

**ÉCOLE NORMALE SUPÉRIEURE
LETTRES ET SCIENCES HUMAINES – LYON**

Laboratoire d'accueil : ICAR (UMR 5191 Université de Lyon - CNRS)
ED 484 – « Lettres, Langues, Linguistique & Arts » (3LA)

N° attribué par la Bibliothèque

**THÈSE DE DOCTORAT
en Sciences du Langage**

présentée et soutenue publiquement par

Alexei LAVRENTIEV

le 16 juillet 2009

**Tendances de la ponctuation
dans les manuscrits et incunables français en prose,
du XIII^e au XV^e siècle**

VOLUME I

Directrice de thèse

Mme Christiane MARCHELLO-NIZIA, Professeure émérite à l'École normale
supérieure lettres et sciences humaines – Lyon

Jury

Mme Marie-Luce DEMONET, Professeur à l'Université de Tours

Mme Elena LLAMAS POMBO, Professeure à l'Université de Salamanque, Espagne

Mme Lorenza MONDADA, Professeur à l'Université Lyon II

Mme Gabriella PARUSSA, Professeur à l'Université Paris III

M. Achim STEIN, Professeur à l'Université de Stuttgart, Allemagne

Résumé

La ponctuation du français médiéval reste de façon générale un domaine peu étudié, malgré un certain nombre de recherches menées dans le domaine depuis une trentaine d'années. Généralement « masquée » par les pratiques d'édition « scientifique », elle constitue pourtant une donnée linguistique non négligeable : elle révèle les structures dont le texte était constitué pour le « sujet écrivain » et donne au lecteur une orientation dans l'interprétation du message écrit.

La présente thèse vise à contribuer à la connaissance des grandes tendances de l'évolution des pratiques de la ponctuation dans les manuscrits et incunables français grâce d'une part au développement d'une méthodologie de représentation des données primaires et d'analyse linguistique de ces données tenant compte de multiples facteurs de variation et, d'autre part, grâce à l'apport de nouvelles données factuelles et à leur interprétation.

L'analyse linguistique effectuée se base sur une approche fonctionnelle de la langue, qui considère l'oral et l'écrit comme des systèmes de marques remplissant un certain nombre de fonctions. Sont donc étudiés non seulement les signes de ponctuation, mais aussi les marques linguistiques pouvant l'accompagner dans la structuration du texte à ses divers niveaux.

L'étude a été menée à partir d'un corpus de transcriptions « multi-facettes » de 29 extraits de manuscrits et d'incunables (28 000 occurrences-mots) exhaustivement annoté en « unités ponctuables ». Les transcriptions encodées en XML sont conformes aux recommandations de la TEI <<http://www.tei-c.org>> et sont intégrées à la Base de Français Médiéval <<http://bfm.ens-lsh.fr>>.

Mots-clés

ponctuation, langue française médiévale, manuscrit médiéval, incunable, corpus, transcription diplomatique, XML-TEI, philologie numérique

Abstract

Medieval French punctuation still remains an underexplored field in historical linguistics, although some research has been carried out since 1970ies. Medieval punctuation is almost completely ignored by scholarly editions. However it does contain some useful linguistic data, as it allows to see what was the “natural” text structure for medieval scribes and gives the reader some guidelines for interpretation of a written message.

The goal of this thesis is to contribute to the knowledge of major trends in the evolution of the practice of punctuation in medieval French manuscripts and incunabula in two ways. On the one hand a methodology of representation of primary sources and their linguistic analysis is elaborated. On the other hand, new data are introduced and analysed.

The linguistic analysis in this thesis is based on the functional approach of the language. Its spoken and written forms are considered to be systems of marks having a number of functions. Not only conventional punctuation marks are studied, but also all the other linguistic marks that can “work together” with punctuation in organising text structures at different levels.

The research has been based on corpus of “multi-level” transcriptions of 29 manuscript and incunabula fragments (28 000 words) fully marked up for “punctuable units”. The XML transcriptions are TEI conformant <<http://www.tei-c.org>> and are integrated into the BFM Old French corpus <<http://bfm.ens-lsh.fr>>.

Keywords

punctuation, Old French, medieval manuscript, incunabula, corpus, diplomatic transcription, XML-TEI, digital philology

Remerciements

Je tiens à remercier ma directrice de thèse Christiane Marchello-Nizia pour son soutien constant tout au long des années de recherche qui ont abouti à la rédaction de cette thèse. Sans son éclairage scientifique, son soutien moral et matériel, sa très grande disponibilité, ce travail n'aurait jamais pu voir le jour.

Je remercie Liselotte Biedermann-Pasques qui m'a guidé dans mes premières tentatives de recherche sur les systèmes graphiques et la ponctuation médiévale et qui m'a aidé énormément lors de mon premier séjour en France. Je remercie Fabrice Jejeic pour ses conseils sur les premières transcriptions de textes anciens.

Je remercie Céline Guillot et Serge Heiden avec qui j'ai eu le plaisir de travailler en équipe depuis plusieurs années et qui m'ont toujours soutenu aux moments où j'en avais le plus besoin. Le logiciel Weblex de Serge Heiden a été mon outil essentiel pour l'analyse des données du corpus.

Je remercie Sophie Prévost et Achim Stein, directeurs du projet qui m'emploie actuellement, pour leurs encouragements et leur compréhension.

Je remercie Benoît Habert pour ses conseils dans le domaine de la linguistique de corpus.

Mes remerciements vont à Elena Llamas Pombo et Nicolas Mazziotta, passionnés de la ponctuation médiévale. Mes échanges avec eux ont beaucoup enrichi mes connaissances et leurs encouragements m'ont été précieux.

Je remercie la Bibliothèque municipale de Lyon et, tout particulièrement le conservateur du fonds ancien M. Guinard, pour m'avoir gracieusement offert les reproductions numériques de haute qualité des manuscrits lyonnais de mon corpus.

Je tiens à rendre hommage à la mémoire du professeur Kirill Timofeev, mon premier directeur de recherche qui m'a donné goût de la linguistique, et du professeur Karl D. Uitti. La participation à son projet *Charrette* m'a ouvert la porte au monde de la philologie numérique.

Avertissement

Nous appliquons dans le présent mémoire les Rectifications orthographiques recommandées par l'Académie française et le Conseil supérieur de la langue française (*Journal Officiel*, Documents administratifs, 6/12/1990).

Cette thèse a été rédigée à l'aide du logiciel libre OpenOffice.org Writer <<http://www.openoffice.org>>.

La représentation des « caractères spéciaux » dans les transcriptions des sources primaires (marques de ponctuation médiévales, marques d'abréviations, variantes de lettres, etc.) est assurée par l'utilisation de la police Andron Scriptor Web (v. 2) développée par Andreas Stötzner et proposée gracieusement à des fins d'édition scientifique de textes médiévaux dans le cadre du projet *Medieval Unicode Font Initiative* (MUFI) <<http://www.mufl.info/fonts/>>.

Toutes les images des manuscrits du corpus utilisées dans ce mémoire sont issues des reproductions réalisées par les services photographiques des bibliothèques correspondantes.

Table des matières

Volume I

Résumé.....	3
Mots-clés.....	3
Abstract.....	5
Keywords.....	5
Remerciements.....	7
Avertissement.....	8
Introduction.....	15
0.1. Ponctuation dans l'oubli.....	15
0.2. La « nouvelle philologie » : une attention nouvelle portée à la matérialité du texte.....	19
0.3. « Réhabilitation » de l'écrit.....	21
0.4. État de l'art : les études antérieures sur la ponctuation des manuscrits français médiévaux.....	22
0.5. La nécessité des corpus.....	24
0.6. Un corpus raisonné et formaté pour être réutilisable.....	25
0.7. Critères de constitution de notre corpus.....	27
0.8. La ponctuation, un « système de marques ».....	29
0.9. Le concept d'« unité ponctuable ».....	30
Partie I	
Aspects théoriques, méthodologiques et techniques.....	33
Chapitre 1	
État des recherches sur la ponctuation.....	35
1.1. Approches de la ponctuation dans les grammaires traditionnelles.....	36
1.1.1. Antiquité et haut Moyen Âge.....	36
1.1.2. Humanistes et premiers imprimeurs français.....	38
1.1.3. Étienne Dolet et ses successeurs.....	42
1.1.4. Vision normative de la ponctuation au XXe siècle.....	44
1.2. Théories « scientifiques » de la ponctuation.....	49
1.2.1. Théories tenant compte des rapports entre l'oral et l'écrit.....	50
1.2.2. Approche autonomiste.....	54
1.3. Ponctuation médiévale française : état des recherches.....	60
Chapitre 2	
Cadre théorique et méthode pour l'analyse linguistique.....	73
2.1. Approche fonctionnaliste.....	74
2.2. Marques de ponctuation : noyau formel et élargissement fonctionnel.....	78
2.2.1. « Balises » graphiques : marques de ponctuation au sens propre.....	78
2.2.2. Marques segmentales de structuration textuelle : connecteurs.....	79

2.3. Fonctions de la ponctuation.....	80
2.3.1. Séparation des unités intratextuelles.....	81
2.3.2. Marquage de relations entre unités textuelles.....	84
2.3.3. Hiérarchisation des unités textuelles.....	85
2.3.4. Mise en relief d'éléments de la structure textuelle.....	87
2.3.5. Marquage du discours rapporté.....	88
2.3.6. Marquage de l'interrogation ou de l'exclamation.....	90
2.3.7. Vue d'ensemble.....	91
2.4. La notion d'« unité ponctuable ».....	92
2.5. « Phrase » : notion inapplicable.....	96
2.6. Classement des frontières ponctuables.....	98
2.6.1. Macro-unités textuelles (A).....	100
2.6.2. Frontières liées au discours rapporté (B).....	104
2.6.3. Propositions juxtaposées ou coordonnées (C).....	106
2.6.3.1. Propositions autonomes (Ca).....	106
2.6.3.2. Propositions avec des éléments communs (Cb).....	108
2.6.4. Propositions principales et subordonnées (D).....	112
2.6.5. Syntagmes coordonnés ou juxtaposés et énumérations (E).....	114
2.6.6. Syntagmes particuliers « ponctuables » (Fa).....	116
2.6.7. En dehors des frontières ponctuables (Fb).....	119
2.7. Analyse formelle des marques de ponctuation.....	120
2.7.1. Forme de la marque.....	121
2.7.2. « Force » de la ponctuation.....	126
2.8. Méthode pour l'analyse.....	129

Chapitre 3

Méthodologie de corpus.....131

3.1. La notion de corpus en linguistique.....	132
3.2. Contraintes sur la constitution d'un corpus.....	133
3.3. Typologie textuelle.....	135
3.3.1. Importance de la typologie textuelle pour la linguistique de corpus.....	135
3.3.2. Typologie textuelle appliquée dans la Base de Français Médiéval.....	137
3.3.3. Variables externes.....	138
3.3.3.1. Date.....	138
3.3.3.2. Espace.....	140
3.3.3.3. Auteur.....	141
3.3.3.4. Type de document.....	142
3.3.3.5. Forme.....	143
3.3.3.6. Domaine.....	144
3.3.3.7. Genre.....	145
3.3.3.8. Séquence textuelle.....	146
3.3.3.9. Variables « conceptionnelles » et « médiales ».....	147
3.3.3.10. Caractérisation du manuscrit.....	148
3.3.4. Variables internes.....	150
3.4. Constitution d'un corpus pour l'étude de la ponctuation médiévale.....	151
3.4.1. Grille des variables externes.....	151

3.4.2. Échantillonnage et représentativité du corpus	153
3.4.3. Caractéristiques du corpus BFM MSS.....	156
3.5. Conclusion : notre corpus comparé aux autres.....	160

Chapitre 4

Représentation des données primaires :

transcription de manuscrits médiévaux.....163

4.1. Éditions modernes comme source de données linguistiques.....	164
4.2. Tradition « diplomatique ».....	168
4.3. Éditions comportant des transcriptions imitatives.....	184
4.4. Présentation des données « sensibles » dans les éditions « multi-facettes ».....	189
4.4.1. Typologie des données.....	189
4.4.1.1. Abréviations.....	191
4.4.1.2. Caractères ambigus.....	196
-x = -us ?.....	196
u/v, i/j et des séries de jambages.....	197
ci/ti et ct/tt.....	199
diacritiques modernes (ç, é, ï, etc.).....	200
4.4.1.3. Diacritiques médiévaux.....	201
4.4.1.4. Variantes calligraphiques.....	202
4.4.1.5. Ligatures.....	203
4.4.1.6. Grandes initiales (lettrines).....	204
4.4.1.7. Marques de ponctuation.....	205
4.4.1.8. Segmentations particulières.....	207
4.4.1.9. Marques de corrections scribales.....	208
4.4.2. Synthèse.....	209
4.4.3. Exemples d'application.....	211
4.4.3.1. Transcription multi-facettes des <i>Serments de Strasbourg</i>	211
4.4.3.2. Format des citations du corpus.....	214
4.4.4. Conclusion.....	217

Chapitre 5

Aspects techniques.....219

5.1. Niveau de diffraction : mot ou caractère ?.....	220
5.2. Choix technologique : XML-TEI.....	222
5.3. Extension de la TEI.....	225
5.4. « Caractères spéciaux » dans les manuscrits médiévaux.....	228
5.5. Syntaxe « compacte » : saisie et corrections.....	231
5.6. Conclusion.....	234

Partie II

Analyse des données texte par texte et synthèse.....237

Chapitre 6

Manuscrits antérieurs à 1400.....239

- 6.1. Paris, BnF, lat. 9768 (*Serments de Strasbourg*).....239
- 6.2. Lyon, BM, Palais des Arts 77 (*Lancelot dou Lac*
et la *Queste del saint Graal*).....241
- 6.3. Paris, BnF, n.a. fr. 1119 (*Queste del saint Graal*).....248
- 6.4. Lyon, BM, 828 (*Livre d'Éracles* de Guillaume de Tyr).....252
- 6.5. Paris, BnF, fr. 12203 (*Conquête de Constantinople*
de Geoffroi de Villehardouin).....258
- 6.6. Paris, Bibliothèque Sainte-Geneviève, 587 (*Image du monde*).....263
- 6.7. Paris, BnF, fr. 11652 (*Coutumes de Beauvaisis* de Philippe de Beaumanoir)...267
- 6.8. Paris, BnF, fr. 574 (*Image du monde*).....272
- 6.9. Paris, BnF, fr. 13568 (*Vie de saint Louis* de Jean de Joinville).....280
- 6.10. Lyon, BM, 768 (*Prophéties* du maître Joachim).....289
- 6.11. Oxford, Bodleian, Laud Misc 587 (*Conquête de Constantinople*
de Geoffroi de Villehardouin).....296
- 6.12. Paris, BnF, fr. 840 (*Art de dictier* d'Eustache Deschamps).....301

Chapitre 7

Manuscrits du XVe siècle.....313

- 7.1. Vatican, Reg. Lat. 869 (*Chroniques* de Jean Froissart).....313
- 7.2. Paris, BnF, fr. 126 (*Quadriologue invectif* d'Alain Chartier).....319
- 7.3. Paris, BnF, fr. 619 (*Livre de chasse* de Gaston Phébus).....329
- 7.4. Lyon, BM, 765 (*Livre de chasse* de Gaston Phébus).....336
- 7.5. Paris, BnF, fr. 13318 (*Sermon pour la fête de la Sainte Trinité*
de Jean Gerson).....342
- 7.6. Lyon, BM, 769 (Fragment d'une description de l'Orient).....355
- 7.7. Paris, BnF, fr. 11610 (*Roman du Comte d'Artois*).....360
- 7.8. Paris, BnF, fr. 2701 (*Proposition... par devant... le conte d'Eu*
de Jean Juvenal des Ursins).....367
- 7.9. Rouen, BM, 1052 (*Quinze joies de mariage*).....377
- 7.10. Paris, BnF, fr. 2682 (*Chroniques* d'Enguerrand de Monstrelet).....383
- 7.11. Paris, BnF, fr. 15217 (*Le Saint voyage de Jherusalem*
du seigneur Ogier d'Anglure).....390
- 7.12. Paris, BnF, fr. 1357 (*Recueil des plus celebres astrologues*
de Simon de Phares).....399

Chapitre 8

Livres imprimés.....409

- 8.1. *Lancelot du Lac*, tome 2 (Paris, Jehan du Pré, 1488).....409
- 8.2. *Image du monde* (Paris, Jean Trepperel, entre 1492 et 1498).....417
- 8.3. *Image du monde* (Paris, Jean Janot, vers 1522).....422

8.4. <i>La manière de bien traduire d'une langue en aultre</i> (Lyon, Estienne Dolet, 1540).....	427
Chapitre 9	
Synthèse des tendances de la ponctuation.....	435
9.1. Observations générales.....	435
9.1.1. Variations du taux général.....	437
9.1.2. Ponctuation forte vs. ponctuation faible.....	438
9.1.3. Forme de la marque.....	439
9.2. Ponctuation selon le type de frontière ponctuable.....	442
9.3. Éditeurs scientifiques face à la ponctuation médiévale.....	447
9.4. Vers une typologie des ponctuations médiévales.....	451
Conclusion.....	461
Bibliographie.....	467
Travaux cités.....	467
Éditions de référence des textes du corpus.....	490
Sources des transcriptions.....	492
Listes et indexes.....	497
Liste des tableaux.....	498
Liste des figures.....	499

Volume II

Annexes

Annexe I

Manuel d'encodage de transcriptions de manuscrits.....	507
Introduction : Principes du codage XML, application et extension de la TEI.....	511
1. Descripteur (entête TEI).....	511
2. Niveaux de représentation des données (diffraction).....	520
3. Corrections sribales et éditoriales ; passages difficilement lisibles ou illisibles.....	524
4. Définition et description des éléments.....	526
5. Saisie et correction des transcriptions (syntaxe compacte).....	539
6. Relations dans la base des descripteurs BFM.....	540
7. Tableaux d'encodage des caractères spéciaux.....	541
8. Projets cités :	541
9. Index des éléments (entête).....	541
10. Index des éléments (hors entête).....	541

Annexe II	
Tables des « caractères spéciaux » et des raccourcis.....	543
II.1. « Caractères spéciaux » de la représentation allographique.....	543
II.2. Marques de corrections sribales.....	546
II.3. Raccourcis de la « syntaxe compacte ».....	548
Annexe III	
Fiches descriptives des textes du corpus.....	553
III.1. Fiches des manuscrits.....	553
III.2. Fiches des imprimés.....	567
Annexe IV	
Extraits du corpus.....	569
IV.1. Transcription XML (syntaxe compacte).....	569
IV.2. Transcription XML (syntaxe diffractée).....	574
IV.3. Présentation allographique.....	578
IV.4. Présentation graphématique.....	585
IV.5. Présentation normalisée.....	592
IV.6. Présentation pour les citations.....	598
Annexe V	
Tableau des fréquences de présence et d'absence de marques de ponctuation selon le type de frontière ponctuable.....	727

CD-ROM

- Edition électronique de la thèse (texte + annexes)
- Images numériques des manuscrits
- Corpus et outils : 30 extraits (transcription XML + schéma + feuilles de style + scripts de conversion)
- Edition multi-facettes des transcriptions (HTML statique)
- Base de descripteurs des manuscrits
 - Ouvrir le fichier index.html dans le dossier racine pour afficher la table des matières

Corpus interrogeable sur Internet via Weblex

- <http://weblex.ens-lsh.fr/wlx/>
- Nom du corpus : bfmms
- Accès réservé aux utilisateurs de la Base de Français Médiéval.
 - S'adresser au gestionnaire de la BFM (bfm@ens-lsh.fr) pour plus d'informations.

Introduction

La ponctuation dans les manuscrits médiévaux...¹ Quel est l'intérêt d'étudier ces petites marques, parfois à peine visibles, dont l'usage semble irrégulier à un lecteur moderne et que les médiévistes n'ont pris que très rarement en considération en préparant les éditions scientifiques de textes anciens ?

0.1. Ponctuation dans l'oubli

De la fin du XIX^e siècle et jusqu'à la fin des années 1970, les études consacrées à la ponctuation médiévale européenne ont été bien rares. A peine quelques pages étaient-elles consacrées à la ponctuation dans les ouvrages généraux de paléographie ou d'histoire de la langue, et les éditeurs de textes n'ont presque jamais hésité à substituer tacitement leur propre ponctuation à celle des scribes médiévaux².

Ainsi, Jacques Stiennon dans son manuel par ailleurs remarquable sur la *Paléographie du Moyen Âge* (1991), ne consacre que cinq lignes aux problèmes de la ponctuation (p. 154). Dans une annexe de cet ouvrage sont proposées des plaques de manuscrits accompagnées de transcriptions diplomatiques. Les transcriptions respectent

1 Notre travail est consacré à l'étude d'un aspect de l'histoire de la langue française. Il se situe dans le contexte de la recherche linguistique française et ne vise pas à proposer de généralisations de type universel. Toutes nos analyses et réflexions s'appliquent donc, sauf précision contraire, au français et à la linguistique française.

2 Les exceptions sont très rares. Christiane Marchello-Nizia (1978, p. 33) mentionne à ce sujet Siméon Luce et Heinrich Oskar Sommer, éditeurs de la fin du XIX^e et du début du XX^e siècle. Elle-même respecte la ponctuation du scribe dans son édition de la *Queste del saint Graal* (édition en ligne, en cours de préparation, dont un prototype est accessible à ce jour (2 mai 2009) à l'adresse suivante : <<http://weblex.ens-lsh.fr/pub/kq/>>).

les sauts de ligne du manuscrit et mettent en évidence les résolutions des abréviations, mais la ponctuation y est modernisée, ainsi que l'usage des majuscules.

Et pourtant, la ponctuation scribale est bien présente sur les folios reproduits, et elle est loin d'être incohérente. Par exemple, sur une page du *Sermon de saint Bernard* (ms. Paris, BnF, fr. 24768, f. 32 – fin du XII^e siècle) transcrite à la page 260 du manuel en question, nous avons relevé 32 marques de ponctuation dans la transcription et 30 dans le manuscrit³. Dans 23 cas une marque est présente au même endroit dans le manuscrit et dans la transcription, ce qui montre que la ponctuation de ce très ancien manuscrit n'est pas aussi éloignée de la ponctuation moderne que l'on pourrait l'imaginer. Notons que 19 de ces marques « communes » sont situées à une frontière entre deux propositions indépendantes, la position la plus « favorable » à l'apparition d'une ponctuation, comme nous le verrons par la suite.

Sur 10 occurrences de ponctuation ajoutée dans la transcription, 8 concernent la mise en relief d'éléments à l'intérieur d'une proposition, comme, par exemple, des compléments circonstanciels ou adjectifs épithètes :

mestiers iert, **lo matin**, ke nos ... soiens longement⁴ ensoniit (*l. 10*) ;

cil qui en sa citeit, **granz et haltismes**, fait grant joie as citains (*l. 25*).

L'absence de ponctuation dans ce type de groupe syntaxique semble être régulière dans les manuscrits médiévaux.

Dans 2 cas, Stiennon ajoute une virgule devant l'adverbe *si* :

quant nos ceu creons fermement, **si** avons nos jai lo merite (*l. 7*) ;

por ceu, **si** me samblet il ke nen iert mies senz exploit (*l. 13*).

Dans le premier cas, *si* est placé au début d'une proposition principale qui suit une subordonnée temporelle. Dans le second, cet adverbe est situé entre un connecteur causal *pour ce* et le noyau d'une proposition. Il s'agit donc toujours d'une frontière syntaxique moins forte qu'entre des propositions indépendantes. Selon une hypothèse de Marchello-Nizia (1985, p. 24), «*si* ne se trouve [jamais] en début absolu d'un texte

3 Nous n'entrons pas pour le moment dans l'analyse des formes et du fonctionnement des marques de ponctuation médiévales et nous nous contentons de noter la présence dans le manuscrit d'un signe ressemblant à une marque de ponctuation.

4 Les italiques servent dans cette transcription « diplomatique » à mettre en évidence les caractères restitués par l'éditeur à la place des marques d'abréviation dans le manuscrit.

Introduction

ou d'une partie d'un texte s'ouvrant après une pause ». C'est en effet plutôt une marque de connexion que de séparation, l'absence de ponctuation dans le manuscrit paraît donc tout à fait logique.

En ce qui concerne les 6 cas où une marque de ponctuation du manuscrit n'a pas été maintenue dans la transcription, ils concernent tous des frontières de groupes syntaxiques distincts (de propositions, en effet) qu'un auteur moderne pourrait bien vouloir ponctuer : une relative dans le premier exemple ou une construction corrélatrice avec la répétition du connecteur dans le deuxième :

car deus ne volt *que* nos nule chose aussiens **∴ ke** parmei les mains marie ne trespessest. (l. 5) ;

& assi cum une fontaine de vie k'espuesiee ne puet estre et ke **de tant** sort plus & habondet **∴ de tant** cum om l'espueset plus. (l. 17).

La suppression de ces marques de ponctuation est donc un choix arbitraire du transcripteur, et il n'est pas évident que ce choix serve vraiment à faciliter la compréhension du texte.

Bien sûr, d'autres manuscrits médiévaux peuvent paraître beaucoup moins « bien ponctués » que celui dont nous venons d'examiner un petit passage, mais le fait que nous ne savons jamais, en lisant une édition, si une marque de ponctuation a une correspondance dans le manuscrit ou si elle est introduite par l'éditeur, ne facilite pas l'analyse linguistique.

Le choix d'une ponctuation peut en effet avoir des conséquences importantes pour l'analyse syntaxique et discursive. L'ambiguïté, omniprésente dans le manuscrit médiéval, est fortement réduite dans un imprimé ponctué par un éditeur moderne.

Dans certains cas, la négligence de la ponctuation scribale a abouti à des erreurs de structuration de texte dans les éditions. Christiane Marchello-Nizia (1985) cite l'édition du *Lancelot en prose* par Alexandre Micha (1978), où le neuvième chapitre commence par un adverbe *si*, ce qui semble contredire son hypothèse (*cf.* ci-dessus). En réalité, ce chapitre fait partie des dix divisions factices que Micha a introduites dans son édition pour fragmenter les chapitres du manuscrit « démesurément longs » (Introduction, p. XIX).

Cependant, on peut remarquer que sur le folio 264 du ms. BnF, fr. 118, qui est reproduit avant la page de garde de l'édition Micha, l'adverbe *si* se trouve au début d'une section marquée par une grande initiale ornée de sept lignes. C'est donc une grande articulation marquée par un copiste, ce qui constitue un vrai contrexemple à l'hypothèse de Marchello-Nizia. Il convient toutefois de préciser que cette hypothèse est formulée sur la base de textes du premier tiers du XIII^e siècle, tandis que le manuscrit BnF, fr. 118 est daté du XIV^e siècle.

Dans l'édition Micha, cette occurrence de *si* n'est même pas placée en début d'une phrase. Peut-être Micha a-t-il suivi son manuscrit de base (Cambridge, Corpus Christi College Library, 45)... Il n'y a aucun moyen de le savoir sans consulter le manuscrit à Cambridge.

La modification (autrement dit, la perte) des données primaires des sources manuscrites dans les éditions imprimées n'a pas été considérée comme un problème pendant très longtemps. A la fin du XIX^e et au début du XX^e siècle les éditeurs cherchaient avant tout à « reconstruire » l'original, le plus souvent perdu, d'une œuvre médiévale en combinant les données de différents manuscrits et en corrigeant sans hésiter les « fautes » des copistes. Cette approche, inspirée de la méthode du philologue allemand Karl Lachmann (1793-1851) pour l'édition des textes de l'Antiquité (le Nouveau Testament, Lucrèce) et des poèmes allemands médiévaux, a été introduite en France par Gaston Paris et a dominé la philologie française pendant plusieurs décennies, même si un certain nombre d'exceptions notables doivent être retenues (par exemple, la célèbre collection des *Plus anciens monuments de la langue française* d'Eduard Koschwitz). En 1928, Joseph Bédier publie un article dans lequel il justifie le principe de fidélité au « meilleur manuscrit » que l'éditeur ne corrige qu'en cas de lacune ou de coquille indubitable. La grande majorité des éditeurs de textes médiévaux français se sont rangés depuis dans le courant bédieriste, mais les sondages effectués dans certaines éditions prétendues « fidèles » au manuscrit de base montrent que toutes les corrections par rapport au manuscrit ne sont pas signalées, et il est parfois difficile de les expliquer. L'édition de la *Mort Artu* par Jean Frappier (dernière révision en 1964) passée au crible par May Plouzeau (1994) en est un exemple des plus frappants.

0.2. La « nouvelle philologie » : une attention nouvelle portée à la matérialité du texte

Dans les années 1980, l'intérêt porté aux données primaires fournies par les manuscrits médiévaux se répand progressivement dans la communauté des linguistes et des philologues. Un mouvement de « nouvelle philologie » se forme, qui se caractérise par la valorisation de la tradition manuscrite des œuvres médiévales dans toute sa complexité et sa variabilité (Nichols 1990, p. 7). *L'Éloge de la variante* de Bernard Cerquiglini (1989), que l'on peut considérer comme le manifeste de ce mouvement, a suscité une grande polémique, mais cette polémique n'a fait qu'accentuer l'intérêt pour les sources primaires dans les études médiévales. Cerquiglini cite deux influences qui ont favorisé la formation de ce nouveau paradigme dans la philologie médiévale : la Nouvelle Critique (Roland Barthes, Jacques Derrida, etc.) et l'informatique. Le rôle de cette dernière n'a fait que croître ces dernières années.

A la fin des années 1980, la révolution que l'informatique est actuellement en train de produire non seulement dans le monde de l'édition et des recherches linguistiques et philologiques, mais dans toute la vie de la société contemporaine, n'en était qu'à ses débuts. Quelques projets de corpus et d'éditions électroniques de textes médiévaux avaient déjà commencé, mais aussi leur ampleur était limitée et les sceptiques étaient nombreux. Les défenseurs de la philologie « traditionnelle » ne nient pas l'utilité des éditions électroniques. Keith Busby (1993, p. 41) reproche même à Cerquiglini de rester très flou en parlant de la forme que pourraient prendre les « éditions écraniques » et de ne pas mentionner les projets visant à faire profiter les études médiévales des progrès de l'informatique : le corpus d'Anthonij Dees et le projet *Charrette* de Karl D. Uitti. Tout en reconnaissant l'intérêt que les éditions de ce type peuvent présenter pour les médiévistes, Busby affirme que celles-ci ne rendront pas les éditions critiques traditionnelles redondantes. Quinze ans après, le développement spectaculaire des nouvelles technologies (l'essor des bibliothèques numériques, l'« aplatissement » des écrans et l'invention de *e-paper*, pour ne citer que quelques exemples des avancées technologiques qui changent radicalement les habitudes de lecture et d'écriture dans la société moderne) fait penser que les éditions traditionnelles

sur papier disparaîtront entièrement dans un avenir relativement proche, au moins dans le domaine de la recherche. Cela ne veut pas dire pour autant que les changements technologiques rendent inutile le travail philologique nécessaire à la préparation d'une édition critique.

Les projets d'éditions électroniques et de bases de données textuelles se sont multipliés ces dernières années, mais il semblerait que souvent une petite partie seulement des possibilités ouvertes par les nouvelles technologies soit exploitée. L'informatisation des éditions constitue sûrement une avancée non moins importante que l'invention de l'imprimerie. Cette avancée est rapide à l'échelle historique, mais elle s'étend tout de même sur plusieurs décennies. De la même façon que les incunables cherchaient dans un premier temps à rester proches de l'écriture manuscrite, les éditions électroniques ne font encore souvent que reproduire à l'écran les éléments qu'on trouve dans les éditions papier, en y ajoutant quelques outils supplémentaires (comme une recherche de mots ou de chaînes de caractères)⁵. Dans le domaine de l'édition scientifique de textes médiévaux un grand travail d'adaptation de la méthodologie aux nouveaux outils disponibles doit être fait pour que les éditeurs et les lecteurs puissent profiter des possibilités offertes par l'informatique et pour assurer une certaine pérennité du travail investi dans une édition face à l'évolution très rapide des technologies.

Les initiatives de standardisation des pratiques de codage des textes et de mutualisation des outils d'édition – telles que la *Text Encoding Initiative* (TEI)⁶ – sont très importantes de ce point de vue, mais si les normes adoptées ne s'appuient pas sur une analyse scientifique de l'objet encodé, leur pérennité risque fortement d'être compromise. Si le manuscrit est reconnu comme la source primaire de l'édition d'un texte médiéval (à laquelle s'ajoute tout le travail d'analyse, d'interprétation et de correction éditoriales), une réflexion profonde sur les données du manuscrit qu'il convient d'encoder est nécessaire. Quel niveau de finesse de distinction de variantes calligraphiques des lettres et des marques de ponctuation faut-il choisir ? Comment représenter les abréviations ? Que faire des mots distincts (du point de vue du linguiste moderne) qui sont « collés » dans le manuscrit ou, au contraire, d'un mot unique coupé

5 La ressemblance entre le développement des éditions numériques et l'introduction de l'imprimerie a été remarquée par plusieurs chercheurs dès les années 1970 (Lowry 1979 ; Faulhaber 1991).

6 <<http://www.tei-c.org>>.

Introduction

par un espace blanc ? Un choix d'encodage opportuniste, non réfléchi, risque de rendre le travail de transcription non réutilisable pour des recherches futures. Nous espérons que notre travail d'analyse scientifique de la ponctuation et d'autres éléments des systèmes graphiques des sources textuelles médiévales contribuera à l'élaboration des normes d'encodage durables pour les futures éditions de textes médiévaux : tel est en effet l'un des buts du travail que nous présentons ici.

0.3. « Réhabilitation » de l'écrit

Le « retour aux sources » dans les études médiévales est donc bien dans l'air du temps. Il a commencé dans les années 1980 (et même 1970), et le développement de l'informatique n'est pas la seule raison qui l'explique.

Un autre facteur qui a joué dans le réveil de l'intérêt porté aux graphies et à la ponctuation médiévales à partir des années 1980 est le développement de la linguistique de l'écrit. Les travaux sur la linguistique du texte (Teun Van Dijk, Jean-Michel Adam, Bernard Combettes, etc.), sur les systèmes graphiques (Nina Catach, Jacques Anis, etc.) ont « réhabilité » l'écrit en tant qu'objet autonome d'études linguistiques. Le degré de l'autonomie de l'écrit par rapport à l'oral reste toujours sujet de discussions en ce qui concerne la langue moderne. Certains, comme Jacques Anis ou Geoffrey Nunberg, considèrent l'écrit comme un système complètement autonome, d'autres, comme Nina Catach, parlent des relations complexes entre l'écrit et l'oral. A l'époque médiévale ces relations n'étaient certainement pas moins complexes que de nos jours. Une petite partie seulement de la société médiévale savait lire et écrire, et tous les manuscrits ou presque étaient destinés à la lecture à haute voix : la lecture silencieuse ne semble tout simplement pas avoir existé avant le XIII^e siècle (Zumthor 1984, p. 37). Mais cela ne signifie pas pour autant que l'écriture médiévale soit une simple transcription de l'oral. L'abondance d'éléments décoratifs, d'abréviations, de variantes calligraphiques de lettres en témoignent. La diglossie franco-latine qui caractérisait l'activité de la plupart des scribes et des copistes médiévaux (sinon de tous) doit également être prise en compte.

Les textes sont de toute façon la seule source de données sur la langue médiévale disponible, et ce n'est qu'à travers les manuscrits qu'on peut étudier l'oralité de l'époque. Il est donc particulièrement important d'exploiter toutes les données dont on dispose, et la ponctuation scribale peut fournir de précieux indices aux chercheurs.

