent globalement rares (au mieux quelques dizaines par texte), il est nécessaire de pouvoir s'appuyer sur un large corpus pour mettre au jour l'ensemble de leurs caractéristiques, et l'évolution de celles-ci. La mise à disposition de la communauté des linguistes, fin 2011, d'un corpus annoté syntaxiquement où seront encodées les informations permettant d'extraire automatiquement les constructions détachées (projet *SRCMF*[75]) permettra de

[75] Projet ANR franco-allemand « Syntactic Reference Corpus of Medieval French » (SRCMF) que je codirige avec Achim Stein (université de Stuttgart) ; le projet est prévu pour 3 ans (2009-2011), et il a reçu, côté français, une dotation de 160 000 euros, qui prévoit l'emploi d'un post-doctorant ou d'un ingénieur de recherche pendant 3 ans. Le projet réunit douze personnes : Julie Glikman, Céline Guillot, Serge Heiden, Alexei Lavrentiev, Christiane Marchello-Nizia, Sophie Prévost, Tom Rainsford et Bernard Victorri à titre d'expert du côté français ; Béatrice Bischoff, Nicolas Mazziotta et Achim Stein,

reprendre et compléter les études déjà menées en systématisant l'examen des constructions détachées. C'est là un bel exemple d'une avancée couplée entre « technique » et « science », les envies scientifiques (en l'occurrence l'étude des constructions détachées) devant se plier aux possibilités techniques (ici l'impossibilité – bientôt levée – de les repérer automatiquement). Assurément les objectifs scientifiques évoluent, dans une plus ou moins large part, en même temps que les techniques.

Fournir un outil de travail « utile » à la communauté signifie qu'il faut – c'est en tout cas ma position – que l'enrichissement soit exploitable par un grand nombre de chercheurs, et qu'il soit donc aussi peu marqué que possible du point de vue des choix théoriques et terminologiques (les deux étant en partie liés). Même si les choix sont explicités et documentés dans un manuel, il me semble nécessaire que les utilisateurs puissent se familiariser assez vite avec le modèle, que sa compréhension ne dissuade pas d'utiliser le corpus. J'ai la conviction que là où la tradition peut être maintenue (que ce soit sur le plan théorique ou simplement terminologique) il faut la maintenir : elle constitue un socle commun pour la communauté.

Adopter une relative neutralité théorique est plus facile dans le cadre d'un étiquetage morpho-syntaxique que dans celui d'une annotation syntaxique : le consensus est plus grand pour identifier un nom que pour distinguer un complément d'un circonstant… Les deux projets auxquels je participe (étiquetage morphosyntaxique[76] et projet *SRCMF* d'annotation syntaxique) témoignent de cet écart : les choix opérés pour le jeu d'étiquettes morpho-syntaxiques correspondent dans une large mesure à une tradition consensuelle, tant dans la nomenclature que dans l'affectation des étiquettes aux formes ; le modèle élaboré pour l'annotation syntaxique, de type dépendanciel, est en revanche beaucoup plus marqué du point de vue théorique.

Il reste que, dans les deux cas, l'on a essayé de « prévoir » les études qui pourraient tirer profit d'un corpus enrichi, et fait en sorte que les informations encodées dans les

---

du côté allemand, ainsi que Fernande Dupuis (Québec).Voir présentation du projet < https://listes.cru.fr/wiki/srcmf/>.

[76] Il s'agit d'un projet autour de la BFM qui a débuté avant même la délocalisation de l'ENS et de la BFM à Lyon. Faute de moyens substantiels et de forces humaines, ce projet n'a pas avancé autant que nous l'aurions souhaité. Il connaît depuis plusieurs mois un nouvel essor et des avancées notables, en partie relancé par le projet ANR *Corptef* (Corpus représentatif des premiers textes français), dirigé par Céline Guillot, ainsi que par le projet d'annotation syntaxique *SRCMF*. Par commodité, je réfèrerai parfois à ce projet d'étiquetage morphosyntaxique par le nom de son jeu d'étiquettes CATTEX. Il réunit principalement Céline Guillot, Alexei Lavrentev, Christiane Marchello-Nizia, Serge Heiden et moi-même.

textes puissent répondre au mieux à ces attentes, tant dans leur conception que dans les modalités pour les extraire. En ce qui concerne le projet d'annotation syntaxique, l'enrichissement est réalisé à l'aide du logiciel *NotaBene*, outil d'annotation manuelle développé par Nicolas Mazziotta[77], mais nous utilisons le logiciel *TIGERSearch*[78] pour les requêtes, que ne permet pas d'effectuer *NotaBene*. Dès à présent, alors que le projet est à plus de deux tiers de son parcours, nous éprouvons la validité du modèle en opérant diverses requêtes « tests » à l'aide de *TIGERSearch* : sujets postverbaux, verbes en seconde position, relatives avec leur antécédent… Cela nous permet de vérifier si l'on peut effectivement formuler des requêtes pertinentes pour les objets en question, de mesurer le degré de complexité de ces requêtes, et, chemin faisant, de repérer les erreurs d'annotation (annotation qui fait par ailleurs l'objet d'une vérification manuelle).

Dans les deux cas, il s'agit de concevoir un enrichissement qui puisse faciliter le plus grand nombre de recherches possible.

4.1.2. L'évolution des classes

Enrichir un corpus, c'est ajouter de l'information à des textes. Dans les projets *CATTEX* et *SRCMF*, cette information, morpho-syntaxique et syntaxique, est encodée dans des étiquettes apposées sur des formes ou sur des relations. Il s'agit donc de classer, selon différents critères. C'est une activité complexe en soi, mais qui l'est davantage encore lorsque l'on travaille sur une langue ancienne : le danger de la posture anachronique guette toujours (voir la discussion à ce sujet en 1.3.). Nous avons fait le choix de partir des catégories[79] « modernes », mais en veillant à ce que leur projection sur des formes anciennes soit cohérente avec ce que nous savons de la valeur desdites formes à l'époque, et avec ce que nous indique leur fonctionnement contextuel. Ainsi, les catégories de déterminant et de pronom démonstratifs sont-elles pertinentes en ancien français, mais elles ne recouvrent pas exactement la même réalité linguistique qu'en français moderne : « celle » peut être pronom, mais aussi déterminant, comme c'est le cas dans l'exemple suivant :

[77] *NotaBene* est un logiciel libre d'annotation. Voir https://sourceforge.net/projects/notabene/

[78] *TIGERSearch* est un logiciel libre qui permet d'interroger des textes annotés linguistiquement. Plus spécifiquement il permet d'interroger des bases de données structurées en arbres (graphes orientés, connexes et sans cycles). <http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/oldindex.shtml>.

[79] J'utilise les termes de « classe » (de mots) et de « catégorie » de manière synonyme.

(22) **Cele nuit** fu Boort plus richement serviz et aiesiez qu'il ne vousist (*Queste del Saint Graal*)

Aux précautions que suppose la projection sur une langue ancienne de catégories pensées à travers le filtre moderne, s'ajoute la possible évolution des formes ou constructions dans les classes, et celle des classes elles-mêmes. A cet égard, les projets *SRCMF* et *CATTEX* diffèrent. Le premier inclut un corpus de textes d'ancien français, alors que le second prévoit l'étiquetage de textes d'ancien et de moyen français. Même si la priorité est donnée aux textes d'ancien français, la prise en compte de textes plus tardifs (trois d'entre eux sont déjà étiquetés[80]) a conduit à concevoir un jeu d'étiquettes qui tienne compte de l'évolution de la langue entre l'ancien et le moyen français. Ainsi nous avons prévu une étiquette spéciale pour le déterminant *ledit*, qui apparaît en moyen français et n'est pas assimilable au simple déterminant défini *le*. Nous avons par ailleurs conçu une dizaine d'étiquettes spécifiques pour les formes contractées qui existent en ancien français, et qui commencent à se réduire en moyen français. En effet, alors que le français moderne ne connaît que le type « préposition + déterminant » (*des*, *au, dudit, auquel*…), l'ancien français multiplie les combinaisons : « préposition + pronom personnel » : *Et tant fu desirranz* ***del*** [de+le] *veoir* ; « pronom personnel + pronom personnel » : *plus que* ***jel*** [je+le] *ne porroie* ; « adverbe + pronom personnel » : *car autrement* ***nel*** [ne+le] *demant je pas.*

Au final, il a été conçu un jeu de 54 étiquettes, structurées en deux champs : <catégorie> et <type>. Les *catégories* correspondent pour la plupart aux classiques parties du discours : verbe, nom, adjectif, pronom, déterminant, adverbe, préposition, conjonction, interjection, auxquelles s'ajoutent les catégories : ponctuation, mot étranger et abréviation. Les *types* correspondent quant à eux aux sous-classes des catégories : (nom) commun, (verbe) conjugué, (déterminant) possessif [81] …

Une fois conçu le jeu d'étiquettes (élaboration qui s'est faite sur le long cours et parallèlement à l'étiquetage des premiers textes[82]), se posent les problèmes liés à l'affectation des étiquettes aux formes. Je distinguerai pour cela deux cas principaux.

[80] *Quinze joyes de mariage*, *Jehan de Paris*, et le livre 1 des *Mémoires* de Commynes

[81] Pour une présentation détaillée du jeu d'étiquettes CATTEX, voir les versions provisoires du manuel de référence : <http://bfm.ens-lyon.fr/article.php3?id_article=323> et les principes d'étiquetage : < http://bfm.ens-lyon.fr/article.php3?id_article=173>

[82] Ce qui a donc impliqué des révisions de l'étiquetage au gré des changements dans le jeu d'étiquettes.