0.4. État de l'art : les études antérieures sur la ponctuation des manuscrits français médiévaux

Nous présenterons un aperçu général de l'état des recherches sur la ponctuation française médiévale dans la troisième section de notre premier chapitre, après avoir passé en revue les théories de la ponctuation significative pour notre recherche. Les excellentes synthèses réalisées par P. Rafti (1988) et par E. Llamas Pombo (2008) nous dispensent de l'obligation de présenter un compte-rendu détaillé des études sur la ponctuation médiévale. Ici, dans l'introduction, nous ne poserons que quelques jalons.

Les articles de Christiane Marchello-Nizia (1978) et de Claude Buridant (1980) ont été les premières véritables études linguistiques consacrées à la ponctuation française médiévale⁷. Ayant effectué une analyse fine des pratiques de ponctuation dans un nombre restreint de manuscrits, les deux auteurs s'accordent pour affirmer l'intérêt de la ponctuation pour la compréhension de la compétence linguistique du scribe médiéval, et pour souhaiter que les recherches sur ce sujet continuent.

Il a fallu attendre une dizaine d'années pour que ces souhaits commencent à se réaliser. C'est dans les années 1990 et 2000 que les recherches sur la ponctuation et les systèmes graphiques du français médiéval se multiplient.

En 1996, Elena Llamas Pombo soutient une thèse, dont une grande partie (plus de 150 pages) est consacrée à la ponctuation dans les traductions françaises d'Ovide (1996b). Tout récemment, deux autres thèses traitant de la ponctuation dans des textes de différents types ont été soutenues (Mazziotta 2007b ; Li 2007). Une dizaine d'articles sur la ponctuation et la structuration graphique dans les manuscrits français médiévaux ont vu le jour ces dix dernières années. Nous reviendrons sur les résultats des études de

⁷ La ponctuation avait été abordée auparavant dans quelques travaux de paléographes, de philologues et d'historiens, par exemple, dans l'article de Mario Roques sur le scribe Guiot (1952).

Introduction

la ponctuation française médiévale dans notre premier chapitre et ne mentionnerons ici que quelques exemples (Barbance 1995 ; Andrieux-Reix & Monsonégo 1996 ; Baddeley 2001).

Certaines tendances générales semblent se dessiner dans les pratiques de ponctuation analysées par tous les chercheurs, mais beaucoup de questions restent ouvertes. Le problème le plus important consiste à notre avis en une très grande variabilité de ces pratiques. Si l'éventail des marques utilisées est relativement stable⁸, la fréquence de leur emploi varie de moins de deux à plus de quinze marques pour cent mots. La singularité du système de ponctuation de chacun des manuscrits étudiés a été remarquée par tous les spécialistes qui ont mené des études comparatives de plusieurs manuscrits. Cependant, le nombre de manuscrits étudiés par chaque chercheur a toujours été limité à moins d'une dizaine, ce qui rend difficile toute tentative de généralisation.

Dans ces conditions, il est légitime de se poser la question de savoir s'il est possible, dans le cadre d'un projet de recherche de durée raisonnable, de tracer l'histoire de la ponctuation française médiévale, en faisant la distinction entre les grandes tendances de l'évolution et les phénomènes singuliers dus aux habitudes personnelles d'un scribe particulier. Nous sommes convaincu que la réponse à cette question peut être positive, à condition de disposer d'une méthodologie adaptée. Dans le présent travail nous tenterons de proposer une telle méthodologie et de l'appliquer à un corpus construit sur des critères précis représentant une certaine catégorie des textes de la période choisie : tel est le second but de cette thèse.

Trois chapitres de ce mémoire seront consacrés aux différents aspects de la méthodologie de la constitution du corpus (chapitre 3), de la représentation des données des sources primaires (manuscrits et incunables) dans les différents types d'éditions et de transcriptions (chapitre 4) et de la technologie de transcriptions « multi-facettes » appliquée dans notre travail (chapitre 5). Nous n'en évoquerons ci-dessous que les points cardinaux.

⁸ Nous évoquerons les marques les plus fréquentes à la fin de cette introduction et en présenterons une analyse détaillée dans le premier chapitre.

0.5. La nécessité des corpus

De nos jours, une grande partie des recherches linguistiques s'appuie sur des données textuelles informatisées. Même dans la linguistique générative les données issues de corpus jouent un rôle de plus en plus important. Beaucoup de recherches utilisent des corpus « génériques » censés représenter un certain état de langue (comme le *British National Corpus*), d'autres utilisent des collections de textes moins organisées (comme Frantext), d'autres encore profitent des technologies de « fouille de textes » (*text mining*) sur le Web, mais souvent les linguistes sont obligés de constituer leurs propres corpus pour des raisons diverses.

Pour le français médiéval, il n'existe pas à ce jour de corpus de référence représentatif⁹. L'une des raisons à cela est que la typologie textuelle n'est pas bien étudiée pour cette période, et en l'absence d'une grille typologique reconnue, tout jugement sur la représentativité d'un corpus ne peut être que subjectif et impressionniste. Par ailleurs, la quantité des textes parvenus jusqu'à nos jours est limitée (surtout pour la période la plus ancienne) et le nombre des domaines où le français écrit fonctionnait à l'époque était très restreint (mais a progressé au cours des siècles). Nous avons déjà publié certaines propositions concernant la grille typologique pour les textes de français médiéval (Lavrentiev 2007c ; 2007e), celles-ci seront développées et appliquées dans le présent travail.

L'une des bases textuelles les plus riches et diversifiées pour la période ancienne de l'histoire du français est la Base de Français Médiéval (BFM) développée à l'ENS LSH à Lyon. Elle comporte à ce jour près de 80 textes entiers, soit 3 000 000 occurrences-mots. La BFM a été constituée à partir de textes d'éditions critiques modernes fiables, dans le souci d'atteindre rapidement une taille importante en termes d'occurrences-mots. Cette base est malheureusement inutilisable directement pour l'étude de la ponctuation médiévale, car, comme nous l'avons déjà indiqué, aucune trace de la ponctuation scribale n'est conservée explicitement et de façon régulière dans les éditions. Cependant, la comparaison entre la ponctuation de l'éditeur d'un texte médiéval et celle de son manuscrit de base peut se révéler un outil intéressant, car elle

9 C'est par ailleurs aussi le cas pour le français moderne, mais pour des raisons différentes.

Introduction

met en évidence la spécificité des usages médiévaux. Il est par ailleurs plus facile de transcrire un passage d'un manuscrit en se basant sur le texte d'une édition, même si dans ce cas il est nécessaire de veiller à ne pas se laisser influencer trop fortement par les choix parfois arbitraires de l'éditeur.

Un exemple nous a été donné par Christiane Marchello-Nizia, qui a préparé une édition quasi-diplomatique d'un manuscrit de la *Queste del Saint Graal* (Lyon, BM, Palais des arts 77). Le nom de l'un des personnages de ce roman est régulièrement graphié *lyon* (ou *lyons* au cas sujet) dans une partie du manuscrit (quelques épisodes au début et vers la fin du roman où ce personnage apparaît). Il n'y a aucune marque d'abréviation, et la terminaison *-s* au cas sujet ne devrait pas être utilisée si le mot était abrégé. Néanmoins, Albert Pauphilet, qui a édité ce même manuscrit en 1923, a choisi de transcrire ce nom comme *Lyonel*, graphie attestée mais largement minoritaire dans le manuscrit. Sous l'influence de ce choix, Christiane Marchello-Nizia a également utilisé la graphie *Lyonel* dans une première version de son édition avant de rétablir la forme du manuscrit, car la nécessité de la correction n'est pas évidente dans ce cas, et en tout cas la correction aurait dû être signalée. Le changement de graphie dans le manuscrit peut correspondre au changement de main ou à un autre facteur pertinent pour une analyse linguistique.

0.6. Un corpus raisonné et formaté pour être réutilisable

Notre corpus de transcriptions de manuscrits a été conçu pour être complémentaire de la Base de Français Médiéval (BFM). Dans le choix des textes, nous avons privilégié ceux qui sont présents dans la BFM et avons travaillé, quand cela était possible, sur les manuscrits de base des éditions correspondantes.

Les dimensions du corpus que nous avons pu constituer pour la présente étude sont relativement modestes (29 extraits, soit 28 693 occurrences-mots au total). Nous avons préféré investir dans la finesse des transcriptions d'un corpus restreint mais adéquatement choisi, plutôt que chercher à constituer un corpus de grande taille mais moins riche en représentation des données primaires du manuscrit. Nous avons déjà

évoqué, avec Serge Heiden, ce dilemme de qualité vs. quantité des données dans les corpus de textes médiévaux (Heiden & Lavrentiev 2004) : selon les types de recherches auxquels un corpus donné est destiné, on est obligé soit de sacrifier la taille du corpus mesurée en nombre de mots, soit de simplifier la représentation des données¹⁰. Au stade actuel des connaissances sur la ponctuation médiévale une analyse qualitative fine semble encore plus importante que la masse de données quantitatives.

Si les caractéristiques des corpus sont déterminées par les types de recherches auxquelles ils sont destinés au moment de leur création, il est important d'assurer la possibilité de leur réutilisation pour des recherches d'autres types, d'enrichissement qualitatif et quantitatif, d'intégration dans des corpus et bases textuelles plus grands. Un corpus « jetable » ou difficilement réutilisable n'a pas beaucoup d'intérêt.

Le moyen le plus sûr d'y parvenir est de suivre dans la mesure du possible des standards ou des normes internationales reconnues dans le domaine de l'encodage des textes, de documenter les choix opérés là où les standards permettent plusieurs alternatives d'encodage ou là où les solutions normatives sont insuffisantes vis-à-vis des objectifs du corpus. La TEI, que nous avons déjà évoquée, cherche précisément à proposer des standards d'encodage basées sur le langage XML (*eXtensible Markup Language*) et applicables à toutes sortes de textes numérisés (ou numériques « de naissance »). La dernière version (P5) des recommandations de la TEI comporte un dispositif spécifique destiné à l'encodage des transcriptions de sources primaires. Le protocole d'encodage de notre corpus est basé sur ces recommandations, mais dans certains cas nous avons été amené à utiliser des balises supplémentaires ou à modifier les propriétés des balises recommandées pour encoder les données des manuscrits pour lesquelles les solutions proposées par la TEI nous ont paru insuffisantes. Nous avons en outre profité des extensions de la TEI proposées dans le cadre du projet Menota¹¹, qui répondent particulièrement bien aux besoins d'encodage de notre corpus. Dans tous les cas où notre protocole d'encodage va au-delà des recommandations de la TEI, nous en

10 La situation est d'ailleurs similaire dans le domaine des corpus oraux : plus un corpus est riche en représentation des données phonético-prosodiques ou interactionnelles moins sa taille est importante en termes d'occurrences-mots ou heures d'enregistrement.

11 *Medieval Nordic Text Archive*, <<http://www.menota.org>>.

expliquons les raisons. La conversion de nos transcriptions vers un format TEI standard est toujours possible, avec la perte d'une certaine partie des données.

0.7. Critères de constitution de notre corpus

En sélectionnant les textes pour notre corpus, nous avons cherché un équilibre entre la diversité des types représentés et la cohérence de l'ensemble.

Les types de textes ont été définis en fonction d'un certain nombre de variables. Ces variables sont les mêmes que dans la Base de Français Médiéval (date de composition, dialecte, type de document, forme, domaine et genre du texte, etc.), auxquelles s'ajoutent quelques variables spécifiques aux manuscrits (date du manuscrit, type d'écriture, etc.).

Pour préserver la cohérence du corpus, nous avons été amené à le limiter en fonction de plusieurs variables. Ainsi, nous avons choisi de travailler uniquement sur des textes en prose de type « œuvre » (littéraires et non-littéraires) et d'exclure les documents d'archive (chartes, etc.).

Les oppositions entre les documents d'archive et les œuvres et, pour les œuvres, entre les textes en prose et en vers, jouent en effet un rôle tellement important dans la structuration du texte que les facteurs à prendre en compte dans l'étude de la ponctuation sont en partie différents pour ces types de textes. Les documents d'archives sont en général de taille réduite et sont composés d'éléments stéréotypés (par exemple, le protocole, la notification, l'exposé, le dispositif, etc. dans les chartes). En analysant la ponctuation des chartes, il faut tenir compte de ces « parties du discours diplomatique » (autrement dit, de leur structure spécifique) (Mazziotta 2007a).

Dans les textes en vers, la division même en vers est une forme de ponctuation qui peut remplacer totalement ou en partie la ponctuation « syntaxique » que l'on trouve dans les textes en prose.

Cela ne signifie pas qu'il n'existe aucun rapport entre les pratiques de ponctuation dans les documents d'archives, dans les textes en vers et dans les textes en prose. Il serait, au contraire, intéressant de les comparer afin de révéler les points communs éventuels et de mieux comprendre les divergences. Les résultats des travaux

d'Elena Llamas Pombo sur des textes en vers et de Nicolas Mazziotta sur des chartes nous permettront de faire un premier pas dans cette direction, mais la constitution et l'analyse d'un corpus beaucoup plus important seront nécessaires pour aboutir à des conclusions définitives.

Le choix du XIII^e et du XV^e siècle comme limites temporelles de notre corpus n'est pas dû au hasard. La littérature française ne connaît pas de textes en prose avant le XIII^e siècle. Les textes non littéraires étaient à l'époque généralement écrits en latin (à l'exception des *Serments de Strasbourg* et quelques « traductions » d'écrits bibliques). L'apparition de la prose au XIII^e siècle et sa progression aux XIV^e et XV^e siècles sont liées à l'évolution générale de la culture de la société médiévale et, en particulier, au changement progressif du mode de lecture (développement de la lecture individuelle et silencieuse). A partir du XIII^e siècle de plus en plus de textes non littéraires (chroniques, traités didactiques, chartes et autres documents juridiques) sont rédigés (ou traduits) en français. Il est donc intéressant d'étudier comment ces changements se sont répercutés dans les systèmes d'écriture des manuscrits, et dans la ponctuation en particulier.

Dans la deuxième moitié du XV^e siècle un changement encore plus important intervient dans le monde de l'écriture : l'invention de l'imprimerie. En France, les premiers incunables voient le jour dans les années 1470. Dans un premier temps, leurs systèmes graphiques restent proches de ceux des manuscrits, mais une tendance à la normalisation se renforce rapidement : le nombre et la fréquence des abréviations et des variantes calligraphiques de caractères baissent progressivement. Le système de la ponctuation se stabilise parallèlement : les principes énoncés et leur application dans la pratique dans le premier traité de la ponctuation française d'Estienne Dolet (1540) sont étonnamment semblables au système moderne.

Si les manuscrits constituent l'objet principal de notre analyse, nous avons inclus dans notre corpus quelques incunables et imprimés du début du XVI^e siècle (ainsi que le traité de Dolet), ce qui nous permet de suivre cette évolution rapide et de mieux distinguer les traits du système ancien.

Ces limites formelles et temporelles posées, nous avons cherché à diversifier les types de textes représentés dans le corpus. Premièrement, nous avons cherché à équilibrer le corpus sur le plan diachronique : le même nombre de manuscrits a été

Introduction

sélectionné pour chacun des siècles. La plupart des manuscrits étant datés très approximativement, une plus grande granularité du classement chronologique n'est pas envisageable.

La deuxième variable prise en compte dans la diversification du corpus est le domaine de la typologie des textes. Cette variable, utilisée dans la description des textes de la Base de Français Médiéval, est définie en fonction de la destination principale d'une œuvre donnée. Les domaines littéraire, religieux, historique, juridique et didactique sont distingués dans la BFM. Ils sont tous représentés dans notre corpus, même si le « poids relatif » de chacun des domaines n'est pas le même. Ainsi, le domaine juridique est peu représenté, car les documents d'archives, qui y occupent une place importante, ont été exclus de notre corpus.

D'autres paramètres potentiellement pertinents pour l'analyse des systèmes graphiques ont été pris en compte dans l'organisation de notre corpus, même s'il n'a pas été possible d'équilibrer leur représentation. Il s'agit de la région de la réalisation du manuscrit, du genre du texte, du type de discours dans le passage transcrit¹², de la « qualité » du manuscrit, etc.

0.8. La ponctuation, un « système de marques »

Après avoir tracé les grandes lignes de la démarche que nous avons suivie dans la constitution du corpus, nous présenterons ici les principes fondamentaux à la base de notre analyse de la ponctuation dans les textes que nous avons transcrits et étudiés. Dans le chapitre 2, le cadre théorique que nous avons adopté et notre méthode d'analyse linguistiques seront exposés plus en détail.

Notre étude de la ponctuation médiévale s'inscrit dans le cadre théorique de la linguistique fonctionnelle et typologique. Selon l'approche fonctionnelle que nous adoptons, l'oral et l'écrit sont considérés comme des systèmes de marques remplissant un certain nombre de fonctions spécifiques. Certaines marques sont communes à l'oral et à l'écrit, d'autres sont spécifiques à l'un ou à l'autre. Toutefois, la plupart des fonctions sont remplies par tout un ensemble de marques hétérogènes, il existe donc des

12 d'après la grille de Jean-Michel Adam (1999).

liens fonctionnels entre des marques de nature différente. Les marques de ponctuation sont spécifiques à l'écrit, mais dans leur fonctionnement elles sont étroitement liées avec les marques verbales (telles que les connecteurs), et on peut trouver dans l'oral des marques ayant des fonctions similaires. Au Moyen Âge, où la connaissance de l'écrit était le privilège d'une petite partie des sujets parlants et où les manuscrits (surtout littéraires) étaient destinés à la lecture à haute voix, la fonction d'aide à l'« oralisation » que la ponctuation pouvait avoir, était sans aucun doute importante, sinon principale, comme l'affirment certains chercheurs :

[Les systèmes de ponctuation] auront avant tout pour but non d'aider l'analyse logique du discours, mais d'indiquer les périodes. Et, fait caractéristique, ils s'inspireront parfois des systèmes de notation musicale (Martin 1977, p. 584).

Nous étudierons donc non seulement les signes de ponctuation proprement dits, mais aussi toutes les marques linguistiques qui jouent un rôle dans la structuration du texte à tous les niveaux : celui des mots, celui des unités syntaxiques et celui de l'organisation globale du texte. Il convient de mentionner parmi ces marques l'emploi des majuscules et des lettrines de différentes tailles et couleurs.

0.9. Le concept d'« unité ponctuable »

L'analyse de la ponctuation dans notre corpus sera basée sur un balisage et une annotation des « unités ponctuales » dans les transcriptions de manuscrits. Tout segment de texte qui peut potentiellement être démarqué par un signe de ponctuation est pour nous une unité ponctuable. Il ne s'agit donc pas d'une unité syntaxique d'un type précis, mais d'un éventail d'unités de niveaux différents aux frontières desquelles des marques de ponctuation apparaissent dans certaines conditions, d'après des observations empiriques. Sa définition est en quelque sorte négative : un ensemble d'unités-mots au sein duquel on ne trouve pas de ponctuation.

Cette définition opportuniste nous semble justifiée, car les facteurs qui déterminent l'emploi de la ponctuation sont multiples (syntaxiques, sémantiques, pragmatiques, stylistiques, etc.), mais ce sont toujours des unités syntaxiques d'un niveau plus ou moins élevé qui sont séparées par une ponctuation, et il n'est pas très

Introduction

difficile d'établir une liste exhaustive de ces niveaux et des conditions qui favorisent, à chaque niveau, l'apparition d'une ponctuation.

Dans la pratique, il est nécessaire que la grille du classement des unités ponctuables repose sur des critères formels, facilement identifiables (pour assurer l'homogénéité de l'annotation du corpus), et indépendants de l'usage de la ponctuation (pour permettre des analyses croisées avec celui-ci).

Le niveau central dans le classement des unités ponctuables est celui de la proposition. La notion de phrase est en revanche inapplicable, car la définition formelle de la phrase s'appuie justement sur la ponctuation. Des doutes sur la pertinence de la notion de phrase pour les textes médiévaux ont par ailleurs été exprimés par plusieurs chercheurs (Marchello-Nizia 1978, p. 36 ; Nunberg 1990, p. 130 ; Baddeley 2001, p. 141). La définition formelle de la proposition comme une unité élémentaire associant un sujet (groupe nominal) à un prédicat (groupe verbal) est, au contraire, indépendante de la ponctuation, même si le degré d'autonomie varie considérablement entre les propositions de types différents (comme par exemple entre les propositions indépendantes d'une part et les propositions infinitives ou participiales d'autre part).

Il existe des frontières d'unités ponctuables de niveaux supérieur et inférieur à une proposition. Au niveau supérieur, il s'agit de titres, chapitres, épisodes et autres unités de la macrostructure du texte portant des marqueurs formels autres que la ponctuation. Au niveau inférieur, on peut trouver des termes multiples (coordonnés ou juxtaposés), des groupes syntaxiques isolés (compléments circonstanciels initiaux, appositions, etc.).

A chaque niveau, un classement plus fin est ensuite opéré. Outre les facteurs purement syntaxiques (comme la distinction entre la coordination et la subordination), nous prenons en compte la progression thématique, le changement du plan de discours, etc.

L'analyse des résultats d'annotation à l'aide des outils textométriques (Weblex de Serge Heiden, en premier lieu) devra nous permettre de révéler dans notre corpus des types de textes en fonction des pratiques de ponctuation et de les mettre en rapport avec les variables externes prises en compte lors de la constitution du corpus. Chaque type sera ensuite étudié de façon plus approfondie.

Nous nous pencherons notamment sur les rapports entre les marques de ponctuation et les connecteurs, sur la façon dont les éditeurs modernes ponctuent les textes médiévaux : tiennent-ils ou non compte dans la pratique de la ponctuation des manuscrits ? Les résultats obtenus seront comparés à ceux obtenus par les autres chercheurs, ce qui permettra d'émettre les premières hypothèses concernant les tendances générales.

Partie I

Aspects théoriques, méthodologiques et techniques

Chapitre 1

État des recherches sur la ponctuation

Il y a quelques dizaines d'années encore la ponctuation constituait la *terra incognita* de la théorie linguistique et de la linguistique historique (Roques 1952, p. 196 ; Marchello-Nizia 1978, p. 33 ; Catach 1979, p. 283 ; Barbance 1995, p. 505). Ce n'est plus le cas, comme le remarque très justement E. Llamas Pombo (2007a, p. 11).

Nous ne saurons mieux faire que de nous référer aux excellents travaux de recensement bibliographique et de synthèse publiés par P. Rafti (1988) et E. Llamas Pombo (2008) pour apprécier le nombre de publications et le panorama des problématiques abordés par les chercheurs. Deux ouvrages fondamentaux (Parkes 1992 ; Mortara 2008) permettent de tracer les grandes lignes de l'histoire de la ponctuation dans les différents pays d'Europe.

Dans ce chapitre, nous présenterons un résumé très bref du développement des théories de la ponctuation de l'Antiquité latine jusqu'à nos jours en France et tenterons de formuler les principales conclusions qui ressortent des études des pratiques de la ponctuation dans les manuscrits français médiévaux.

1.1. Approches de la ponctuation dans les grammaires traditionnelles

1.1.1. Antiquité et haut Moyen Âge

Dans un certains sens, la ponctuation existe dès qu'apparaît le premier texte écrit. Le fait qu'un texte donné commence et finit à un endroit donné constitue déjà une sorte de balise graphique. De grandes divisions sont marquées d'une façon ou d'une autre dès le II^e siècle avant J.-C. (Parkes 1992, p.10). Un système hiérarchisé allant de la distinction des mots jusqu'aux unités syntaxiques de différentes dimensions met plus longtemps à s'établir.

Dans les manuscrits latins en *scriptio continua* les mots n'étaient pas séparés et des marques de divisions d'unités logiques (*distinctiones*) pouvaient être introduites par les lecteurs afin de faciliter la lecture et la compréhension du texte (Parkes, p.9-12 ; Geymonat 2008, p. 27).

La ponctuation était considérée avant tout comme un outil d'aide à un lecteur inexpérimenté et n'intéressait pas beaucoup les philosophes et les grammairiens antiques. Chez Platon et Aristote, on peut trouver quelques mentions des problèmes de la division des textes (Geymonat 2008, p. 29), Cicéron évoque l'usage de signes spéciaux pour couper des phrases chez Démosthène (Hubert 1970, p. 14), mais il s'agit de toute façon de passages très brefs et allusifs.

Le premier traité mentionnant les signes de ponctuation proprement dits (στιγμή) est la *Grammaire* de Denys le Thrace¹³. Les trois signes décrits (τελεία, μέση et ὑποστιγμή) servent à marquer « un sens complet », une pause « pour respiration » et une toute petite pause signalant un « sens incomplet » respectivement. Le sens (qui est nécessairement véhiculé par une structure syntaxique) et l'expression (une pause respectée dans la prononciation) sont ici indissociables, et cet amalgame ressort en effet

13 Denys le Thrace vécut dans la deuxième moitié du II^e siècle avant J.-C. L'ensemble de sa *Grammaire* est aujourd'hui considéré comme une compilation tardive (III^e – V^e siècle après J.-C.), mais les premiers chapitres, dont le quatrième consacré à la ponctuation, remontent très probablement à Denys le Thrace (Lalot 1998, p. 25).

État des recherches sur la ponctuation

à chaque fois que la grammaire traditionnelle parle des marques de ponctuation¹⁴. Le grammairien grec ayant consacré le plus d'attention à la ponctuation est Nicanor (II^e siècle après J.-C.). On lui doit un traité en six livres intitulé *De la ponctuation en général* (Περὶ στιγμῆς καθόλου) (Geymonat 2008, p. 43). Sous l'influence de la logique des Stoïciens, il a proposé de distinguer huit marques de ponctuation sur des critères très fins. Le système de Nicanor s'est pourtant révélé trop complexe et artificiel pour se répandre dans la pratique¹⁵.

Dans la tradition latine, Quintilien, au I^{er} siècle de notre ère, évoque, lorsqu'il donne des consignes pour la lecture à haute voix, la nécessité d'observer des pauses et d'appliquer une intonation convenable aux endroits où le sens du texte le demande (*Institutio oratoria*, 1,8,1-2¹⁶ et 11,3,33-35¹⁷). Donat, dans son *Ars grammatica* (IV^e s.) utilise le terme de *positura* au sens « technique » de signe de ponctuation (Keil 1981, p. 372). A la suite de Denys le Thrace, il distingue trois *positurae*, ou *distinctiones* (*distinctio*, *subdistinctio* et *media distinctio*) et explique leur usage dans pratiquement les mêmes termes que le grammairien alexandrin.

Cette distinction tripartite est reprise par pratiquement tous les grammairiens de l'Antiquité tardive et du Moyen Âge qui traitent de la ponctuation (sans jamais y consacrer plus de quelques pages, voire quelques lignes)¹⁸, même si les termes utilisés, la forme des marques et le sens qui leur est donné varient selon les auteurs (Hubert 1972). En aucun cas, il ne s'agit d'une théorie au sens propre, mais de consignes pratiques plus ou moins raisonnées.

Isidore de Séville (c. 560 – 636) a établi un lien entre les marques de ponctuation et les parties du discours rhétorique : *subdistinctio* avec le *comma*, *media distinctio* avec le *colon* et *distinctio* avec le *periodus* (Parkes 1992, p. 21). Les termes de la rhétorique

14 Il s'agit bien d'un amalgame et non d'une confusion, car les grammairiens étaient bien conscients, dès l'Antiquité, des rapports complexes entre l'oral et l'écrit (Desbordes 1988)

15 Selon (Geymonat 2008, p. 45) des traces plus ou moins importantes de tentatives d'application de ce système sont perceptibles uniquement dans les *Scholia vetera* pour Homère et dans certaines gloses de la *Grammaire* de Denys le Thrace.

16 (Cousin 1975, p. 124).

17 (Cousin 1979, p. 231).

18 Les extraits des textes grecs et latins (des plus anciens jusqu'au XV^e siècle) qui parlent de la ponctuation ont été rassemblés par M. Hubert (1970). F. Desbordes (1990, p. 237) a fait pourtant remarquer le caractère non critique de ce recueil et le manque de précision dans la chronologie.

ont ensuite remplacé les anciennes dénominations des signes. En un sens, Isidore peut être considéré comme le fondateur de la conception « autonomiste » de la ponctuation.

Dans la pratique courante, cependant, la ponctuation du « niveau syntaxique » se limitait le plus souvent, tout au long du Moyen-Âge, à l'usage dominant d'un seul signe, dont la « force » pouvait changer grâce à la lettre suivante (lettre ordinaire ou *littera notabilior*).

Certaines innovations ont été apportées dans le répertoire des marques de ponctuation à l'époque de la « Renaissance » carolingienne, suite notamment au besoin de mieux mettre en forme des textes liturgiques. C'est à cette époque qu'on commence à utiliser comme *litterae notabiliores* des caractères « empruntés » à des écritures anciennes (distinction majuscule/minuscule) et que les nouvelles marques, le *punctus elevatus* et le *punctus interrogativus*, font leur apparition (Parkes 1992, p. 34-36). Ces innovations n'ont pourtant pas donné lieu à des spéculations théoriques.

1.1.2. Humanistes et premiers imprimeurs français

Une étape importante dans l'évolution des pratiques de la ponctuation est liée à l'activité des humanistes italiens qui utilisaient pleinement l'éventail des marques disponibles et l'ont enrichi de quelques nouveaux éléments, comme les parenthèses et le point d'exclamation (*punctus admirativus*) (Parkes 1992, p. 48-49).

Dans leur désir de restaurer la langue latine dans sa pureté, les humanistes ont prêté un soin particulier à la clarté matérielle de l'expression orale et écrite, d'où l'attention portée à la calligraphie, l'orthographe et la prosodie (Beltran 1985, p. 285).

C'est à un humaniste italien, Gasparin de Bergame (ou *Gasparino Barzizza*, ca. 1370 – 1431) que l'on doit un premier traité depuis l'Antiquité consacré spécialement à la ponctuation (*Doctrina ponctandi*). Ce texte n'a malheureusement jamais été publié et n'est accessible à la communauté scientifique que grâce au résumé qu'en a fait G. Ouy (1987, p. 182-187).

En France, le premier « traité » consacré à la ponctuation voit le jour dans le deuxième livre imprimé en France, à l'atelier de la Sorbonne. Il s'agit d'un petit *Compendiosus de arte punctandi dialogus* attribué à Jean Heynlin (Lapide), fondateur

État des recherches sur la ponctuation

avec Guillaume Fichet, de la première imprimerie française. Ce texte de quatre pages de petit format (in 4°)¹⁹ a été ajouté à la fin de l'*Orthographia* de Barzizza, imprimé dans les premiers jours de 1471, quelques semaines après le premier incunable français (*Epistolae* de ce même Gasparino Barzizza) (Veyrin-Forrer 1987, p. 173).

Selon E. Beltran (1985), ce dialogue est une sorte de résumé du traité de Guillaume Fichet *De caracteribus institutio* qui se présente sous la forme d'une longue lettre adressée à Étienne Piard. Ce dernier est conservé dans le manuscrit BnF lat. 7555, compilé, toujours selon E. Beltran, après 1473 par un étudiant. Il ne s'agit donc pas d'un manuscrit autographe, et la copie est décrite comme « médiocre » par le même chercheur. La date d'envoi du traité à son destinataire se situerait entre 1474 et 1480. Même si la date de composition du traité est forcément plus ancienne, il n'est pas évident qu'elle soit antérieure à celle du *Compendiosus dialogus*. Le dialogue n'est pas signé dans son édition originale, mais le nom de *Johanne de Lapide* (Jean Heynlin) apparaît dans les éditions dès la fin du XV^e siècle²⁰. Nous préférons donc nous en tenir à l'attribution traditionnelle de la paternité du *Compendiosus dialogus*²¹.

Le fait que les premiers imprimeurs français se soient intéressés de près à la ponctuation n'est pas dû au hasard. Comme tous les humanistes, ils cherchaient à rétablir la clarté de l'expression écrite. L'introduction de l'imprimerie se présentait en effet comme une excellente occasion de mettre de l'ordre dans les pratiques médiévales, plutôt « anarchiques » en matière de la ponctuation. Le succès que le *Dialogus* a connu à la fin du XV^e siècle²² témoigne de l'intérêt que le problème de la ponctuation suscitait à l'époque. Ce texte mérite donc un examen attentif²³.

Comme son titre l'indique, le traité se présente sous la forme d'un dialogue entre un élève et son maître qui explique la forme et la fonction des signes de ponctuation. La

19 Dans les rééditions de plus grand format, le dialogue entier ne dépasse pas une seule page.

20 E. Beltran (1985, p. 286) cite l'édition de Arnold von Köln (Leipzig, 1493).

21 Quoi qu'il en soit Guillaume Fichet et Jean Heynlin étaient des amis, et le traité peut très bien être le fruit de leurs discussions.

22 Ce texte a été réédité à plusieurs reprises (à côté de l'*Ars diphtongandi* de Guarino de Verone) à Bâle et à Lyon entre 1478 et 1488, comme en témoigne le catalogue de la BnF. Un facsimilé (de très mauvaise qualité malheureusement) de l'édition lyonnaise de 1482 est disponible sur Gallica <<http://gallica.bnf.fr/ark:/12148/bpt6k535510>>.

23 Une transcription du début de ce dialogue effectuée selon le même protocole que celles de notre corpus est proposée à la fin de l'annexe 4. Une reproduction photographique en couleur est fournie sur le CD ROM accompagnant la présente thèse.

fonction des signes de ponctuation est triple²⁴ : ils séparent les unités du discours (*disiugendas lectionis partes separant*), permettent au locuteur de respirer (*pronuntiantis spiritum recreant*) et aux auditeurs de saisir le sens de l'énoncé (*auditoribus orationis senso distinguunt*). La ponctuation est donc toujours perçue comme un outil d'aide à la lecture, mais sa fonction primaire consiste à structurer le discours et à bien en exprimer le sens.

Le Maître du *Dialogue* présente, tour à tour, la *virgula* (petite barre oblique située en bas d'une ligne), le *comma* (point avec une petite barre oblique au-dessus), le *colon* (point tout simplement) et le *periodus* (point avec une barre oblique ou une virgule au-dessous). La liste comprend également *punctus interrogativus* (point surmonté d'une virgule renversée), *parenthesis* (parenthèses rondes), *gemipunctus* (deux points disposés horizontalement) et *semipunctus* (petite barre oblique située vers le haut d'une ligne).

La virgule est utilisée après les expressions sans verbe (*ponitur post dictiones verbo ac perfectione carentes*) que nous appellerions « syntagmes non propositionnels »²⁵. Dans la prononciation (évoquée tout à la fin du dialogue) la virgule doit être représentée par une très courte pause. Le *comma* est utilisé après les expressions dotées d'un verbe mais dépourvues de « perfection » (*post dictiones habentes verbum : sed perfectionis expertes*), ce qui correspond dans notre terminologie aux frontières « subordonnée/principale » et « propositions avec des éléments communs ». Dans la prononciation, elle correspond à une pause plus longue et à une « intonation suspendue » (*cum suspensa tamen voce*). Le *colon* marque la fin d'une expression dotée du verbe et de la « perfection » du sens, ce qui peut être interprété comme une « phrase » (ou une « proposition autonome » dans notre terminologie). Aucune indication n'est donnée quant aux conditions d'emploi du *periodus*. Dans le texte du traité, il n'est utilisé qu'à la fin du dialogue (devant *Amen*), ce qui fait supposer qu'il est perçu comme une marque de « fin de section ». Les consignes de prononciation sont les mêmes pour le *colon* et pour le *periodus* : respecter une longue pause (*prolixius*

24 Il est étonnant de voir à quel point certaines théories du XX^e siècle sont proches de ces idées !

25 Cf. le classement des « frontières ponctuables » dans notre deuxième chapitre.

État des recherches sur la ponctuation

intervallum desiderant). L'usage du point d'interrogation n'est pas expliqué, mais il doit être facile à déduire du nom de la marque et d'un exemple cité.