4.1.2.1. La polycatégorisation des formes

Une même forme peut changer de catégorie, changement qui peut s'opérer en diachronie et en synchronie, et se décliner selon différentes modalités. Ainsi les formes démonstratives en ancien français fonctionnent pour la plupart comme déterminant ET comme pronom démonstratif pendant plusieurs siècles ; mais comme, nous l'avons vu à propos de l'exemple (22), le contexte morphosyntaxique de la forme permet de l'interpréter avec certitude comme l'un ou l'autre, selon qu'elle est suivie ou non d'un nom[83]. La forme *plaisir* constitue un cas un peu différent dans la mesure où elle a subi une évolution durant la période de l'ancien français : verbe à l'origine, elle endosse très tôt le statut de nom commun ; les deux valeurs se chevauchent un temps, avant que le nom commun l'emporte et que la forme verbale soit refaite en *plaire*. Pendant la période de chevauchement des deux valeurs, le contexte morpho-syntaxique permet cependant de trancher entre les deux interprétations, comme en témoignent les deux occurrences empruntées à la *Vie de Saint Thomas Becket* de G. de Pont St-Maxence, éloignées seulement d'une vingtaine de vers :

(23a) Itels briefs enveieient al saint humme ultre mer,
Plusurs altres asez, dunt rien ne sai conter,
Pur mielz **plaisir** al rei e pur lur sens mustrer (*Saint Thomas Becket*, fin 12ème)

(23b) Car ço n'est pas aumosne, s'il en fait sun **plaisir** :
Quant l'a dunee a Deu , ne li puet retolir (*Saint Thomas Becket*, fin 12ème)

Il est par ailleurs un cas de figure non rare, et qui n'est pas propre à la langue ancienne : il s'agit de celui où une unité linguistique, dans certains contextes morpho-syntaxiques, endosse une valeur morphologique différente de sa valeur d'origine. Par exemple, dans :

(24) Et cele qui estoit la plus **dame** le menoit par la main et ploroit mout tendrement (*Queste del Saint Graal*),

le nom commun *dame* est assimilable à un adjectif qualificatif (présence de l'adverbe superlatif *plus*). C'est un changement analogue qui caractérise les substantivations (encore en français moderne : dans *les méchants ont toujours tort*, « méchants » est-il adjectif ou substantif ?)

[83] C'est pour cette raison que nous considérons CATTEX comme un jeu d'étiquettes morpho-syntaxiques, et non simplement morphologiques.

Nous avons choisi, dans des cas comme celui-ci, de restituer la double valeur de l'unité linguistique, nom commun et adjectif qualificatif. Dans la mesure où nous avons par ailleurs décidé de n'attribuer qu'une seule valeur de propriété aux unités, nous avons mis en place, parallèlement à la propriété morphologique CATTEX, une propriété morphosyntaxique, CATTEX bis, dont les valeurs, identiques à celle de CATTEX, sont attribuées selon des principes syntaxiques. Par défaut, une unité a la même valeur pour les deux propriétés, et donc deux étiquettes identiques, mais dans certains cas (*dame* en 24 par exemple) les deux valeurs de propriété diffèrent.

Les infinitifs sont particulièrement complexes à traiter. Ainsi dans :

(25) A cel **encontrer** furent toutes lances brisiees (*Queste del Saint Graal*),

la présence du déterminant démonstratif *cel* incite à analyser la forme, du point de vue morpho-syntaxique, comme un nom. A l'inverse, dans d'autres cas, la présence d'un complément favorise le maintien de l'interprétation comme infinitif. Mais il est deux situations dans lesquelles il est difficile de trancher : c'est lorsqu'il n'y a pas d'indice (ni déterminant, ni complément), comme en (26) :

(26) Ore dist li contes que a heure de **disner**, quant les tables... (*Merlin*, début. 13ème)[84],

ou au contraire lorsque les indices sont contradictoires (présence conjointe d'un déterminant et d'un complément), comme en (27)

(27) Ostes, ja fustes vous compaing,
Puisque che vient **au dire voir** (Bodel, *Jeu de Saint Nicolas*, fin 12ème)

Nous avons finalement décidé que, en l'absence d'indices, la forme est analysée morpho-syntaxiquement comme un infinitif (26), et qu'en présence d'indices contradictoires, elle est au contraire interprétée comme un nom (27). L'étiquette morphologique conserve la valeur 'infinitif' dans les deux cas.

Une difficulté d'un autre ordre surgit pour les formes qui ont changé de catégorie, sans pour autant changer de contexte morpho-syntaxique. C'est le cas pour *om* (et ses différentes graphies : *(l')on, (l') om,(l')en, hom.…*), originellement substantif, puis passé dans la catégorie de ce que nous appelons aujourd'hui les pronoms indéfinis. Dans ce cas, nous nous sommes en partie appuyés sur la graphie, analysant la forme

[84] Il existe 'heure de vespres'.

en ‘h’ (*hom*) – qui disparaît d’ailleurs assez vite – comme nom, et toutes les autres comme pronoms indéfinis. Pour les adverbes de renforcement de la négation, *pas, point et mie*, substantifs à l’origine, le fait qu’ils aient fonctionné très tôt comme particules de renforcement nous a conduit à les analyser systématiquement comme telles (sauf cas avéré de substantif : *Onques Tristran ne fist* ***un pas****,/ Qant il fu pris, qu'il dut estre ars, Tristan*, de Beroul).

Nous avons donc fait, pour les exemples de ce type, un choix assez radical, en décrétant des bornes chronologiques en partie arbitraires, décidant que, au sein de la période que couvre les textes du corpus étiqueté, les lexèmes ont telle valeur de catégorie : c’est effectivement nier la possible variation qu’ils peuvent encore connaître, c’est perdre la richesse de cette variation, et donc en partie les clés du changement (classer, c’est discrétiser, au sein du continuum aussi bien temporel que catégoriel).

C’est pour des raisons matérielles de temps, mais aussi parce que certains cas étaient indécidables, que nous avons opté pour cette solution, sachant qu’elle ne concerne que peu de formes, et qu’elle est de toute façon explicitée dans la documentation. Le but de ce corpus enrichi est de fournir à la communauté un outil de travail, que chacun est libre de compléter ou de réviser en fonction de ses besoins spécifiques et/ou de ses convictions linguistiques : aussi impartiale qu’elle se veuille, une démarche d’enrichissement linguistique des textes est toujours un parti-pris sur la langue, qui ne convient pas forcément à tous dans son intégralité.

Une autre difficulté liée à *on, om, hom*, est l’existence des formes *l’en/l’on*,... traces de son statut de substantif, qui se maintient bien après que la forme ne puisse plus être analysée comme telle (et qui s’est d’ailleurs maintenue jusqu’au français moderne : *l’on*). Il s’agit ici de savoir si l’on a deux formes : *l’* + *en*, ou bien une seule. Nous avons choisi la deuxième option, et analysé l’ensemble comme un pronom indéfini, choix qui se justifie du fait de l’invariabilité de la forme *l’* : on ne rencontre pas *un, les*,...

Ce dernier exemple me conduit à aborder le second type de problèmes liés à l’affectation des étiquettes aux formes : ceux qui touchent à la segmentation graphique et à la segmentation des unités linguistiques.

4.1.2.2. Unités graphiques et unités linguistiques

En général, chaque unité graphique est caractérisée par une étiquette qui lui est propre, l'unité graphique correspondant à une unité linguistique. Mais il est quelques cas, moins rares en français médiéval qu'en français moderne, où la segmentation graphique ne reflète pas la segmentation linguistique des unités. Nous avons donc été amenés à dissocier dans ces cas précis, les unités linguistiques des unités graphiques[85], et l'étiquetage opère alors sur des unités linguistiques qui ne correspondent pas aux formes graphiques apparentes. Ainsi, il arrive qu'à deux unités graphiques soit attribuée une seule étiquette (fusion). Cela ne se produit que dans quelques cas : par exemple, dans *en quel que lieu que il alast (Queste del Saint Graal)*, il a été décidé de fusionner 'quel' et 'que' en une seule unité (déterminant relatif), en raison non seulement des habitudes fluctuantes des scribes et des éditeurs, mais aussi – et surtout – de la difficulté à analyser isolément le premier 'que'.

A l'inverse, à un seul mot graphique peuvent – très rarement – être associées deux étiquettes. Nous avons ainsi choisi de segmenter systématiquement 'parce' en deux unités linguistiques : 'par' préposition, et 'ce' pronom démonstratif. En effet, le figement de ce genre de locution est progressif et il est difficile de dire à quel moment le groupe prépositionnel suivi d'une complétive en *que* devient une conjonction unie[86]. Il arrive par ailleurs, exceptionnellement, que la graphie *sil* résulte de ce que l'éditeur a graphié la construction *se + il* en une seule unité, sans apostrophe comme s'il s'agissait d'un cas d'enclise (*si + le*). Lorsque l'erreur d'interprétation de l'éditeur était manifeste, elle a été corrigée en segmentant la forme graphique *sil* en deux unités linguistiques : conjonction de subordination + pronom personnel. Nous avons procédé pareillement pour *qui* ou *quil* correspondant à des élisions non respectées (*qu'i* /*qu'il*) et donc à deux unités linguistiques, c'est-à-dire que nous avons segmenté la forme. Dans tous ces cas, la segmentation informatique effectuée en amont de l'étiquetage permet de dissocier les deux éléments pour l'étiquetage et pour les outils de requêtes tout en préservant au niveau de l'affichage l'aspect de l'édition source. Pour cela on ajoute un attribut rend= « aggl » à la valeur des deux étiquettes.

[85] Mais la forme graphique et la surface des textes n'ont pas été modifiées.

[86] Par ailleurs, face à la difficulté d'analyser – *a fortiori* automatiquement – certains *ce* (pronom démonstratif, impersonnel, présentatif…), nous avons choisi de l'étiqueter systématiquement comme 'pronom démonstratif', décision explicitée dans la documentation.

Il est à noter que, indépendamment des évolutions qui ont pu se produire dans la langue (tendance à la fusion de formes graphiques : *par mi > parmi*, *par ce > parce*, *ja mes > jamais...*), les habitudes graphiques des scribes étaient très fluctuantes pour certaines unités, de même que le sont les pratiques de transcription des éditeurs. Dans la mesure où notre étiquetage est associé à une édition référencée d'un texte, nous avons choisi de respecter les choix opérés par l'éditeur de ce texte (à l'exception des quelques cas mentionnés ci-dessus), même si certains sont de toute évidence discutables. Par conséquent, la segmentation de formes telles que *par mi/parmi* ; *ja mes / jamais, sicom/si com.…* est variable selon les textes, voire dans un même texte, et le nombre d'étiquettes associées (une ou deux) l'est aussi.

Il arrive enfin qu'à un mot soit associée une seule étiquette, mais une étiquette complexe qui réunit les valeurs de deux étiquettes de notre jeu (formes contractées, qui correspondent à des phénomènes d'enclise ; voir les exemples donnés plus haut, à la fin de 4.1.2 : par exemple « préposition + pronom personnel » : *Et tant fu desirranz* ***del*** [de+le] *veoir*).