Les parenthèses servent à marquer une expression insérée dans une structure commencée et non encore terminée (*ab incepta clausula nondum perfecta clauditur*). Cela correspond aux structures que nous analysons comme des incises, incidentes ou précisions non propositionnelles.

Avant de présenter les deux autres marques, le Maître du *Dialogue* précise que celles-ci ne servent pas à séparer les structures syntaxiques (*clausulas non recipiunt*). On dirait qu'elles relèvent de la ponctuation « du niveau lexical » (*cf.* la théorie des trois niveaux de la ponctuation).

Le *gemipunctus* sert à « mettre en valeur » les noms de personnes, de lieux ou de titres de noblesse. Il joue donc le même rôle que la majuscule du nom propre dans le système actuel. On peut également le rapprocher des points utilisés dans les manuscrits devant et après les chiffres et certaines abréviations.

Le *semipunctus*, enfin, est une marque de coupure de mots en fin de ligne.

Le système de ponctuation décrit dans le *Compendiosus dialogus* est donc orienté vers l'expression du sens du discours et s'interprète facilement en termes de structures syntaxiques. Les pauses et l'intonation évoquées à la fin ne sont que des marques utilisées dans un autre médium de communication linguistique pour remplir les mêmes fonctions. Cette vision de la ponctuation est en effet très proche de celle que nous adoptons dans notre modèle de description linguistique présenté dans le chapitre 2.

La pratique de la ponctuation dans les incunables français est néanmoins restée loin des principes logiques déclarés dans le traité de Heynlin. Même les imprimeurs de la Sorbonne ne semblent pas avoir rigoureusement suivi ces directives (Beltran 1985, p. 287), sans parler des autres ateliers qui se sont installés en France dans les dernières décennies du XV^e siècle. Nous verrons cela en analysant la ponctuation dans les deux incunables et l'imprimé de 1522 intégrés dans notre corpus.

1.1.3. Étienne Dolet et ses successeurs

Une vraie rupture dans les pratiques de la ponctuation est liée à l'activité d'Estienne Dolet (1509 – 1546) qui a publié son célèbre traité *De la ponctuation de la langue françoise* en 1540 à Lyon. Nous avons intégré dans notre corpus un fragment d'un autre traité de Dolet (*La maniere de bien traduire...*) publié dans le même livre et nous analyserons dans la section 8.4 la façon dont Dolet applique dans la pratique ses propres consignes.

Le traité de Dolet sur la ponctuation est très connu et souvent cité dans des ouvrages normatifs jusqu'à nos jours (Drillon 1991, par exemple)²⁶. Un bon résumé de sa théorie est proposé notamment par M.-L. Demonet (2000, p. 41-43). Nous nous limiterons donc ici à une présentation sommaire.

Dolet commence son traité en affirmant le caractère universel de la ponctuation²⁷ :

Si toutes langues generalement ont leurs differences en parler, *et* escripture, toutesfoys non obstant cela elles n'ont qu'une punctuation seulement : *et* ne trouueras, qu'en ycelle les Grecs, Latins, François, Italiens, ou Hespaignolz soient differents (Dolet 1540, p. 17).

L'inventaire des marques utilisées par Dolet se rapproche de celui dont dispose la typographie moderne. Pour les marques principales, Dolet cite les appellations utilisées par « les Grecs » et « les Latins », mais on ne sait pas sur quelles sources il s'appuie, car le vocabulaire des traités de ponctuation médiévaux était très instable (Hubert 1972). Apparemment, la confusion était toujours grande au XVI^e siècle, car Dolet consacre une vingtaine de lignes pour prévenir d'éventuelles accusations de mauvaise connaissance des « sources d'autorité ».

La première marque qu'il distingue, la *virgule* (ou *point à queue*) a la forme de la virgule moderne < , >. Dolet mentionne également une forme alternative de cette marque (un allographe dans notre terminologie) qui est représentée par une petite barre oblique en haut de ligne. Cette variante est qualifiée d'ancienne.

26 Un facsimilé (de qualité médiocre) de l'ouvrage de Dolet est accessible sur Gallica <<http://gallica.bnf.fr/ark:/12148/bpt6k106073c>>.

27 Nous appliquerons aux citations de Dolet dans cette section les règles d'édition correspondant à la transcription « diplomatique » (*cf.* section 4.4.2).

État des recherches sur la ponctuation

La deuxième marque est celle qu'on appellerait aujourd'hui « deux-points » « : ». Dolet mentionne le nom grec *comma* et dit que « les Latins ne luy ont baillé aultre nom. Mais il faut entendre, que toutes ces sortes de punctuer n'ont leur appellation, *et* nom à cause de leur forme, *et* marcque, ains pour leur effect, *et* propriété » (*ibidem*, p. 18).

La troisième marque est le point (ou *point rond*). Dolet remarque que la différence d'« effet » entre le point et le *comma* est minime, « sinon que l'ung (qui est comma) tient le sens en partie suspens,²⁸ Et l'aultre (qui est le colon) conclud la sentence »²⁹ (*ibidem*).

La quatrième marque est le point d'interrogation « ? », la cinquième est le *point admiratif* « ! ». Cette dernière marque est une innovation. Inventée par les humanistes italiens au XIV^e siècle, elle n'apparaît en France que dans les imprimés du milieu du XVI^e siècle (Parkes 1992, p. 49). Quant à son utilisation, Dolet donne la consigne suivante :

L'admiratif n'a si grand' uehemence : *et* eschet en admiration prosedante de ioye, ou de detestation de uice, *et* meschanceté faicte. Il conuient aussi en expression de soubhait, *et* desir. Brief : il peult estre par tout, ou il ya interiection (*ibidem*, p. 23).

On peut donc faire un rapprochement avec le *comma* médiéval qui était souvent utilisé après les interjections (Marchello-Nizia 2007), mais l'*admiratif* se met à la fin de l'unité « affectée » et non directement après l'interjection, comme c'était le cas du *comma*.

La dernière marque présentée est une paire de parenthèses qui servent, comme c'était déjà le cas dans le *Dialogus* de Heynlin, à introduire des incisives et des incidentes (notamment des gloses).

L'usage des marques est illustré par des exemples concrets. Dolet reprend la notion de *periode* (ou de *clausule*) des grammaires grecques et latines, sans la définir précisément. Il dit que le *periode* ne doit avoir que deux ou trois membres (qui sont à notre sens des propositions), car « si par sa longueur il excède l'alaine de l'homme, il est uicieux » (*ibidem*, p. 19). C'est en effet la seule indication « phonétique » dans le traité de Dolet, qui est de façon générale très « syntaxique ». Même s'il ne recourt pas

28 *Sic* : virgule devant une majuscule !

29 C'est cette affirmation que Dolet défend ensuite sur une vingtaine de lignes contre d'éventuelles attaques de détracteurs.

toujours à des termes grammaticaux dans ses explications, son usage est très fortement déterminé par les structures formelles et peut même être qualifié de « mécanique ». Par exemple, il ponctue très régulièrement les subordonnées et les syntagmes coordonnés et juxtaposés même si ceux-ci sont très courts et très étroitement liés à leur entourage.

Quant à l'usage que Dolet fait lui-même de la ponctuation dans son traité, on constate que la densité d'usage de ces signes de ponctuation (le « taux de ponctuation ») chez lui dépasse sensiblement celui des livres en français moderne. Nous reviendrons plus en détail sur ces chiffres dans la section 8.4.

Le traité de Dolet a dessiné en quelque sorte les contours des conventions de la ponctuation française moderne, ses idées ont été reprises par plusieurs grammairiens de la fin du XVI^e siècle et plus tard, même si son approche radicalement syntaxique a été progressivement « tempérée ».

Nous ne nous arrêtons pas sur ces aspects théoriques, car ils ne sont applicables ni aux données primaires de notre corpus ni aux pratiques des éditeurs scientifiques modernes et ils n'ont pas influencé l'élaboration du cadre théorique de notre recherche³⁰.

Nous passerons donc directement aux traités prescriptifs et descriptifs plus récents qui permettent d'apprécier la conception « normative » de la ponctuation française moderne et de comprendre pourquoi la ponctuation médiévale paraît si surprenante et désordonnée à un lecteur moderne.

1.1.4. Vision normative de la ponctuation au XX^e siècle

Nous commencerons notre survol des traités modernes de la ponctuation par un livre qui porte précisément ce titre : *Traité moderne de ponctuation* (Damourette 1939)³¹. Cet ouvrage porte une trace très nette du phonocentrisme qui régnait dans

30 Signalons tout de même qu'un bref résumé de ces théories est présenté dans (Tournier 1980), ainsi que dans (Catach 1994, p. 29-46).

31 Ce traité fait partie, selon C. Tournier (1980, p. 31), de la « première vague » d'ouvrages du XX^e siècle consacrés à la ponctuation et en est sans doute le représentant le plus connu et le plus intéressant.

État des recherches sur la ponctuation

certaines branches de la linguistique de la première moitié du XX^e siècle. L'écrit est perçu comme une représentation imparfaite de la richesse de la langue orale :

Si nous écoutons un Français parler, nous nous rendons compte de toutes les ressources que lui offrent la mélodie et la cadence de sa langue. Un texte écrit manque de toutes ces ressources, et le rôle de la ponctuation est de suppléer à ce manque en donnant des indications aussi précises que possible, permettant de reconstituer ce mouvement vivant de l'élocution orale. Il ne faut pas se dissimuler, d'ailleurs, que ces renseignements donnés par la ponctuation sont sommaires et même assez souvent insuffisants. C'est un défaut que la ponctuation partage avec l'ensemble de l'écriture, dont elle est une partie (Damourette 1939, p. 6-7).

Les signes de ponctuation sont classés par leur valeur à l'oral : signes paux (la virgule, le point et le point-virgule)³², signes mélodiques (le deux-points, les guillemets, les points d'interrogation et d'exclamation, les points de suspension, les parenthèses, les crochets, le tiret, le trait-d'union et le point d'ironie³³). L'astérisque, l'alinéa, les majuscules et les différents procédés de mise en relief typographique sont présentés comme des moyens accessoires et n'occupent que quelques pages dans le traité. Il convient de noter que dans le classement des emplois des différents signes (de la virgule, en premier lieu) Damourette prend néanmoins en compte les structures syntaxiques.

Il a fallu attendre plusieurs décennies pour qu'une « nouvelle vague » de recherches et d'ouvrages entièrement consacrés à la ponctuation voient le jour en France. Nous aborderons les recherches scientifiques correspondantes dans la section suivante et ne mentionnerons ici que des traités sans ambition théorique qui concernent les problèmes pratiques de la « ponctuation efficace » de l'écrit quotidien (Colignon 1975 ; 272 Doppagne 1978 ; 223 Drillon 1991). Les deux premiers des ouvrages cités ne dépassent d'ailleurs pas une centaine de pages chacun.

Le *Traité de la ponctuation française* de Jacques Drillon est un ouvrage plus ambitieux et sans doute le plus connu. Il s'adresse au grand public et ne prétend pas être

32 Ces signes ne sont pas pour autant privés d'une certaine valeur mélodique (Damourette 1939, p. 54).

33 Il s'agit d'une marque imaginée par Alcanter de Brahm et que Damourette juge comme « un signe superflu, et qui n'a point réussi ».

le fruit d'une recherche scientifique. Son objectif est de raconter au lecteur, sous une forme captivante, l'histoire des idées sur la ponctuation (avant tout, chez les écrivains et les grammairiens français du XVIII^e siècle à nos jours) et de donner des conseils pratiques pour l'usage des signes de ponctuation basés sur les « bonnes pratiques » des écrivains et appuyés par de nombreux exemples.

La description de chaque signe débute par une série de définitions proposées par les grands grammairiens (à commencer par Dolet), viennent ensuite les différentes fonctions et cas d'usage particuliers (où il faut et où il ne faut utiliser une marque donnée) présentés sans tentative de classement logique. Les conventions typographiques (emploi d'espace blanc avant et après la marque) ne sont pas oubliées. La plus grande section est consacrée à la virgule, marque dont l'usage peut être motivé par une multitude de facteurs syntaxiques, sémantiques, pragmatiques, stylistiques et conventionnels et qui génère le plus grand nombre d'hésitations dans la pratique de l'écrit.

Si on essaie de dégager la conception théorique de la ponctuation qui constitue le fondement des consignes de Jacques Drillon, on peut parler d'une vision plutôt autonomiste (la ponctuation est propre à la langue écrite et ne sert pas forcément à « traduire » les phénomènes de l'oral) et orientée vers les fonctionnalités conventionnelles (présenter d'une façon régulière et uniforme les structures linguistiques similaires) et stylistique (utiliser le potentiel expressif des marques pour augmenter l'effet que le texte produit sur le lecteur).

Même s'il ne s'agit pas d'une recherche linguistique, le traité de Jacques Drillon constitue grâce à ses nombreux exemples un excellent recueil des pratiques de ponctuation classiques et modernes et permet d'apprécier le potentiel sémantique et stylistique des différentes marques de ponctuation.

Parmi les grammaires traditionnelles du français qui font à ce jour le plus autorité, *Le bon usage* de Maurice Grevisse (2007)³⁴ s'impose comme une référence incontournable. La dernière édition que nous avons consultée inclut les références de

34 14^e édition. La première édition date de 1936, le texte a été refondu par André Goosse à partir de la 12^e édition de 1986.

État des recherches sur la ponctuation

certaines travaux linguistiques récents, mais ne semble pas tenir compte de certaines avancées des connaissances³⁵.

La ponctuation est définie comme « l'ensemble des signes conventionnels servant à indiquer, dans l'écrit, des faits de la langue orale comme les pauses et l'intonation, ou à marquer certaines coupures et certains liens logiques » (Grevisse & Goosse 2007, p. 121). Il s'agit donc d'une conception mixte, phonocentrique au départ, mais reconnaissant l'existence de fonctions propres à l'écrit. Dès le premier paragraphe, l'accent est mis sur l'importance sémantique de certaines marques de ponctuation et sur les effets stylistiques utilisés par certains écrivains. Les signes de ponctuations sont énumérés³⁶ et ensuite présentés un par un. Les blancs et les alinéas sont également traités dans la même section, mais ils ne sont pas considérés comme des signes de ponctuation³⁷. A chaque fois, on formule brièvement la fonction (ou la valeur) principale de la marque et on discute ensuite en détail des cas particuliers et des affinités de l'usage. Selon les marques, la fonction principale peut être « logique » (comme le marquage de « fin d'une phrase interrogative » pour le point d'interrogation³⁸) ou « prosodique » (« une pause de peu de durée » pour la virgule). Tous les emplois d'une marque, parfois très hétérogènes du point de vue linguistique, sont présentés sur le même plan. Par exemple, parmi les fonctions du point, sont mentionnés le marquage d'abréviations, la séparation entre le jour, le mois et l'année dans les dates (par exemple, 28. 5. 1999) et la connexion entre les titres et les numéros de pages dans les tables de matières (« points de conduite »).

La présentation de la ponctuation dans *Le bon usage* ne s'appuie donc pas sur un modèle théorique précis et ne suit pas un schéma méthodologique rigoureux. Elle a néanmoins le mérite de décrire d'une façon précise les cas de difficultés d'application des règles de ponctuation du français moderne et peut être utile dans l'explication de la ponctuation dans les éditions scientifiques des textes médiévaux.

35 Ainsi la remarque selon laquelle l'usage de la ponctuation dans les textes ne remonte qu'au IX^e siècle et qu'il était fort irrégulier jusqu'au XVI^e siècle (Grevisse & Goosse 2007, p. 122, note historique) n'a pas été retirée malgré la critique de Nina Catach (1994, p. 13) pourtant citée dans les références.

36 Même le *pied-de-mouche* médiéval est mentionné dans une note.

37 Il convient de noter que l'essentiel des remarques concernant l'usage du blanc consiste à présenter les règles typographiques de son usage avant et après les signes de ponctuation.

38 On sait qu'il n'y a pas en français d'intonation interrogative universelle : tout dépend de la présence de morphèmes interrogatifs dans la phrase.

La deuxième (et la dernière) grammaire « générique » du français dont il nous semble indispensable de tenir compte pour avoir une idée de la perception de la ponctuation par des linguistes qui ne sont pas engagés dans tel ou tel camp de conception théorique est la *Grammaire méthodique du français* (Riegel *et al.* 1996). Cet ouvrage vise en effet à donner une description scientifique de la grammaire du français tenant compte des avancées des différentes branches de la linguistique sous une forme accessible à un public plus large que la communauté des chercheurs et des étudiants en linguistique. Même si elle peut être critiquée sur certains points, la *Grammaire méthodique* est un succès indéniable, et elle est devenue une référence incontournable. C'est d'ailleurs sur les catégories et les classements proposés dans cet ouvrage que nous avons basé notre jeu d'étiquettes de « frontières ponctuables » décrit dans la section 2.6.

Voici la définition de la ponctuation qui y est proposée :

La ponctuation est le système des signes graphiques qui contribuent à l'organisation d'un texte écrit et qui apportent des indications prosodiques, marquent des rapports syntaxiques ou véhiculent des informations sémantiques (Riegel *et al.* 1996, p. 83).

Cette définition s'inspire en effet des travaux de N. Catach (que nous aborderons dans la section suivante), mais est formulée dans des termes moins restrictifs et met en avant la « fonction prosodique », sous l'influence de la tradition phonocentriste.

Les trois fonctions de la ponctuation sont ensuite présentées d'une façon plus détaillée. Ainsi, la fonction prosodique se subdivise en marquage des pauses et en indication du rythme et de l'intonation. La fonction syntaxique opère à trois niveaux : séparation des mots, délimitations intraphrastiques et délimitations interphrastiques. La fonction sémantique consiste en une indication modale, en une analyse sémantique de la phrase (séparation en thème et propos) et en une structuration en plans d'énonciation distincts. Les signes de ponctuation sont ensuite présentés par catégories correspondant à leur fonction principale. Le trait-d'union et l'astérisque sont classés dans la catégorie « autres », et les signes typographiques (l'alinéa et les variations de caractères) sont traités à part.

État des recherches sur la ponctuation

Il s'agit en somme d'une présentation éclectique qui mêle plusieurs approches théoriques et cache des contradictions importantes derrière une apparence logique et structurée.

Cela est sans doute inévitable lorsqu'on essaie de trouver un « dénominateur commun » aux différentes théories linguistiques. Le cadre théorique que nous avons adopté pour notre analyse de la ponctuation médiévale n'est d'ailleurs pas exempt de ce défaut, car nous avons cherché également à dépasser les clivages théoriques et à décrire les données observées en prenant en compte les multiples facteurs qui peuvent entrer en jeu.

Quoi qu'il en soit, le modèle de la description de la ponctuation présenté dans la *Grammaire méthodique* n'est pas applicable « tel quel » aux données de notre corpus pour plusieurs raisons. Premièrement, il est difficile d'étudier les fonctions prosodiques sur les états de langue pour lesquels nous ne disposons pas d'accès direct à la production orale. Deuxièmement, la notion de phrase qui joue le rôle central dans la définition des fonctions syntaxiques est difficilement applicable au français médiéval (*cf.* la section 2.5). Troisièmement et enfin, la démarche qui consiste à partir de la forme de la marque pour décrire ensuite ses fonctions nous paraît inadéquate par rapport aux pratiques de la ponctuation médiévales.

Nous avons donc été amené à développer notre propre « théorie » basée sur la notion de l'« unité ponctuable » et une typologie de frontières entre celles-ci (voir le chapitre 2). Bien entendu, ce cadre théorique ne s'est pas construit *ex nihilo*. Il se base dans la mesure du possible sur les notions consensuelles de la grammaire traditionnelle, mais aussi sur les théories scientifiques de la ponctuation et de la langue écrite en général les plus connues dans la communauté linguistique française et surtout sur les recherches linguistiques consacrées à la ponctuation française médiévale. Ces deux « piliers » de notre analyse linguistique seront présentés dans les sections qui suivent.

1.2. Théories « scientifiques » de la ponctuation

Il convient de signaler avant tout les travaux de synthèse consacrés aux théories de la ponctuation actuelles. Un grand chapitre du livre de S. Pétillon-Boucheron (2002,

p. 13-75) fait le point sur les différentes théories en classifiant les approches et en présentant les structures terminologiques sous la forme de schémas graphiques.

Dans un article d'un numéro thématique de l'*Information grammaticale*, Jacques Anis (2004) a confronté les théories scientifiques et les points de vue des grammaires traditionnelles tout en défendant sa propre théorie (même si sa position semble moins radicale que dans les publications antérieures). Dans ce même numéro, I. Serça (2004) adopte une perspective plus large et présente les théories linguistiques à côté des approches littéraires et artistiques.

Sans entrer dans le détail des théories, nous présenterons les concepts clés des deux principales approches qui s'opposent dans la linguistique française actuelle et chercherons à comprendre dans quelle mesure ces concepts sont applicables dans le cadre de notre recherche.

1.2.1. Théories tenant compte des rapports entre l'oral et l'écrit

La ponctuation et plus généralement la langue écrite en tant que telle n'ont pas été au centre d'intérêt des grands courants théoriques de la linguistique du XX^e siècle. L'écrit était considéré comme une représentation secondaire et imparfaite de l'activité langagière. Des traces de ce phonocentrisme sont très perceptibles dans le traité de Jacques Damourette (1939) dont nous avons déjà parlé, mais le plus souvent la ponctuation n'intéressait tout simplement pas les linguistes.

Un certain regain d'intérêt est lié au développement des premières méthodes du traitement automatique du langage naturel (TALN) dans les années 1960, mais il n'a pas amené à la formation d'une théorie de la ponctuation au sens propre.

En France, le développement de la théorie de l'écrit en général et de la ponctuation en particulier est lié à l'activité du groupe de recherche animé par Nina Catach dans les années 1970 et 1980. Des tables rondes internationales consacrées à l'analyse linguistique de l'orthographe (1973), de la ponctuation (1978) et à la théorie de la langue écrite (1986) ont été organisées par cette équipe, et les publications préparatoires et subséquentes à ces réunions scientifiques (par exemple, Catach 1979 ;

État des recherches sur la ponctuation

1988) ont constitué une avancée très importante dans le développement de la théorie de la ponctuation. Un numéro de la *Langue française* coordonné par Nina Catach (1980) est particulièrement riche en contributions significatives (surtout les articles de Nina Catach, de Claude Tournier et de Ludmilla Védénina).

L. Védénina a sans doute été la première à soulever le problème de la théorie linguistique de la ponctuation en présentant une communication à la table ronde de 1973. La même année, elle publie un article consacré au rôle de la ponctuation dans les rapports entre l'oral et l'écrit (Védénina 1973). En s'appuyant sur des données expérimentales, la chercheuse confirme l'existence des rapports entre la ponctuation et la prosodie, mais insiste sur leur caractère complexe. Dans son article de 1980, L. Védénina présente sa vision de la triple fonction de la ponctuation (syntaxique, communicative et sémantique) inspirée de la distinction des trois plans linguistiques par Ch. Bally. Chaque signe de ponctuation peut selon les cas remplir une ou plusieurs de ces fonctions à la fois en interaction avec les autres marques linguistiques, comme l'ordre des mots, les conjonctions et d'autres mots grammaticaux. Le problème de la relation entre l'écrit et l'oral n'a plus de grande importance dans cette perspective d'analyse fonctionnelle.

Dans la première partie de son livre paru en 1989, la chercheuse étudie de façon détaillée chacun des signes de ponctuation traditionnels, en distinguant les signes uni- et bilatéraux et en définissant leur valeur principale (modale, disjonctive, etc.). La deuxième partie est consacrée aux différentes techniques typographiques de la présentation du texte. La troisième et dernière partie traite des rapports de la ponctuation avec la syntaxe, la sémantique et l'intonation. Il convient de noter que l'intonation n'est considérée qu'en dernier lieu et n'occupe que quatre pages de l'ouvrage.

C. Tournier (1980), après avoir passé en revue les principales approches de la ponctuation dans la tradition linguistique française du XV^e au XX^e siècle, introduit plusieurs définitions et distinctions importantes. Ainsi la ponctuation est définie de la façon suivante :

La ponctuation est l'ensemble des graphèmes purement plérémiques (unités dotées d'un sens, selon la terminologie hjelmslevienne – *A.L.*), non décomposables en unités de rang inférieur, et à caractère discret (*ibidem*, p. 36).

Cette définition a l'avantage de s'inscrire dans un système terminologique précis et d'éviter le recours à des notions floues et intuitives³⁹. Son adoption nécessite tout de même l'adhésion à un cadre théorique particulier. Elle ne permet par ailleurs pas de distinguer les signes de ponctuation des signes logographiques, comme les marques d'abréviation, très fréquentes dans les manuscrits médiévaux.

Dans la suite de son article, C. Tournier propose de distinguer les signes *liés* (comme par exemple les majuscules) des signes *autonomes* (les signes de ponctuation au sens traditionnel) ; les signes *ponctuels* des signes *linéaires* (comme les italiques). Le chercheur établit un classement de 4 catégories des signes de ponctuation : (1) la ponctuation de mot, (2) la ponctuation de phrase, (3) la ponctuation métaphrastique et (4) la ponctuation spécifique. La dernière sert à signaler au lecteur le caractère particulier de certains mots ou syntagmes (par exemple, la mise en italique de mots étrangers). Il formule enfin trois « lois d'occurrences » qui ont été reprises par plusieurs chercheurs par la suite. Il s'agit de la « loi d'exclusion » (la majuscule d'un nom propre « exclut » la majuscule de début de phrase), la « loi de neutralisation » (le point d'abréviation « neutralise » le point de fin de phrase)⁴⁰ et la « loi d'absorption » (l'apostrophe « absorbe » le trait d'union dans *chef-d'œuvre*).

Des trois auteurs évoqués au début de la section, c'est pourtant N. Catach qui a développé la théorie de l'écrit et de la ponctuation la plus complète. Cette théorie a subi plusieurs ajustements au cours des années, et nous nous référerons à son dernier état résumé dans un « Que sais-je » sur la ponctuation. Voici ce qui y est dit sur les rapports entre l'oral et l'écrit :

Selon nous, l'oral et l'écrit sont, pour l'adulte cultivé moderne, deux médias différents, ayant des rapports privilégiés, mais égaux et relativement indépendants, présentant chacun leurs caractéristiques propres et complémentaires. Ils participent tous deux

39 C'est pour cette raison qu'elle a été adoptée comme point de départ par N. Mazziotta (2007b, p. 54).

40 Il convient de noter toutefois que la différence entre l'exclusion et la neutralisation ne ressort pas clairement des définitions et des exemples cités.

État des recherches sur la ponctuation

à l'évolution et l'enrichissement de la langue, appelée L, de telle sorte qu'elle se transforme sous l'influence de l'écriture et acquiert de nouvelles capacités (nous l'appelons alors L Prime). De plus... nous considérons que l'écrit, comme l'oral, présente sous certains aspects un signifiant et en signifié, d'une part en tant que processus d'encodage et de décodage directs (média qui devient *premier*), d'autre part par certaines de ses parties constitutives et de ses unités ("plérèmes"), ce qui est le cas pour la ponctuation (Catach 1994, p.97).

La ponctuation « au sens large » est définie comme :

Système de renfort de l'écriture, formé de signes syntaxiques, chargés d'organiser les rapports et la proportion des parties du discours et des pauses orales et écrites. Ces signes participent ainsi à toutes les fonctions de la syntaxe, grammaticales, intonatives et sémantiques (*ibidem*, p. 7).

Le terme de *ponctème* est introduit pour désigner le signe de ponctuation en tant qu'unité de première articulation (selon Martinet) de la langue présentant à la fois un signifiant et un signifié.

En parlant des fonctions du ponctème, N. Catach établit constamment un parallèle avec l'*intonème* de l'oral et va même jusqu'à dire qu'il s'agit dans les deux cas « d'unités fondamentalement suprasegmentales et syntaxiques » (Catach 1994, p. 105). Ce point suscite une critique de N. Mazziotta (2007b, p. 53) qui accuse la chercheuse de « confondre les propriétés distributionnelles de la ponctuation et celles de l'intonation sur la base d'une ressemblance fonctionnelle ». Il remarque, par ailleurs, à juste titre que le qualificatif de *suprasegmental* est difficilement applicable aux signes de ponctuation⁴¹.

A notre avis, la complexité de la théorie de N. Catach reflète la complexité des rapports réels qui existent entre les différents médiums d'expression et le système « abstrait » de la langue. Faire l'économie de ces rapports et considérer l'écrit comme un système entièrement autonome permet de créer un modèle plus simple et logique (comme nous le verrons dans la section suivante), mais un tel modèle risque d'être inadéquat et de ne pas permettre d'expliquer les phénomènes observés dans des corpus.

Le problème avec la théorie de N. Catach est qu'elle ne permet pas non plus de toujours expliquer les données observées, surtout lorsqu'il s'agit des usages

41 Notons que le terme de *graphème suprasegmental* apparaît également chez J. Anis (1983, p. 33) avant d'être abandonné dans les publications ultérieures.

« prénormatifs » des manuscrits médiévaux. On voit mal comment la notion de « L prime » peut être utile pour analyser les cas d'usage (et d'absence !) de marques de ponctuation à certains points de la chaîne textuelle de nos sources primaires.

1.2.2. Approche autonomiste

Une autre approche théorique de la langue écrite, la *graphématique autonome*, a été développée en France dans les travaux de Jacques Anis (1983 ; 1988 ; Anis *et al.* 1988). Le chercheur défend une conception entièrement autonomiste de l'écrit dans la lignée des travaux de J. Vachek et de E. Pulgram. La langue existe, selon ce point de vue, « sous deux formes, entre lesquelles la linguistique ne postule ni hiérarchie ni dépendance » (Anis 1988, p. 214). Il reconnaît néanmoins que « la graphématique autonome n'épuise pas l'analyse de la langue écrite : un certain nombre de phénomènes, par exemple certaines utilisations expressives de la ponctuation, nécessitent la mise en relation avec la phonie » (*ibidem*, p. 220).

Le système de l'écrit est composé de trois types de *graphèmes* : les *alphagrammes* (les lettres, pour simplifier), les *topogrammes libres* (les signes de ponctuation et les blancs) et *liés* (majuscules et autres variantes de caractères), ainsi que les *logogrammes* (les chiffres, les sigles, les logos, etc.).

La définition précise du topogramme est donnée dans le *Glossaire* :

graphème punctuo-typographique, qui contribue à la production du sens, en tant qu'organisateur de la séquentialité et indicateur syntagmatique et énonciatif (Anis *et al.* 1988, p. 246).

Cette définition est en quelque sorte circulaire, car elle renvoie à la notion de ponctuation, définie elle-même comme un système de topogrammes. L'introduction de l'aspect typographique n'est pas non plus suffisamment justifiée (*cf.* Mazziotta 2007b, p. 54).

Deux fonctions des topogrammes sont identifiées : la fonction syntagmatique (organisation syntaxique « remodelée » par la dynamique communicationnelle) et la fonction « polyphonique » qui regroupe le marquage de la modalité, du discours rapporté, la hiérarchisation discursive et le marquage expressif.

État des recherches sur la ponctuation

J. Anis introduit une distinction importante liée à la *portée* des topogrammes :

La règle générale est que la portée d'un topogramme s'arrête là où intervient un topogramme de même niveau ou de niveau supérieur, mais pour certaines unités – par exemple la capitale initiale de la phrase –, on va *de gauche à droite* (c'est l'auteur qui surligne – *A.L.*), pour d'autres – par exemple le point –, on va de droite à gauche, pour d'autres encore – par exemple la virgule –, les deux sens sont utilisés (Anis *et al.* 1988, p. 121).

Ce système sophistiqué reposant sur la hiérarchie et sur la « direction » des marques est malheureusement difficilement applicable à l'analyse des manuscrits médiévaux : les marques ne sont pas toujours clairement hiérarchisées, et la direction de leur portée n'est pas toujours évidente. Nous avons donc préféré adopter pour nos recherches une division « plate » en « unités ponctuables », et nous considérons que, sauf évidence contraire, la portée des marques est toujours orientée *à droite*⁴². Une certaine hiérarchisation et directionnalité sont introduites grâce aux concepts de la *force de la marque* et du *type de la frontière* ponctuable (*cf.* les sections 2.6 et 2.7.2 du chapitre suivant).

Le fait d'affirmer l'autonomie totale⁴³ de l'écrit par rapport à l'oral n'a pas forcément une grande incidence sur l'analyse pratique des données des sources primaires, surtout lorsqu'il s'agit de la période pour laquelle les données directes sur l'oral sont absentes. Pour le reste la théorie de J. Anis n'est pas fondamentalement incompatible avec celle de N. Catach, une fois les différences terminologiques apparentes évacuées. Les innovations terminologiques de J. Anis sont certes intéressantes, mais l'absence de rigueur dans les définitions sème le doute en ce qui concerne leur utilité par rapport aux termes traditionnels.

Une autre théorie autonomiste centrée spécifiquement sur la ponctuation demande à être évoquée ici. Il s'agit de *la linguistique de la ponctuation* de G. Nunberg (1990). Même si elle se situe en dehors du domaine de la linguistique française et

42 Il ne s'agit bien entendu que d'un principe méthodologique préalable à l'analyse fine. Les marques comme le point d'interrogation ou la parenthèse droite (pratiquement absentes de notre corpus) portent évidemment à gauche.

43 Avec toutefois des réserves, comme nous l'avons vu.

s'appuie majoritairement sur les données empiriques de l'anglais, il s'agit d'une théorie majeure qui a influencé les recherches en France, surtout dans le domaine du TALN.

Le chercheur défend une autonomie absolue du système des « indicateurs textuels » qui n'entretient avec l'oral que des liens historiques :

From the historical perspective, the point can be made by saying that the system of text-category indicators is an *application* of natural language. By this I mean the system has arisen through a kind of application of the principles of natural-language design, as a means of imposing a certain organization on the lexical content of written texts. Viewed as a whole, the system has no analogue in the spoken language (though certain of its functions overlap with those of various spoken language devices). Rather, the system has arisen, in the course of written-language standardization and intellectualization, as a response to the particular communicative requirements of written language texts, and as an exploitation of the particular expressive resources that graphical presentation makes available (*ibidem*, 1990, p. 7)⁴⁴.

L'ambition principale de G. Nunberg consiste à proposer un modèle formel de la « génération » de la structure textuelle et de sa représentation par des marques de ponctuation. La *grammaire textuelle* ainsi décrite est opposée à la *grammaire lexicale* commune à la langue orale et à la langue écrite (*cf.* notre conception des marques segmentales, suprasegmentales et des « balises graphiques » présentée dans les sections 2.1 et 2.2). La grammaire textuelle est dans une grande mesure indépendante de la grammaire lexicale. Comme la grammaire lexicale, elle se base sur la « connaissance tacite » (*tacit knowledge*) : la plupart de ses caractéristiques intéressantes ne sont pas accessibles à la « réflexion banale » (*casual reflection*) et ne sont pas décrites dans les manuels d'écoliers (*ibidem*, p. 23).

44 Notre traduction : « Dans une perspective historique, on peut faire le point en disant que le système des indicateurs des catégories textuelles est une *application* de la langue naturelle. Je veux dire par cela que le système s'est développé à travers une forme d'application des principes de la conception de la langue naturelle, en tant que moyen d'imposer une certaine organisation au contenu lexical dans les textes écrits. Dans son ensemble, le système n'a pas d'équivalent dans la langue orale (même si certaines de ses fonctions se superposent avec celles des différents outils de la langue orale). En effet, ce système s'est développé au cours de la standardisation et de l'«intellectualisation» de la langue écrite, comme une réponse aux exigences communicative particulières aux textes écrits et comme une exploitation des ressources expressives particulières que la présentation graphique rend accessibles ».

État des recherches sur la ponctuation

Les unités centrales de la structure textuelle sont pour G. Nunberg le paragraphe et la *phrase textuelle*, qui correspondent aux catégories structurelles de base dans la représentation de la structure argumentale de l'interprétation du texte (*ibidem*, p. 25).

Le fonctionnement de la ponctuation est décrit comme l'application d'une série de règles de *linéarisation* (*linearization*) et de *versation* (*pouring*). Les premières concernant la projection des structures syntaxiques dans les séquences linéaires de caractères et des analogies, ne serait-ce que grossières, peuvent être établies entre ces règles et les règles phonologiques. Les règles de versation n'ont pas d'équivalence à l'oral. Il s'agit de la projection des séquences linéaires de caractères dans un espace bi-dimensionnel de la page du texte.

A la fin de son ouvrage, G. Nunberg consacre quelques pages à l'aspect historique de la formation de la « grammaire textuelle ». Il reconnaît que son système est inapplicable aux textes anciens et que l'unité « phrase textuelle » ne fait son apparition que dans les imprimés du XVII^e siècle. L'imprimerie a d'ailleurs joué à son avis un rôle primordial dans la formation de cette grammaire. L'absence de « phrases » dans les textes anciens ne signifie pas que la langue et la pensée était « désorganisées » à l'époque. Il s'agit simplement d'une autre organisation qui reste à découvrir. Le chercheur se prononce d'ailleurs contre les pratiques de la « modernisation » de la ponctuation dans les éditions scientifiques des textes anciens : cette pratique entraîne nécessairement l'altération du sens et repose sur une « fausse présupposition » selon laquelle la ponctuation moderne marque les unités d'information d'une façon neutre et qu'elle est applicable à n'importe quel texte qui est organisé d'une manière cohérente sur des bases indépendantes (*ibidem*, p. 131). Il exprime enfin le souhait que l'histoire de la « grammaire textuelle » de l'anglais ou « d'une autre langue développée » soit étudiée afin de révéler les influences externes qui ont « façonné » son développement (*ibidem*, p. 132).

La théorie de G. Nunberg apporte d'importants éléments nouveaux dans la modélisation des mécanismes linguistiques qui régissent l'emploi des marques de ponctuation. La notion de *phrase textuelle* rappelle les « unités de lecture » imaginées par Ch. Marchello-Nizia pour la langue médiévale (*cf. infra*), et notre concept d'« unité ponctuable » n'en est pas loin. L'application de cette théorie demande toutefois une

adhésion à des postulats méthodologiques discutables, et elle n'est de toute façon pas adaptée aux données « prénormatives » des manuscrits médiévaux et des incunables.

La seule tentative, à notre connaissance, d'application systématique d'un cadre méthodologique rigoureux à l'analyse de la ponctuation médiévale a été réalisée par Nicolas Mazziotta (2007b). Tout en étant conscient des différents facteurs qui peuvent intervenir dans la décision du scribe de mettre (ou de ne pas mettre) une marque de ponctuation à un endroit donné de la chaîne textuelle, le chercheur « fait le pari » de la syntaxe en affirmant que « la syntaxe, si elle ne peut tout expliquer, suffit à décrire la grande majorité des signes rencontrés » (*ibidem*, p. 12). Sa démarche consiste (1) à définir le corpus employé, (2) à définir exactement les termes de *ponctuation* et de *syntaxe*, (3) à décrire le système graphique de la façon la plus indépendante possible de la syntaxe (et inversement) et (4) à évaluer les corrélations entre les deux systèmes décrits.

Nous ne nous arrêterons pas sur la théorie de la syntaxe, innovante à plusieurs égards, développée N. Mazziotta (elle mérite certainement une analyse et une discussion beaucoup plus importantes que celles que nous pouvons développer dans ces pages)⁴⁵, mais tenterons de résumer en quelques lignes les points cardinaux de sa théorie de la langue écrite.

En se positionnant dans un cadre résolument autonomiste, N. Mazziotta définit la langue écrite comme une langue à part entière caractérisée par son médium spatial (*ibidem*, p. 36). En s'inspirant du système terminologique hjelmslevien, il forge toute une hiérarchie de termes définis par des critères de plus en plus fins. Ainsi, l'unité linguistique minimale de la langue écrite est nommée *scriptème* (propriétés : unité linguistique minimale et médium spatial). Les scriptèmes sont divisés en *grammèmes* (non organisateur spatial) et *topèmes* (organisateur spatial). Sur un critère fonctionnel, les topèmes sont classés en *plérétopes* (significatif) et *cénétopes* (distinctif). Le même critère permet de subdiviser les grammèmes en *plérégrammes* et *cénégrammes*. Il convient de noter que « minimal » dans la définition du scriptème ne signifie pas « non décomposable » sur le plan segmental ou purement analytique. Ainsi un plérégramme

45 Nous reviendrons tout de même, dans la section 2.5, sur sa définition de la phrase.

État des recherches sur la ponctuation

« conu » est décomposable en cénégrammes « c », « o », etc. Les topèmes font également leur réapparition à un certain niveau de la décomposition analytique des grammèmes. Au bout de la chaîne terminologique (que nous désignerons par une succession de barres obliques), un *ponctogramme* est défini comme un scriptème / grammème / plérégramme / *autogramme* (indépendant, axe de combinaison unique) / *nébulogramme* (topème de construction non orientée) avec la propriété définitoire « non paraphrasable » (*ibidem*, p. 51).

Ce dispositif de définitions logique mais très lourd permet en fin de compte de retrouver des équivalences parmi les termes « flous » de la grammaire traditionnelle. Ainsi, un *ponctogramme* correspond à un *signe de ponctuation* et un *linéogramme* correspond à une *lettre*. Notons que la « variante majuscule » d'une lettre sera considérée comme un *hétérogramme* et n'est pas assimilée à des ponctogrammes.

L'ambition d'utiliser un appareil terminologique fondé sur des définitions précises et non contradictoires ne peut être que saluée. On peut néanmoins se poser la question de savoir si le système très complexe et abondant en néologismes obtenu constitue une réelle avancée dans les connaissances et s'il est susceptible de favoriser la dissémination des résultats des recherches dans la communauté scientifique.

Ceci dit, le travail accompli par N. Mazziotta est sans aucun doute novateur et convaincant. La confrontation de l'annotation syntaxique exhaustive avec les emplois de la ponctuation sur un corpus de dimensions importantes lui a permis de démontrer statistiquement les affinités entre la ponctuation et certaines structures syntaxiques. Cette piste de recherche mérite certainement d'être poursuivie sur un corpus plus diversifié en prenant en compte différentes variables de la typologie textuelle.

La démarche que nous avons adoptée et qui consiste à tenir compte de critères supplémentaires, en plus de la structuration syntaxique, et à utiliser dans la mesure du possible l'appareil terminologique consensuel de la grammaire traditionnelle (avec tous ses défauts, mais aussi l'avantage d'être intuitif et accessible à une plus grande communauté scientifique) ne nous semble pas contradictoire, mais complémentaire par rapport aux recherches de N. Mazziotta, qui nous amènent d'ailleurs au cœur de la problématique de la ponctuation médiévale.

1.3. Ponctuation médiévale française : état des recherches

Comme nous l'avons déjà indiqué, les excellentes synthèses (Rafti 1988) et (Llamas Pombo 2008) nous permettent de faire l'économie d'un compte-rendu détaillé des recherches entreprises et des résultats obtenus dans le domaine de la ponctuation médiévale européenne. Nous ne nous arrêterons que sur les travaux qui présentent un intérêt particulier dans l'optique de l'analyse linguistique que nous avons adoptée dans notre travail. Ni les recherches sur les théories de la ponctuation (brièvement évoquées au début du présent chapitre), ni les aspects purement paléographiques (l'histoire de l'apparition d'une forme particulière d'un signe de ponctuation) ne concernent directement notre recherche (même si nous avons fait de notre mieux pour décrire correctement et tracer l'histoire générale des marques attestées dans nos sources). Nous nous intéressons aux unités linguistiques que les marques de ponctuation permettent de mettre en évidence et de structurer. Ce sont donc avant tout les travaux de linguistes consacrés à l'analyse des pratiques de la ponctuation du « niveau syntaxique » qui seront évoqués ci-dessous. Les recherches sur la « ponctuation des mots » et sur la « mise en page » ne seront abordées que dans la mesure où ces « niveaux » de la ponctuation interagissent avec celui qui nous intéresse.

Les pratiques de la ponctuation dans les manuscrits français médiévaux n'ont intéressé qu'un petit nombre de paléographes et de philologues avant le dernier quart du XX^e siècle. Les recherches menées étaient essentiellement consacrées aux traités de ponctuation et à l'histoire de l'apparition de telle ou telle marque.

L'article de Mario Roques (1952) où il procède à un examen minutieux des habitudes graphiques du célèbre scribe Guiot fait partie des rares exceptions. La ponctuation de Guiot est effectivement remarquable, d'autant plus qu'il s'agit d'un texte en vers. M. Roques distingue les emplois qui servent à « donner de la netteté à la présentation et à la lecture du texte » (comme des points de fin de page, points d'abréviations ou de chiffres, points de « coupure de vers », etc.) des ponctuations « à

État des recherches sur la ponctuation

valeur expressive » (« points exclamationnels »⁴⁶, points « rythmiques » au milieu d'un vers, points de fin d'enjambements, etc.). A la fin de l'article le philologue établit quelques parallèles entre la ponctuation de Guiot et celle de certains manuscrits plus anciens⁴⁷ et exprime le souhait qu'une étude générale soit menée sur les essais ou les habitudes de ponctuation du Moyen Âge :

Le scribe Guiot mériterait ainsi une place dans l'histoire de la ponctuation française, qui reste encore à écrire et qui devra tenir compte non seulement de la nature et de la valeur des signes, mais aussi des habitudes et des nécessités de pensée, de lecture, de diction, auxquelles ils correspondent, et qui ont varié suivant les époques, les orateurs et les lecteurs.

Il serait bien souhaitable que quelque jeune érudit entreprit l'étude des essais ou des habitudes de ponctuation du moyen âge selon un ordre chronologique et en examinant parallèlement manuscrits liturgiques et manuscrits littéraires, textes latins et textes français (*ibidem*, p. 196).

Ce n'est qu'un demi-siècle plus tard qu'un tel projet commence à se réaliser, même si la ponctuation médiévale est encore loin d'avoir livré tous ses secrets.

Un véritable travail pionnier d'analyse linguistique de la ponctuation médiévale a été réalisé par Christiane Marchello-Nizia (1978). A partir d'un corpus composé de 6 versions (5 manuscrits du XV^e siècle et un incunable) du chapitre VII du *Jouvencel* de Jean de Bueil elle a cherché à dégager les structures linguistiques (les « unités de lecture ») que les marques de ponctuation servent à démarquer.

La chercheuse commence par présenter l'inventaire complet des marques attestées, en soulignant le lien étroit entre les signes de ponctuation et les majuscules. Elle fait une remarque très juste sur la difficulté d'appliquer la notion de phrase aux manuscrits médiévaux (*cf.* notre section 2.5) et préfère baser son découpage syntaxique sur les *propositions* « centrées autour d'un verbe ». Une première estimation montre qu'à l'exception d'un seul manuscrit, la grande majorité des marques (entre 89% et 98%) apparaissent au début d'une proposition. La linguiste remarque ensuite que seule

46 Il s'agit en effet soit du *comma* (ou *punctus elevatus*), soit du *punctus interrogativus* selon la terminologie désormais établie grâce notamment à l'ouvrage de M.B. Parkes (1992).

47 Il convient de noter que de nouvelles observations sur la ponctuation de Guiot comparée aux autres manuscrits de la tradition du *Chevalier de la Charrette* ont récemment été publiées par S.-J. Murray (2007) et C. Pignatelli (2007).

la moitié des propositions sont ponctuées en moyenne dans son corpus, ce qui l'amène à la conclusion que « la proposition n'est donc pas, ou au moins elle n'est pas la seule unité de discours que servent à séparer les marques de ponctuation » (*ibidem*, p.37). Cette affirmation paraît aujourd'hui un peu trop catégorique : un simple découpage en propositions est insuffisant, car il semble nécessaire d'analyser la « force » des liens formels (subordination, présence d'éléments communs), thématiques (continuité du sujet) et sémantique entre des propositions voisines avant de procéder à des généralisations. Rappelons tout de même qu'il s'agit d'une recherche pionnière qui visait à explorer les différentes pistes d'analyse possibles et il était impossible d'avancer dans tous les sens à la fois.

Une autre piste proposée consistait à étudier la cooccurrence des signes de ponctuation avec des « connecteurs » (conjonctions ou certains adverbes) placés en tête de propositions. Il en est ressorti que certains d'entre eux (*lors, car, mais, or*) sont très couramment précédés d'une ponctuation. Les termes qui introduisent une proposition subordonnée placée après la principale sont très rarement précédés d'une ponctuation, tandis que si ces mêmes termes introduisent la subordonnée précédant la principale (*et quant, et se*), ils sont bien plus souvent accompagnés d'une ponctuation.

Ch. Marchello-Nizia note également que la ponctuation forte est souvent utilisée quand le sujet au début de la proposition est exprimé par un nom propre précédé d'un déterminant (*Le Jouvencel, Le capitaine de Crator*).

Un autre « lieu » repéré où la ponctuation apparaît quasi systématiquement est le début d'un discours au style direct et le changement de locuteur dans un dialogue.

La ponctuation « intrapositionnelle » est étudiée en dernier lieu, car elle est extrêmement rare « sauf dans quelques manuscrits exceptionnels ».

En comparant les données des différents manuscrits, Ch. Marchello-Nizia insiste sur la *singularité* de chaque manuscrit :

Même lorsque deux ou plusieurs manuscrits offrent des versions presque identiques (elles ne le sont jamais totalement) du même « texte », jamais la ponctuation n'est la même, et elle diffère toujours largement d'une leçon à l'autre, d'un copiste à l'autre (*ibidem*, p. 42).

État des recherches sur la ponctuation

La linguiste arrive néanmoins à dessiner une certaine « typologie de lectures » qui oppose d'une part des « manuscrits qui présente un texte... comme une pure suite d'unités non organisées » et, d'autre part, des manuscrits qui « organisent le texte en regroupant et en hiérarchisant les unités ».

Elle conclut sur l'insuffisance des « outils théoriques » de la linguistique actuelle pour aborder les langues mortes et appelle à appréhender les « unités de lecture » qui se forment entre les marques de ponctuation présentes dans les manuscrits « en oubliant le primat de la syntaxe ».

Malgré sa conclusion quelque peu pessimiste, l'article de Ch. Marchello-Nizia a contribué au regain progressif de l'intérêt à la ponctuation médiévale. Il a ouvert les différentes pistes de recherche possibles, qui se sont toutes révélées intéressantes. Notre démarche d'analyse en « frontières ponctuables » est en grande partie inspirée par le procédé suivi dans cette recherche pionnière. Même si le lien entre la ponctuation médiévale et la syntaxe nous paraît aujourd'hui plus fort que ce qui ressort de la conclusion de l'article, l'appel à renouveler l'apparat méthodologique et à ne pas se laisser guider par des présuppositions dogmatiques est toujours d'actualité.

Plus récemment, Ch. Marchello-Nizia (2007) est revenue à la problématique de la ponctuation médiévale en s'intéressant à l'emploi du *comma* dans un manuscrit de la *Queste del saint Graal*. Ce manuscrit fait partie du corpus que nous avons étudié dans le cadre de la présente recherche, et nous reviendrons sur l'article en question lorsque nous procéderons à l'analyse de la ponctuation dans ce manuscrit (section 6.2).

A la même époque que le premier article de Ch. Marchello-Nizia, H. Naïs (1979) a présenté, dans le cadre de la table ronde organisée par N. Catach et C. Tournier, une recherche sur la ponctuation dans le manuscrit B de Villehardouin⁴⁸. Sa conclusion est que la fonction principale de la ponctuation dans le manuscrit était de « mettre en valeur les qualités esthétiques des pages, souligner les éléments intéressants du texte » (*ibidem*, p. 54). Même si les facteurs décoratifs et pragmatiques (« ponctuation du respect », mise en relief d'informations importantes) jouent certainement un rôle dans la ponctuation médiévale, les résultats des autres recherches sur le sujet et les données de notre propre

⁴⁸ Il s'agit du manuscrit Paris, BnF, fr. 2137 (XIII^e s.). Cf. les manuscrits D (Paris, BnF, fr. 12203) et O (Oxford, Bodleian, Laud. misc. 587) qui font partie de notre corpus.

corpus semblent infirmer l'hypothèse d'une absence de valeur syntaxique de cette ponctuation.

Un an plus tard, Claude Buridant (1980) publie un article consacré au rapports entre la conjonction⁴⁹ *et* et la ponctuation. Il s'agit donc d'une exploration approfondie de l'une des pistes esquissées par Ch. Marchello-Nizia. Le choix du connecteur étudié n'est pas dû au hasard, car il s'agit de la conjonction de loin la plus fréquente, dont l'abondance apparente frappe toujours le lecteur moderne de textes médiévaux (*cf.* l'exemple (2.1) dans la section 2.3.1. pour une illustration).

Le corpus étudié par C. Buridant est composé d'un document principal (le manuscrit Copenhague, Bibl. Royale 487, comprenant *La Conquête de Constantinople* de Robert de Clari) et de deux documents annexes, dont *L'Estoire des contes de Flandre* du manuscrit Paris, BnF fr. 12203⁵⁰.

Le document principal est soumis à un dépouillement systématique selon une « grille d'analyse » établie, et des sondages sont effectués dans les documents annexes.

Les résultats de son enquête rejoignent dans une grande mesure les conclusions de Ch. Marchello-Nizia. Il constate que « si la ponctuation médiévale est mouvante, elle est loin d'être inconstante et anarchique » (*ibidem*, p. 45). Le chercheur souligne le rôle des lettrines qui « qui jalonnent le plus fréquemment les articulations temporelles du récit ». En ce qui concerne les rapports avec le « strument » *et*, il remarque qu'« au niveau des attaques de phrase en *et* + sujet nominal ou pronominal, la ponctuation accuse volontiers la dramatisation, l'opposition des protagonistes, le parallélisme des actions, l'accumulation énumérative ». Au niveau de la coordination de syntagmes, le linguiste trouve que la ponctuation permet le repérage des schèmes fondamentaux distingués par G. Antoine (1958), à savoir *a et b*, *a et b et c* et *et a et b*. Il remarque tout de même que « pour le scribe médiéval, ces schèmes ne sont pas rigides : ils offrent un jeu souple de variantes occulté par la transposition moderne ».

49 Le chercheur utilise le terme de *strument* emprunté à l'*Essai de grammaire* de Damourette et Pichon pour traiter ensemble les emplois conjonctionnels et adverbiaux.

50 Un autre extrait de ce même manuscrit fait partie de notre corpus (*cf.* la section 6.5).

État des recherches sur la ponctuation

Les rapports entre la ponctuation et la pratique de la lecture orale, ainsi qu'avec les caractéristiques rhétoriques propres au genre de la chronique sont à son avis indubitables.

La recherche de C. Buridant reste à ce jour un remarquable exemple d'analyse profonde et systématique d'un aspect particulier (et très important) de la ponctuation médiévale sur un corpus précis.

Une nouvelle publication significative consacrée à la ponctuation française médiévale voit le jour au milieu des années 1990. Il s'agit de l'article de Céline Barbance (Guillot) (1995) qui traite de la ponctuation dans cinq manuscrits du *Des cas des nobles hommes et femmes* de Laurent de Premierfait (XV^e s.). Les pratiques de ponctuation varient considérablement d'un manuscrit étudié à l'autre, et la chercheuse établit des liens d'une part entre la densité de la ponctuation et la variété des signes et, d'autre part, entre ces deux paramètres et la « qualité » des manuscrits (*ibidem*, p. 511). C. Guillot affirme que tous les manuscrits présentent une partition entre ponctuations forte et faible et parle également des « distinctions moyennes, suspensives et finales », sans indiquer pour autant le sens précis qu'elle donne à tous ces termes. L'analyse des rapports entre la ponctuation et la syntaxe la mène à la conclusion suivante :

... les manuscrits les plus ponctués sont ceux dans lesquels la ponctuation semble le mieux épouser la syntaxe. Les propositions sont délimitées de façon systématique, les mots ou groupes de mots juxtaposés sont le plus souvent séparés par un point. De même, la reprise du pronom personnel après une série de propositions circonstancielles s'accompagne en général de l'emploi d'un signe de ponctuation (*ibidem*, p. 515).

Aucune « règle » ne semble toutefois être respectée rigoureusement. Les incohérences apparentes dans les rapports entre la ponctuation et la syntaxe semblent s'expliquer dans certains cas pour des raisons sémantiques. Ainsi, le fait que les frontières entre les subordonnées et les principales soient plus souvent ponctuées lorsque la subordonnée précède la principale peut être expliqué par la nécessité de séparer la partie thématique du reste de la phrase. C. Guillot remarque également l'emploi « assez systématique » de la ponctuation devant les connecteurs *car* et *mais*. Les usages du *comma*, du *periodus* et du pied-de-mouche sont rangés dans la

« ponctuation métalinguistique », car « ils apparaissent toujours lorsqu'on change de séquence textuelle », par exemple au passage du style direct au style indirect (*ibidem*, p. 521).

A la fin de l'article, un parallèle est établi entre la ponctuation médiévale et celle des jeunes enfants, ce qui fait penser aux mécanismes psycholinguistiques dont la ponctuation constitue une trace (Fayol 1989).

L'article de C. Guillot est très riche en exemples et analyses fines des cas d'usage observés dans son corpus. Malheureusement, les données quantitatives sont totalement absentes de cette publication, ce qui rend impossible toute comparaison précise de ses résultats avec ceux obtenus par d'autres chercheurs sur d'autres corpus.

Quelques années plus tard, Micheline Stasse (1998) fait un aperçu très sommaire de la ponctuation dans les deux textes français les plus anciens (*Les Serments de Strasbourg* et *La Séquence de Sainte Eulalie*), ainsi que dans le manuscrit d'Oxford Digby 23 (*Chanson de Roland*). Elle présente par ailleurs des résumés des conclusions de M. Roques sur la ponctuation du scribe Guiot et de Ch. Marchello-Nizia sur le *Jouvencel* que nous venons d'évoquer. La conclusion que l'auteur de l'article tire de ses observations et qui consiste en une pauvreté et en une extrême individualité de la ponctuation médiévale paraît aujourd'hui prématurée. Malgré les divergences apparentes, des tendances générales se dessinent à travers l'ensemble des manuscrits étudiés jusqu'à présent, et les variations concernent un cercle précis de phénomènes. En même temps, l'appel à étudier les indications de la ponctuation pour « délimiter non seulement des "unités de lecture" [de la langue médiévale] mais aussi des "unités signifiantes", des articulations logiques du discours même si – et surtout – celles-ci diffèrent des nôtres » (*ibidem*, p. 184) est tout-à-fait légitime.

Une analyse plus profonde de la ponctuation dans les textes du « très ancien » français est proposée dans un article de Susan Baddeley (2001). Il s'agit de trois textes en prose : la *Séquence de Sainte Eulalie*, la *Passion de Clermont* et la *Vie de Saint Alexis*. L'analyse est effectuée d'abord par marque et puis, pour la marque la plus répandue (le point), texte par texte. Les marques étudiées sont les « grandes lettres »

État des recherches sur la ponctuation

(lettrines), le passage à la ligne (ce qui est une nouveauté), les majuscules, le « point-virgule »⁵¹ et le point.

L'analyse détaillée des occurrences amène la chercheuse à découvrir un phénomène jusque là peu connu et à formuler une hypothèse intéressante.

Dans la *Passion*, elle a identifié deux zones correspondant à chaque fois à un « changement de main », au passage d'un copiste à l'autre⁵². On trouve dans ces zones une concentration exceptionnelle de points que S. Baddeley a baptisée « environnement à points ». Une hypothèse selon laquelle il s'agit dans ces passages d'un « usage expérimental du point, une ébauche de ponctuation interne qui est abandonnée ensuite », même si d'autres explications possibles ne sont pas exclues (*ibidem*, p. 144).

Quelle que soit l'explication du phénomène, le fait que certains passages dans des manuscrits (il s'agit souvent des premières pages des textes) sont ponctués d'une façon particulière, plus dense que le reste du texte, est important et doit être pris en compte dans les analyses quantitatives, surtout lorsqu'on travaille sur des échantillons. Nous avons observé plusieurs manifestations d'un phénomène semblable dans notre corpus.

Une autre hypothèse intéressante concerne la fonction du point dans *Sainte Eulalie* :

A notre avis, le point, au lieu de signaler une rupture, serait un moyen d'indiquer la solidarité entre les deux éléments qui se trouvent de part et d'autre de ce signe, et à les relier ensemble (*ibidem*, p. 145).

L'hypothèse de S. Baddeley se base sur l'usage du point pour signaler, en fin de ligne physique, que le vers n'est pas fini (la fin du vers étant inscrit en interligne) et sur l'absence du point entre deux vers disposés sur une ligne dans le cas unique où ces deux vers n'ont aucun lien direct entre eux. Bien entendu, une hypothèse formulée à partir d'une occurrence isolée ne peut être acceptée sans vérification sur des données plus

51 Consciente de l'anachronisme de ce terme, S. Baddeley le préfère au *periodus* ou *punctus versus*, puisqu'il a l'avantage de bien décrire la forme. Ce choix nous paraît légitime dans la mesure où il s'agit de la dénomination de la forme du signe indépendamment de toute interprétation fonctionnelle. La même logique nous amène à préférer la dénomination « barre oblique » à la *virgula*. La situation est différente pour le *comma*, dont la forme ne ressemble de près à aucune des marques modernes.

52 Ce manuscrit de la fin du X^e siècle a la particularité de présenter une nette différence entre les « mains » de trois scribes différents qui ont participé à son exécution.

importantes, mais la fonction de connexion que les marques de ponctuation peuvent assumer aussi bien que celle de séparation mérite sans aucun doute d'être étudiée.

E. Llamas Pombo a consacré à la ponctuation des manuscrits en vers une partie importante de sa thèse (1996b) qui portait sur l'oralité et l'écriture dans la tradition « ovidienne » du XII^e siècle en France. Dans les années qui ont suivi, elle a publié plusieurs articles très intéressants sur les différents aspects de la ponctuation médiévale (Llamas Pombo 1996a ; 2001 ; 2003 ; 2007a ; 2007b ; 2008). Outre la synthèse de l'état des recherches sur la ponctuation que nous avons déjà citée à maintes reprises (Llamas Pombo 2008), c'est l'article (Llamas Pombo 2007b) qui nous paraît le plus significatif du point de vue méthodologique.

L'article commence par l'évocation des objectifs d'une étude de la ponctuation médiévale. Le premier d'entre eux dépasse le cadre de la linguistique et consiste à enseigner à lire les textes médiévaux pour mieux en éclairer la teneur et l'histoire. Il s'agit là d'une contribution à la connaissance du patrimoine historique et culturel. Un deuxième objectif relève de la linguistique diachronique et vise à élaborer une typologie des ponctuations médiévales pour appréhender, derrière la variation apparente des pratiques, les grandes lignes de l'évolution de ce phénomène linguistique. Le troisième objectif ressortit au champ de la linguistique générale : expliciter la compétence et le savoir intuitifs des locuteurs.

Nous espérons que notre mémoire pourra contribuer à son tour à la réalisation, à long terme, de ces objectifs.

La chercheuse présente ensuite quatre aspects qui concernent la « reconstruction » de la compétence linguistique du scribe médiéval en matière de ponctuation : (1) la recherche de la logique des signes, (2) la nécessité des études comparées (panromanes), (3) l'étude des rapports entre l'oral et l'écrit et (4) la prise en considération du facteur stylistique.

En guise d'exemple de recherche de la logique des signes, E. Llamas Pombo analyse les différentes fonctions du point et affirme, en reprenant l'hypothèse de S. Baddeley, que c'est aussi bien un « signe d'union » qu'un « signe de séparation ». Elle

État des recherches sur la ponctuation

met la valeur de connexion du point en rapport avec la notion de vers dans la littérature médiévale :

Ce type de ponctuation, apparemment inutile et redondante, rattache comme des chaînons les unités de versification autant qu'elle les distingue. Si le propre du vers est "de ne pas se suivre en continu comme le ruban de la prose"⁵³, les points représentent le lien des unités mesurées. Points et virgules de la césure deviennent fonctionnels, si nous considérons qu'ils signifient l'enchaînement, tout en balisant une frontière. Un peu comme nos traits d'union, ils soulignent simultanément une coupure et un renvoi. Et c'est cette valeur qui rendrait pertinente et logique l'absence de point terminal dans les anciens poèmes, *Eulalie* ou le *Roland* d'Oxford (*ibidem*, p. 20).

La même valeur de connexion se manifeste, selon l'analyse de la linguiste, dans les points d'abréviation et dans la coordination ou juxtaposition de syntagmes.

La nécessité de la prise en compte des traditions panromanes se justifie par le fait que le latin était une langue commune dans laquelle se réalisait une grande partie de la production écrite médiévale et que par conséquent « les ponctuations du latin médiéval et celles des langues romanes du Moyen Âge entretiennent un rapport de plus grande proximité entre elles que la ponctuation de l'ancien et du moyen français par rapport à celle du français moderne » (*ibidem*, p. 24).

Des travaux d'ordre général (Parkes 1992 ; Mortara 2008) constituent des avancées très importantes dans cette direction, mais l'augmentation de la base de connaissances empiriques concernant chacune des traditions nationales d'écriture en latin et en langue vernaculaire nous semble encore nécessaire avant qu'on puisse procéder à des généralisations à ce niveau.

En ce qui concerne les rapports entre l'oral et l'écrit, E. Llamas Pombo souligne, sans nier l'existence de la ponctuation « rythmique », l'importance, dès le Moyen Âge, de la « ponctuation pour l'œil » sans aucune correspondance rythmique possible. Elle montre ainsi que le *comma*, qui apparaît souvent après les interjections, ne peut pas correspondre à une pause compte tenu de la structure rythmique du vers (*ibidem*, p. 29). Un autre phénomène de la « ponctuation pour l'œil » concerne le « jeu des segmentations graphiques particulières » qui relève de la « ponctuation des mots ».

53 Citation de (Bourgain 1989, p. 253).

La prise en compte du facteur stylistique paraît importante, enfin, compte tenu des fonctions expressive et décorative que les marques de ponctuation peuvent assumer.

Dans sa communication récente au colloque « Diachro IV », E. Llamas Pombo (à paraître) s'est intéressée au phénomène de la « rubrique énonciative » et à la façon dont la mise en page du manuscrit permet de « compenser » l'absence d'une marque de ponctuation spécifique pour signaler des passages au discours direct, explorant ainsi une nouvelle piste de recherche.

C. Pignatelli (2007) quant à elle a étudié dans un corpus de l'ensemble des manuscrits conservés à ce jour du *Chevalier de la Charrette* de Chrétien de Troyes⁵⁴, la fréquence de la ponctuation selon les « lieux de ponctuation », lieux qui correspondent dans une grande mesure à ce que nous appelons des « frontières ponctuables ». Elle constate que

La plupart [des scribes] apparaissent capables de reconnaître et de souligner les lieux les plus dramatiques de changement intonatif (interjections, interrogations, exclamations), tout en privilégiant dans le rythme la mesure réglée de l'octosyllabe qui met en évidence la richesse de la rime ; seul le scribe du ms. T (et en moindre mesure celui du ms. E) semble introduire dans l'utilisation de la ponctuation des critères qui sont davantage syntaxiques (séparation des éléments d'une proposition ou d'une phrase) : ce faisant il distend la régularité métrique et subvertit le rythme de l'octosyllabe, et tire le vers du côté de la prose (*ibidem*, p. 101).

Ceci ne dénote pas pour autant un manque d'esprit poétique chez le scribe du manuscrit T, qui se montre par ailleurs « sensible aux brisures du rythme » de Chrétien.

Tout récemment, deux thèses ont été soutenues (2007) qui étaient tout entières consacrées à l'étude de la ponctuation médiévale.

Dans sa thèse de doctorat (2007b) et dans une série de publications (2007a ; 2008a ; à paraître_a ; à paraître_b), N. Mazziotta a mené l'analyse linguistique de la ponctuation dans un corpus de chartes liégeoises du XIII^e siècle. Nous avons déjà

54 Ce corpus a pu être constitué grâce aux transcriptions « ultra-diplomatiques » réalisées dans le cadre du projet *Charrette* <<http://lancelot.baylor.edu>>.

État des recherches sur la ponctuation

évoqué son modèle théorique dans la section précédente et nous nous référerons aux autres aspects de ses travaux par la suite.

Une autre thèse consacrée à l'analyse de la structuration « globale » du texte en français médiéval a été soutenue par Huei-Chen Li (2007). La recherche est menée à partir d'un corpus de transcriptions des trois versions du chapitre I de la 4^e partie du *Roman de Perceforest* fournies par deux manuscrits du XV^e siècle et par un imprimé de 1528 (entre 21 000 et 28 000 occurrences-mots chacune). En s'inspirant en partie de notre méthode (Lavrentiev 2007c ; 2007e), la chercheuse a calculé le « taux de ponctuation » et qui varie de 4,6% à 10% dans les manuscrits et s'élève à 8,3% dans l'imprimé. Elle a également étudié le rapport entre le taux de ponctuation et la fréquence relative des abréviations, ce qui apparaît comme une piste intéressante pour l'établissement de la « typologie interne » des systèmes graphiques des textes médiévaux. Les résultats obtenus par Huei-Chen Li apportent sans aucun doute une contribution précieuse à la connaissance des pratiques de la ponctuation à la fin du Moyen Âge. Malheureusement, il n'est pas certain que ses transcriptions effectuées en Microsoft Word et ses annotations puissent être réintégrés dans des bases textuelles plus grandes et réutilisées par d'autres chercheurs.

Enfin, des ouvrages plus généraux ont consacré quelques pages ou chapitres à la ponctuation en français médiéval. On trouve ainsi des remarques intéressantes concernant l'emploi de la ponctuation dans l'*Album de manuscrits français du XIII^e siècle* (Careri *et al.* 2001). Par ailleurs, la récente *Storia della punteggiatura in Europa* (Mortara 2008) consacre à l'histoire de la ponctuation française deux chapitres rédigés par plusieurs auteurs (Careri 2008 ; 257 Ferrari 2008 ; 256 Colombo 2008). La présentation est organisée dans l'ordre chronologique : les sections correspondent le plus souvent à des siècles ou à des moments jugés importants, comme par exemple le passage du XII^e au XIII^e siècle. Les observations présentées sont le plus souvent basées sur des extraits d'un nombre limité de manuscrits, et les auteurs ne cherchent pas à généraliser les résultats. Il s'agit donc d'un témoignage intéressant concernant les pratiques de ponctuation dans un certain nombre de manuscrits représentant chaque période étudiée, mais cette contribution ne peut pas être considérée comme un véritable

résumé de l'histoire de la ponctuation en France. Ceci n'est pas la faute des auteurs, car malgré le nombre considérable des recherches sur la ponctuation médiévales qui ont été menées dans les dernières décennies, la quantité des données accumulées reste encore insuffisante pour qu'on puisse en tirer des conclusions sûres et définitives. L'heure est donc encore et toujours au dépouillement et à l'analyse des données primaires – et c'est à cette tâche que notre étude est en grande partie consacrée.