Rappelons que la forme *sil* peut être analysée de deux façons différentes selon les contextes. Il peut s'agir d'une élision (avec absence de l'apostrophe) : *se* ('si') + *il*, *sil* correspondant à deux unités linguistiques ayant chacune leur propre étiquette (voir ci-dessus), ou bien, plus souvent, *sil* correspond à une enclise. Dans ce cas aussi on a deux unités linguistiques, mais l'analyse morphologique des unités est différente – *si + le* – ainsi que la façon de coder cette analyse : on a une seule étiquette complexe (qui rend compte de la contraction de l'adverbe *si* et du pronom personnel).

La question de la classification se pose à différents niveaux. Je l'ai évoquée ci-dessus à propos des constructions simples (les unités linguistiques et leur composition interne) dans une perspective synchronique et diachronique.

Il convient maintenant de reprendre ce fil directeur à propos des constructions plus complexes que j'ai étudiées ces dernières années, dans une perspective diachronique.

4.2. Evolution des constructions et changement de classes

L'évolution du français s'est caractérisée, depuis l'ancien français jusqu'au 17ème siècle, d'une part par des changements d'ordre phonétique et morphologique (achevés pour beaucoup au 16ème siècle), d'autre part par une modification importante des

principes de linéarisation des syntagmes et de la phrase, qui ont entraîné une profonde restructuration de ceux-ci. Ces principes se caractérisent par une hiérarchisation et une spécialisation accrue des éléments (voir Combettes 2006 et Marchello-Nizia 2003). Au niveau de la phrase, qui nous intéresse ici, on a coutume de considérer que l'on est passé d'une structuration dominée par la contrainte du verbe en seconde position (contrainte « V2 ») et par un principe d'organisation majoritairement informationnel pour les autres éléments, à une structuration où l'ordre est prioritairement régi par les fonctions grammaticales (Sujet-Verbe-Complément).

C'est évidemment un schéma simplifié : « V2 » n'est pas une contrainte absolue en ancien français, on trouve des occurrences de verbe en première ou en troisième position ; de même l'organisation n'est pas uniquement informationnelle, certaines séquences morpho-syntaxiques ne sont pas possibles (l'objet pronominal séparé du verbe, par exemple), et l'objet nominal tend à se placer derrière le verbe dès le 13ème siècle (voir Marchello-Nizia 1995). A l'inverse, des considérations informationnelles interviennent dans l'organisation de la phrase en français moderne (ainsi de la mise en relief d'un élément par le recours à une construction clivée).
Une autre caractéristique de l'évolution de la « phrase[87] » en français, en partie liée à la disparition de la contrainte V2, est la modification du rapport entre le « noyau » (défini ici comme le verbe et ses arguments : sujet et compléments) et la « périphérie » (ce qui entretient une relation lâche avec le verbe, un de ses arguments ou l'ensemble de la phrase ; ce qui n'est pas le noyau) : d'un côté, le noyau a vu se restreindre ses possibilités syntaxiques (fixation de l'ordre des mots et expression croissante puis obligatoire du sujet), d'un autre côté, la périphérie s'est développée, à gauche surtout[88] dès lors qu'elle n'a plus été contrainte par le verbe en seconde position (voir Combettes 1988) : le rapport entre les deux s'est modifié. La différence est particulièrement frappante entre l'ancien et le moyen français : outre la possible multiplication des éléments devant le verbe (conséquence du « recul » du verbe dans la proposition), on voit se développer certains éléments qui entretiennent un rapport assez lâche – en termes de rection – avec les éléments du noyau.

[87] Concept qui ne devient à vrai dire pleinement pertinent qu'au 18ème siècle (voir Seguin 1994).
[88] Il n'a jamais pesé de contrainte sur la zone postverbale.

Mes travaux s'inscrivent depuis ma thèse dans une perspective principalement diachronique, portant sur l'évolution de certaines constructions, simples ou complexes, depuis le français médiéval jusqu'au 16ème ou 17ème siècle, période à laquelle les constructions considérées ont acquis les caractéristiques syntaxiques, sémantiques et pragmatiques majeures qu'elles ont en français moderne. Les aspects phonétiques et morphologiques (et graphiques) n'ont été abordés que dans la mesure où ils avaient une incidence sur les caractéristiques précédemment mentionnées[89].

Les constructions à l'évolution desquelles je me suis intéressée participent à la restructuration de la phrase et mettent en jeu des changements de classes de différents ordres. C'est selon cet axe directeur de la classification que je les évoquerai, en distinguant deux cas de figure principaux : les constructions dont les caractéristiques se modifient sans pour autant qu'elles changent de classe, et celles qui, au contraire, changent de classe.

4.2.1. Evolution des caractéristiques des éléments d'une classe

4.2.1.1. Les sujets : d'une valeur sémantico-pragmatique à un statut d'indice grammatical

Les expressions concernées appartiennent à une même classe prédéfinie, au sens où elle est héritée de la tradition : ainsi de la catégorie « sujet », définie ici en termes morpho-syntaxiques comme l'ensemble des expressions, exprimées ou non, nominales ou pronominales, qui donnent leurs marques d'accord au verbe.

Il s'agit ici de décrire l'évolution de certaines des caractéristiques de la catégorie « sujet » à travers les différentes expressions qui peuvent l'instancier, la classe elle-même n'étant pas mise en cause dans son identité. Ces caractéristiques sont de différents ordres. Elles sont en premier lieu syntactico-positionnelles : le sujet peut être en position préverbale ou postverbale, exprimé ou non, ces possibilités s'étant réduites au fil des siècles au profit de l'antéposition au verbe et de l'expression obligatoire. Elles sont aussi, conjointement, de nature sémantico-pragmatique, des valeurs spécifiques étant attachées à la position et à la (non-)expression du sujet ; elles sont enfin d'ordre phonétique, le sujet pronominal s'étant progressivement cliticisé.

[89] Autant la cliticisation du pronom sujet à la fin du moyen français a eu une incidence forte sur sa syntaxe, autant la stabilisation morphologique et graphique de *propos* (vs. *propous, propoz, propost…*) n'a pas eu d'influence sur l'évolution des expressions *à (ce) propos / à propos de*.

J'avais, dans ma thèse (remaniée dans Prévost 2001), étudié le recul de la postposition du sujet, nominal et pronominal, dans six textes s'étalant de 1450 à 1550, cela dans une perspective informationnelle et sémantico-pragmatique. J'ai repris cette question depuis[90], en élargissant le nombre de textes considérés et en remontant jusque l'ancien français, et en me concentrant sur les sujets pronominaux (pronoms personnels plus spécifiquement). Ce choix a eu plusieurs motivations. En premier lieu, il me semble qu'il y a encore beaucoup à dire sur la syntaxe du pronom personnel sujet en ancien français et sur son évolution, dans un cadre autre que celui de la grammaire générative, théorie qui a donné lieu à une production importante sur la question[91], mais dont les explications proposées (purement syntaxiques) ne sont que partiellement satisfaisantes, et dont les travaux sur le sujet ne se sont en outre jamais appuyés sur un corpus conséquent. Les autres travaux, anciens pour beaucoup[92] consistent pour la plupart en l'étude synchronique d'un ou quelques textes, mais ne présentent pas d'étude approfondie sur un corpus large (ce qui n'empêche pas que plusieurs d'entre eux fournissent de précieuses informations et proposent des remarques très pertinentes[93]). Dans les grammaires sur la langue ancienne, la question n'est guère développée : la non-expression du sujet serait signe de continuité référentielle, tandis que l'expression dénoterait au contraire une discontinuité référentielle. Or les textes montrent que c'est souvent plus complexe. Quant à la postposition du sujet, rare, elle est (trop) souvent réduite à un automatisme : la présence d'un élément devant le verbe provoque, en raison de la contrainte « V2 », l'inversion du sujet ; or les occurrences de verbe en 3ème position existent.

Par ailleurs, la question de la non-expression du sujet reste encore largement à explorer, et il est difficile, pour l'aborder, de ne pas traiter le sujet pronominal, dans la mesure où il est admis de manière consensuelle que le sujet non exprimé correspond à un pronom potentiel (l'accord est moindre en revanche en ce qui concerne sa position). Jusqu'ici je n'ai que peu exploré la question de l'expression du sujet, au sens où je n'ai pas mené d'étude exhaustive, me contentant de sondages (Prévost, 2011a), et ce en raison de l'investissement important que constitue la collecte des

[90] Voir Prévost 2002, 2009, 2010, et 2011a.
[91] Voir en particulier Adams 1987, Rouveret 2004, Vance 1997.
[92] Voir entre autres Foulet, L. 1930/1965, Martin 1979, Price 1966, Zwanenburg 1978.
[93] C'est en particulier le cas de Foulet, qui, au début du 20ème siècle, dans *la Petite syntaxe de l'ancien français*, propose quelques réflexions assez innovantes sur les genres à privilégier pour l'étude de certains phénomènes (voir Foulet 1930/1965 : 345-356).

données, nécessairement manuelle en l'état actuel d'enrichissement des corpus. Rappelons que, de ce point de vue, il est bien plus simple de travailler sur les sujets pronominaux – liste fermée de formes identifiées – que sur les sujets nominaux.

L'inversion[94] du sujet pronominal est de prime abord une énigme dans un état de langue qui privilégie une organisation informationnelle de l'énoncé : aussi bien 'il' que 'je' constituent des topiques (ce dont on parle) ou bien des thèmes (charge informative peu élevée) par excellence, et on les attendrait de ce fait en tête d'énoncé, devant le verbe. Cela explique probablement leur rareté en position postverbale (rarement plus de 20% de l'ensemble des pronoms personnels sujets exprimés) dès l'ancien français, à une époque où le sujet nominal, dans certains textes, est davantage postposé qu'antéposé au verbe.

Il ressort des textes que j'ai étudiés que l'inversion pronominale, qui n'a cessé de se raréfier au fil des siècles au fur et à mesure que s'est généralisée la position préverbale du sujet en général, a de plus en plus été associée à des ruptures d'ordre logico-pragmatique. Aujourd'hui, l'inversion du sujet pronominal n'est plus possible que derrière quelques adverbes argumentatifs ou épistémiques, et elle dénote toujours une forme de mise en balance de la relation prédicative (*Paul est fatigué :* ***peut-être ne viendra-t-il pas*** *ce soir*)[95], même si celle-ci n'est pas toujours perceptible : l'inversion pronominale est en effet ressentie par de nombreux locuteurs comme une tournure figée, une structure archaïque.

En perdant sa relative liberté en matière de position et d'expression, le sujet pronominal a perdu son aptitude à véhiculer des valeurs sémantico-pragmatiques : il est devenu un pur indice grammatical.

Il reste encore beaucoup à explorer en ce qui concerne cette évolution : il convient de la corréler plus étroitement à celle de l'expression accrue du sujet, et, pour l'une et l'autre, de prêter une attention particulière aux évolutions respectives des différentes personnes (première, deuxième et troisième personne : je/nous, tu/vous, ils/ils). Il a en effet été suggéré (Detges 2003) que l'expression accrue du sujet aurait d'abord affecté le pronom de première personne, pour des raisons d'expressivité. C'est une piste intéressante, mais qui exige d'être mise à l'épreuve d'un corpus important et

[94] Dans la mesure où il s'agit d'un phénomène minoritaire, le terme d' « inversion » est légitime. Il ne l'est pas lorsqu'il s'agit du sujet nominal, dont la position postverbale peut dépasser les 50% : dans ce cas nous parlons de « postposition », terme neutre.

[95] voir (Guimier 1997) pour une étude complète de cette question.

diversifié. La prise en compte d'un tel corpus permettra d'affiner les chronologies respectives des différentes évolutions. De plus, les études que j'ai menées ont mis au jour un phénomène assez inattendu : en matière de répartition des sujets préverbaux et postverbaux, chaque texte a pour ainsi dire un micro-système qui lui est propre. Il n'est par conséquent pas aisé de dégager des tendances qui subsument ces fonctionnements individuels, et seule la multiplication des textes analysés, et leurs regroupements raisonnés, pourra faire avancer dans cette direction. A bien des égards il reste encore de nombreuses pistes à explorer du côté de la syntaxe du sujet pronominal : c'est la raison pour laquelle j'ai décidé d'y consacrer la recherche inédite qui accompagne ce mémoire de synthèse.

L'appartenance des expressions à la classe « sujet » est restée inchangée au fil des siècles : la classe préexistait à l'analyse moderne, et elle n'a pas été mise en cause dans sa définition[96]. On observe en revanche l'évolution de certaines des caractéristiques des éléments qui la composent (la perte d'une valeur sémantico-pragmatique au profit du statut de simple indice grammatical), changement qui n'est pas susceptible de remettre en cause l'appartenance des éléments à ladite classe (la définition reste inchangée, seules certaines caractéristiques ont évolué.)

A ce cas de figure on peut ajouter celui où une classe définie d'un point de vue strictement formel voit sa ou ses fonction(s) se modifier : l'évolution des constructions détachées illustre ce cas.

4.2.1.2. Les constructions détachées

Les constructions dites détachées sont identifiées et définies par certains critères syntaxiques : présence d'un élément « détaché », qui n'a pas de fonction dans la proposition qui suit, mais qui est lié par coréférence à un élément de reprise qui, lui, a une fonction dans la proposition (respectivement 'noz cumpaignons' et 'il' dans l'exemple (28) ci-dessous).

[96] Dans l'*une* de ses définitions, à savoir celle que nous avons retenue, de nature morpho-syntaxique : on sait par ailleurs que les définitions de la notion de sujet sont nombreuses.

Ces constructions sont attestées dès les plus anciens textes (la première occurrence que nous connaissons se trouve dans *La vie de Saint Léger* (début du 11ème siècle) et elles sont fréquentes dans la *Chanson de Roland* :

(28) *Noz cumpaignuns*, que oümes tanz chers,
Or sunt ***il*** morz, nes i devuns laiser. (*Chanson de Roland*, fin 11ème)

[*Nos compagnons* que nous avons tant aimés,
Ils sont maintenant morts, mais nous ne devons pas les laisser ainsi.]

Le fonctionnement syntaxique de ces constructions ne s'est guère modifié, hormis en ce qui concerne les formes possibles de reprise, qui se sont diversifiées. Deux points essentiels ont en revanche évolué : d'une part, le statut cognitif du référent détaché est moins libre qu'il ne l'était en ancien français, période à laquelle on pouvait trouver des référents identifiables mais inactifs[97] :

(29) « Orgoill oi e folage.
Ço set hom ben, n'ai cure de manace;
Mai *saives hom**, il*** deit faire message. (*Chanson de Roland*, fin 11ème)

[J'ai entendu les paroles d'un homme arrogant ou hors de son esprit.
Cela tout le monde le sait, je ne crains pas la menace
Mais c'est un homme sage qui doit porter le message]

Cela n'est plus possible aujourd'hui, où le référent doit être au moins accessible[98]. D'autre part (et c'est partiellement lié au point précédent), la fonction pragmatique de ces constructions a changé. En ancien français, et dans une moindre mesure en moyen français (période à laquelle s'opère le changement), la construction peut correspondre à une topicalisation (explicitation ou mise en relief d'un topique[99]) ou à une focalisation (mise en relief du focus[100]), le contexte permettant d'opter pour telle ou

[97] Je reprends ici la classification utilisée, entre autres, dans (Prévost 2001), et empruntée à Lambrecht (1994). Y sont distinguées, pour caractériser les référents d'un point de vue cognitif, les catégories de l'identifiabilité et celle de l'activation : « the first [information-structure category] is identifiability, which has to do with a speaker's assessment of whether a discourse representation of a particular referent is already stored in the hearer's mind or not. The second is activation, which has to do with the speaker's assessment of the status of the representation of an identifiable referent as already « activated », as merely « accessible », or as « inactive » in the mind of the hearer at the time of the speech act » (Lambrecht, 1994 : 76).

[98] Muller (1999) a montré que « un porte-manteau, vous en avez un là-bas » n'est possible que dans un contexte de reprise (« où puis-je trouver un porte-manteau ? ») : le référent n'est donc que faussement non identifiable/ inactif : il est en réalité accessible.

[99] Au sens de « ce dont on parle ».

[100] Elément dont la mise en relief avec le reste de l'énoncé rend celui-ci informatif.

telle interprétation[101]. Dans l'exemple 29 ci-dessus, la construction détachée correspond à une structure clivée moderne, dédiée à l'expression de la focalisation.

En français moderne, la construction détachée correspond toujours à une topicalisation, opération de « promotion du topique » qui permet, comme l'a montré Lambrecht (1994), de séparer l'expression de la référence et celle du rôle[102]. Ce n'était pas le cas en français médiéval, où une même construction pouvait remplir des fonctions pragmatiques différentes. Nous manquons malheureusement d'informations précises sur l'oral de l'époque : peut-être le syntagme « détaché » était-il accentué différemment (avait-il un accent ou deux ?) selon sa fonction pragmatique de topique ou de focus. Mais pour le locuteur moderne qui examine la construction ancienne, celle-ci est porteuse d'ambiguïté et seul le recours au contexte permet de déterminer sa valeur pragmatique.

L'évolution de la langue a tendu, au moins pour cette construction, vers la spécialisation fonctionnelle d'une classe formelle et vers la levée d'une possible ambiguïté.

Les études que j'ai menées des constructions détachées (Prévost 2003b et Prévost 2009), qui se sont en partie appuyées sur celles de Härma (1990) et de Marchello-Nizia (1998a et 1998b) sont restées d'une portée limitée en raison de la lourde collecte des données. Comme je l'ai souligné en 4.1.1., il n'est pas possible de repérer automatiquement ces structures en l'absence de corpus annoté syntaxiquement. Même si les données généreusement fournies par C. Marchello-Nizia m'ont permis d'enrichir ma propre collecte, il serait nécessaire d'explorer de nouveaux textes pour préciser et quantifier les caractéristiques morphosyntaxiques (forme, fonction…) de ces constructions, ainsi que l'évolution de leurs propriétés fonctionnelles. La mise à disposition d'ici fin 2011 d'un corpus enrichi syntaxiquement et dans lequel ce type de structures sera repéré devrait permettre de reprendre cette étude en lui donnant une plus grande ampleur.

4.2.1.3. Les éléments initiaux

Il est possible aussi de définir et de regrouper certaines expressions selon un critère strictement positionnel. Ainsi en est-il de la classe des éléments initiaux, qui

[101] On aimerait savoir si l'intonation était la même….

[102] Le traitement des deux étant trop complexe, au moins à l'oral ; par conséquent on commence par introduire le référent, ensuite on en dit quelque chose.

rassemble les expressions qui précèdent le sujet (ou le verbe si le sujet est postverbal). Ni la définition de la classe ni les éléments qui la composent ne se discutent (sauf à contester la borne gauche de la « phrase »…). L'intérêt porte sur la description des caractéristiques (possiblement de tous ordres : morpho-syntaxiques, syntaxiques, sémantiques, fonctionnelles…) des éléments de cette classe, et sur l'évolution diachronique de celles-ci. Cause ou conséquence de la modification de ces caractéristiques, certaines expressions ont cessé d'appartenir à cette classe. C'est en relation avec l'évolution de la syntaxe du sujet pronominal que je me suis intéressée aux éléments initiaux, afin de mettre au jour les modalités selon lesquelles la classe des éléments apparaissant avec un sujet postverbal s'est progressivement réduite, pour ne plus contenir aujourd'hui que quelques expressions adverbiales dénotant une mise en balance de la relation prédicative (*peut-être, sans doute, à peine…*). A l'inverse, la classe des éléments initiaux qui peuvent apparaître avec un sujet pronominal préverbal s'est considérablement développée, en raison de la disparition de la contrainte du verbe en seconde position.

Les caractéristiques d'une classe prédéfinie, celle des éléments initiaux, ont évolué, entraînant la modification des éléments qui la composent : il s'agit donc d'un cas en partie différent de celui de la classe des sujets et des constructions détachées, dont les éléments sont restés formellement inchangés, même si leurs caractéristiques ont évolué.

4.2.2. Changements de classes et grammaticalisation

4.2.2.1. Migrations entre classes prédéfinies

Dans l'évolution d'une langue, des constructions, simples ou complexes, peuvent non seulement endosser de nouvelles caractéristiques, mais aussi changer de classe. Si l'on considère la traditionnelle classification en parties du discours, on constate que l'histoire du français est riche en migrations d'expressions depuis des catégories dites « lexicales » vers des catégories dites « grammaticales ». C'est le processus typique, en français et semble-t-il dans la plupart des langues du monde, de la grammaticalisation.