Dans la perspective de cette thèse, nous avons également réalisé une contribution modeste au recensement des pratiques de la ponctuation médiévale dans un article consacré à la ponctuation dans l'*Image du monde* (Lavrentiev 2000). par la suite, en juin 2005, nous avons eu le plaisir d'organiser à Lyon, à l'ENS LSH, une journée d'étude consacrée à la ponctuation et aux graphies médiévales. La journée a permis des échanges très intéressants entre les spécialistes du domaine et a donné à la publication des actes (Lavrentiev dir. 2007d), dont nous avons déjà cité plusieurs contributions. Plus récemment, enfin, nous avons publié nos réflexions sur les rapports entre les connecteurs et la ponctuation dans un certain nombre de textes (Lavrentiev 2007b) et sur les aspects méthodologiques de l'étude linguistique de la ponctuation médiévale (2007c ; 2007e). Les résultats de ces recherches sont développés et approfondis dans le présent mémoire.

Chapitre 2

Cadre théorique et méthode pour l'analyse linguistique

Dans ce chapitre nous présenterons notre cadre théorique et la méthode d'analyse linguistique des données de notre corpus. Nous n'avons pas visé à élaborer un nouveau modèle de la syntaxe ou d'un autre sous-système de la langue, ni à appliquer rigoureusement un modèle proposé par une école linguistique donnée. Ce que nous avons cherché à construire, c'est un cadre théorique adéquat pour essayer de comprendre quel rôle est réservé aux marques de ponctuation dans les textes de notre corpus et quels mécanismes linguistiques « se cachent » derrière leur emploi.

Nous utiliserons dans la mesure du possible l'apparat terminologique consensuel de la grammaire traditionnelle (Grevisse & Goosse 2007) en y introduisant des précisions et des éléments supplémentaires apportés par plusieurs branches de la linguistique de la deuxième moitié du XX^e siècle : la syntaxe sémantique (Tesnière 1965 ; Lazard 1994), la grammaire fonctionnelle (Bondarko 1991), la grammaire textuelle (van Dijk 1977), etc. Cette approche « éclectique » s'inspire en grande partie de la *Grammaire méthodique du français* (Riegel *et al.* 1997).

Dans les sections qui suivent, nous définirons les concepts-clefs et décrirons précisément notre démarche d'analyse linguistique des données. Nous commencerons par présenter l'approche fonctionnelle de la langue que nous avons adoptée dans notre travail. Nous présenterons ensuite notre définition du « noyau formel » de la

ponctuation et de ses rapports avec les marques segmentales. Nous aborderons ensuite les fonctions principales de la ponctuation et la façon dont la ponctuation interagit avec les connecteurs en comparant le système graphique des manuscrits médiévaux à celui du français moderne. Nous présenterons ensuite la notion d'« unité ponctuable » qui joue un rôle central dans notre méthodologie d'analyse linguistique et nous expliquerons pourquoi la notion de phrase nous paraît inapplicable dans le cadre de notre recherche. Nous procéderons ensuite à la présentation de notre classement des « frontières ponctuables » qui prend la forme d'un jeu d'étiquettes utilisé lors de l'annotation de notre corpus. Nous présenterons également les étiquettes employées pour l'annotation de la forme et de la force des marques de ponctuation. Le chapitre se terminera par la description de la procédure appliquée lors de l'analyse de la ponctuation dans chacun des textes de notre corpus.

Les aspects liés à la représentation des données et à l'interprétation des résultats obtenus à partir d'un corpus seront exposés dans les chapitres suivants.

2.1. Approche fonctionnaliste

Comme nous l'avons vu dans le chapitre précédent, les théories de la ponctuation peuvent se ranger grossièrement dans trois grandes catégories : phonocentristes, autonomistes et mixtes. Une approche purement phonocentriste est certainement trop radicale (l'écrit n'est pas une simple transcription de l'oral !) et de toute façon inapplicable aux textes anciens, pour lesquels nous disposons de trop peu de données concernant le rythme et l'intonation. Nous avons constaté par ailleurs qu'aucune théorie ne peut être qualifiée de totalement phonocentriste⁵⁵, car les « pauses » orales sont toujours motivées par des considérations « logiques » (syntaxiques ou sémantiques) ou « pragmatiques ». Une approche purement autonomiste permet de bâtir une théorie qui se tient, mais la séparation totale des deux formes d'expression de la langue est tout de même artificielle et contre-intuitive, surtout dans le contexte de la culture européenne médiévale où la représentation orale jouait un rôle central (même si la lecture et l'écriture silencieuses étaient peut-être plus répandues

55 A l'exception peut-être de celle de Damourette (1939), mais des critères sémantiques y « pénètrent » également dans une certaine mesure.

Cadre théorique et méthode pour l'analyse linguistique

qu'on ne le pensait auparavant)⁵⁶. Une approche mixte paraît dans ces conditions la plus raisonnable et la plus équilibrée, mais l'absence quasi totale de données sur l'oral de l'époque la rend difficile à mettre en œuvre sans tomber dans le cercle vicieux d'explication des phénomènes de l'écrit par des phénomènes de l'oral, dont ces traces écrites sont les seuls témoins.

Pour sortir de ce cercle, il faut baser l'analyse des données sur des critères indépendants du médium d'expression et considérer que l'oral et l'écrit sont des manifestations de la même activité mentale. L'écrit est secondaire par rapport à l'oral sur le plan historique et au cours de l'acquisition du langage par un individu, il n'est donc pas déraisonnable de chercher dans l'oral l'origine de tel ou tel phénomène écrit, mais on ne pourra jamais dépasser le stade d'hypothèses non confirmées faute de données expérimentales.

Des expériences psycholinguistiques auraient sans doute été l'outil de recherche le plus adéquat dans cette optique. Des résultats très intéressants ont d'ailleurs été obtenus par ces méthodes dans les études de la production de l'écrit en français moderne. Ainsi, Michel Fayol arrive à la conclusion suivante (1989, p. 27) : « Tout semble se passer comme si la ponctuation (et les connecteurs) constituait la *trace* en surface d'opérations de mise en relation d'événements (ou d'états) que la *mise en texte* "traduit" par des propositions concaténées ».

Les méthodes de recherches psycholinguistiques sont malheureusement aussi inapplicables aux états de langue anciens que l'exploitation directe des données de l'oral, car ces états ne sont « naturels » pour aucun sujet parlant actuel, et les textes conservés sont les seuls témoins de la production linguistique. Cela nous oblige à adopter une démarche plus abstraite qui consiste à décrire les fonctions de la ponctuation attestées dans les textes anciens et modernes et à étudier comment elle remplit ces fonctions dans les textes de notre corpus, en interaction avec les autres marques d'expression linguistique.

Cette démarche s'inscrit dans la lignée des travaux du cercle linguistique de Prague, du fonctionnalisme d'André Martinet, reprise plus récemment par la

⁵⁶ Cf. (Zumthor 1984) et (Saenger 1987).

« grammaire fonctionnelle » de A. Bondarko et ses collaborateurs et par la linguistique typologique « de la Côte Ouest » américaine.

Du point de vue fonctionnel, la langue peut être décrite comme un système de marques⁵⁷ qui remplissent les différentes fonctions lui permettant de servir dans son ensemble de moyen de communication, de véhicule d'information, de cognition, etc.⁵⁸ Physiquement, ce système se réalise sous deux formes : orale et écrite. Pour la majorité des marques, que l'on peut appeler *segmentales*, il est possible d'établir une correspondance directe entre les deux formes d'expression pour une unité de sens ou une fonction donnée. Ce sont les mots, la morphologie (ou les lexèmes et les morphèmes, dans la terminologie structuraliste) et les constructions syntaxiques et éventuellement textuelles.

Dans les systèmes d'écriture alphabétiques, il existe une correspondance plus fine entre les composantes distinctives des marques phoniques et graphiques (phonèmes et graphèmes respectivement), mais en réalité ces correspondances peuvent « se brouiller » soit à cause de l'évolution de la prononciation que les graphies ne suivent pas, soit suite au développement d'abréviations écrites.

Il existe par ailleurs des marques spécifiques à chaque forme de la langue, qui remplissent souvent des fonctions équivalentes, mais entre lesquelles il n'existe pas de correspondance directe. A l'oral, ces marques *suprasegmentales* incluent, entre autres, le rythme, l'intonation, les gestes et la mimique. Pour l'écrit, il s'agit de toutes sortes d'« organisateurs textuels » ou de « balises » *graphiques* qui facilitent la reconnaissance et l'appréciation des relations entre les marques segmentales⁵⁹. Celles-ci incluent les signes de ponctuation proprement dits, les espaces blancs, les variantes de caractères (dont les majuscules), la variation de la taille des caractères, la mise en page, etc. Cela correspond plus ou moins à la ponctuation « au sens large » (Catach 1994, p.8), une notion que nous développerons dans la section suivante.

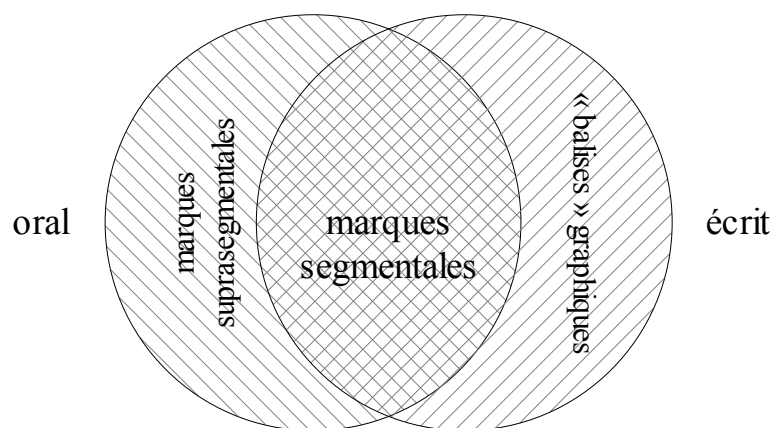
57 Nous utilisons ce terme pour désigner un objet linguistique quelconque portant une certaine fonction (cf. Luscher 1994, p. 175).

58 Cf. les fonctions du langage d'après R. Jakobson (1963, p. 213-220) : référentielle (dénotative), émotive, poétique, conative, phatique et métalinguistique.

59 Cf. les notions de *topogramme* (Anis 1988) ou *topème* (Mazziotta 2007b).

Cadre théorique et méthode pour l'analyse linguistique

La typologie « formelle » des marques linguistiques peut ainsi être représentée par le schéma suivant⁶⁰ :



Marques segmentales : mots et morphèmes, paradigmes morphologiques, constructions syntaxiques, ordre des constituants...

Marques suprasegmentales : intonation, rythme, gestes...

« Balises » graphiques : signes de ponctuation, espaces blancs, variantes de caractères, taille des caractères, mise en page...

Fig. 1 : Typologie formelle des marques linguistiques

Avant de procéder à l'analyse des fonctions des marques de ponctuation et à l'identification des marques segmentales et suprasegmentales, avec lesquelles celles-ci « coopèrent », il convient de donner une définition précise de ce que nous entendons par la ponctuation et les différents types de marques qui en font partie.

⁶⁰ Ce schéma reproduit, avec certaines modifications terminologique, celui que nous avons publié dans un article consacré à l'interaction entre la ponctuation et les connecteurs (Lavrentiev 2007b).

2.2. Marques de ponctuation : noyau formel et élargissement fonctionnel

2.2.1. « Balises » graphiques : marques de ponctuation au sens propre

Nous adoptons dans la définition de la ponctuation une approche large. Nina Catach (1994, p.8) a proposé la formulation suivante : « un système de renfort de l'écriture... chargé d'organiser les rapports et la proportion des parties du discours et des pauses orales et écrites ». De façon plus restrictive, nous pouvons définir la ponctuation comme un système de marques graphiques sans correspondance directe segmentale à l'oral (*topogrammes* selon Anis *et al.* 1988) destinées à la structuration et à l'organisation visuelle du texte. Ce système est composé de caractères spéciaux (*signes de ponctuation*), de *variantes fonctionnelles* de caractères alphabétiques (letrines, majuscules, italiques, etc.), d'*espaces blancs horizontaux* de différentes tailles et de sauts de ligne. Suivant Claude Tournier (1980) et Nina Catach (1994), nous distinguons trois niveaux de ponctuation : lexical (marquage des limites et de l'unité des mots, mise en relief de mots et de locutions) ; syntaxique (marquage des limites et des rapports entre les unités syntaxiques et communicatives) ; macro-textuel (mise en page, division des paragraphes, des chapitres, des parties, etc.).

La ponctuation du niveau lexical constitue un domaine à part qui mérite une étude spécifique : les premiers résultats obtenus (Andrieux-Reix & Monsonégo 1997 ; 1998 ; Baddeley 1998 ; Biedermann-Pasques 1998 ; Baddeley & Biedermann-Pasques 2004 ; Llamas Pombo 2003) sont très intéressants, mais beaucoup reste à découvrir. Dans le cadre de la présente étude notre analyse porte avant tout sur la ponctuation du niveau syntaxique et dans une certaine mesure macro-textuel. Nous verrons toutefois qu'il n'existe pas de frontière infranchissable entre les niveaux et que certaines marques apparemment « syntaxiques » servent en réalité à marquer des mots individuels. Nous signalerons de tels cas quand ils se présenteront. Notons par ailleurs que les données représentées dans nos transcriptions de sources primaires incluent les informations sur les segmentations particulières, sur l'emploi des variantes positionnelles des lettres et

des marques d'abréviation, ce qui permettrait dans l'avenir d'utiliser notre corpus pour l'étude de la « ponctuation lexicale ».

2.2.2. Marques segmentales de structuration textuelle : connecteurs

De toutes les marques segmentales, les connecteurs sont sans doute les « compagnons » les plus fidèles de la ponctuation dans l'ensemble de ses fonctions, ou presque.

Le terme de « connecteur » a été largement utilisé dans les travaux des linguistes depuis une trentaine d'années, et plusieurs définitions en ont été proposées. Christian Touratier, qui fait le point sur les différentes acceptions utilisées par les chercheurs, distingue deux « pôles opposés » : un pôle « large et flou » et un pôle « plus restreint qui correspondrait à une certaine spécificité du terme » (2006, p. 19).

Nous adoptons dans notre recherche une vision plutôt large de cette notion, en y incluant, suivant la *Grammaire méthodique du français* de M. Riegel, J.-C. Pellat et R. Rioul (1997, p. 617), « tous les termes qui assurent l'organisation d'un texte », et notamment les conjonctions de coordination et de subordination, des adverbes (*si, alors, puis...*), des groupes prépositionnels (*d'une part...*), des présentatifs (*c'est, voilà...*), des locutions (*c'est à savoir...*), etc⁶¹. Ces termes peuvent fonctionner au niveau de la phrase complexe aussi bien qu'à l'intérieur d'une proposition ou encore au niveau macro-textuel. A notre avis, le critère principal dans la définition des connecteurs est fonctionnel (l'organisation du texte), ce qui nous permet de regrouper des marques linguistiques segmentales de nature différente mais utilisées, au moins en partie, à des fins similaires. C'est justement cette approche fonctionnaliste qui nous permet d'étudier les rapports entre les connecteurs et la ponctuation.

Le terme de « connecteur » n'est sans doute pas parfait, car il ne met en valeur que l'une des fonctions de ces marques linguistiques. B. Schneuwly, M.-C. Rosat et J. Dolz (1989) préfèrent parler d'« organisateurs textuels », en insistant notamment sur leur rôle dans la segmentation aussi bien que sur celui qu'ils jouent dans les

61 Cf. la définition de Jacques Jayez : « éléments qui sont interprétables comme associant plusieurs entités sémantiques » (1988, p. 22).

phénomènes de connexion. Le terme de connecteur est cependant de loin celui qui est le plus largement utilisé et accepté par la communauté des chercheurs, et nous l'adoptons donc ici.

Il convient cependant de poser certaines limites formelles à l'éventail de marques linguistiques que nous considérons comme des connecteurs. Premièrement, comme nous l'avons déjà indiqué, il s'agit d'objets segmentaux (morphèmes, lexèmes ou locutions). Deuxièmement, ces objets doivent avoir pour fonction principale la connexion/segmentation de structures textuelles, plutôt que la cohésion textuelle, assurée généralement par les anaphores, la thématization, etc⁶².

Contrairement à (Schneuwly *et al.* 1989) et à (Riegel *et al.* 1997), nous admettons que les connecteurs puissent faire partie intégrante d'une proposition. C'est le cas de pronoms et d'adverbes relatifs qui occupent la place d'un actant ou circonstant dans une proposition tout en remplissant la fonction de connexion/séparation. C'est également le cas des pronoms démonstratifs antécédents de propositions relatives (*ce que, ce... qui, celui qui*, etc.). A notre avis, tout l'intérêt de la notion de connecteur vient du fait qu'elle permet de regrouper des objets linguistiques de nature différente (mots « pleins » et mots grammaticaux).

Les connecteurs peuvent se situer à la frontière ou au milieu de l'unité syntaxique qu'ils servent à introduire dans la structure textuelle (connecteurs *externes* et *internes* respectivement). Les connecteurs internes (comme, par exemple, les adverbes *donc* et *alors*) ont la particularité de remplir uniquement la fonction de connexion et non celle de segmentation ; seuls les connecteurs externes peuvent avoir un rapport direct avec les marques de ponctuation.

2.3. Fonctions de la ponctuation

Pour cerner les fonctions des marques de ponctuation dans les manuscrits médiévaux et dans les incunables, nous prendrons comme point de départ les fonctions de la ponctuation dans les textes modernes, car ces fonctions sont relativement bien connues et étudiées. Nous verrons par la suite dans quelle mesure ces fonctions sont

62 La distinction entre la cohésion et la connexion/segmentation au niveau psycholinguistique est introduite dans (Schneuwly *et al.* 1989, p. 42-43).

d'actualité pour les textes de notre corpus en nous basant sur les résultats publiés par nos prédécesseurs et sur nos propres observations⁶³.

Dans le numéro de la *Langue française* consacré à la ponctuation (Catach 1980), plusieurs articles ont traité des fonctions de la ponctuation en français moderne (Catach 1980b ; 222 Perrot 1980 ; 229 Védénina 1980). Derrière les différences apparentes de terminologie, les auteurs s'accordent à distinguer trois types de fonctions de la ponctuation : syntaxique (marquage formel de certaines unités syntaxiques), pragmatique (marquage ou mise en relief d'éléments focalisés, thématisés, disloqués, etc.) et sémantique (codage d'information sur le statut énonciatif d'une unité syntaxique ou lexicale).

A notre avis, les facteurs pragmatiques ont dans tous les cas une influence sur la structure syntaxique de l'énoncé : une phrase qui comporte la focalisation ou la dislocation d'un élément n'a pas la même structure syntaxique qu'une phrase « plate ». Il est donc possible de considérer la fonction pragmatique de la ponctuation comme un complément de la fonction de marquage syntaxique qu'elle assure dans le continuum discursif.

En nous inspirant des travaux cités ci-dessus, nous avons dressé une liste des principales fonctions que les connecteurs et les marques de ponctuation peuvent partager dans un texte. Nous analyserons ces fonctions tour à tour, en précisant à chaque fois les rapports que la ponctuation entretient dans cette fonction avec les connecteurs.

2.3.1. Séparation des unités intratextuelles

La démarcation des unités de différents niveaux est sans aucun doute l'une des fonctions principales de la ponctuation. Les connecteurs « externes », comme nous l'avons indiqué, n'y sont pas non plus étrangers. Le rôle des connecteurs dans la séparation des unités paraît par ailleurs nettement plus important dans les manuscrits médiévaux que dans la langue moderne.

En effet, en l'absence de codification explicite de la ponctuation, l'usage « abondant » des connecteurs était parfois le seul moyen pour exprimer la bonne

⁶³ Nous reprenons et développons ci-dessous les remarques que nous avons présentées dans une communication au XI^e colloque sur le moyen français (Lavrentiev 2007b).

structuration du texte (Catach 1994, p. 7). L'abondance des connecteurs (et notamment de la conjonction *et*) dans les textes médiévaux (surtout en prose) a été signalée par plusieurs chercheurs, même si cette observation est difficile à confirmer statistiquement (Antoine 1958, p. 465 ; Buridant 1980, p. 16). Les traductions en français moderne conservent le plus souvent les connecteurs de la langue médiévale⁶⁴. Pour comparer les fréquences relatives des connecteurs dans les textes anciens et modernes, il faudrait comparer des textes très proches sur le plan de la typologie textuelle pour neutraliser les facteurs de variation autres que le temps. Or il n'existe pas encore de typologie textuelle consensuelle allant du français médiéval au français contemporain⁶⁵.

Nous avons toutefois tenté d'obtenir une estimation grossière de la différence de fréquence d'un connecteur très employé – la conjonction *et* – sur un corpus relativement important. En profitant des bases textuelles accessibles en ligne, nous avons recherché les occurrences des graphies *Et* et *et* dans les textes littéraires de la Base de Français Médiéval <<http://bfm.ens-lsh.fr>>⁶⁶ et dans les romans des années 1950-1955 de la base Frantext <<http://www.frantext.fr>>⁶⁷. Le résultat obtenu est assez convaincant : la fréquence relative de *et* est plus de deux fois supérieure dans les textes littéraires médiévaux à celle de *et* dans les romans des années 1950 (3,90 contre 1,63 occurrences en moyenne pour 100 mots).

Pour illustrer l'abondance des connecteurs dans un texte médiéval, nous avons choisi un petit extrait du manuscrit Paris, BnF, fr. 12203 de la *Conquête de Constantinople* de Geoffroy de Villehardouin dans lequel nous avons surligné les occurrences de la conjonction qui nous intéresse⁶⁸ :

(2.1) **S**Achiez que .M. **et** || C. **et** quatre+uinz anz || apres l'incarnation || *no*stre sengnor
iesucrist / || As+tens Innocent A||postoille de Rome / **et** || phelippe Roy de France . **et**

64 A l'exception, dans certains cas, de l'adverbe *si* initial, qui « contrairement à des morphèmes tels que *et*, *ou*, *certes*... a subi des changements nombreux et profonds dans sa syntaxe » (Marchello-Nizia, 1985, p. 4).

65 Cf. les variables de la typologie textuelles de la Base de Français Médiéval présentées dans la section 3.3.3.

66 La requête a été effectuée avant la suspension de l'accès public aux corpus de la BFM en octobre 2007.

67 Le choix arbitraire des critères de sélection de textes modernes a été dicté par la nécessité d'obtenir un corpus de taille comparable à celle de la période médiévale. Les corpus utilisés comptent 1 672 055 mots pour le français médiéval et 2 025 768 mots pour le français moderne.

68 Cet extrait de notre corpus est présenté selon les conventions de citation décrites dans la section 4.4.3.2. et basées sur les principes de la transcription exposés en détail tout au long du chapitre 4.

Richart || Roy d'engleterre **et** un saint home || en france / qui ot nom folques de || nuilli . Cil nuillis si est entre li||gni sor marne e paris . **et** il ere *pre*||stres **et** tenoit la parroche de+la+uil||le . **Et** cil folques dont ie vos di / *con*||menca / a parler de dieu par fran||ce / **et** par les autres terres entor . || **et** *nostre* sires fist maintes miracles || por lui . Sachiez que la renomee || de cel saint home ala tant qu'ele || uint a+l'apostoille de Rome Inno||cent . **et** l'a _postoilles enuoia en fran||ce / || **et** manda as prodome / *que* il pen||chast || des+croiz par s'autorite . **et** *apres* || i+enuoia un sien chardonal mai||stre perron de chappes croisie . **et** **†** || manda par lui le pardon tel *con* ie || uos dirai .
(villehardouin587⁶⁹, 1r, 1-26).

Dans ce passage de près de 150 mots, il y a 14 occurrences de la conjonction *et*, ce qui correspond à une fréquence moyenne de 9,3 occurrences pour 100 mots. Bien entendu, de tels passages « farcis » de connecteurs sont rares, mais leur présence est significative.

2.3.2. Marquage de relations entre unités textuelles

L'expression des relations entre unités textuelles est une fonction sémantique. C'est la fonction principale des connecteurs, et la ponctuation ne peut jouer dans ce domaine qu'un rôle périphérique. Dans la ponctuation moderne, il existe cependant deux marques qui peuvent renforcer ou, dans certains cas remplacer, les connecteurs dans cette fonction.

La première de ces marques est la paire de parenthèses. Les parenthèses indiquent clairement la relation de l'unité ponctuée à son entourage : c'est une précision, une reformulation, un complément d'information, un commentaire. Cette relation peut également être exprimée grâce aux connecteurs, tels que *ou*, *c'est-à-dire*, *à savoir*, *autrement dit*, etc. :

- (2.2) J'ay regardé en moy, dit le curé, pource que je vouldroye le bien et avancement de vostre filz, que la fille d'un tel (c'est assavoir sa dame) seroit trop bien sa charge...
(*Cent nouvelles nouvelles*, éd. Sweetser, 1966, p. 297).

Les manuscrits médiévaux français que nous avons étudiés ne comportent pas de marque de ponctuation équivalente aux parenthèses modernes. Les éditeurs scientifiques des textes médiévaux introduisent parfois les parenthèses pour marquer les syntagmes

69 Nous utiliserons dans les références des citations du corpus les identifiants de nos transcriptions présentées dans les chapitres 6 à 8 et dans la liste des sources dans la section B de la bibliographie.

ou propositions apportant une précision au texte précédent, comme c'est le cas dans l'exemple (2.2) ci-dessus ou encore autour de *Cil nuillis si est entre ligni sor marne e paris* dans (2.1), selon l'édition de E. Faral (1961). L'usage de ce signe de ponctuation est tout de même rare de façon générale dans les éditions de textes médiévaux (854 occurrences seulement dans la BFM contre, par exemple, 16 265 occurrences du deux-points)⁷⁰. La présence de connecteurs explicitant la relation de reformulation ou de commentaire explique en partie l'absence de nécessité de recourir aux parenthèses.

Dans l'histoire de la ponctuation, les parenthèses ont été introduites par les humanistes italiens à la fin du XIV^e siècle pour démarquer les incisives et incidentes (Parkes 1992, p. 49), mais en France leur usage ne se répand qu'avec le développement de l'imprimerie. Il est notamment recommandé dans le *Dialogus compendiosus* de Heynlin (1471). Dans notre corpus, les seules occurrences indubitables⁷¹ des parenthèses sont attestées chez Etienne Dolet (1540).

L'autre signe de ponctuation avec un potentiel sémantique est le deux-points. D'après *Le Bon usage* de Maurice Grevisse⁷², ce signe « annonce : a) la citation d'un texte, la reproduction des paroles ou des pensées de quelqu'un... ; b) l'analyse, l'explication, la cause, la conséquence, la synthèse de ce qui précède » (2007, p. 136-137). Jacques Drillon va plus loin en affirmant que « le deux-points a un pouvoir logique très puissant. Il équivaut à “donc”, à “parce que”, à “bien que”, à mille et une de ces charnières qui permettent d'articuler le raisonnement » (1991, p. 394). Bien entendu, le signe de ponctuation ne fait dans ces cas que marquer (ou « cristalliser ») la relation sémantique inférée par le contexte.

En français médiéval, le deux-points n'existait pas dans sa forme actuelle avant la fin du XV^e siècle. Le *comma* (ou *punctus elevatus*) < ; >, qui est souvent considéré comme l'ancêtre du deux-points, n'a jamais été, d'après les données de notre corpus, spécialisé dans la fonction remplie par le deux-points en français moderne. Pour Estienne Dolet (1540), le *comma* est représenté par la marque graphique de deux-points

70 C'est à cause de cette rareté d'usage des parenthèses par les éditeurs que nous avons été amené à chercher un exemple en dehors de notre corpus.

71 Cf. toutefois les occurrences « douteuses » relevées dans **tyr828** (section 6.4.) et **qjm1052** (section 7.9.).

72 Nous citons la 14^e édition, refondue par André Goosse.

⟨ : ⟩. Le grammairien l'utilise comme une ponctuation « moyenne » (l'équivalent du point-virgule moderne), et il met une petite barre oblique devant une énumération (où, de nos jours, on utiliserait le point-virgule).

On peut donc conclure que d'après les données dont nous disposons les marques de ponctuation médiévales ne remplissent pas la fonction « sémantique » de spécification des relations entre des unités syntaxiques.

2.3.3. Hiérarchisation des unités textuelles

Le texte d'un manuscrit ou d'un livre moderne forme une structure hiérarchique. Les grandes unités qui le composent (par exemple, les chapitres) sont à leur tour composées d'unités plus petites (paragrapes), etc.

La mise en page et la ponctuation au sens large ont toujours joué le rôle principal dans la représentation de cette hiérarchie dans un texte écrit. Les signes de ponctuation au sens propre participent dans une certaine mesure à la hiérarchisation des unités textuelles. Si leur importance relative pour cette fonction n'a pas changé considérablement des manuscrits médiévaux aux textes modernes, le fonctionnement de la ponctuation hiérarchisante est différent aux deux époques. En français moderne, on parle souvent de ponctuations fortes (point, point d'interrogation ou point d'exclamation suivi d'une majuscule), moyennes (point-virgule, deux-points) et faible (virgule)⁷³. Dans les manuscrits et incunables médiévaux, il n'est pas rare, comme l'analyse des données de notre corpus le montrera, de voir la dominance (voire l'usage exclusif) d'un seul signe de ponctuation (le point ou la barre oblique le plus souvent) à tous les niveaux de frontières syntaxiques. Ce signe peut être suivi aussi bien d'une minuscule que d'une majuscule ou d'une lettrine. Par ailleurs, les majuscules peuvent fonctionner en tant que marques de séparation et de hiérarchisation des unités de la structure textuelle, sans « appui » de signes de ponctuation. Ce n'est donc pas la forme du signe de ponctuation, mais sa combinaison avec la variante fonctionnelle (lettrine, majuscule ou minuscule) de la première lettre du mot suivant qui permettait de hiérarchiser les « unités de

⁷³ Cf., par exemple, l'affirmation suivante : « Ces trois signes [le point, le point-virgule et la virgule] marquent des pauses d'importance décroissante » (Riegel *et al.* 1996, p. 87). N. Catach (1994), quant à elle, distingue 7 « forces » à 3 « niveaux » de la ponctuation.

lecture » dans ces textes. Dans d'autres manuscrits, une certaine « spécialisation » des signes semble se profiler : le comma et la barre oblique sont rarement suivis d'une majuscule, mais le point reste le plus souvent une marque polyvalente. La tendance à la spécialisation des signes de ponctuation sera pour nous l'un des critères principaux permettant de caractériser le type de ponctuation dans un texte.

Par ailleurs, certains manuscrits médiévaux utilisaient un signe spécial permettant de marquer des unités macro-textuelles (des « paragraphes », en quelque sorte) sans passage à la ligne ou de renforcer la mise en relief graphique d'une unité déjà marquée par un saut de ligne. Selon les manuscrits, ce signe prend parfois la forme d'une grande lettre grecque *gamma* < Γ > (« crochet adlinéaire » dans la tradition paléographique française) ou d'un < C > latin (« pied-de-mouche » ou *paraph*⁷⁴), comme par exemple à la quatrième ligne de l'illustration suivante :

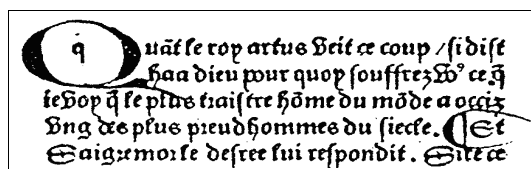


Fig. 2 : « Pied-de-mouche » (*artu1488*, p. 50).

Ces marques étaient souvent dessinées en couleur non pas par le scribe, mais par le rubricateur (la personne qui réalisait les grandes lettrines). Elles occupent donc dans le système graphique une position intermédiaire entre les signes de ponctuation et les éléments de la décoration. Il est d'ailleurs parfois difficile d'expliquer l'emploi d'un pied-mouche par des raisons autres que décoratives.

Les marques segmentales sont assez rarement utilisées à des fins de hiérarchisation des unités textuelles, à l'exception des cas de mention explicite du début

74 Ce terme est couramment utilisé pour désigner la marque en question dans les recherches consacrées à la ponctuation française médiévale (par exemple, Barbance 1995, p. 509 ; Llamas Pombo 1996a, p. 138) et dans les albums paléographiques (Careri *et al.* 2001, p. 215). Il est par ailleurs mentionné dans *Le bon usage* de M. Grevisse (2007, p. 122). J. Stiennon (1991, p. 348) lui donne pourtant un autre sens : « petit trait de couleur surchargeant la première lettre d'un mot pour la mettre en valeur ». M. B. Parkes (1992) et les auteurs de la récente *Storia della punteggiatura in Europa* (Mortara Garavelli éd. 2008) utilisent, pour la marque en forme de < C >, le terme de *paraph* (ou *paraffo*). Nous utiliserons le terme de « pied-de-mouche » selon la tradition des recherches linguistiques françaises, en ajoutant *paraph* entre parenthèses pour lever toute ambiguïté à chaque fois que le terme sera utilisé pour une première fois dans une section. Nous ferons de même pour tous les termes où les différentes traditions scientifiques divergent.

ou de la fin d'une unité (*Ici commence le premier chapitre...*). Dans les textes médiévaux, certaines locutions (contenant des connecteurs) avaient tendance à apparaître plus ou moins régulièrement au début de grandes unités textuelles. Par exemple, dans l'incunable **artu1488**, 20 chapitres sur 22 commencent par une formule contenant l'expression *dist le compte*, dont 13 occurrences de *Or dist le compte*.

2.3.4. Mise en relief d'éléments de la structure textuelle

En français moderne, la mise en relief de certains mots ou syntagmes dans un texte s'effectue le plus souvent grâce à des marques typographiques telles que les italiques, le gras, la mise en majuscules, etc. Les signes de ponctuation ne jouent qu'un rôle périphérique dans cette mise en relief (par exemple, un point d'exclamation entre parenthèses permet de marquer un mot important).

Dans les manuscrits médiévaux, les signes de ponctuation jouaient un rôle plus important en tant que marques de mise en évidence. Les points sont par exemple utilisés pour entourer, et ainsi mettre en relief graphiquement, les chiffres romains (*cf.* les deux premières lignes dans (2.1)). Certains noms propres abrégés pouvaient aussi être entourés de points, qu'on peut considérer dans ce cas comme des marques de mise en relief graphique plutôt que comme simples marques d'abréviation. On peut remarquer également que dans certains manuscrits les noms propres ou titres de noblesse étaient plus ou moins systématiquement marqués par une ponctuation forte (un point suivi d'une majuscule). Ce type de ponctuation est parfois appelé « ponctuation de respect » (Parkes 1992, p. 42 ; *cf.* 70 Careri *et al.* 2001, p. XXXIV et 137). Nous verrons plusieurs exemples de ce phénomène dans notre corpus.

2.3.5. Marquage du discours rapporté

Dans les textes modernes, un dispositif particulier est utilisé pour délimiter et mettre en relief les passages au discours direct et les citations. Il inclut un signe spécifique, une paire de guillemets, qui entourent le passage ; le deux-points qui le précède, s'il existe un lien syntaxique ou sémantique fort avec le contexte gauche ; le tiret et les alinéas dans les dialogues, etc.

Un signe spécifique signalant un discours direct ou une citation n'apparaît que tardivement dans le « répertoire » des marques de ponctuation. Selon M. B. Parkes (1992, p. 58), les guillemets ont pour origine des *notae* marginales utilisées dès le haut Moyen Âge pour attirer l'attention sur un passage important, en particulier dans le cas des citations bibliques dans des textes religieux. C'est à partir de la fin du XVI^e siècle que les imprimeurs commencent à introduire ces marques dans le corps du texte.

Ceci ne signifie pas que les frontières du discours direct n'étaient pas ponctuées : une ponctuation forte ou faible y apparaît le plus souvent, mais, comme le souligne S. Marnette, « ces signes n'ont pas pour fonction principale de signaler le discours rapporté » (2006, p. 42).

Certains manuscrits ont développé des techniques très intéressantes de signalisation de parole rapportée que E. Llamas Pombo (à paraître) appelle la « rubrique énonciative » : une « inscription qui fonctionne comme marque visuelle d'aperture d'un discours rapporté, très souvent redondante par rapport aux moyens strictement linguistiques ». Ce sont des marques segmentales, mises en relief graphiquement (retour à la ligne, encre colorée, etc.), qui ne portent guère plus d'information utile que de « doubler les signes de parole ». Il s'agit donc d'un phénomène qui se situe à la frontière entre les marques segmentales et les « balises graphiques ».

Notre corpus ne présente pas d'occurrences de la « rubrique énonciative », en revanche, d'autres marques segmentales de signalisation du discours rapporté y apparaissent régulièrement. Comme dans le cas des connecteurs, l'abondance de ces marques dans les textes médiévaux n'a pas échappé à l'attention des chercheurs. Bernard Cerquiglini a remarqué que « le discours direct, en prose, au cours de son déroulement, a tendance à reprendre des marques de signalisation, à réitérer son énonciation. » (1981, p. 86). Il ne s'agit pas seulement des verbes de parole, mais aussi d'« une série de morphèmes, coordonnants et adverbes... Ces morphèmes apparaissent en effet à la même place que l'analepse, et souvent en relation d'équivalence avec elle... on qualifiera à bon droit ces morphèmes d'« énonciatifs » » (Cerquiglini 1981, p. 92). Parmi ces morphèmes « énonciatifs » le chercheur cite en particulier les connecteurs suivants : *or, si, car, mais car, et, donc*. S. Marnette (2006, p. 40-41) signale également

la transposition des embrayeurs de personne, temps, etc., la modalisation, les adverbes énonciatifs *oui*, *certes*, etc., les termes d'adresses, etc.

Il convient d'ajouter à cette liste les interjections *ha(a)* et *hé* qui sont très souvent placées au début d'un discours direct dans certains textes. Ces interjections sont par ailleurs très fréquemment accompagnées d'un *comma* < ʔ >, comme dans l'exemple suivant :

- (2.3) *et+se+passe desus lui || menu et souent . et+ausi fait li+nains et les+autres || damoiseles qui estoient .ii . ha ʔ sire cheualiers fait li || nains tant mal auez exploitie . ha ʔ sire cheualiers fait || la+dame tant m'aeuz traie . (lancelot77, 8c:4-8).*

Selon C. Marchello-Nizia (2007), le *comma* dans ce manuscrit (comme sans doute dans beaucoup d'autres) peut avoir une correspondance prosodique (sans doute une intonation montante). Cette hypothèse est tout-à-fait plausible, mais difficile à prouver faute de données directes sur la prononciation médiévale. En revanche, le fait que la séquence « *ha* + *comma* » est un marqueur fort du discours direct ne fait aucun doute.

2.3.6. Marquage de l'interrogation ou de l'exclamation

Le *punctus interrogativus* faisait partie des *positurae* introduites dans l'écriture de certains textes liturgiques dès le 8^e siècle (Parkes 1992, p. 35), et il est mentionné dans la majorité des traités médiévaux sur la ponctuation. Or son usage dans les manuscrits médiévaux français est rare. Certains manuscrits ne connaissent pas du tout l'usage de ce signe, d'autres l'utilisent parfois, mais jamais de façon systématique. Dans son édition de la *Queste del saint Graal* (ms. Lyon, BM, Palais des arts 77) C. Marchello-Nizia a relevé 72 occurrences du point d'interrogation dans le manuscrit, sur 182 phrases interrogatives. E. Llamas Pombo (2001, p. 163) note la polyvalence de ce signe qui peut marquer l'interrogation aussi bien que l'exclamation, mais dans notre corpus nous n'avons relevé aucune occurrence du *punctus interrogativus* exclamatif. Le terme même d'exclamation fait référence à un phénomène prosodique, sans correspondance verbale, il est par conséquent très difficile de qualifier d'« exclamatif » l'emploi de telle ou telle marque de ponctuation dans un texte ancien.

Un signe spécifique destiné à marquer l'exclamation, le *punctus admirativus*, a été inventé par les humanistes italiens au XIV^e siècle et ne se répand en France que dans les imprimés du XVI^e siècle (Parkes 1992, p. 49). Il est notamment mentionné dans le traité d'E. Dolet (1540).

En ce qui concerne les marques segmentales d'interrogation (pronoms et adverbes interrogatifs, structures syntaxiques spécifiques), leur rôle n'a pas considérablement changé au cours des siècles, même si dans les textes anciens ces marques signalaient parfois à elles seules l'interrogation, tandis que dans les textes modernes elles sont toujours « appuyées » par la ponctuation.

2.3.7. Vue d'ensemble

Dans le *Tableau 1* ci-dessous, nous avons tenté de résumer le rôle des marques segmentales et des « balises » graphiques en général, et des connecteurs et signes de ponctuation en particulier dans l'organisation du texte dans les manuscrits médiévaux et dans les imprimés modernes. Nous avons utilisé le signe « + » si les marques d'un type donné remplissent systématiquement telle ou telle fonction. Le signe « - » est utilisé quand les marques de ce type ne jouent pas de façon générale un rôle considérable pour une fonction donnée. Enfin, si un type de marque peut occasionnellement remplir une certaine fonction sans devenir le moyen principal de son expression, nous avons mis le signe « ± ».

Si nous n'avons pas constaté de changement considérable entre les systèmes ancien et moderne, nous n'avons mis qu'un seul signe. Si, en revanche, il y a eu un changement, nous indiquons d'abord le signe correspondant à l'état ancien, puis une flèche (→) et enfin le signe correspondant à l'état moderne.

Fonction	Segmental	Connecteur	« Balise » graphique	Signe de ponctuation
Marquage d'unités intratextuelles				
– séparation	±	+ → ±	+	± → +
– relations	±	+	–	– → ±
– hiérarchie	±	± → –	+	± → ±
– mise en relief	±	–	+	+ → ±
Marquage énonciatif				
– discours direct / citation	+ → ±	± → –	– → +	– → +
– interrogation	±	–	– → +	±/– → +
– exclamation	–	–	– → +	±/– → +

Tableau 1 : Fonctions des marques segmentales et graphiques dans les manuscrits médiévaux et dans les textes modernes

Le rôle joué par les marques de ponctuation est devenu plus important dans les textes modernes par rapport aux manuscrits médiévaux dans toutes les fonctions, à l'exception de la mise en relief de mots ou de syntagmes particuliers, où la ponctuation au sens propre a été remplacée par d'autres marques typographiques (italiques, gras, etc.). Les marques segmentales ont gardé sensiblement le même rôle dans l'expression des relations entre des unités textuelles, dans la mise en relief d'éléments textuels et dans l'interrogation. Mais elles jouent dans les textes modernes un rôle moins grand dans la séparation des unités textuelles, dans l'établissement de leur hiérarchie et dans le marquage du discours direct.

2.4. La notion d'« unité ponctuable »

La fonction de marquage des frontières des unités intratextuelles étant centrale pour les marques de ponctuation, l'essentiel de notre travail sera consacré à l'analyse de la nature des unités qui sont effectivement ponctuées dans notre corpus.

Pour commencer à cerner l'objet de l'« unité de lecture » (Marchello-Nizia 1978), ou, autrement dit, de l'*unité ponctuée*, nous nous proposons de partir d'une notion plus large, celle de l'*unité ponctuable*.

L'unité ponctuable est un objet artificiel à vocation méthodologique. Son but est de permettre le repérage des frontières syntaxiques sur lesquelles on peut s'attendre à l'apparition d'une marque de ponctuation quelconque. Une marque de ponctuation réelle peut être présente ou absente à une frontière ponctuable.

On définit l'unité ponctuable comme le plus petit segment textuel susceptible d'être délimité à gauche et/ou à droite par une marque de ponctuation. Les limites des unités ponctuables correspondent toujours à des limites d'unités syntaxiques d'un niveau plus ou moins élevé.

Cette notion se rapproche de celle de *chunk* ('gros morceau') largement utilisée en linguistique informatique : pour la reconnaissance et la compréhension de la parole, l'analyse syntaxique automatique, etc. (Schadle *et al.* 2004). Le terme a été proposé par S. Abney (1991) qui prenait comme point de départ la segmentation intuitive de phrases au cours de la lecture :

These chunks correspond in some way to prosodic patterns. It appears, for instance, that the strongest stresses in the sentence fall one to a chunk, and pauses are most likely to fall between chunks. Chunks also represent grammatical watershed of sorts. The typical chunk consists of a single content word surrounded by a constellation of function words, matching a fixed template (Abney 1991, p. 257)⁷⁵.

Le chercheur s'appuie dans ses réflexions sur la notion de « structures de performances » découverte grâce aux expériences psycholinguistiques (Gee *et al.* 1983) et essaie de lui donner une dimension plus syntaxique (dans le cadre de la grammaire générative).

Notre notion d'unité ponctuable est plus pragmatique : elle s'appuie sur l'analyse des emplois des marques décrits dans les recherches publiées et/ou attestés

75 Notre traduction : « Ces *chunks* correspondent dans une certaine mesure à des unités prosodiques. Il semble, par exemple, que les accents toniques les plus forts dans une phrase se distribuent un à un par *chunk*. Les *chunks* représentent également des ensembles de traits grammaticaux. Un *chunk* typique est composé d'un mot lexical entouré d'une constellation de mots grammaticaux, le tout étant organisé selon un modèle défini ».

Cadre théorique et méthode pour l'analyse linguistique

lors de nos premières expériences d'annotation du corpus⁷⁶. Nous avons tâché de qualifier de la façon la plus adéquate possible la nature des unités que la ponctuation sert à séparer en prenant en compte la structure syntaxique formelle, aussi bien que les facteurs énonciatifs, sémantiques et pragmatiques.

Les facteurs rythmiques et intonatifs ne sont pas pris en compte. Ceci se justifie par fait que notre corpus comprend uniquement des textes en prose, pour lesquels nous ne possédons aucun indice (mis à part la ponctuation elle-même) qui puisse nous renseigner sur le rythme et la structure mélodique de l'énoncé. Bien entendu, pour les textes en vers ou en prose rythmique, ces facteurs jouent un rôle capital.

Ce mélange de divers facteurs dans un jeu d'étiquettes unique se justifie selon nous, d'une part, par la simplicité technique de l'annotation « plate » et, d'autre part, par la nature même des marques de ponctuation qui remplissent de multiples fonctions et qui « naissent » au croisement de structures concurrentes.

Après avoir dressé la liste des unités, dont les frontières sont potentiellement ponctuables, nous annotons systématiquement ces frontières dans notre corpus, en indiquant le type de frontière et la présence ou l'absence effective d'une marque de ponctuation.

Dans la pratique, l'unité ponctuable est assez souvent supérieure à un *chunk*. Ainsi, nous faisons l'hypothèse qu'une « phrase simple » composée d'un verbe, d'un sujet et des compléments essentiels ne forme qu'une seule unité ponctuable.

Les unités de la structure textuelle forment une certaine hiérarchie, et nous admettons qu'une frontière d'un niveau supérieur « absorbe »⁷⁷ la ou les frontières de niveaux inférieurs situées à un même endroit dans le texte. Dans le passage suivant, par exemple, la frontière devant *Et il monte* correspond à la fin d'une deuxième relative coordonnée et au début d'une nouvelle proposition autonome⁷⁸ :

(2.4) *et troeue cheualiers et escuiers ius_qu'a .xv. qui || l'atendoient . et+estoient uenuz o+lui*
| *Et il monte et se || part de+la+cort (qgraal77, 161d:18-20).*

76 Une approche similaire a été adoptée par C. Pignatelli (2007) dans son étude de la ponctuation dans les manuscrits du *Chevalier à la charrette* de Chrétien de Troyes. Il s'agit d'identifier des « lieux de ponctuation » et d'étudier les pratiques des différents scribes en matière de marquage de ces structures textuelles.

77 Nous empruntons ce terme à C. Tournier (1980, p. 39) et à G. Nunberg (1990, p. 57).

78 Nous évitons l'utilisation du terme « phrase » pour des raisons que nous expliquerons dans la section 2.5. ci-dessous.

La frontière entre propositions autonomes est supérieure à celle d'une subordonnée, c'est donc l'annotation « frontière entre propositions autonomes » qui sera utilisée.

Toutefois, puisque notre classement prend en compte plusieurs types de facteurs, il n'est pas toujours possible d'affirmer quel type de frontière est supérieur dans un cas concret. Voici quelques exemples :

(2.5) Veraiement le+sachiez uos fet boort || que ce est il (*qgraal77, 162a:14*) ;

(2.6) tuit li huis dou palais ou il men||ioient . ||[fol. 161c] et+les fenestres clostrent (*qgraal77, 161b:40-161:c1*) ;

(2.7) Et+qui que || il soit ou cil que uos dites ou autres ꝛ ie uoldroie || que bien li+uenist (*qgraal77, 161c:27*).

Dans (2.5), la fin d'une incise *fet Boort* coïncide avec le début d'une complétive. On ne peut pas dire *a priori* que l'incise soit une unité de niveau supérieur à une complétive (d'autant plus que les données du corpus montrent que la ponctuation est beaucoup plus fréquente aux frontières des subordonnées qu'à celles des incises). Dans (2.6), nous avons une coordination de sujets, dont le premier est doté d'une relative circonstancielle : des frontières entre propositions sont normalement supérieures à celles entre syntagmes coordonnés, mais ici la subordonnée est gouvernée par l'un des syntagmes coordonnés. Dans (2.7), nous avons un cas semblable avec une coordination de deux pronoms, dont le premier est antécédent d'une relative. Dans ces cas de doute ou d'ambiguïté, nous avons utilisé une étiquette double, en choisissant tout de même une frontière dominante, mais en nous réservant le droit de revenir sur cette analyse lors de l'examen plus fin des occurrences.

Au début de notre travail d'annotation, nous avons élaboré une « pré-grille »⁷⁹ de classement des frontières ponctuables basée sur les grammaires descriptives, et en particulier (Riegel *et al.* 1998), ainsi que sur les recherches antérieures consacrées à la ponctuation française médiévale et sur notre propre expérience.

Nous avons commencé par distinguer les « grandes catégories », à l'intérieur desquelles nous avons ensuite dressé la liste des types de frontières précis. Cette liste

⁷⁹ Terme proposé par S. Heiden.

« fine » a dû être complétée et précisée au cours du travail. L'analyse des données nous a permis par ailleurs de rassembler plusieurs « petites » catégories dans des groupes relativement homogènes.

De façon générale, nous avons tâché de respecter le principe hiérarchique dès l'étape de l'attribution des étiquettes : les frontières « supérieures » viennent d'abord. Dans les cas où les relations hiérarchiques ne pouvaient pas être établies, nous avons suivi l'ordre de présentation de notre ouvrage de référence (Riegel *et al.* 1996).

Avant de passer à la description des catégories retenues dans notre annotation, il convient de justifier notre choix de ne pas recourir à la notion de phrase.

2.5. « Phrase » : notion inapplicable

La pertinence de la notion de phrase par rapport aux textes médiévaux a été mise en doute par plusieurs chercheurs. Ainsi, C. Buridant n'y recourt pas dans sa *Grammaire nouvelle de l'ancien français* (2000). C. Marchello-Nizia précise dans une note de son étude pionnière de la ponctuation dans un cadre linguistique :

J'utilise le mot "proposition" plutôt que "phrase simple" ou "phrase minimale" qui, en linguistique, désigne proprement une unité "abstraite", et non une portion d'énoncé découpée empiriquement, comme c'est le cas ici ; ces propositions que je repère d'après les critères de la grammaire traditionnelle (présence d'un verbe conjugué), correspondent à des "phrases minimales" sans doute, soit autonomes, soit reliées les unes aux autres, soit enchâssées les unes dans les autres ; mais aussi longtemps qu'on n'a pas élaboré une grammaire générative de ces états anciens de la langue, il semble prudent d'éviter d'employer un terme qui suppose cette élaboration achevée (Marchello-Nizia 1978, p. 36).

B. Combettes ne situe qu'au XVI^e siècle l'émergence de cette unité dans l'usage et dans la pensée linguistique (2000, p. 63). Une observation semblable est faite par G. Nunberg au sujet de textes anglais : une « phrase textuelle » ne se forme en tant que catégorie grammaticale que sous l'influence de l'imprimerie, sans doute pas avant le XVII^e siècle :

... whenever the text-sentence can be said to have made its official debut, certainly there is no analogous unit in the syntax of medieval and Renaissance texts (1990, p. 129-130)⁸⁰.

Il est en effet très difficile de définir de façon formelle cette unité dans un manuscrit médiéval, car dans les textes modernes c'est justement la ponctuation qui permet à l'auteur d'exprimer et au lecteur de repérer les frontières des phrases. L'intonation est souvent évoquée comme un critère définitoire de la phrase en langue parlée. Ces critères sont bien entendu inapplicables à notre travail où nous n'avons aucun accès aux données prosodiques et où les unités doivent être annotées d'une façon entièrement indépendante de la ponctuation.

N. Mazziotta (2007b, p. 97) propose de définir la phrase à partir de la présence d'un « verbe principal » (et un seul). Cela l'amène à éliminer tout ce qu'on considère traditionnellement comme des phrases « non verbales » et à poser des limites de phrases entre deux verbes, même si ceux-ci partagent l'ensemble du schéma actanciel et le cadre spatio-temporel, comme dans l'exemple suivant :

(2.8) Si le servent et honorent de quan qu' il pueent (*qgraal77, 161d:34*).

Si la définition de N. Mazziotta a l'avantage d'être précisément formulée et facilement applicable dans l'annotation du corpus, elle s'éloigne considérablement de la notion de phrase traditionnelle et impose un découpage souvent contre-intuitif et, comme l'analyse du corpus nous le montrera, contraire aux données attestées.

Nous avons donc décidé de renoncer à la notion de phrase dans notre système de frontières ponctuables et de nous arrêter, au niveau des structures purement syntaxiques, à la *proposition* que nous définissons de la façon suivante, en nous inspirant des travaux dans le domaine de la syntaxe sémantique :

Une structure linguistique composée d'un *prédicat* et d'un certain nombre de *participants* (imposés par le prédicat) et de *circonstants* (facultatifs).

80 Notre traduction : « quel que soit le moment à partir duquel on peut considérer que la phrase textuelle a fait son début officiel, il est certain qu'il n'existait pas d'unité équivalente dans la syntaxe des textes médiévaux et de la Renaissance ».

Le prédicat est typiquement exprimé par un verbe, mais il peut être omis dans le cas où la juxtaposition des participants est suffisante pour exprimer leurs relations⁸¹ ou s'il est « emprunté » à la proposition précédente. Les participants et les circonstants peuvent être déterminés ou modifiés par des mots grammaticaux, lexicaux ou encore par des propositions subordonnées.

Une proposition peut être autonome ou entretenir des liens plus ou moins forts avec d'autres propositions. Ces liens peuvent être formels (subordination, partage d'éléments) ou sémantiques.

Nous donnerons des exemples précis des différents cas possibles dans la section suivante qui débutera par la présentation générale des « grandes catégories » des frontières ponctuables.

2.6. Classement des frontières ponctuables

Parmi les grandes catégories, nous avons distingué :

- (A)⁸² les frontières entre les unités de la macrostructure textuelle ;
- (B) les frontières liées au changement de plan énonciatif et, en général, tout ce qui est lié au discours direct et à la citation (les incises, les incidentes, les interjections et les mots-phrases *oïl* et *non*, etc.) ;
- (C) les frontières entre les propositions indépendantes (juxtaposées ou coordonnées) ;
- (D) les frontières entre les propositions subordonnées et leurs principales ;
- (E) les frontières entre les syntagmes coordonnés ou juxtaposés à l'intérieur d'une proposition ; les débuts et fins d'énumération ;
- (F) les syntagmes particuliers potentiellement « ponctuables » à l'intérieur de propositions (compléments circonstanciels, appositions, reformulations, etc.).

81 Ce cas est extrêmement rare en français médiéval et se limite pratiquement aux proverbes du type *apres le sangler le mire . et apres le cerf la biere (chasse619, 4r:8)*.

82 Les lettres qui figurent dans cette liste sont celles que nous avons utilisées dans les étiquettes. Il s'agit d'une simple liste alphabétique qui facilite les tris automatiques.

La dernière catégorie a aussi servi à marquer les occurrences « atypiques », parfois difficiles à expliquer. Il s'agit des emplois de marques de ponctuation à des endroits « inattendus », en dehors des frontières ponctuables définies *a priori*. Nous verrons que les occurrences de ce type sont très rares (sinon, nous devrions revoir notre grille d'analyse !) et s'expliquent le plus souvent par des facteurs indépendants de la structure linguistique (la « ponctuation de respect », la justification des fins de ligne, etc.).

Dans les sections suivantes, nous présenterons les catégories « fines » à l'intérieur de chaque grande catégorie (avec éventuellement des regroupements que notre analyse a permis d'effectuer), en fournissant des exemples concrets tirés de notre corpus.

Un seul détail reste à préciser avant de commencer la présentation des catégories fines : selon la nature des relations entre les unités séparées par une frontière, on peut distinguer d'une part des frontières « horizontales », quand les deux unités se situent à un même niveau hiérarchique (par exemple, des propositions juxtaposées ou des syntagmes coordonnés) et qu'il s'agit d'une simple démarcation, et d'autre part des frontières « verticales » quand l'une des unités est d'une certaine façon inférieure à une autre (proposition subordonnée, incise, syntagme particulier). Dans le deuxième cas, il convient de distinguer les occurrences où l'unité inférieure suit l'unité supérieure des occurrences où elle la précède. Nous utilisons une étiquette simple quand l'unité inférieure suit la supérieure (par exemple, **b3** au début d'une incise après le discours direct) et une étiquette « suffixée » **-f** dans le cas contraire (par exemple, **b3f** à la fin d'une incise, avant la reprise du discours direct). Dans la présentation générale, ces étiquettes « suffixables » seront dotées d'un astérisque (par exemple, **b3***).

2.6.1. Macro-unités textuelles (A)

Les unités macro-textuelles relèvent du « troisième niveau » de la ponctuation, selon C. Tournier (1980) et N. Catach (1994). Contrairement aux autres grandes catégories, l'annotation des frontières de ce niveau ne peut être entièrement indépendante de la ponctuation : les grandes divisions du texte ne se révèlent que grâce

Cadre théorique et méthode pour l'analyse linguistique

à la mise en page décidée par l'auteur ou par l'éditeur (le scribe joue pleinement le rôle d'éditeur dans ce cas).

Il convient de noter qu'un grand nombre d'éditeurs modernes n'hésitent pas à intervenir à ce niveau macrotextuel de la ponctuation et restructurent les paragraphes, voire les chapitres et les divisions encore plus importantes de leurs manuscrits de base. Il est vrai que différents manuscrits dans la tradition d'une même œuvre peuvent parfois présenter des variantes importantes dans la structuration du texte, mais il est rare que les grandes divisions graphiques d'un « bon » manuscrit soient complètement aberrantes par rapport à la logique de l'œuvre.

Théoriquement, la hiérarchie des divisions textuelles peut entrer en conflit avec celle des unités linguistiques. L'exemple le plus célèbre est le sixième chapitre de *Don Quichotte* de Cervantes, qui commence par une relative subordonnée au dernier syntagme nominal du chapitre précédent. Il convient toutefois de noter que le pronom relatif *el qual* ('lequel') utilisé en l'occurrence ne crée pas forcément une relation de subordination forte, et les deux propositions sont relativement autonomes sur le plan purement syntaxique. Certains éditeurs modernes tentent de rétablir la « phrase complexe », en mettant une virgule à la fin du chapitre V et une minuscule au début du chapitre 6 (par exemple, Sevilla 1999), ce qui fait qu'on peut parler dans ce cas d'indépendance entre les niveaux de ponctuation. Dans le système graphique médiéval, une telle autonomisation paraît impossible : une frontière de division textuelle implique forcément une ponctuation forte (usage d'une lettrine ou d'une majuscule) au niveau « syntaxique ». C'est d'ailleurs le cas de la première édition de *Don Quichotte* (Madrid, 1605), comme nous pouvons le voir sur l'illustration suivante (Fig. 3)⁸³.

83 L'image provient du portail Internet <<http://www.cervantesvirtual.com>>.

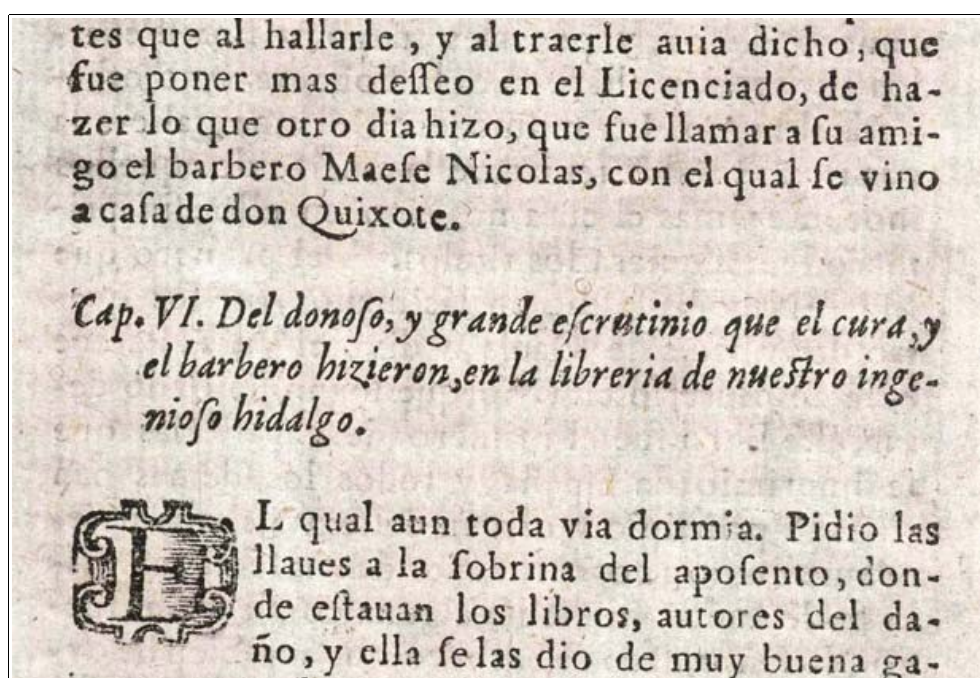


Fig. 3 : Début du chapitre 6 de *Don Quichotte* (Madrid, Juan de la Cuesta, 1605, f. 15).

*Traduction*⁸⁴ : C'était donner au licencié plus de désir encore de faire ce qu'en effet il fit le lendemain, à savoir : d'aller appeler son ami le barbier maître Nicolas, et de s'en venir avec lui à la maison de don Quichotte...

[chapitre 6]

... Lequel dormait encore.

L'analyse extensive de la ponctuation macrotextuelle dépasse l'objet de la présente étude. Elle aurait d'ailleurs nécessité la constitution d'un corpus très différent du nôtre : les petits extraits de quelques centaines ou milliers de mots ne contiennent que très peu de frontières de ce type. Néanmoins, nous annotons ces frontières quand nous les rencontrons et donnons un bref aperçu de la structuration globale du texte au début de l'analyse détaillée de chaque extrait dépouillé.

A côté des grandes divisions textuelles, la catégorie « macrotextuelle » comprend les frontières des « unités de service » qui sont elles-même en quelque sorte des « balises graphiques » : il s'agit des titres, des rubriques de toutes sortes, des *explicit*, etc.

⁸⁴ La traduction provient de la bibliothèque virtuelle du projet WikiSource <<http://fr.wikisource.org>>.

Cadre théorique et méthode pour l'analyse linguistique

Les catégories « fines » que nous distinguons sont les suivantes : division textuelle majeure, y compris le début du texte (**a1**) ; nouvel épisode (**a2**) ; introduction, titre descriptif (**a3***) ; conclusion (**a4***) ; titre court (**a5***) ; explicit ou signature (**a6***) ; didascalie ou identification du personnage qui parle (**a7***).

Le choix de l'étiquette **a1** est dicté par la mise en page de notre source : on l'utilise pour les divisions marquées par de grandes lettrines, souvent dotées d'un titre et d'un nom de division, comme c'est le cas du début de chapitre suivant dans l'*Image du monde*⁸⁵ :

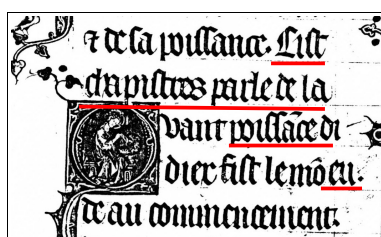


Fig. 4 : Début d'un chapitre dans l'image du monde (im574, 4v).

- (2.9) et de sa poissance . [a5] Cist
chapistres parle de la || poissance di||eu .
[a1] Qvant
diex fist le monde
de au commencement . (im574, 4c:7-11)⁸⁶.

Le choix de l'étiquette **a2** est aussi influencé par la mise en page de la source (la présence d'une lettrine ou d'un pied-de-mouche), mais il doit être appuyé par l'analyse de la séquence textuelle : un nouvel épisode doit correspondre à un changement du cadre spatio-temporel dans une narration ou à un changement de thème dans une séquence argumentative.

Les étiquettes **a3** et **a4** prévues par notre schéma initial n'ont en réalité pratiquement pas été utilisées dans l'annotation du corpus.

L'étiquette **a5** a été utilisée pour marquer la frontière d'un titre où la ponctuation est plus faible (la frontière fortement ponctuée est « absorbée » par celle de la division textuelle). Contrairement aux imprimés modernes, le titre dans les manuscrits peut être

85 Nous avons souligné le titre (écrit en rouge dans l'original).

86 Notons que dans la transcription nous rétablissons la linéarité « logique » du texte, même si dans le manuscrit le titre « se croise » avec le début du chapitre.

graphiquement rattaché davantage à la division textuelle précédente qu'à celle qu'il annonce (cf. Fig. 4 ci-dessus).

L'étiquette **a6** a été appliquée une seule fois dans notre corpus : devant *Amen* à la fin du texte de Jean Juvenal des Ursins (**ursins2701**). Cette rareté s'explique non pas par l'absence des unités de ce type dans les manuscrits médiévaux, mais par le fait que notre corpus ne comprend pas d'extraits tirés des fins de textes (à l'exception précisément de **ursins2701**, une petite œuvre transcrite en entier).

L'étiquette **a7**, elle aussi, n'est représentée que par une seule occurrence dans notre corpus :



Fig. 5 : Mise en forme de l'identification d'un personnage parlant (« rubrique énonciative ») (*quadrilogue*126, 192b).

- (2.10) **[a5]** Incipit quadrilogium inuectiuum et .
 comicum ad morum gallicorum correctionem
[a7] ¶ L'Atteur commence .
[a1] Environ l'aube du Jour (*quadrilogue*126, 192b:34).

L'exemple cité permet également d'observer l'usage des étiquettes **a5** et **a1**.

La dernière étiquette placée dans cette catégorie ne relève pas, à proprement parler, de la structure macrotextuelle, mais la place particulière que les unités concernées occupent dans la chaîne textuelle justifie à notre avis ce rapprochement (qui n'a de toute façon aucune incidence sur l'analyse des données où ces unités seront traitées à part). Il s'agit des références des sources citées (les textes bibliques, par exemple) dans des textes religieux ou didactiques (**a8***). Les citations elles-mêmes sont assimilées au discours direct. Une double étiquette est utilisée si la citation suit la référence, car il est difficile de déterminer la « supériorité » d'une frontière par rapport à l'autre dans ce cas :

Cadre théorique et méthode pour l'analyse linguistique

- (2.11) car malice empesche a || congnoistre dieu et verite **[a8]** sapiencie .ij. **[b1-a8f]** Excauit
|| eos malicia eorum. (*gerson13318, 167r:12*)⁸⁷ ;
- (2.12) Et pour ce+dit calciodorus **[a8]** libro iij^{ro}. *epistolarum* || **[b1-a8f]** decet cunctos patrie
sue augmenta cogitare / (*ursins2701, 121c:4*).

Dans notre corpus, deux textes seulement (**gerson13318** et **ursins2701**) présentent des occurrences de ce type de frontière.

2.6.2. Frontières liées au discours rapporté (B)

Cette grande catégorie regroupe des frontières de niveaux hiérarchiques différents. Les étiquettes **b1** et **b2** sont utilisées pour marquer les débuts et les fins de passages au discours direct (ou de citations) respectivement⁸⁸. Selon les textes, les passages au discours direct sont plus ou moins nombreux : nous avons relevé 133 occurrences de ce type de frontières dans notre corpus.

L'étiquette **b3**⁸⁹ est utilisée pour marquer les frontières des incises insérées au milieu du discours direct (2.13) ou le suivant immédiatement (2.14) :

- (2.13) Certes **[b3]** fet mes || sires *Gauuains* **[b3f]** ce+me plest mout . (*qgraal77, 160d:9-10*) ;
- (2.14) or+uos+en+souuiegne **[b3]** fet *Keu*
[a2] **E**Ndementres qu'il parloient einsi (*qgraal77, 161a:5*).

Une centaine de frontières d'incises ont été relevées dans notre corpus.

L'étiquette **b5*** est utilisée aux frontières de propositions incidentes. Les incidentes ne sont pas forcément liées au discours direct ou à un changement du plan énonciatif, mais elles sont souvent rapprochées des incises dans les grammaires descriptives, ce qui a motivé le choix de l'étiquette. Nous définissons les incidentes d'une façon restrictive : ce sont des propositions qui s'insèrent, sans marque segmentale de dépendance, au milieu d'une autre proposition (2.15) ou entre une principale et subordonnée (2.16) :

87 Le soulignement rouge provient du manuscrit.

88 Il aurait pu être logique de remplacer l'étiquette **b2** par **b1f**, mais nous avons préféré garder l'étiquette créée avant l'introduction du suffixe, parce l'« infériorité » du discours direct par rapport aux structures syntaxiques (propositions) est loin d'être évidente.

89 L'étiquette **b3f** a remplacé **b4**, qui était utilisée initialement et qui est devenue non attribuée dans l'état du corpus actuel.

- (2.15) En || l'autre annee [b5] bien m'en souvient [b5f] Je parlay de trois mirouers
(*gerson13318, 168r:8*) ;
- (2.16) L'ystoire dit que le noble conte d'artois || aprez ce qu'il eubt par longue espasse || de
tempz poursieuy lez armez et que || plus ne trouuoit a soy employer / [b5] Car aussy ||
ne puet paz tousiours estre guerre / [b5f] Il se mist || au Retour vers sa conte d'artois
(*artois11610, 2v:22-27*).

Les précisions et reformulations non propositionnelles portent l'étiquette **f4*** (cf. 2.6.6. ci-dessous). Les frontières d'incidentes sont peu fréquentes dans notre corpus, on en compte à peine une quinzaine d'occurrences.