La grammaticalisation constitue un type de changement linguistique, défini comme l'évolution d'une construction, simple ou complexe, lexicale ou grammaticale, vers un statut (plus) grammatical, et caractérisé par certains mécanismes, phonétiques,

morpho-syntaxiques, sémantiques et pragmatiques, parmi lesquels les processus de réinterprétation sémantique et de réanalyse syntaxique sont très fréquents. De nombreux phénomènes de ce type ont été décrits dans des langues très différentes (voir Heine et Kuteva 2000), ce qui a permis de mettre au jour les mécanismes précités. Ainsi, dans des langues typologiquement éloignées, plusieurs prépositions (ou locutions prépositionnelles) proviennent de substantifs dénotant des parties du corps (en français : *en face de, au dos de, à la tête de, aux pieds de, à côté de…*). La migration s'opère ici d'une catégorie pré-identifiée, celle des noms, vers une autre catégorie pré-identifiée, celle des prépositions. Il s'agit d'un cas relativement simple, l'évolution du lexical vers le grammatical est avérée, même si la locution qui résulte du processus peut être aussi interprétée comme une « lexicalisation », le phénomène de coalescence se retrouvant commun dans une large mesure aux deux processus (voir Brinton et Traugott 2005, et Prévost 2006). Cette ambivalence, ainsi que d'autres questions polémiques touchant à la grammaticalisation, ont été largement développées dans Prévost (2003a[103] et 2006), en particulier pour ce qui est de la question de l'unidirectionnalité du processus, que je considère comme un faux-débat. En effet, l'unidirectionnalité de la grammaticalisation étant inscrite dans sa définition (évolution d'une construction vers un statut plus grammatical), elle n'a pas lieu d'être discutée : les prétendus « contre-exemples » (évolution vers un statut moins grammatical : *des hauts et des bas*) n'en sont pas, ils correspondent simplement à un autre type de changement. Seul le caractère réversible d'une grammaticalisation mérite d'être envisagé, de même que le caractère directionnel ou non du changement linguistique en général, lequel ne se réduit assurément pas aux seuls cas de grammaticalisation, même si ceux-ci sont fréquents. Parmi les autres questions problématiques et/ou polémiques liées à la grammaticalisation, j'ai largement discuté celle des relations complexes entre grammaticalisation, dégrammaticalisation et lexicalisation, et, intimement liée, celle de la délimitation entre classes lexicales et grammaticales[104].

La grammaticalisation est considérée comme un type de changement linguistique

[103] Qui se veut une réponse aux articles très critiques sur la grammaticalisation parus dans *Langage Science* n°23, 2001.

[104] Ainsi certaines catégories sont mixtes : parmi les verbes, les auxiliaires sont à classer du côté grammatical, par opposition aux verbes précisément qualifiés de lexicaux. De même, parmi les adverbes, certains sont plutôt grammaticaux (*pas, guère, beaucoup, jamais…*) tandis que d'autres constituent une classe lexicale (*stupidement, agréablement*…).

généralisé à toutes les langues, et, en dépit de la difficulté à classer de manière consensuelle les différents types de changement, il semble que ceux relevant de la grammaticalisation soient les plus fréquents. La description d'un nombre sans cesse accru de cas de grammaticalisation, dans des langues différentes, a permis, par le repérage de régularités, d'ériger la grammaticalisation en un cadre d'analyse. Celui-ci est pourvu non seulement d'une définition des phénomènes concernés – évolution d'une construction vers un statut plus grammatical – mais il met aussi en avant des « prérequis » : les constructions susceptibles de se grammaticaliser sont généralement dotées d'un sémantisme assez général et, corollairement, d'une fréquence assez élevée : ainsi c'est le verbe *aller* qui s'est grammaticalisé en semi-auxiliaire, et non pas le verbe *sautiller*. Sont aussi mis en avant un certain nombre de mécanismes, relevant de différents domaines, comme je l'ai indiqué ci-dessus. Outre la réinterprétation sémantique et la réanalyse syntaxique, on peut citer les phénomènes de réduction phonétique, de décatégorialisation, de hausse de la liaison structurelle et de fixation de la position, d'affaiblissement sémantique – souvent compensé par un renforcement sur le plan pragmatique…, bien que tous n'entrent pas systématiquement en jeu[105].

Type de changement linguistique et cadre d'analyse, la grammaticalisation ne constitue en revanche pas, selon moi, une théorie, ne serait-ce que parce que son caractère prédictible s'avère limité : elle prédit surtout ce qui est peu susceptible de se grammaticaliser. En effet, si les mécanismes sont désormais bien décrits, de même que les prérequis, ce qui touche aux motivations (qu'est-ce qui fait qu'une grammaticalisation va se mettre en route ou non ?) reste encore largement mystérieux : le besoin d'expressivité, qui conduit au renouvellement des formes, mis en avant dès le début du 20ème siècle par Meillet (1912), ne suffit en effet pas à expliquer tous les changements. Assurément l'un des vastes chantiers qu'il reste à défricher est bien celui des motivations de la grammaticalisation.

[105] Parmi les nombreuses publications de référence sur la grammaticalisation, nous retiendrons : Heine 2003, Hopper 1991, Hopper et Traugott 2003/1993, Lehmann 1995/1982, Marchello-Nizia 2006, Traugott 2003.

4.2.2.2. Emergence de classes inédites : marqueurs de topicalisation et marqueurs discursifs

La grammaticalisation consiste typiquement en migrations entre catégories bien identifiées, même si la délimitation entre classes lexicales et grammaticales n'est pas consensuelle.

La migration d'une construction peut cependant s'opérer, aussi, vers une catégorie, une classe inédite, qui est progressivement identifiée et caractérisée par les propriétés communes (d'ordre syntaxique, sémantique, pragmatique…) de constructions, qui, précisément, vont constituer cette classe[106]. L'idée qu'une classe soit progressivement identifiée pose la question du moment de cette identification. Est-elle contemporaine, opérée, sinon par les locuteurs, au moins par les « grammairiens » de l'époque ? Est-elle au contraire postérieure au moment de la constitution de ladite classe ? La situation est à cet égard très différente selon les époques : en français médiéval l'activité métalinguistique est encore très réduite, elle est souvent prescriptive, et peu encline à repérer les changements, *a fortiori* naissants. La situation est différente en français moderne (et dans de nombreuses autres langues contemporaines), où la veille linguistique est constante, à l'affût de tout changement émergent, que seuls le temps et le recul permettront cependant de confirmer ou d'infirmer comme tel.

Certaines classes sont bien identifiées, de manière relativement consensuelle, depuis longtemps : il en est ainsi des classes morphosyntaxiques que sont les parties du discours (voir 1.3.2.), et, plus récemment, des catégories syntaxiques telles que sujet, verbe, complément d'objet (voir 1.3.3.). L'identification de certaines autres classes est en revanche plus récente, soulève un consensus moindre, et repose sur des critères pragmatico-discursifs (ce qui explique leur caractère moins consensuel, ces critères et leur application laissant une place plus grande – qu'en morphologie ou en syntaxe – à la subjectivité du locuteur/linguiste).

C'est le cas des marqueurs de topicalisation (*quant à X, à propos de X, en ce qui concerne X, pour ce qui est de X…*), expressions qui ont pour fonction de (ré)introduire un référent accessible du point de vue cognitif, qui fera l'objet d'une

[106] « ... deux questions distinctes bien que très liées sont soulevées : celle de la définition théorique et pratique des classes, et celle du codage, c'est-à-dire du travail concret d'affectation d'un cas à une classe. Un des apports de ces recherches [étude des nomenclatures dans le domaine sociologique et historique] est précisément de montrer que la première phase (définition des classes) ne peut jamais être pensée indépendamment de la seconde. » (Desrosières, 2001 : 116). On voit, dans l'exemple qui nous intéresse, combien les deux phases sont étroitement liées.

prédication dans l'énoncé qui suit, ainsi que des marqueurs discursifs (*à propos, à ce propos, au fait…*), expressions qui ont pour fonction d'indiquer des relations entre des unités discursives séquentiellement dépendantes, et qui spécifient la façon dont ce qui suit est connecté à ce qui précède (voir Fraser 1990 et 1999). J'ai réalisé plusieurs études de ces constructions : certaines sont consacrées à l'évolution de *quant à X* seul (Prévost 2003d et 2010a), une est dédiée à l'émergence comparée des deux marqueurs de topicalisation *à propos de X* et *quant à X* (Prévost 2008a), une autre (Prévost 2007) aux trois expressions bâties sur *propos* (*à propos, à ce propos, et à propos de*), et la dernière enfin est consacrée à l'émergence de *à propos* comme marqueur discursif (Prévost 2011b). Indépendamment de la proximité lexicale qui rapproche certaines des expressions auxquelles je me suis plus spécifiquement intéressée, les marqueurs de topicalisation et les marqueurs discursifs présentent des affinités à différents égards.

La première tient à ce que leur émergence a contribué au développement de la périphérie gauche de la phrase, et que leur processus de formation présente en outre de fortes similarités. Pour les marqueurs de topicalisation, hormis *quant à* qui remonte au latin et qui fonctionne dès l'ancien français comme une sorte d'introducteur de cadre (absence d'un élément de coréférence dans la proposition qui suit), il s'agit d'expressions qui développent une telle fonction à partir de celle de complément verbal, et ce selon un mouvement en trois étapes majeures : dans un premier temps, l'expression occupe une position postverbale et fonctionne comme un complément régi, proche d'un circonstant de domaine, comme en (30) :

(30) Cy dit exemple de princes vertueux et de vie bien ordenencée, **ramenant à propos du roy Charles** comment en touteschoses estoit bien riglé. (C. de Pizan, *Le livre des fais et bonnes meurs du sage roy Charles V,* 1401)

Dans un second temps, l'expression « migre » en position préverbale, et même souvent initiale. La fonction y est plus ou moins analogue au cas précédent, mais s'y ajoute une valeur textuelle de liaison avec le contexte précédent :

(31) Et pour ce, **à propos de telz gens dit trop bien Orace** que ilz sont aucuns qui se cuident avoir les yeulx plus fors que le spere du souleil, mais en eulx efforçant d'y regarder s'avuglent eulx mesmes. (C. de Pizan, *Le livre de la paix*, 1412)

Enfin, toujours en position initiale, l'expression endosse le statut de marqueur de topicalisation, comme en (32) et (33) (voir aussi (12) et (13) ci-dessus) :

(32) **Au regard de ses disciples**, **ils** se sauvèrent où ils purent (Rabelais, 1538)

(33) **Quant à la bataille à pié, elle** est tout au contraire de la bataille à cheval ; car gens à pié ne doivent jamaiz querir ne requerir leurs ennemiz (Jean de Bueil, *Le Jouvencel*, 1461)

Ce type de progression s'apparente tout à fait au mouvement en trois étapes mis au jour par Traugott (1982) pour un certain nombre d'expressions, qui évoluent du niveau propositionnel vers le niveau pragmatico-énonciatif, en passant par un niveau textuel. Précisons qu'il s'agit d'une succession logique, et non chronologique. On observe en effet fréquemment la coexistence des deux premières étapes, et si la troisième est plus tardive, elle n'exclut nullement le maintien des deux précédentes, comme on le constate en français moderne (exemples construits) :

(34a) Paul est inquiet *à propos de son avenir*

(34b) *A propos de son avenir*, Paul est inquiet

(34c) *A propos de Paul*, il est inquiet pour son avenir

Un mouvement analogue a présidé à l'émergence des marqueurs discursifs : des expressions qui avaient à l'origine un statut de complément régi postverbal ou préverbal, comme *a ce propos* en (35) :

(35) A laquele chose vault moult ajouster exemples manifestes, selonc le dit de Varron : « Tres clere maniere de enseignier est ajouster exemples ». **A ce propos dist Aristote** ou premier livre de Methafisique : « La chose dont nous sommes acoustumés doit on au corage appliquier, che qui est au dehors est inconvenient ». (J. Daudin, *De la erudition*, 1360)

se sont autonomisées vis-à-vis du verbe, comme en (36) :

(36) On voit coustumierement qu'à ces foires et marchez sont plusieurs coupeurs de bourses, qui ne font autre chose qu'espier leur belle, et regarder les moyens d'en avoir. **A ce propos**, un jour de marché, à Lyons, estoit un bon simple homme baissé assez bas, lequel marchandoit des naveaux estant contre terre sur du foirre, comme on les estalle. (P. d'Alcrippe, *La Nouvelle fabrique des excellents traicts de verité*, 1580-1596).

Cette autonomisation a pu se produire, semble-t-il, par l'intermédiaire d'énoncés ambigus qui autorisent une double lecture, avec ou sans rection verbale, comme en (37), où il est difficile de décider si *à ce propos* dépend ou non de 'réciter' :

(37) Mais au contraire il en y a d'aultres qui le sont par leur propre malice, comme vous avés ouy en la LXVIIIe nouvelle et de pluseurs aultres, comme cy devant ait esté dit. Et **à ce propos** je vous en vueil *reciter* encor une pour tousjours multiplier le

> nombre laquelle parle d'une jeune femme qui …(P. de Vigneulles, *Les Cent Nouvelles nouvelles*, 1515)

L'émergence de *à propos* comme marqueur discursif est plus difficile à appréhender, dans la mesure où l'expression ne s'est pas trouvée, initialement, en position de complément régi par le verbe. L'hypothèse avancée dans (Prévost 2011b), appuyée sur une étude détaillée de l'évolution des fréquences, est celle d'une analogie avec *à ce propos*.

L'autonomisation progressive de ces constructions vis-à-vis de l'énoncé a pu être mise en avant pour refuser de les considérer comme des cas de grammaticalisation. En effet, la hausse de la liaison structurelle et la réduction de la portée sont considérées par certains comme des paramètres essentiels de la grammaticalisation (voir Lehmann 1982/1995). En outre, dans la mesure où l'évolution de ces constructions les a conduites vers des emplois pragmatiques, il a été suggéré[107] d'en faire des cas de « pragmaticalisation » plutôt que de grammaticalisation, ce qui a pour corollaire de ne pas considérer la pragmatique comme partie prenante de la grammaire. J'ai défendu (Prévost 2011b) leur caractérisation en termes de grammaticalisation, et cela pour deux raisons. D'une part, je considère la pragmatique comme intégrée à la grammaire : quel que soit l'intérêt théorique qu'il puisse y avoir à maintenir le domaine des fonctions grammaticales et celui des fonctions pragmatiques clairement distincts, il a été montré (Traugott 1995 : section 2) que cela n'est pas toujours possible. Par exemple, alors que les expressions modales, temporelles, aspectuelles sont souvent considérées comme des catégories syntaxiques ou sémantiques (c'est-à-dire grammaticales), elles ont aussi des fonctions pragmatiques. Il a d'autre part été montré (Tabor et Traugott 1998 et Traugott 1995) que des cas de grammaticalisation avérés par ailleurs peuvent effectivement s'accompagner d'un relâchement syntaxique.

Autant il n'est pas souhaitable de voir la grammaticalisation devenir une notion vague susceptible d'accueillir tous les types de changement présentant une forme de rigidification, autant il ne me semble pas souhaitable de multiplier les termes et les concepts sans motif solide : on ne gagne rien selon moi à introduire la notion de pragmaticalisation, qui n'est au mieux qu'un sous-type de grammaticalisation.

[107] Voir en particulier : Aijimer 1997, Dostie 2004, Erman et Kostinas 1993, Güenthner et Mutz 2004.

Marqueurs de topicalisation et marqueurs discursifs ont par ailleurs en commun que les expressions n'ont pas changé de catégorie « morpho-syntaxique » en accédant à ce statut : il s'agit toujours de locutions prépositionnelles ou adverbiales. C'est du point de vue fonctionnel que s'est produit un changement : les constructions sont passées d'une classe identifiée (analysée en termes de « complément verbal » par le locuteur-linguiste moderne), à une fonction doublement nouvelle : nouvelle pour l'expression, et inédite dans la langue écrite de l'époque. L'expression ne vient pas s'insérer dans un paradigme préexistant : elle contribue, avec d'autres, à le créer, sur la base de caractéristiques communes, en l'occurrence syntactico-pragmatiques.

Mais c'est évidemment notre regard moderne et rétrospectif qui opère cette analyse, repérant l'accession de *à propos de* ou à *ce propos* à de nouvelles fonctions, qui leur sont désormais bien familières, et repérant du même coup l'émergence d'une classe, bien identifiée en français moderne.

Outre la difficulté à définir une classe et à identifier les éléments qui y entrent, se pose la question de projeter sur un état antérieur de la langue une classe conçue pour des états langagiers plus récents (voir 1.3.4.). Il s'agit du problème du décalage anachronique de l'analyse : quelle est la pertinence de la notion de marqueur discursif ou de marqueur de topicalisation au 15ème siècle ?

Plusieurs questions se posent à cet égard : quelle a pu être l' « analyse » de ce fonctionnement nouveau, à une époque où on peut le considérer comme stabilisé ? A ma connaissance, il n'en est pas question dans les grammaires du 16ème siècle. Mais comment donc faisait-on auparavant, sans marqueurs discursifs, sans marqueurs de topicalisation ? Les constructions détachées sont certes anciennes (4.2.1.2.), mais la possibilité d'introduire un référent à l'aide d'une expression dédiée – *quant a, à propos de* – avant qu'il ne devienne objet de prédication, est en revanche plus récente. Dans la mesure où ces constructions ne semblent pas venir en remplacer d'autres, faut-il en conclure qu'elles traduisent une nouvelle manière de penser ? Ou bien s'incrivent-elles dans une restructuration plus globale de l'organisation de la phrase, avec un élargissement de la « périphérie », tandis que le « noyau » se resserre ?

Mes recherches ne m'ont fourni jusqu'ici que peu de réponses à ces questions, mais il me semble important, néanmoins, de continuer à mettre en avant ces interrogations : outre leur intérêt intrinsèque, elles nous aident à relativiser notre positionnement

moderne et nos affirmations sur la langue ancienne, elles nous gardent d'une posture *trop* anachronique.

Conclusion

Se placer dans une perspective diachronique, c'est étudier des changements. Un changement peut correspondre à une construction qui apparaît, disparaît, ou change de fonction et/ou de classe ; il peut aussi s'agir d'un paradigme qui naît ou s'éteint (ainsi de l'évolution des temps en français), ou bien d'une classe, quelle qu'elle soit, qui surgit, s'éclipse, se reconfigure (ainsi de la nature de l'opposition entre les deux séries de démonstratifs : de sémantique, elle est devenue morpho-syntaxique). Ainsi donc des classes apparaissent, disparaissent, se modifient, et des constructions changent de caractéristiques et/ou de classes.

Peut-on déterminer certaines des motivations de ces différents changements ? Quelles sont les modalités de ces changements ? Quelles en sont les conséquences éventuelles ? Existe-t-il des corrélations entre différents changements (qu'il s'agisse d'un « effet domino », d'une macro-évolution, ou bien encore d'un rééquilibrage du système[108]) ? Peut-on déterminer des régularités ?

Ce sont ces questions, et la perspective de pouvoir y répondre, qui constituent à la fois les enjeux et l'intérêt des recherches diachroniques. J'ai essayé au mieux, dans mes travaux, de proposer des éléments de réponse, bien consciente cependant qu'il reste encore beaucoup à explorer.

J'ai la conviction que le développement des corpus numérisés, en nous permettant un accès facilité à un nombre croissant de données diversifiées nous permettra peu à peu de compléter les réponses en partie déjà apportées à ces questions. Il ne s'agit certainement pas de faire table rase des études passées. Même si beaucoup sont fondées sur un nombre de textes qui nous semble parfois, aujourd'hui, trop peu élevé, elles nous frappent souvent par leur perspicacité et leur intelligence de la langue.

Sans doute est-ce là désormais l'enjeu majeur pour nous, diachroniciens du français médiéval : savoir tirer parti et profit de la massification des données sans pour autant perdre le contact ni avec elles, ni avec les textes dont elles sont issues. Les données langagières sont produites dans des contextes, les changements s'initient dans des contextes : à tous égards la contextualisation des données est essentielle, et elle l'est

[108] Ainsi, typiquement, des créations analogiques :

le	*les*	Par analogie avec *le*, la langue a créé *ce* :	*le*	*les*
x	*ces*		*ce*	*ces*

d’autant plus que nous travaillons sur une langue dont nous n’avons pas la compétence.