L'étiquette **b6** est utilisée dans les dialogues, en cas de changement de locuteur :

- (2.17) [Boort] dist [b1] biau frere sauez uos qui cist *cheualiers* || est qui siet ou siege
perilleux . [b6] Je+nel sai+mie+bien+ [b3] fet || lion ! (*qgraal77, 162a:9*).

L'étiquette **b7*** est utilisée à la frontière entre un syntagme nominal en apostrophe et le reste du discours direct :

- (2.18) Certes [b7] sire [b7f] ne ele n'est moie ne ie n'auroie le+corage || de metre+i+main
(*qgraal77, 161a:26*).

L'étiquette **b8*** était destinée à marquer les frontières d'interjections (*Ha, O*, etc.), mais dans notre corpus, ces frontières (déjà peu nombreuses) sont toujours « absorbées » par des frontières « supérieures », comme celles d'incises ou d'apostrophes :

- (2.19) [b6] ha : [b7] sire *cheualiers* fait || la+dame tant m'auez traie (*lancelot77, 8c:7*).

Ce choix de hiérarchisation n'est pas incontestable, mais le nombre de formes graphiques des interjections étant limité, il est toujours possible dans notre corpus de faire des requêtes sur ces formes et d'étudier la ponctuation qui les entoure (ce que nous avons fait pour *Ha*, que nous considérons comme une « balise graphique » du discours direct).

La dernière étiquette dans cette catégorie, **b9***, est utilisée aux frontières de « mots-phrases » 'oui' et 'non' si celles-ci ne sont pas « absorbées » par des frontières supérieures. Dans l'exemple suivant, où ces deux mots sont utilisés comme objets du verbe *dire* dans une expression figée, l'étiquette **b9** n'apparaît qu'une seule fois, car la

relation dominante pour la frontière après *vvy* est la coordination et la frontières apres *nanil* est celle du début de discours direct :

(2.20) A+quoy jl Respondy / sans dire ||[b9] vvy / [e2] ou nanil / [b1] par foy sire jl souffira [...](*artois11610, 4r:11*).

2.6.3. Propositions juxtaposées ou coordonnées (C)

Deux groupes distincts se forment à l'intérieur de cette grande catégorie : les propositions relativement autonomes (dotées de tous les éléments structurels obligatoires) et les propositions qui soit partagent un élément avec une autre proposition (qui se situe habituellement dans le contexte gauche), soit sont subordonnées à une même proposition principale ou à l'un de ses éléments.

2.6.3.1. Propositions autonomes (Ca)

Pour des raisons théoriques et en nous basant sur des observations empiriques, nous avons décidé de baser le classement fin des propositions autonomes sur le degré de la continuité thématique exprimé par la forme du sujet. Nous analysons la forme du sujet dans les deux propositions qui se trouvent immédiatement à gauche et à droite de la frontière, quelle que soit la « profondeur » de leur imbrication dans des constructions subordonnées éventuelles. La plus grande autonomie concerne les propositions avec un sujet nominal, dont le référent n'est pas présent sous une forme ou une autre dans la proposition qui la précède immédiatement (**c1**). C'est une frontière assez proche de la limite de phrase dans la perception du lecteur moderne. Plus de la moitié des occurrences (255 sur 498) de cette frontière portent une ponctuation forte (point, point d'interrogation ou point d'exclamation) dans les éditions modernes des textes de notre corpus, sans compter les 75 occurrences de ponctuation moyenne (point-virgule ou deux-points)⁹⁰. Dans les manuscrits cette frontière est également très « favorable » à la présence d'une ponctuation, mais des exceptions existent. Dans l'exemple suivant, la

⁹⁰ Il convient de souligner qu'il s'agit d'éditions modernes. Dans les manuscrits, ce n'est pas la forme de la marque, mais la variante de la lettre qui suit qui définit selon nous la force de la ponctuation. La plupart des manuscrits de notre corpus ne connaissent pas la « ponctuation moyenne ».

frontière de ce type n'est pas ponctuée dans le manuscrit, mais l'éditeur (Micha 1979) a choisi d'y mettre une ponctuation forte :

- (2.21) *et quant il le+uoient si l'acolent et besent [c1][>.]*⁹¹ *et lors com||ence la ioie que li+cosin firent li .i. de+l'autre .* (*qgraal1119, 138b:3*).

Dans certains cas, les éditeurs ont jugé que le lien entre les propositions correspondantes était suffisamment fort pour n'utiliser qu'une ponctuation faible, voire ne mettre aucune ponctuation :

- (2.22) *bien le || pooit len ueoir [c1][> ,] quar || ses cheuax en fu en||core touz tressuez* (*qgraal1119, 138a:14*) ;
 (2.23) *lors monte la dame [c1][.>] || et+li+nains destent les pauillons* (*lancelot77, 8d:12*).

L'étiquette **c2** est utilisée si le référent du sujet nominal (2.24), pronominal (2.25) ou nul (2.26) de la proposition qui suit la frontière est présent dans la proposition précédente :

- (2.24) *mes il n'ot cheualier laienz || qui seust par ou+il [=li preudons] i+entra .[c2]*
et+li+preudons uenoit || a+pie (*qgraal77, 161c:12*) ;
 (2.25) *ne ele n'est moie [c2] ne ie n'aurioie le+corage || de metre+i+main* (*qgraal77, 161:26*) ;
 (2.26) *et il monte et la+damoiselle ausint [c2] si se+par||tent de laienz sanz autre congie*
(qgraal1119, 138a:35).

Dans les éditions modernes, les frontières de ce type portent une ponctuation forte dans un tiers des cas environ (63 occurrences sur 191).

L'étiquette **c3** s'applique aux frontières où le sujet pronominal personnel de la deuxième proposition renvoie au même référent que celui de la proposition précédente :

- (2.27) *Si met la main a l'espee et+tret .[c3] mes il || ne+la+puet auoir* (*qgraal77, 161b*).

Cette étiquette est également utilisée, sous certaines conditions précises, dans les cas où le sujet de la deuxième proposition n'est pas exprimé. Nous reviendrons sur ces cas lorsqu'il s'agira de l'étiquette **c5**.

91 Conformément à la convention que nous avons adoptée pour la présentation des citations du corpus, nous indiquons, la ponctuation « normalisée » des éditeurs après la marque scribale suivie du symbole >, le tout placé entre crochets.

Cadre théorique et méthode pour l'analyse linguistique

Les éditeurs scientifiques placent une ponctuation forte dans près de 40% des occurrences de ce type de frontière (100 occurrences sur 258).

L'étiquette **c4** est posée si la frontière qui précède une proposition est impersonnelle :

- (2.28) *et+li+rois comande que les tables soient mises .[c4] || car il est tens de mengier (qgraal77, 160d:14) ;*
- (2.29) *n'i+metrai ia la+main .[c4] car ce+seroit || folie se ie tendoie a auoir la (qgraal77, 161a:29).*

Ce type de frontière est nettement moins fréquent que les trois précédents (68 occurrences seulement dans notre corpus). La ponctuation forte y est utilisée dans près d'un tiers des cas par les éditeurs.

Dans la pratique, notre analyse a montré que les quatre types de frontières présentés ci-dessus sont très proches en ce qui concerne l'emploi de la ponctuation dans les sources primaires de notre corpus (tandis qu'une légère différence se dessine dans les éditions), nous les avons donc traités en bloc lors de la description des données observées.

2.6.3.2. Propositions avec des éléments communs (Cb)

La première catégorie fine de ce groupe est celle que nous nous avons dénommée « propositions avec un sujet commun » (**c5**). Il s'agit d'une frontière entre deux propositions, dont la deuxième n'a pas de sujet exprimé et où tout (la forme du verbe et le sens) porte à croire que le sujet est le même que dans la proposition précédente.

Puisque l'expression du sujet en français médiéval n'est pas obligatoire, on peut se poser la question de savoir s'il existe une réelle différence entre cette catégorie et la **c3** (où le sujet, qui renvoie toujours au même référent, est exprimé par un pronom). Prenons l'exemple suivant, où nous avons indiqué nos choix d'étiquetage avec un numéro séquentiel en indice :

(2.30) D'ileuques [Homar]⁹² se traist uers damas [c5₁] *et* assist || la cite [c5₂] *et*+assailli [c5₃] *et* la prist par force [c3₄] Car il || auoit si grant plante de gent . (*tyr*828, *Ib*:24-26).

L'intuition, appuyée par la ponctuation scribale, nous dit que la division à la quatrième frontière est plus forte que les précédentes. La présence du pronom sujet et du connecteur *car*, qui à la différence de *et* ne s'utilise qu'au niveau interpropositionnel, ainsi que le changement du temps verbal sont des marques segmentales de cette autonomie. On peut également apercevoir que le lien entre les propositions *assist la cite* et *assailli* est plus fort qu'entre les autres, car celles-ci partagent l'ensemble du schéma actanciel (l'objet *la cite*, en plus du sujet) et les événements qu'elles décrivent sont simultanés ou presque.

Les facteurs qui peuvent intervenir dans l'appréciation de la proximité entre deux propositions « avec un sujet commun » sont multiples. Pour assurer une procédure d'étiquetage stable et efficace, nous avons élaboré l'algorithme suivant :

On utilise l'étiquette **c3** :

- 1) si un sujet pronominal est présent dans la deuxième proposition (2.27) et (2.30₄) ;
- 2) si le cadre spatio-temporel de la deuxième proposition est différent de celui de la première (ce qui est signalé par des adverbes, des circonstants exprimés par des syntagmes nominaux ou par des propositions subordonnées) (2.31) ;
- 3) si les connecteurs *car* ou *si* sont utilisés (2.30₄), (2.32) et (2.33₂) ;
- 4) si la conjonction *et* reliant les deux propositions est suivie d'un adverbe à valeur temporelle, causale, finale ou concessive (*lors*, *donc*, *pour ce*, *neporquant*, etc.) (2.33₁).

L'étiquette **c5** s'applique dans les autres cas, et notamment :

- 1) si les deux propositions partagent l'ensemble du schéma actanciel (2.34) ;
- 2) si les prédicats des deux propositions n'ont pas d'autres arguments que le sujet (prédicats « nus ») (2.35) ;

92 Le sujet est exprimé 5 lignes plus haut, dans une principale dotée d'une longue série de subordonnées que nous ne citons pas ici pour ne pas complexifier inutilement l'exemple.

Cadre théorique et méthode pour l'analyse linguistique

- 3) si le prédicat de la deuxième proposition est « nu » et peut être interprété soit comme partageant l'objet de la première, soit comme un emploi absolu (2.36) ;
- 4) les deux propositions ont (en plus du sujet) des participants, des circonstants ou des subordinées en commun (2.37) ;
- 5) les deux prédicats ont un ou plusieurs arguments différents, mais aucun indice suffisant pour classer la frontière en **c3** (2.38).

- (2.31) Et l'a||postoles || manda en france au saint || home . **que** il preechast des crois || par s'auctorite **[c3]** Et apres i+enuoia || .ii. sien cardonnal (*villehardouin12203, 69b:24*) ;
- (2.32) il regarde l'enfant **[c3]** si || le+uoit garniz de+totes biautez (*qgraal1119, 138b:19*) ;
- (2.33) si uoit gesir ses .ii. cosins bohort **et** lionel || en .ii. liz **[c3₁]** **et** lors est a+merueilles liez **[c3₂]** si les esueil||le (*qgraal1119, 138b:2*) ;
- (2.34) Si+le+seruent **[c5]** **et** honorent || de quan qu'il pueent (*qgraal77, 161d:34*) ;
- (2.35) la terre tremble **[c5]** crole **[c5]** et fent . (*im1495, 21v:30*)⁹³ ;
- (2.36) [messire Gauvain] prent l'espee par le heut || **[c5]** **et** sache (*qgraal77, 161b:6*) ;
- (2.37) **et**+por+ce || m'en tendrai ie **[c5]** ne n'i+metrai ia la+main (*qgraal77, 161a:29*) ;
- (2.38) Et il monte **[c5]** **et** se || part de+la+cort (*qgraal77, 161d:19*).

Les critères précis choisis pour trancher entre les deux étiquettes appartenant à des groupes différents peuvent se discuter, mais dans la majorité des cas ils permettent d'obtenir des résultats qui s'accordent avec les tendances de la ponctuation dans nos sources primaires aussi bien que dans les éditions modernes. Ainsi, dans les sources primaires, les frontières du type **c3** sont ponctuées dans 75% des cas, tandis que dans la catégorie **c5** ce taux tombe à 36%. Dans les éditions, le pourcentage des frontières ponctuées est plus élevé dans les deux cas, mais la différence entre les catégories est aussi bien marquée (90% des frontières ponctuées pour **c3** contre 48% pour **c5**).

L'étiquette **c6** est utilisée à la frontière entre deux subordinées coordonnées :

- (2.39) cil de+la+sale uirent seoir le+cheualier ou siege || que tant preudome auoient redoute **[c6]** **et** ou || tantes granz auentures estoient auenues par || maintes foiz (*qgraal77, 161d:23*) ;

93 Exemple tiré d'un extrait hors corpus principal.

- (2.40) QVant li+rois fu [uenu el]⁹⁴ mostier .[c6] et il+uit || que lance*lot* fu+uenuz [c6]
et+il+ot+amene boort et+lion || si+lor fet mout grant ioie (*qgraal77, 160d:1-3*).

Cette étiquette est posée uniquement s'il n'y a aucun doute sur le fait que les deux propositions à gauche et à droite de la frontière sont subordonnées à une même principale, par exemple en cas de renouvellement du marqueur de subordination (2.39) ou si les deux subordonnées précèdent la principale ou s'insèrent dans une structure plus grande (2.40).

L'étiquette **c7** est utilisée dans les cas où la deuxième proposition ne contient pas de verbe conjugué (ellipse du verbe, selon la grammaire traditionnelle) :

- (2.41) Et des||dictes liquides les vnes sont conso||nans [c7] les aut*res* de _mi voyeux [c7] et
les || aut*res* mutes (*dictier840, 396b:5-8*) ;
- (2.42) Et li .ii. mistrent les cierges sus la table .[c7] et li tierz || la toaille lez le saint vessel .
[c1] et li quarz tint la lance tote || droite sus le saint vessel (*qgraal77, 224a:23-25*)⁹⁵ ;
- (2.43) Et venismes au+giste ced*ict* soir a+plaisance qui || est grande et+belle cite [c7] De
plaisance a+cremonna || qui est cite grande et belle /[c7] D'illec a+peticolle [c7] apres
|| a+wastala au pont d'ueil (*anglure15217, 28r:3-6*).

Dans (2.41), nous avons une simple omission de la copule, ce qui est le cas le plus fréquent. Dans (2.42), c'est le verbe 'mettre' qui est omis, tandis que tous les participants sont différents dans les deux propositions. Notons que dans la suite de cet extrait, nous avons une frontière **c1**, car il s'agit d'une proposition autonome dotée d'un verbe et d'un sujet nominal (adjectif ordinal substantivé). Dans (2.43), qui est un exemple tiré d'un récit de voyage, nous avons toute une série de séquences composées de syntagmes prépositionnels indiquant le point de départ et d'arrivée d'un déplacement (ou juste le point d'arrivée) éventuellement complétées par des relatives. Nous considérons chacun de ces groupes comme une proposition elliptique, puisqu'il s'agit à chaque fois d'un événement indépendant localisé dans le temps et dans l'espace.

Notons que selon l'analyse de N. Mazziotta (2007b), il ne s'agirait dans aucun de ces cas d'une « phrase » (notion qu'il utilise dans un sens proche de notre « proposition non subordonnée ») : les schémas actanciels démultipliés seront

94 Fragment illisible dans le manuscrit suite à une déchirure.

95 Exemple tiré d'un extrait hors corpus dépouillé.

Cadre théorique et méthode pour l'analyse linguistique

considérés comme des « groupes coordonnés » à l'intérieur d'une même phrase/proposition. Une telle analyse est bien-sûr possible, mais l'hypothèse de l'« effacement » ou du « partage » du verbe nous paraît plus plausible dans les cas comme (2.42) et (2.43).

L'étiquette **c8** est réservée pour les cas, théoriquement possibles, où des propositions avec des sujets différents partagent un même objet ou un autre complément essentiel :

(2.44) *Les journalistes écrivent et le public lit les articles.

Notre corpus ne présente toutefois aucune occurrence de ce type de construction.

Comme c'était le cas des propositions autonomes, les différents types de propositions avec des éléments communs présentent, d'après les résultats de l'analyse de notre corpus, plus de traits communs que de différences.

2.6.4. Propositions principales et subordonnées (D)

Les frontières entre propositions subordonnées et principales ne présentent pas de problèmes particuliers pour un étiquetage. Nous avons subdivisé cette grande catégorie selon les types de subordonnées distingués par la grammaire traditionnelle :

- les circonstancielles (**d1***) ;
- les relatives adjectives (**d2***) ;
- les relatives substantives (**d3***) ;
- les relatives circonstancielles (**d4***) ;
- les complétives (**d5***) ;
- les complétives interrogatives (**d7***)⁹⁶ ;
- les infinitives (**d6***) ;
- les participiales et les gérondifs (**d8***).

⁹⁶ La numérotation des étiquettes ne correspond pas toujours à l'ordre de classement que nous proposons ici, car certaines étiquettes ont été intégrées dans le jeu au cours du travail. La liste alphabético-numérique des étiquettes est disponible dans l'Annexe.

Les constructions infinitives et participiales occupent, en quelque sorte, une position intermédiaire entre les propositions subordonnées et les syntagmes particuliers à l'intérieur d'une proposition.

En ce qui concerne les infinitives, la grammaire scolaire traditionnelle ne les reconnaît en tant que propositions qu'à trois conditions que M. Riegel *et al.* (Riegel *et al.* 1996, p. 497) résumant ainsi : l'infinitif doit avoir un sujet propre et exprimé ; ni l'infinitif ni son sujet ne doivent être prépositionnels ; l'infinitif ne doit pas exprimer le but de l'action. Nous adoptons une définition moins restrictive basée essentiellement sur une analyse sémantique : un groupe infinitif est considéré comme une proposition si on peut parler de deux « situations » (participants et leurs relations) qui se forment autour du verbe principal et de l'infinitif. C'est bien entendu le cas des constructions où l'infinitif a un sujet exprimé indépendamment de celui du verbe principal (avec ou sans préposition), mais aussi de la majorité des infinitifs exprimant le but (2.45). En revanche, ne sont pas traités comme propositions les groupes infinitifs liés à un verbe auxiliaire ou à un verbe principal exprimant l'aspect, la modalité ou la phase de l'action d'un verbe à l'infinitif (2.46), ni *a fortiori* les infinitifs substantivés jouant le rôle du sujet ou autre participant d'une proposition.

(2.45) il+dist || que si+fera+il uolentiers [d6] por fere a+mon *Seignor Gauuain* com||paignie (*qgraal77, 161b:16*) ;

(2.46) Lors comence la damoisele [d6] a faire+le+greig||nor duel+dou monde (*lancelot77, 8c:3*).

Des critères similaires peuvent être appliqués aux constructions participiales : on considère comme propositions les groupes participiaux exprimant des situations distinctes (avec au moins un participant différent par rapport à la proposition principale) (2.47). Les participes apposés sans complément (2.48), ni *a fortiori* les participes utilisés dans des formes temporelles ou aspectuelles composées (2.49) ne sont en revanche traités comme des propositions. Tous les gérondifs sont considérés comme des propositions (2.50).

Cadre théorique et méthode pour l'analyse linguistique

- (2.47) Logique est apres vne sci||ence d'arguer choses faintes et sub||tiles [d8] coulourees de faulx argumens [d8f-d6] pour discerner et mieulx congnoistre || la+verite (*dictier840, 394a:27-30*) ;
- (2.48) Selom son enten||dement en vraies parolles [d8] declae||rees (*joachim768, 97c:18*) ;
- (2.49) [goliast] aloit [d8] disant chascun jour (*anglure15217, 29r:22*) ;
- (2.50) Item [d8] en alant || d'an en an jusques a+l'an .l. || [d8f] sera grant guerre amenuise||ment et mortalite de peuple (*joachim768, 97d:23-26*).

De façon générale, nous avons constaté que toutes les frontières du type principale/subordonnées forment un groupe assez homogène en ce qui concerne les conditions d'emploi des marques de ponctuation, et si un manuscrit présente des variations importantes dans leur traitement, celles-ci ne sont pas forcément liées au type des subordonnées. Nous allons donc, au cours de notre analyse, d'abord étudier cette grande catégorie dans son ensemble et ensuite essayer de dégager les facteurs qui peuvent expliquer la variation éventuelle.

2.6.5. Syntagmes coordonnés ou juxtaposés et énumérations (E)

Nous partons de l'hypothèse selon laquelle la coordination ou la juxtaposition d'éléments de même nature à l'intérieur d'une proposition amène « automatiquement » la création de frontières ponctuables : soit seulement entre les éléments coordonnés ou juxtaposés, soit également à gauche et à droite de l'ensemble des éléments, en cas d'énumération.

La subdivision de cette grande catégorie est conditionnée par deux facteurs principaux : la fonction syntaxique pour les syntagmes nominaux concernés, la nature pour les groupes infinitifs et participiaux (qui ne sont pas des propositions). Une étiquette spéciale est par ailleurs réservée pour les débuts et fins d'énumérations, quelle que soit la nature et la fonction des composants.

Les étiquettes utilisées sont les suivantes :

- sujets (**e1**) ;
- compléments (**e2**) ;
- circonstants (**e3**) ;

- attributs du sujet ou d'objet (**e4**) ;
- épithètes ou modificateurs du syntagme nominal (**e5**) ;
- infinitifs (**e6**) ;
- participes (**e8**) ;
- frontière d'énumération (**e7***).

L'étiquette **e7** (ou **e7f**) est utilisée dans les cas suivants : devant le premier et après le dernier (si la frontière n'est pas « absorbée ») élément d'une série d'au moins trois éléments coordonnés ou juxtaposés (2.51) ; en cas de présence d'un terme généralisant (2.52) et devant une construction binaire avec la répétition du marqueur de coordination (*et... et, ne... ne, ou... ou, que... que, etc.*) (2.53). Les étiquettes correspondant à la nature ou à la fonction des termes énumérés sont utilisées entre ces termes.

- (2.51) Quar aucu||nes gens chassent .[e7] lions .[e2] liepars .[e2] che||uaux .[e2] || et buefs sauuaiges . (*chasse619, 1c:12-13*) ;
- (2.52) si uoit gesir ses .ii. cosins [e7] bohort [e2] et lionel || [e7f] en .ii. liz (*qgraal1119, 138b:1*) ;
- (2.53) si+saluarai+eo • || cist meon fradre karlo .[e7] et in a _iudha .[e3] || et in cad _huna cosa (*strasb9768, 13b:8*).

L'analyse du corpus étiqueté a permis de constater que la longueur des énumérations et la présence éventuelle de noms propres ou de termes importants sur le plan informationnel jouent un rôle plus important dans l'usage de la ponctuation que la fonction des éléments coordonnés.

2.6.6. Syntagmes particuliers « ponctuables » (Fa)

Cette catégorie regroupe les éléments internes à la proposition qui peuvent sous certaines conditions être ponctués. En français moderne, en dehors des emplois « stylistiques » et de la coordination/juxtaposition/énumération, la ponctuation peut apparaître à l'intérieur d'une proposition dans les cas suivants (*cf. Riegel et al. 1996, p. 89-90*) :

- apposition ;

Cadre théorique et méthode pour l'analyse linguistique

- termes détachés en tête ou en fin de phrase dans une construction segmentée ;
- compléments circonstanciels en tête de phrase ;
- complément de verbe ou de nom placé en tête de phrase ;
- groupes syntaxiques différents rapprochés à la suite de l'ellipse.

Rappelons que les apostrophes et les interjections sont classées dans la catégorie **B** (**b7** et **b8** respectivement), puisqu'elles entretiennent un rapport avec la représentation du discours rapporté.

D'après notre expérience d'analyse de textes en français médiéval, seules deux de ces catégories sont représentées et peuvent être considérées comme ponctuables : les appositions (**f1***) et les compléments circonstanciels initiaux (**f3***).

Les appositions sont assez fréquentes dans notre corpus, surtout dans les formules initiales qui servent d'introduction à certaines œuvres :

(2.54) Je [f1] gaston [d8] par la grace de dieu surnom||me febus [f1-d8f] Comte de foyz [f1] seigneur de be||arn . (*chasse619, 138b:1-2*).

La structure syntaxique de ces formules est très particulière, et nous avons souvent été amené à les analyser au cas par cas. L'emploi de la ponctuation peut s'y expliquer aussi bien par la syntaxe que par la mise en relief des noms propres et des titres de noblesse ou par le soin particulier apporté à la décoration de la première page du texte.

Les circonstants initiaux ne sont pas rares dans notre corpus, mais les occurrences ponctuées se situent presque toutes dans les formules initiales ou devant des noms propres ou titres de noblesse, comme dans l'exemple suivant :

(2.55) Mais en son || tens [f3f] Mahomet auoit ia este qui fu || messages au deable (*tyr828, 1a:5*).

D'autres étiquettes se sont rajoutées progressivement au cours du travail. Nous avons en fait à catégoriser les emplois de la ponctuation attestés aussi bien dans nos sources primaires que dans les éditions de celles-ci.

L'étiquette **f2** est utilisée devant (et très rarement après) certains compléments circonstanciels postposés au verbe qui soit sont complexes (par exemple, comportent

des relatives), soit sont introduits par la préposition *sans* ou *pour* où une légère tendance à l'emploi de la ponctuation apparaît dans certains manuscrits et éditions :

(2.56) Car il est dieu || tout entierement [f2] sans commence||ment *et* sans fin / (*im1522a*, 8a:32).

L'étiquette **f4*** est utilisée autour de précisions ou reformulations non propositionnelles de toutes sortes. Il s'agit assez souvent de compléments circonstanciels qui s'enchaînent :

(2.57) Sire la aual [f4] desoz || uostre pales [f4f] a .i. perron (*qgraal77*, 161a:9).

Chaque complément suivant apporte une précision par rapport au précédent, il ne s'agit donc ici ni d'une juxtaposition d'éléments de la même nature ni de compléments circonstanciels indépendants caractérisant les différents aspects d'une situation.

D'autres cas sont plus difficiles à généraliser. Dans l'exemple suivant, nous avons une explicitation détachée du pronom de la première personne du pluriel (*nous = toy et moy*) :

(2.58) nous || ne sommes [f4] toy [e1] et moy [f4f] blanches ne noires (*gerson13318*, 169v:14).

Nous avons également choisi de classer dans cette catégorie les précisions et reformulations introduites par les expressions *c'est a dire*, *c'est assavoir* ou simplement *c'est* :

(2.59) ilz ont l'apast emp*res* eux dedens la nasse || [f4] C'est assavoir la+f*emme* qui est belle bien pa||ree *et* bien abillee (*qjm1052*, 87r:24) ;

(2.60) ilz sont au ruyt [f4] c'est en leur || amour en forest (*chasse619*, 4c:15).

Ce choix reposait sur l'interprétation selon laquelle *c'est* dans ces expressions s'était transformé en un connecteur grammaticalisé. Le fait que la quasi-totalité des occurrences des reformulations ponctuées dans notre corpus concerne ces expressions, nous amène à croire qu'il s'agit dans ces cas de propositions indépendantes ou incidentes.

Cadre théorique et méthode pour l'analyse linguistique

L'étiquette **f5*** est utilisée pour marquer les frontières des constructions comparatives non propositionnelles. Il s'agit dans la majorité des cas de syntagmes nominaux introduits par *comme*⁹⁷, mais aussi de syntagmes introduits par *fors* ou *fors que* ('à l'exception de') et de la deuxième partie de la construction corrélatrice en *tant... que* :

- (2.61) [les cerfs] se font abayer [f5] *comme* vn sanglier (*chasse619, 4c:10*) ;
(2.62) ne puet estre en null lieu [f5] *fors que* de||dans || le parc (*chasse619, 4d:1*) ;
(2.63) sera *grant* guerre amenuise||ment et mortalite de peuple || tant *par* espee [f5] *que* autrement (*joachim768, 97d:27*).

Deux sortes de « syntagmes complétifs » introduits par *que* sont étiquetés **f6***. Il s'agit des compléments de comparatifs et des compléments du prédicat dans les constructions restrictives (*ne... que*) :

- (2.64) **IT**em || la cite de carie est plus grant [f6] *que* celle de babilloyne (*egipte769, 4v:2*) ;
(2.65) Et n'est pas H . *proprement* *le*tre [c5] mais || n'est [f6] *que* vne aspiracion (*dictier840, 396a:34*).

L'étiquette **f8***, enfin, est utilisée pour marquer les frontières d'un « groupe adjectival étendu ». Il s'agit d'adjectifs épithètes dotés de compléments prépositionnels :

- (2.66) la *premiere* || clarte du soleil [e1] *et* na||ture [f8] *contente* du || Repos de la nuit .[f8f] || nous rappellent aux mondains labours (*quadrilogue126, 192r:38-40*).

La relation entre l'adjectif et son complément est du type actanciel (Riegel *et al.* 1996, p. 367). Ces groupes sont donc très proches sur le plan sémantique et structurel des propositions participiales. Si, en revanche, le groupe adjectival étendu est employé comme attribut du sujet (ce qui est le cas de loin le plus fréquent), il ne constitue pas une unité ponctuable, car sur le plan sémantique il s'agit d'une proposition unique :

- (2.67) Et li+rois est [f8] *mout* liez+de+ceste nouele ! (*qgraal1119, 161c:21*).

97 Nous ne nous prononçons pas sur la catégorie grammaticale de ce connecteur. Il a certainement une fonction différente de celle d'une conjonction de subordination, car il introduit un syntagme nominal et non une proposition. S. Prévost (2001) interprète ce mot grammatical comme une préposition. La même analyse vaut pour *que* introduisant les syntagmes de type **f5** et **f6**. Néanmoins, le lien formel et sémantique avec la conjonction de subordination est également indubitable. C'est donc ce rapprochement qui nous autorise à considérer ces syntagmes comme « ponctuable ».

Il convient de noter que les groupes adjectivaux « ponctuables » sont extrêmement rares dans notre corpus : seules 7 occurrences de frontières sont attestées, dont deux portent une marque de ponctuation.

Les sept types de syntagmes particuliers présentés ci-dessus terminent la liste des frontières ponctuables que nous avons cherché à annoter d'une manière systématique dans notre corpus. Dans la section suivante, nous présentons les étiquettes utilisées pour annoter les occurrences qui n'ont pu être rangées dans aucune de ces catégories.

2.6.7. En dehors des frontières ponctuables (Fb)

Parmi les cas « aberrants », où la ponctuation apparaît dans nos textes aux endroits que nous ne considérons pas comme « ponctuables », nous avons distingué une catégorie particulière : les frontières entre les groupes du sujet et du prédicat (**f7**) Les autres occurrences ont été étiquetées **f9** (« autre »).

Seuls 7 des 29 textes de notre corpus présentent des occurrences de la ponctuation entre le groupe du sujet et le prédicat, et il s'agit d'une seule occurrence pour quatre d'entre eux. Sur l'ensemble du corpus, une dizaine d'occurrences ponctuées de ce type ont pu être identifiées. Nous avons choisi d'attribuer une étiquette spéciale pour repérer plus facilement ces cas intéressants à plusieurs égards. On peut notamment se poser la question de savoir si l'ordre des constituants, la présence de circonstants interposés et l'opposition « logique » entre le thème et le propos ne peuvent jouer un rôle dans l'apparition d'une marque⁹⁸ :

(2.68) lors entra en la || sale a+cheual .[f7] une || moult bele damoisele||le (qgraal1119, 138a:10) ;

(2.69) Car pestilences *et* || famines .[f7] estoient grans sur terre (tyr282, 5b:8).

Nous ne nous arrêterons pas ici sur l'emploi de l'étiquette **f9**, car il s'agit de cas particuliers qu'il convient de présenter individuellement. Nous y reviendrons dans la

98 Il convient de noter que l'emploi d'une virgule entre le sujet et le verbe n'était pas exclu par certains grammairiens du XVIII^e et du XX^e siècle, et on peut en trouver des occurrences chez certains écrivains de l'époque (Dürrenmatt 2004).

partie consacrée à l'analyse du corpus texte par texte. Il convient de donner ici une seule précision d'ordre méthodologique : la majuscule seule n'est pas considérée comme une marque de ponctuation en dehors des frontières ponctuables : ces emplois relèvent en effet de la « ponctuation lexicale » (mise en relief des noms propres ou simplement marquage d'un début de mot) qui ne font pas l'objet de notre étude actuelle.

Le taux des occurrences de ponctuation inexplicables dans le cadre de notre modèle des frontières ponctuables correspond environ à 2,5% (52 sur 2105) des occurrences ponctuées et de 0,9% (52 sur 5937) des frontières ponctuables identifiées. Il est donc suffisamment bas pour ne pas remettre en cause la validité du modèle. Nous prêterons néanmoins une attention particulière à l'analyse de chacune de ces occurrences, et une explication (raisons pragmatiques, décorative, erreur d'interprétation d'un passage difficile) a pu être avancée dans la plupart des cas.

2.7. Analyse formelle des marques de ponctuation

Seules les marques de ponctuation des niveaux syntaxique et macrotextuel sont systématiquement annotées. Les marques incontestablement « lexicales » (points d'abréviations, points entourant les chiffres, marques de coupure de mots en fin de ligne) et décoratives (servant notamment à la justification des fins de ligne), qui se trouvent en dehors des frontières ponctuables, sont exclues de nos décomptes et de notre analyse.

Les marques dont le statut est incertain bénéficient toutes de l'« avantage du doute » et sont étiquetées et analysées au cas par cas.

2.7.1. Forme de la marque

L'inventaire précis des signes de ponctuation médiévaux n'est pas facile à établir. Les recommandations des traités de l'époque (voir les sections 1.1.1 et 1.1.2 pour les références) sont loin d'être suivies dans la pratique. Il n'est pas toujours évident de déterminer si deux occurrences d'un même manuscrit représentent un même signe, sans parler de l'identité des signes à travers les manuscrits représentant des époques et des types d'écriture différents.

L'identification des oppositions significatives dans la forme des signes de ponctuation fait partie du problème plus vaste de la représentation des données des sources primaires dans des corpus électroniques. Nous avons consacré un chapitre entier (chapitre 4) à ce problème, et nous ne présenterons ici que l'application des principes adoptés.

La forme d'un signe de ponctuation est encodée au niveau *allographique* de la transcription. A ce niveau, nous cherchons à distinguer les variantes perceptibles (et potentiellement fonctionnelles) des caractères alphabétiques, des marques d'abréviation et de la ponctuation. Le degré de subjectivité dans la perception de ces variantes n'est pas négligeable, mais la « base de connaissances » accumulée grâce aux travaux de paléographes et de philologues depuis le XIX^e siècle et aux recherches sur l'histoire de la ponctuation menées dans les dernières décennies permet néanmoins de mettre au point des critères précis applicables dans la plupart des situations. La liste des variantes de marques (allographes) distinguées dans nos transcriptions est disponible dans l'Annexe II.1.

Dans la représentation *graphématique*, nous avons fait le choix de nous baser sur la *force* (cf. la section suivante) et non sur la *forme* de la marque de ponctuation, puisque c'est ce paramètre qui, selon notre analyse, est déterminant dans le fonctionnement du système de la ponctuation médiévale. Bien entendu, des manuscrits ou imprimés individuels, voire des traditions d'écriture, peuvent présenter des systèmes plus complexes, où la forme de la marque joue un rôle plus important (comme c'est le cas dans les imprimés à partir du XVI^e siècle).