C’est un usage du corpus aussi raisonné que possible que j’ai tenté de mettre en pratique dans la recherche inédite qui accompagne ce mémoire, et qui est consacrée à une approche quantitative de l’évolution de l’expression et de la position du sujet pronominal du 12ème au 14ème siècle. Elle constitue la première étape d’une étude plus vaste qui développera la dimension qualitative de ces deux phénomènes et accordera donc une place prépondérante à la contextualisation des constructions.

Références bibliographiques

Adams M., 1987. *Old French, Null Subjects and Verb Second Phenomena*, Ph.D. Dissertation, University of California, Los Angeles.

Aijmer, K. 1997. *I think* — an English modal particle. In Toril Swan & Olaf J. Westvik (eds.), *Modality in Germanic languages. Historical and comparative perspectives*, Berlin & New York : Mouton de Gruyter, 1-47.

Auroux, S. 1988. « La grammaire générale et les fondements philosophiques des classements de mots ». *Langages*, 92, 79-92.

Bally, Ch. 1932. *Linguistique générale et linguistique française.* Paris : E. Leroux

Biber, D. 1988. *Variation across speech and writing*. Cambridge : Cambridge University Press.

Biber, D. 1990. « Methodological issues regarding corpus-based analyses of linguistic variation ». *Literary and Linguistic Computing*, 5 (4), 257-270.

Biber, D. 1993. « Using register-diversified corpora for general language studies ». *Computationnal linguistics*, 19, (2), 219-241.

Biber, D. 1995. *Dimensions of register variation : a cross-linguistic comparison*. Cambridge : Cambridge University Press.

Brinton, L. J. et Traugott, E. C. 2005. *Lexicalization and Language Change*, Cambridge : Cambridge University Press.

Brunot, F. 1905-1938, puis 1979-2001. *Histoire de la langue française.* Paris : A. Colin & CNRS Editions, 17 vol.

Buridant, Cl. 2000. *Grammaire nouvelle de l'ancien français.* Paris : SEDES.

Charolles M. 1997. *L'encadrement du discours : univers, champs, domaines et espaces, Cahiers de Recherche Linguistique* 6, Université de Nancy-2, 1-73

Chervel, A. 1977. *Histoire de la grammaire scolaire…. Et il fallut apprendre à écrire à tous les petits Français.* Paris : Petite Bibliothèque Payot.

Chevalier, J.-C. 1994. *Histoire de la grammaire française*. Paris : PUF, collection *Que sais-je ?*

Chevalier, J.-C. 2006. *Histoire de la syntaxe. Naissance de la notion de complément dans la grammaire française (1530-1750).* Paris : Honoré Champion.

Chevalier, J.-C. et Encrevé P. 2006. *Combats pour la linguistique : de Marinet à Kristeva. Essai de dramaturgie épistémologique*. Lyon : ENS éditions.

Colombat, B. 1988a. « Présentation : Eléments de réflexion pour une histoire des parties du discours ». *Langages*, 92, 5-10.

Colombat, B. 1988b « Les 'parties du discours' (partes orationis) et la reconstruction d'une syntaxe latine au XVIe siècle ». *Langages*, 92, 51-64

Combettes, B. 2004. « *Quelque* : aspects diachroniques ». *Scolia*, 18, 9-40.

Combettes, B. 2006. « Grammaticalisation et parties du discours : la différentiation des pronoms et des déterminants en français ». In Guillot C., Heiden S. et Prévost S**.** (éds), *A la quête du sens. Études littéraires, historiques et linguistiques en hommage à C. Marchello-Nizia*, Lyon : ENS Editions.

Combettes, B. à paraître. « La délimitation du français préclassique : aspects syntaxiques », *Diachronie*, 1.

Combettes, B. et Marchello-Nizia C. 2010. « La périodisation en linguistique historique : la cas du français préclassique ». In B. Combettes, C. Guillot, E.

Oppermann-Marsaux, S. Prévost & A. Rodríguez Somolinos (éds), Le *changement en français. Etudes de linguistique diachronique.* Bern : Peter Lang, 131-143.

Corbin, P. 1980. « De la production des données en linguistique introspective », in A.-M. Dessaux-Bertheneau (éd.) *Théories linguistiques et traditions grammaticales*. Villeneuve d'Asq : PU de Lille.

Coseriu, E. 1980. « Vom Primat der Geschichte ». *Sprachwissenschaft* 5, 125-145.
Traduction en ligne par S. Verleyen (2007) http://www.revue-texto.net/1996-2007/Saussure/Sur_Saussure/Coseriu_Histoire.pdf

Coseriu, E. 1973. *Sincronía, diacronía e historia,* Madrid : Gredos
Traduction en ligne par T. Verjans (2007) http://www.revue-texto.net/1996-2007/Parutions/Livres-E/Coseriu_SDH/Sommaire.html

Desrosières, A. 1989. « Comment faire des choses qui tiennent : histoire sociale et statistique ». *Histoire et mesure,* IV, ¾, 225-242.

Desrosières, A. 2001. « Entre réalisme et conventions d'équivalence : les ambiguïtés de la sociologie quantitative ». *Genèses*, 43, 112-127.

Desrosières, A. et Thévenot L. 2000. *Les catégories socio-professionnelles.* Paris : la Découverte, série Repères Sociologie.

Detges, U. 2003. « Du sujet parlant au sujet grammatical. L'obligatorisation des pronoms sujets en ancien français dans une perspective pragmatique et comparative. *Verbum*, XXV, 3, 307-333.

Dostie, G. 2004. *Pragmaticalisation et marqueurs discursifs*. Buxelles : De Boeck-Duculot.

Erman, B. et Kotsinas U. B. 1993. « Pragmaticalization: the case of *ba* and *you know* ». *Studier i modern språkvetenskap* 10, 76–93.

Fillmore, C. J. 1992. « 'corpus linguistics' or 'computeraided armchair linguistics' ». In Svartvik, J. (ed.) *Directions in Corpus Linguistics*, n° 65 in *Trends in Linguistics*, 35-59. Berlin : Mouton de Gruyter,. Proceedings of Nobel Symposium 82, Stockholm, 4-8 août 1991.

Foulet, L. 1919. « *Quelque* ». *Romania*, 45, 220-249.

Foulet, L. 1919/1963. *Petite syntaxe de l'ancien français.* Paris : Champion.

Fraser, B. 1990. « An approach to discourse markers ». *Journal of Pragmatics*, 14, 383-395

Fraser, B. 1999. « What are discourse markers ». *Journal of pragmatics*, 31, 931-952.

Fournier, N. 1998. *Grammaire du français classique.* Paris : Belin.

Guimier, C. 1997. « La place du sujet clitique dans les énoncés avec adverbe initial ». In C. Fuchs (éd.), *La place du sujet en français contemporain*, Louvain : Duculot (Coll. Champs linguistiques), 43-96.

Güenthner, S. et Mutz K. 2004. « Grammaticalization vs. Pragmaticalization? The development of pragmatic markers in German and Italian ». In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes Grammaticalisation and its background: a Look from its Fringes and its Components,* Berlin & New York: Mouton de Gruyter, 77-108..

Habert, B. 2000. « Des corpus représentatifs : de quoi, pour quoi, comment ? ». In M. Bilger (éd), *Cahiers de l'université de Perpignan, 31, 'Linguistiques sur corpus. Etudes et réflexions'*, Perpignan : Presses universitaires de Perpignan, 11-58.

Habert, B et Zweigenbaum, P. 2002. Régler les règles. *TAL,* 43, (3), 83-105.

Härmä, J. 1990. «Les constructions disloquées en ancien français : problèmes de définition et de délimitation», In Kleiber G. & Tyvaert J.-E. (éds), *L'anaphore et ses domaines*. Paris : Klincksieck, 159-183.

Heine, B. 2002. « On the role of context in grammaticalization ». In Wischer et Diewald (éds), *New Reflections on Grammaticalization*, 83-101.

Heine B., 2003, « Grammaticalization ». In Joseph B. & Janda R., *The Handbook of Historical Linguistics*, Malden, MA : Blackwell, 575-601.

Heine, B. et Kuteva, T. 2002. *World Lexicon of Grammaticalization.* Cambridge : Cambridge University Press.

Histoire, Epistémologie, Langages. 1981. vol.3, fasc.1 « Sémantiques médiévales : cinq études sur la logique et la grammaire au Moyen Age »,

Histoire, Epistémologie, Langages. 1984. vol.6, fasc.1 « Logique et grammaire ».

Histoire, Epistémologie, Langages.1990. vol. 12, fasc.2 « Grammaires médiévales ».

Hopper, P. J. 1991. « On some principles of grammaticization ». In E.- C. Traugott & B. Heine (éds.) *Approaches to Grammaticalization*, vol. 1. Amsterdam & Philadelphia: John Benjamins Publishing Company, 17-36.

Hopper, P. J. et Traugott, E. C. 1993/2003. *Grammaticalization,* Cambridge: Cambridge University Press.

Hopper, P. J. 1998 « Emergent Grammar ». In Tomasello M. (éd), *New psychology of language*, 155-175.

Jakobson, R. 1952/1963. *Essais de linguistique générale*, Tome 1, Paris : Editions de Minuit.

Julien, J. 1988. « La terminologie française des parties du discours et de leurs sous-classes au XVIe siècle ». *Langages*, 92, 65-78.

Kroch, A. 1989. « Reflexes of grammar in patterns of language change », *Language Variation and Change 1,* 199-244.

Labov, W. 1976/1973, *Sociolinguistique.* Paris : Minuit.

Lagarde, J.-P. 1988. « Les parties du discours dans la linguistique moderne et contemporaine ». *Langages*, 92, 93-108.

Lallot, J. 1988. « Origines et développement de la théorie des parties du discours en Grèce ». *Langages*, 92, 11-23.

Lambrecht, K. 1994. *Information structure and sentence form : Topic, Focus, and the mental representations of discourse referents,* Cambridge : Cambridge University Press

Language Sciences. 2001. numero 23, *Grammaticalization : a critical assessment.*

Lavrentev, A. 2009. « Tendances de la ponctuation dans les manuscrits et incunables français en prose, du XIIIe au XVe siècle », thèse soutenue en juillet 2009 à l'ENS-LSH de Lyon.

Lehmann, C. 1982/1995. *Thoughts on grammaticalization*, Munich: Lincom-Europa.

Lightfoot, D. W. 1979. *Principles of diachronic Syntax.* Cambridge : Cambridge University Press.

Lusignan, S. 1986. *Parler vulgairement. Les intellectuels et la langue française aux XIII*^{ème} *et XIV*^{ème} *siècles.* Paris-Montréal : Vrin- Presses de l'université de Montréal.

Malmberg, B. 1945. *Système et méthode*, Lund.

Marchello-Nizia, C. 1995. *L'évolution du français : ordre des mots, démonstratifs, accent tonique*. Paris : Armand Colin.

Marchello-Nizia, C. 1979/1997. *La langue française aux XIVe et XVe siècles*. Paris : A. Colin.

Marchello-Nizia, C. 1998a. « Dislocations en ancien français : thématisation ou rhématisation ? » *Cahiers de Praxématique* 30, 162-178.

Marchello-Nizia, C. 1998b. «Dislocations en diachronie : archéologie d'un phénomène de 'français ora l' ». In Bilger M., van den Eynde K. & Gadet F. (éds), *Analyse linguistique et approches de l'oral, Mélanges Claire Blanche-Benveniste*. Paris/Louvain : Peeters, 327-337.

Marchello-Nizia, C. 1999. « Corpus diachroniques ». *Revue française de linguistique appliquée*, IV, (1), 31-39.

Marchello-Nizia, C. 2003. « La "contrainte de contiguïté ordonnée" dans l'évolution du latin au français et aux autres langues romanes », *Presencia y renovacion de la lingïistica francesa*, Actes du Colloque des langues romanes de Salamanque, I. Uzcangar Vivar, E. Llamas Pombo et J.-M. Pérez Velasco éd., Salamanque, Ediciones Universidad, pp. 231-244.

Marchello-Nizia, C. 2006. *Grammaticalisation et changement linguistique*, Bruxelles : De Boeck.

Marchello-Nizia, C. à paraître. « L'oral représenté : un accès construit à une face cachée des langues 'mortes' ». Actes du colloque Diachro-V, ENS Lyon, octobre 2010

Martin R. 1979. « L'ordre des mots dans le *Jehan de Saintré* » in Wilmet M. (éd) *Sémantique lexicale et sémantique grammaticale*. Bruxelles : VUB Centrum voor Taal- Literatuurwetenschap, 305-337.

Meillet, A. 1982. *Linguistique générale et linguistique française.* Paris-Genève : Champion-Slatkine. « L'évolution des formes grammaticales » (1912/1982 : 131-148).

Milner, J.-C. 1989. *Introduction à une science du langage*, édition abrégée, Paris : éditions du Seuil, collection Points Essais

Muller C. 1999, « La thématisation des indéfinis en français : un paradoxe apparent » *in* Guimier C. (éd.) *La thématisation dans les langues*. Bern : Peter Lang, 185-199.

Péry-Woodley, M.-P. 1995. « Quels corpus pour quels traitements automatiques ? ». *TAL*, 36, (1-2), 213-232.

Prévost, S. 1999 « *Aussi* en position initiale : évolution sémantico-syntaxique du 12ème au 16ème siècle », *Verbum*, tome XXI (3), 351-380.

Prévost, S. 2001. *La postposition du sujet aux 15ème et 16ème siècles : approche sémantico-pragmatique*. Paris : éditions du CNRS.

Prévost S. 2002. « Evolution de la syntaxe du pronom personnel sujet depuis le français médiéval : la disparition d'alternances signifiantes », in Lagorgette D. et Larrivée P. (éds) *Représentations du sens linguistique*, Munich : Lincom, *Studies in Theoretical Linguistics*, 22, 309-329.

Prévost S. 2003a. « La grammaticalisation : unidirectionnalité et statut », *Le Français Moderne*, tome LXXI (2), 144-166

Prévost S. 2003b. « Détachement et topicalisation : des niveaux d'analyse différents », *Cahiers de Praxématique*, 40, 97-126

Prévost S. 2003c. « Les compléments spatiaux : du topique au focus en passant par les cadres », *Travaux de Linguistique*, 47, 51-78

Prévost S. 2003d. « *quant a* : analyse pragmatique de l'évolution diachronique (14ème-16ème siècles) ». In Combettes B., Theissen A.& Schnedecker C. (éds) *Actes du colloque 'Ordre et distinction dans la langue et le discours'* (Metz 1999), Paris : Honoré Champion, 443-459.

Prévost, S. 2005. « Exploitation d'un corpus de français médiéval : enjeux, spécificités et apports ». In Condamines A. (éd) *Sémantique et corpus*, Paris : Hermès/Lavoisier (Série «Traité IC2 »; Cognition et traitement de l'information), 147-176.

Prévost S. 2006. « Grammaticalisation, lexicalisation et dégrammaticalisation : des relations complexes », *Cahiers de Praxématique*, 46, 121-139.

Prévost, S. 2007. « *à propos de, à ce propos, à propos* : évolution du 14ème au 16ème siècle», *Langue Française* 156, 108-126.

Prévost, S. 2008a. « *Quant à X* et *à propos de X* du 14ème au 16ème siècle : émergence de deux marqueurs de topicalisation », *L'information grammaticale*, 118, 38-43.

Prévost, S. 2008b. « Contraintes et spécificités de la constitution d'un corpus de français médiéval », *Corpus* 7, 35-64.

Prévost, S. 2009. « Topicalisation, focalisation et constructions syntaxiques en français médiéval : des relations complexes ». In Apothéloz D., Combettes B. et Neveu F. (éds) *Les linguistiques du détachement. Actes du colloque international de Nancy*, Bern : Peter Lang, 427-439

Prévost, S. 2010a. « *Quant à X* : du complément à l'introducteur de topique en passant par l'introducteur de cadre ». In Combettes B., Guillot C., Oppermann-Marsaux E., Prévost S. & Rodríguez Somolinos A. (éds), Le *changement en français. Etudes de linguistique diachronique.* Bern : Peter Lang, 325-343.

Prévost, S. 2010b. « Evolution de la position du sujet pronominal en français médiéval : une approche sémantico-pragmatique ». In Neveu F., Muni Toke V., Durand J., Klingler T., Mondada L., Prévost S. (éds.), *Congrès Mondial de Linguistique Française - CMLF 2010*, Paris : Institut de Linguistique Française, 305-320. [En ligne] http://dx.doi.org/10.1051/cmlf/2010106

Prévost, S. 2011a. « Expression et position du sujet pronominal en français », *Mémoires de la Société de Linguistique de Paris*, Tome XIX, François J. et Prévost S. (éds) 'L'Evolution grammaticale à travers les langues romanes', 13-33.

Prévost, S. 2011b. « *A propos :* from verbal complement to 'utterance marker' of discourse shift », *Linguistics*, 49 : 2, 391-413.

Prévost, S. et Schnedecker, C. 2004. « *Aucun(e)(s)* /*d'aucun(e)(s)* / *les aucun(e)(s)* : évolution du français médiéval au français moderne », *Scolia*, 8, 39-73

Price, G. 1966. « Contribution à l'étude de la syntaxe des pronoms personnels en sujets en ancien français », *Romania* n° 87, 476-504.

Romaine, S. 1989. « The role of children in linguistic change ». In Breivik, L.E. & Jahr, E.H. eds, 199-226.

Romaine, S. 1982. *Socio-historical Linguistics*, Cambridge : Cambridge University Press.

Rosier, I. 1988. « Les parties du discours aux confins du XIIème siècle ». *Langages*, 92, 37-49.

Rosier, I. 1981. « La notion de parties du discours dans la grammaire spéculative », *Histoire Epistémologie Langage*, Tome 3, Fascicule 1, *Sémantiques médiévales: Cinq études sur la logique et la grammaire au moyen âge*, 49-62.

Rouveret, A., 2004. « Les clitiques pronominaux et la périphérie gauche en ancien français », *Bulletin de la Société de Linguistique de Paris*, t. XCIX, fasc. l, 181-237

Saussure de, F. 1916. *Cours de linguistique générale*, Paris : Payot.

Seguin, J.-P. 1993. *L'invention de la phrase au XVIIIème siècle*, BIG, Louvain-Paris : Peeters-Société pour l'Information grammaticale.

Sinclair, J. (1996). *Preliminary recommendations on Coprus Typology*. Technical report, EAGLES, (Expert Advisory Group on Language Engineering Standards). <http://www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/corpustyp.ps.gz>

Skårup, P. 1975. *Les premières zones de la proposition en ancien français . Essai de syntaxe de position*. Etudes romanes de l'Université de Copenhague, *Revue Romane*, numéro spécial 6, Akademisk Forlag.

Smith, J.-C. 2002. « Middle French : When ? What ? Why ? », *Language Sciences*, 24, 423-445

Tabor, W. et Traugott E. C. 1998. « Structural scope expansion and grammaticalization ». In Ramat A. G. & Hopper P. J.. (eds.), *The Limits of grammaticalization* (Typological studies in Language 37). Amsterdam & Philadelphia : John Benjamins Publishing Company, 229-272.

Traugott, E. C. 1982. « From propositionnal to textual and expressive meanings : some semantic-pragmatic aspects of grammaticalization ». In Lehmann W. P & Malkiel Y. (éds), *Perspectives on Historical Linguistics*, Amsterdam, Benjamins, 245-271.

Traugott, E. C. 1995. « The role of the development of discourse markers in a theory of grammaticalization », Paper presented at the 12th *International Conference on Historical Linguistics*, Manchester, August. <www.stanford.edu/~traugott/ect-paperonline.html > (accessed 15 october 2007).

Traugott E. C., 2003, « Constructions in grammaticalization ». In Joseph B. & Janda R. (éds), 624-647.

Vance, B., 1997. *Syntactic Change in Medieval French : Verb-Second and Null Subjects*, Dordrecht-Boston-Londres, Kluwer Academic Publishers.

Weinrich, U., Labov, W. et Herzog, M. 1968. « Empirical Foundations for a Theory of Language Change ». In Lehmann W.P. (éd.) *Directions for Historical Linguistics, A Symposium*, Austin & London : University of Texas Press, 97-195.

Zwanenburg, W., 1978. « L'ordre des mots en français médiéval ». In *Etudes de syntaxe du moyen français*, R. Martin (éd.), Paris : Klincksieck, 153-171.