En ce qui concerne l'étiquetage de la forme des marques, nous avons préféré respecter les distinctions fines du niveau allographique, en réunissant toutefois quelques variantes mineures, très proches tant sur le plan formel que fonctionnel.

L'étiquette **a** est réservée aux signes de ponctuation « macro-textuels », comme le « pied-de-mouche » (*paraph*, voir Fig. 2)⁹⁹ ou le « crochet adlinéaire » (*paragraphus*, voir Fig. 6). Ces deux marques se confondent d'ailleurs dans certains types d'écriture.

99 Cf. la note 74, p. 86.



Fig. 6 : Crochets adlinéaires (gerson13318, 166r et 168v)

Nous n'avons relevé qu'une quarantaine d'occurrences de ces marques dans l'ensemble du corpus¹⁰⁰.

L'étiquette **b** est utilisée pour le point « simple » (*punctus*) quelle que soit sa position sur la ligne (basse, médiane ou haute). La distinction des grammaires antiques basée sur la position de la marque (cf. 1.1.1) n'est plus d'actualité dans la période qui nous intéresse. Dans la grande majorité des textes de notre corpus, le point apparaît au milieu ou dans la moitié basse de la hauteur d'une ligne. C'est la marque de ponctuation la plus fréquente dans notre corpus (près de 1300 occurrences). Il convient de rappeler que les points servant à mettre en relief les chiffres et certaines abréviations ne sont ni étiquetés ni comptabilisés.

L'étiquette **c** sert à marquer la barre oblique (*virgula*)¹⁰¹ et les virgules (« point à queue », pour E. Dolet). C'est E. Dolet lui-même qui affirme l'identité des deux marques dont la deuxième vient remplacer la première :

(2.70) Le premier point est appelé en Latin incisum: || & en Francoys (principalement en L'imprimerie) || on l'appelle ung point à queue, ou uirgule: & le sou||loit marquer ainsi' (dolet1540, p. 18).

M.B. Parkes (1992, p. 303) associe la deuxième marque avec le *comma* « ; » médiéval, mais nous préférons respecter l'interprétation de Dolet, d'autant plus que son traité est le seul texte où le « point à queue » apparaît dans notre corpus.

En ce qui concerne la barre oblique, nous avons distingué au niveau allographique les barres obliques « pleines » (qui occupent, voire dépassent la hauteur moyenne des lettres sur une ligne) et « courtes » (qui se situent dans la moitié haute ou basse d'une ligne, comme dans la citation de Dolet ci-dessus). Cette variante de la barre oblique est attestée dans plusieurs traditions manuscrites (y compris nordique, cf.

100 Près de 29 000 occurrences-mots, 2 800 occurrences de marques de ponctuation.

101 M.B. Parkes utilise le terme *virgula suspensiva* pour les différentes variantes de la barre oblique. Nous réserverons ce terme plus précis à la variante dotée d'un point au milieu.

Haugen éd. 2006b, p. 103), mais elle ne semble avoir aucune spécificité fonctionnelle quelle qu'elle soit, en tout cas dans notre corpus.

Il convient de noter que dans certains manuscrits de notre corpus (**tyr828**, par exemple), les barres obliques ont probablement été ajoutées à une époque postérieure à l'exécution du manuscrit.

Nous avons relevé au total 375 occurrences de la barre oblique et de la virgule dans notre corpus.

Dans quelques cas, nous avons été amené à utiliser une étiquette double **bc**. Il s'agit d'une part des occurrences où une barre oblique semble avoir été ajoutée à un point placé initialement par le scribe (cas extrêmement rare avec 3 occurrences seulement dans notre corpus)¹⁰² et d'autre part en cas d'usage de la barre oblique avec un point au milieu (une variante de la *virgula suspensiva*, selon Parkes ; nous utiliserons néanmoins ce terme pour désigner spécifiquement cette variante¹⁰³), dont nous avons relevé une douzaine d'occurrences dans un seul et même texte (**gerson13318**). Ce texte présente une différence de forme assez nette entre les deux variantes, ce qui nous a poussé à tenir compte de cette opposition dans l'étiquetage.

L'étiquette **d** est utilisée pour marquer les occurrences du *comma (punctus elevatus)* < ∷ > et du deux-points. Comme dans le cas de la virgule et de la barre oblique, nous identifions ces deux marques conformément au traité d'E. Dolet (1540).

Au niveau allographique, nous distinguons deux variantes de cette marque (< ∷ > et < ∷' >), même s'il ne semble y avoir aucune différence fonctionnelle entre celles-ci.

Le *comma* est une marque de ponctuation très intéressante (*cf.* Marchello-Nizia 2007), et nous accordons une attention particulière à toutes les occurrences attestées dans notre corpus. Il n'apparaît pourtant que dans une douzaine des textes de notre corpus, et sa fréquence est très inférieure à celle du point ou de la barre oblique (moins de 70 occurrences dans l'ensemble du corpus).

102 Il ne s'agit tout de même pas d'un cas entièrement isolé. Des occurrences de cette double marque ont été relevées dans plusieurs manuscrits (Marchello-Nizia 1978, p. 35 ; Mazziotta 2007b, p. 57 ; Careri 2008, p. 231).

103 Nous suivons sur ce point le choix de l'initiative MUFI (Haugen éd. 2006b, p. 103).

Cadre théorique et méthode pour l'analyse linguistique

Les autres marques de ponctuation sont encore moins fréquentes : on n'en trouve que quelques occurrences isolées dans un ou deux textes de notre corpus. Le corpus d'échantillons restreints que nous avons constitué ne permet pas d'étudier sérieusement les phénomènes peu fréquents. Nous relevons néanmoins soigneusement ces occurrences isolées dans l'espoir qu'elles pourront ouvrir des pistes pour des recherches à venir.

L'étiquette **e** est utilisée pour marquer les occurrences du point d'interrogation (*punctus interrogativus*). Cette marque, connue depuis le VIII^e siècle, n'est que très rarement employée dans les manuscrits français médiévaux (*cf.* la section 2.3.6. ci-dessus). Les seules occurrences attestées dans notre corpus proviennent du traité d'E. Dolet. Nous savons pourtant que le texte de la *Queste del saint Graal* dans la version du manuscrit K (dont notre transcription **qgraal77** est extraite) présente plusieurs dizaines d'occurrences de cette marque, qui se trouvent tout simplement hors extrait dépouillé. Cela montre bien que l'utilisation des échantillons restreints ne convient pas pour l'étude des phénomènes peu fréquents.

L'étiquette **f** est réservée pour les occurrences du « point-virgule »¹⁰⁴ (*periodus* ou encore *punctus versus*). Notre corpus n'en présente aucune (ce qui ne signifie pas forcément que la marque n'était pas employée ailleurs dans les manuscrits d'où proviennent les extraits dépouillés).

L'étiquette **g** est utilisée pour annoter une marque en forme de *v* suivie de deux points utilisée en fin de paragraphe dans **artois11610**. (voir la section 7.7. pour une illustration et une description détaillée de cette marque).

Deux occurrences isolées de marques en forme de triangle (**dictier840**) et de losange (**phares1357**) de points ont été étiquetées **h**. Nous y reviendrons dans les sections consacrées à l'analyse de ces textes (6.12. et 7.12. respectivement).

L'étiquette **i** marque les occurrences du point surmonté d'une barre verticale (une sorte de petit « point d'exclamation » moderne) que nous avons relevées dans

104 Nous utilisons ce terme, certainement anachronique, en suivant S. Baddeley (2001), car il correspond précisément à la forme de la marque. La fonction du *periodus* n'a bien entendu rien à voir avec celle du point-virgule actuel. L'usage du point-virgule comme ponctuation « moyenne » s'est développé en France à partir du milieu du XVI^e siècle. *Cf.* (Demonet 2000, p. 42) pour une synthèse des points de vue des chercheurs sur la question.

quelques textes (**im587**, **im574** et **joinville13568**). Ces marques apparaissent uniquement en fin de ligne et pas forcément à des frontières ponctuables. Nous les considérons comme des éléments purement décoratifs sans rapport avec la ponctuation « syntaxique ».

Nous utilisons l'étiquette **j** pour annoter la double barre oblique employée une quinzaine de fois dans **phares1357** (*cf.* la section 7.12.).

Les étiquettes **m** et **n** servent à annoter les parenthèses ouvrante et fermante respectivement qui apparaissent uniquement chez E. Dolet.

Les étiquettes **o**, **p** et **q** sont réservées pour l'annotation des guillemets (neutre, ouvrant et fermant respectivement). Ces marques apparaissent dans les imprimés de la fin du XVI^e siècle qui pourraient un jour être intégrés dans un corpus de transcriptions de sources primaires.

L'étiquette **x** permet d'annoter un espace blanc prolongé qui peut parfois remplir la fonction d'une marque de ponctuation. Ces occurrences sont très rares (7 au total dans le corpus) et seront présentées lors de l'analyse détaillée des textes où elles sont attestées.

L'étiquette **y** (« autre ») est réservée pour les cas où l'identification formelle d'un signe de ponctuation serait impossible au moment de l'étiquetage. Dans l'état actuel de notre corpus, cette étiquette n'a pas été utilisée.

L'étiquette **z** correspond, enfin, à l'absence d'un signe de ponctuation. Cela n'équivaut pas à l'absence totale de la ponctuation, car l'usage d'une majuscule ou d'une lettrine après la frontière ponctuable constitue en soi une marque de ponctuation.

Cette liste pourra éventuellement être complétée si de nouvelles marques de ponctuation sont attestées au fur à mesure de l'élargissement du corpus.

2.7.2. « Force » de la ponctuation

N. Catach (1994, p. 51-52) distingue 7 « forces » et 3 « ordres » des signes de ponctuation. Curieusement, leur numérotation est inverse : du plus faible au plus fort d'une part et du plus grand au plus petit d'autre part.

Cadre théorique et méthode pour l'analyse linguistique

Pour préserver un certain parallélisme avec le classement des frontières ponctuables, nous avons préféré la numérotation « du plus fort au plus faible ». La « force » de la ponctuation est définie par un critère purement formel : qu'est-ce qui suit le signe de ponctuation (s'il y a lieu), ou autrement dit : comment se présente le début de l'unité ponctuable à droite de la frontière annotée ?

En premier lieu, il convient de définir les notions-clefs : la lettrine, la majuscule (ordinaire), la « grande » minuscule et la minuscule ordinaire.

La technique qui consistait à écrire d'une façon particulière les premières lettres de divisions textuelles plus ou moins importantes date de l'Antiquité (Geymonat 2008, p. 30). M. B. Parkes (1992, p. 1) utilise le terme générique de *littera notabilior* pour désigner les différentes formes que cette mise en relief pouvait prendre selon les traditions sribales et les manuscrits. La tradition consistant à utiliser pour ces lettres des tracés provenant d'un autre type d'écriture (de la *capitale* romaine ou de l'*onciale*) au milieu des textes écrits en *minuscule insulaire* provient sans doute des monastères irlandais et britanniques du VIII^e ou du IX^e siècle (Geymonat 2008, p. 61 ; cf. Stiennon 1991, p. 99-100) et s'est développée dans la pratique des scribes carolingiens (Parkes 1992, p. 34).

Dans les textes de notre corpus, il est le plus souvent possible d'établir une nette distinction entre trois types de lettres :

- les *minuscules* ordinaires, caractères « non marqués », utilisés au milieu des unités ponctuables dans le corps du texte ;
- les *majuscules* ordinaires, caractérisées par un tracé distinct, généralement plus grandes que les minuscules, utilisées le plus souvent « en tête » d'unités ponctuables et dans tous les cas au début de mots¹⁰⁵ ; elles sont de la même couleur que les minuscules suivantes, mais peuvent être surlignées d'un petit trait de couleur ;
- les *lettrines* en couleur différente du corps du texte (d'habitude, bleue ou rouge) de tailles variées et utilisées le plus souvent au début d'une ligne, réalisées plus

105 La majuscule peut également être utilisée comme deuxième caractère d'un mot commençant par une grande lettrine.

tard par une autre personne (*rubricateur*) dans l'espace de retrait prévu par le scribe.

Pour certains caractères (comme le < j >, par exemple) et dans certains types d'écriture (la cursive), la différence de forme et de taille entre les majuscules et les minuscules peut devenir moins évidente à établir. Dans certains manuscrits, un quatrième niveau semble se profiler dans ce système hiérarchique : des caractères en forme minuscule légèrement plus grands que le reste du mot (Leeuw van Weenen 2000). Dans notre corpus, nous n'avons pas été confronté à des manifestations indubitables de ce niveau hiérarchique supplémentaire, mais nous recourons à la notion de « *grande minuscule* » dans les cas où ne nous pouvons pas déterminer avec certitude s'il s'agit d'une majuscule ou d'une minuscule.

Nous pouvons désormais passer à la description de l'étiquetage de la « force » de la ponctuation dans notre corpus.

La plus grande force (**1**) est attribuée à une division de texte marquée par un saut de page ou simplement commençant au début du manuscrit. La « force **2** » correspond à une division marquée par un retrait qui concerne au moins deux lignes, cet espace étant destiné à une lettrine, qui pouvait pourtant rester non réalisée. La ponctuation de « force **3** » se traduit par l'emploi d'une lettrine qui n'entraîne pas le retrait de plusieurs lignes (et qui peut éventuellement être placée au milieu d'une ligne, *cf.* Fig. 7) ou bien par un saut de ligne avec une majuscule « ordinaire ».



Fig. 7 : « Force 3 » de la ponctuation : lettrine au milieu de ligne (*joinville13568, p. 2*).

La force **4** de la ponctuation correspond à l'emploi d'une majuscule ordinaire à une frontière ponctuable, avec ou sans un signe de ponctuation devant. Ce choix nécessite un bref commentaire.

Cadre théorique et méthode pour l'analyse linguistique

E. Llamas Pombo (2007) préfère établir la hiérarchie des « niveaux de ponctuation » individuellement pour chaque manuscrit, compte-tenu de la grande variabilité des pratiques. Dans un cas au moins (*Le Canarien*, Londres, British Museum, Egerton fund, ms. 2709, ca. 1420), elle place les majuscules seules au même niveau que les signes de ponctuation suivis de minuscules, tandis que les majuscules « appuyées » par les signes de ponctuation se situent à un niveau supérieur (p. 47).

Notre démarche méthodologique nous a conduit à appliquer des critères uniques de définition de la force de la ponctuation pour l'ensemble du corpus, et dans la plupart des textes les frontières ponctuables marquées par des majuscules s'assimilent à celles qui portent des ponctuations « fortes appuyées » plutôt qu'aux frontières dotées de ponctuations faibles. Il convient de noter toutefois, que les occurrences des majuscules seules sont facilement repérables dans notre corpus, grâce à l'étiquette **z4** qui associe le signe « zéro » à la « force 4 ».

La force **5** caractérise les ponctuations « moyennes » liées à l'usage d'une « grande minuscule » (éventuellement précédée d'un signe de ponctuation) au début d'une unité ponctuable. Comme nous l'avons déjà indiqué, notre corpus ne présente pas d'attestations sûres de triple opposition majuscule / grande minuscule / minuscule, mais dans certains cas, l'interprétation d'une lettre dans un endroit donné a été difficile (cf. < a > de *ainz* dans la Fig. 8 ci-dessous). Nous avons alors utilisé l'étiquette « force 5 » pour pouvoir traiter ces cas de doute séparément. Dans la représentation graphématique de nos transcriptions et dans les calculs globaux, nous assimilons ces ponctuations « moyennes » à des ponctuations fortes, mais de toute façon leur taux général (24 occurrences dans tout le corpus) est trop faible pour qu'elles puissent jouer un rôle considérable dans l'ensemble du système. Bien entendu, cette affirmation ne vaut que pour les manuscrits que nous avons étudiés et il se peut qu'elle doive être revue après l'intégration de nouvelles données.

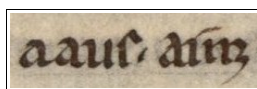


Fig. 8 : « Grande minuscule » (*qgraal77, 161d*).

La force **6** correspond aux frontières ponctuables suivies d'une minuscule. Il s'agit d'une ponctuation faible, si un signe de ponctuation est utilisé, et d'une « ponctuation zéro » (étiquette **z6**) en cas d'absence de toute autre marque.

Nous pouvons désormais passer à la présentation de notre méthode pour l'analyse des données du corpus étiqueté des transcriptions des sources. Les aspects méthodologiques de la représentation des données primaires dans ces transcriptions seront discutés dans le chapitre 3.

2.8. Méthode pour l'analyse

Notre analyse de chaque extrait intégré dans le corpus commence par la description « externe » de l'échantillon (l'identification et la description physique du manuscrit ou de l'imprimé, la date et si possible la région de la rédaction, les informations essentielles concernant l'œuvre d'où l'extrait est tiré et son auteur). Nous reviendrons plus en détail sur ces « variables externes » dans le chapitre 3 consacré à la méthodologie de corpus.

1. L'analyse des données linguistiques commence par une **appréciation globale du « taux de ponctuation »** (le nombre d'occurrences des marques de ponctuation divisé par le nombre d'occurrences-mots dans l'extrait). Ce taux est présenté sous forme de pourcentages (le nombre moyen de marques pour un empan de 100 occurrences-mots)¹⁰⁶.

Nous analyserons ensuite la répartition des marques par force (les forces 1 à 5 d'une part et les ponctuations faibles de l'autre) et par forme (marque dominante, marques occasionnelles, marques exceptionnelles). Le fait qu'un manuscrit privilégie la ponctuation forte ou faible ou qu'il utilise une ou plusieurs marques d'une façon plus ou moins régulière permet d'obtenir des indices intéressants pour l'établissement d'une « typologie interne » des pratiques de la ponctuation.

2. La deuxième étape de notre analyse consiste à étudier de façon précise l'**emploi des marques de ponctuation** à chaque type de **frontière ponctuable** (par

106 Il pourrait donc atteindre les 100%, si chaque mot était accompagné d'une marque de ponctuation.

Cadre théorique et méthode pour l'analyse linguistique

« grande catégorie » ou par groupe de « catégories fines »). Nous essayons de dégager une tendance dominante et d'expliquer les occurrences qui y échappent, en tenant compte de facteurs supplémentaires.

3. La ponctuation modernisée des éditions scientifiques peut également nous servir d'outil d'analyse supplémentaire : elle permet de voir dans quels cas précis les éditeurs ont décidé d'intervenir dans la structuration du texte original et dans quelle mesure ces interventions sont justifiées. Rappelons que selon G. Nunberg toute tentative de « modernisation » de la ponctuation des textes anciens amène nécessairement à l'altération du sens original (Nunberg 1990, p. 131). Si on sait précisément quelles structures sont ponctuées différemment dans les sources primaires et les éditions, on peut mieux évaluer l'utilisabilité des éditions en tant que sources de données pour les recherches linguistiques.

Nous terminons l'analyse de chaque extrait par un bref résumé des traits les plus caractéristiques de la ponctuation qui y sont constatés et par la façon dont le texte se positionne par rapport au reste du corpus.

Chapitre 3

Méthodologie de corpus

Notre étude des tendances de la ponctuation médiévale se situe dans le cadre méthodologique de la linguistique de corpus. Même si les dimensions de notre corpus en termes de nombre d'occurrences-mots (près de 28 700) ne sont pas comparables à celles de grands corpus ou bases textuelles modernes (centaines de millions d'occurrences¹⁰⁷) ou anciens (des millions d'occurrences¹⁰⁸), l'application de la méthodologie de corpus permet d'évaluer les résultats obtenus et de tracer les pistes pour approfondir les recherches.

Nous commencerons par définir la notion du corpus en général et par spécifier les contraintes qui pèsent sur la constitution d'un corpus pour l'étude du français médiéval, et, en particulier, pour des recherches portant sur les systèmes graphiques et la ponctuation. Nous verrons que la typologie textuelle joue un rôle primordial dans l'évaluation de la représentativité d'un corpus. Cela nous amènera à la discussion des variables pertinentes pour l'établissement d'une typologie de textes de français médiéval valable pour toute étude linguistique et, plus spécifiquement, pour une recherche sur la ponctuation. Nous conclurons par une évaluation de la représentativité de notre corpus et par la définition de sa place parmi les corpus qui ont servi jusqu'à présent pour des recherches sur les systèmes graphiques du français médiéval.

107 100 000 000 de mots pour le *British National Corpus* (BNC).

108 3 000 000 de mots pour la Base de Français Médiéval (BFM).

Les réflexions présentées dans le présent chapitre reprennent en partie et développent les idées exposées dans notre article (Lavrentiev 2007e) et s'inscrivent dans le programme du développement de la Base de Français Médiéval, auquel nous avons participé depuis plusieurs années.

3.1. La notion de corpus en linguistique

Les ressources textuelles informatisées (*corpus* ou *bases textuelles*, nous reviendrons sur ces termes dans la section 3.2.) sont devenues depuis plusieurs années un outil incontournable pour les recherches en linguistique. Même parmi les générativistes l'importance des données de corpus semble de plus en plus reconnue¹⁰⁹. Cependant, la question de savoir dans quelle mesure un corpus peut être représentatif de la langue étudiée reste ouverte.

Charles Fillmore (1992) a fait remarquer qu'aucun corpus d'anglais alors disponible, aussi large fût-il, ne contenait d'information suffisante pour tous les domaines de la langue qu'il désirait explorer. Pour lui, l'avantage principal des corpus, qu'ils soient grands ou petits, était de mettre en évidence les phénomènes auxquels un chercheur « n'aurait jamais pensé autrement ». Malgré le développement spectaculaire (quantitatif et qualitatif) des ressources textuelles informatisées proposées aux chercheurs depuis une dizaine d'années, le problème posé par Fillmore reste toujours d'actualité.

Dans l'étude des langues contemporaines, il est toujours possible de « combler les trous » en rajoutant de nouveaux textes (voire enregistrements de pratiques langagières dans différentes situations) et en complétant les données du corpus par des expériences avec les sujets parlants.

Dans les recherches sur l'évolution linguistique, les corpus textuels sont la seule source de données disponible pour les périodes anciennes de l'histoire d'une langue, et le nombre de textes disponibles n'est pas illimité. Sophie Prévost (2005, p. 149) parle d'une « situation circulaire » à laquelle est confronté tout linguiste médiéviste : « étudier des textes avec une connaissance de la langue exclusivement fondée sur ces

109 Tendence constatée notamment par Christiane Marchello-Nizia (2004, p. 58).

mêmes textes ». Elle voit toutefois une sortie de ce cercle vicieux qui consiste en une démarche critique :

... l'analyse des textes déjà étudiés conduit [le linguiste] à enrichir la connaissance de la langue en se penchant sur des faits non encore mis au jour ou approfondis, et la prise en compte de textes encore peu sollicités lui permet de confirmer ou d'infirmer l'existence de phénomènes ou de règles précédemment avancés, et peut-être de confirmer ces dernières. C'est grâce à cet accroissement du corpus que peut progresser la connaissance de la langue... Le corpus est indispensable au linguiste médiéviste, mais ce dernier a la possibilité de relativiser ce que révèlent les textes, par une maîtrise de leurs caractéristiques externes et de leurs conditions de production (*Ibid.*, p. 149-150).

La notion même du corpus demande par ailleurs à être précisée. John Sinclair (1996, p. 4) définit un corpus comme « une collection de manifestations langagières sélectionnées et ordonnées selon des critères linguistiques explicites pour servir d'échantillon du langage »¹¹⁰. Benoît Habert (2000) propose de restreindre cette définition en ajoutant des critères extra-linguistiques et en précisant qu'il s'agit d'échantillons d'*emplois déterminés* d'une langue.

La constitution d'un « corpus de référence » sensé représenter « toutes les variétés pertinentes d'une langue » (Sinclair 1996) relève, selon Sophie Prévost, d'une gageure, même si elle « n'en demeure pas moins un idéal vers lequel il est nécessaire de tendre » (Prévost 2005, p. 155).

Les caractéristiques d'un corpus dépendent donc de l'objet d'étude à laquelle il est destiné et doivent être déterminées ou évaluées au cas par cas.

3.2. Contraintes sur la constitution d'un corpus

Très grossièrement, on peut admettre que plus un corpus est « gros » d'une part et, d'autre part, plus il contient de textes intégraux de types variés, plus il sera représentatif, quel que soit l'objet d'étude¹¹¹. Mais, pour être utilisable, un corpus doit également rester maniable et profiter au maximum des ressources disponibles :

110 Citation originale en anglais : « A *corpus* is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language ».

111 Pour les recherches sur les états de langue disparus, il faut ajouter aux qualités du corpus la fidélité de représentation des sources primaires, un problème auquel nous avons consacré un chapitre entier.

Il convient de veiller... au meilleur équilibre entre représentativité maximale au regard du « but » du corpus, et maniabilité. Le corpus le plus satisfaisant reste un compromis entre l'idéal, le disponible et le gérable (Prévost 2005).

La contrainte de maniabilité signifie qu'il faut pouvoir interpréter et analyser correctement les données afin de dégager des résultats scientifiques significatifs. La bonne « logistique » du corpus (respect des normes internationalement reconnues en matière d'encodage des textes et utilisation des outils appropriés à toutes les étapes de constitution, d'enrichissement et d'exploitation) est indispensable, mais de toute façon la « valeur ajoutée » du corpus dépend en grande partie de l'investissement dans son annotation.

Le recours aux ressources disponibles permet d'optimiser l'effort porté à la numérisation et à l'encodage initial du corpus et d'investir davantage dans son enrichissement et dans l'analyse des données. Là aussi la bonne logistique doit assurer la possibilité de réutiliser les ressources existantes (qui respectent les normes internationales) et de rendre accessibles les enrichissements apportés aux recherches futures.

Les ressources textuelles qui peuvent être utilisées pour la construction de corpus destinés à des recherches diverses et variées sont appelés « réservoirs à corpus » (Habert 2000) ou « bases de textes » (Prévost 2005). Malheureusement, les bases textuelles de français médiéval existantes sont créées entièrement ou essentiellement à partir d'éditions critiques modernes, ce qui les rend inutilisables pour les études sur les systèmes graphiques et la ponctuation (*cf.* Heiden & Lavrentiev 2004). Les ressources offrant des transcriptions diplomatiques fines de manuscrits médiévaux sont rares (*cf.* chapitre 4), ce qui nous a obligé à investir dans les transcriptions primaires et a limité par conséquent les possibilités d'enrichissement quantitatif des données textuelles.

Le coût élevé des transcriptions (en termes de temps de travail) nous a poussé à une réflexion approfondie sur les aspects méthodologiques et techniques de la représentation des données dans le souci d'assurer dans la mesure du possible la pérennité et l'interopérabilité du corpus. Ces réflexions et les solutions retenues sont exposées dans les chapitres 4 et 5 de la présente thèse.

Méthodologie de corpus

Même si la représentativité du corpus, ne serait-ce que pour l'étude de la ponctuation et des graphies, n'a pu être garantie, compte tenu de la lenteur du processus de la transcription (et du temps que la mise au point de la méthodologie et des outils pour ces transcriptions ont nécessité), nous avons cherché à rendre le corpus aussi représentatif, diversifié et équilibré que possible et, surtout, à nous donner les moyens d'évaluer cette représentativité et cette diversité¹¹².

Pour Benoît Habert,

... améliorer la représentativité d'un corpus consiste à préciser la production et la réception de chacun de ses composants, en lien avec les motifs qui ont conduit à la création du corpus, mais aussi à pouvoir déterminer sur des bases objectivables les différents emplois du langage auxquels on s'intéresse (Habert 2000).

Le chercheur qualifie ces deux dimensions de la représentativité du corpus respectivement d'« externe » (définissable *a priori*) et d'« interne » (que l'on ne peut mesurer qu'après l'analyse des textes du corpus), et les utilise pour passer en revue les différents types de classification des textes, ce qui relève selon nous, de la typologie textuelle.

3.3. Typologie textuelle

3.3.1. Importance de la typologie textuelle pour la linguistique de corpus

Nous utilisons le terme de *typologie textuelle* pour désigner le dispositif méthodologique permettant d'organiser les textes (au sens large, comprenant également les transcriptions de l'oral) d'un corpus, en prenant en compte l'ensemble des variables pertinentes pour l'analyse linguistique. Tout comme la typologie des langues, la typologie textuelle utilise la démarche qui consiste à organiser la masse des phénomènes

¹¹² La contrainte de disponibilité a également joué son rôle. Nous avons toujours réutilisé les transcriptions que nous avons réalisées lors des périodes précédentes de nos recherches : celle de *l'Image du monde* (notre projet de thèse initial se limitait à seuls manuscrits en prose de cette œuvre) et de manuscrits en prose de la Bibliothèque municipale de Lyon que nous avons pu effectuer lors d'une mission de recherches en 2001. A partir de 2003, nous transcrivons en priorité les manuscrits de base des textes faisant partie de la Base de Français Médiéval.

observés en mettant en évidence les traits communs et les points de divergence. Un corpus organisé typologiquement doit permettre d'évaluer le degré de généralité des résultats obtenus à partir de celui-ci.

Par ailleurs, il peut exister plusieurs taxonomies de textes établies chaque fois en fonction des phénomènes linguistiques ciblés. Certaines distinctions fondamentales peuvent toutefois être pertinentes pour un grand nombre de phénomènes.

David Lee (2001), dans un article consacré à la définition des notions permettant de classer les textes du BNC, a remarqué que la plupart des études basées sur corpus s'appuyaient implicitement ou explicitement sur la notion de genre ou sur des concepts proches, tels que le registre, le type de texte, le domaine, le sous-langage, la forme du message, etc. En même temps, il a constaté une grande confusion dans l'emploi de ces termes, et cette confusion n'est pas complètement dissipée à ce jour.

Benoît Habert (2000) distingue les classifications basées (1) sur les conditions de production des textes, (2) sur les buts visés par les textes, (3) sur l'inscription dans des « genres » (ou « registres ») et, enfin, (4) sur l'emploi ou non de certaines marques linguistiques, telles que les « embrayeurs » (Jakobson 1963, ch. 9). C'est la troisième catégorie qui retient le plus son attention.

Pour les textes médiévaux qui nous sont parvenus, les conditions de production et les buts visés ne sont pas toujours faciles à déterminer et sont de toute façon étroitement liés aux « genres » (ou « domaines », ou « registres »)¹¹³, c'est donc ce paramètre qui mérite une attention toute particulière. Par ailleurs, d'autres facteurs externes doivent obligatoirement être pris en compte dans un corpus diachronique, notamment le temps et l'espace.

Le corpus de transcriptions de manuscrits et d'incunables que nous avons constitué est destiné à être un complément de la Base de Français Médiéval (Lavrentiev 2007a). Il est donc important qu'il partage avec la BFM un « socle commun » en typologie textuelle tout en tenant compte des spécificités liées à la représentation fine des sources primaires et des recherches auxquelles il est destiné, prioritaires pour cette thèse.

113 Nous donnerons notre définition des termes retenus dans la section 3.3.3.

Nous présentons ci-dessous l'état actuel de la description des textes adoptée dans le projet BFM en général et nous aborderons ensuite les problèmes spécifiques que pose la constitution d'un corpus destiné à l'étude de la ponctuation médiévale et des systèmes graphiques de manuscrits.

3.3.2. Typologie textuelle appliquée dans la Base de Français Médiéval

Dans le projet BFM, l'étude des phénomènes linguistiques s'appuie sur la méthodologie textométrique qui pose un cadre multidimensionnel pour les recherches sur les textes (Guillot *et al.* 2007) : les textes sont caractérisés par un certain nombre de *variables externes* et *internes*. La BFM étant une base textuelle plutôt qu'un corpus, il convient d'utiliser les variables pertinentes pour en sélectionner les textes en vue de la constitution d'un corpus destiné à une certaine recherche linguistique¹¹⁴.

Les variables externes sont les descripteurs (métadonnées) associés à un texte (ou à un passage dans un texte), et les variables internes sont les indices que l'on peut obtenir à partir de l'analyse de certains phénomènes linguistiques dans le texte (*cf.* les dimensions de la représentativité de Benoît Habert (2000) citées plus haut).

Les types de textes peuvent être définis par la mise en évidence de « faisceaux » de variables internes et externes¹¹⁵. Selon les objets linguistiques qui ont été choisis comme variables internes, les types obtenus peuvent être différents. Nous présenterons donc dans un premier temps l'ensemble des variables jugées pertinentes dans le projet BFM et procéderons ensuite à la définition de la grille typologique effective retenue lors de la constitution de notre corpus.

Il convient de souligner que la définition des variables et l'étiquetage de leurs valeurs résulte toujours d'un compromis entre la « qualité » et le « prix » du travail à effectuer, autrement dit, entre d'une part le nombre des facteurs et la précision de leur définition, et d'autre part, la complexité de leur analyse et des ressources requises.

114 Bien entendu, la BFM cherche par ailleurs à améliorer l'équilibre des types de textes présents dans la base et à se rapprocher ainsi de l'« idéal » d'un corpus de référence pour le français médiéval.

115 *Cf.* : « un type de texte est défini par la cooccurrence d'un certain nombre de traits linguistiques (et éventuellement par l'évitement systématique d'autres traits) » (Habert 2000).

3.3.3. Variables externes

Nous présentons ci-dessous la liste des variables « externes » sélectionnées à ce jour pour les textes de la BFM. Nous distinguons les principales dimensions de variation et les variables concrètes, qui peuvent être multiples pour une dimension donnée. Dans ce dernier cas, les variables sont identifiées par des lettres ajoutées au numéro de la dimension (par exemple, 1a, 1b, etc.)¹¹⁶.

3.3.3.1. Date

Cette dimension est extrêmement importante pour les études diachroniques. Sa définition pour les textes médiévaux pose cependant plusieurs problèmes. Tout d'abord, il faut faire un choix entre la date de composition de l'œuvre et la date du manuscrit (manuscrit de base, s'il s'agit d'une édition critique). Dans certains cas, l'écart entre la date de composition et celle du manuscrit de base d'une édition de référence peut être supérieur à un siècle¹¹⁷. On sait par ailleurs que des manuscrits tardifs présentent parfois des formes linguistiques plus archaïques que celles qu'on peut trouver dans des manuscrits plus anciens, mais il est très difficile d'interpréter formellement cette « archaïcité ».

Par ailleurs, la date de composition d'un texte n'est souvent connue que très approximativement. La datation de certains textes se situe parfois dans une fourchette d'un demi-siècle (par exemple, la seconde moitié du X^e siècle pour la *Passion de Clermont*) ou même davantage (par exemple, le XIII^e siècle pour le *Lancelot en prose*).

Les manuscrits sont eux aussi très souvent difficiles à dater avec précision. Souvent, les auteurs de catalogues et les éditeurs se limitent à une datation vague type « écriture du XIV^e siècle ». Certains manuscrits sont en revanche datés ou peuvent être datés à quelques années près grâce à des indices paléographiques ou codicologiques.

116 Une liste complète des descripteurs de textes de la BFM et les règles de leur rédaction, régulièrement mises à jour, sont données dans le *Manuel de description des textes de la BFM* (Guillot & Lavrentiev 2007). Nous n'évoquerons dans la présente section que les descripteurs pertinents pour la typologie textuelle, ainsi que les problèmes méthodologiques connexes.

117 Il est toutefois préférable d'éviter dans la mesure du possible d'utiliser dans des études linguistiques des textes pour lesquels l'écart entre la date supposée de composition et la date du manuscrit est trop important.

Méthodologie de corpus

De façon générale, c'est la (*1a*) *date de composition de l'œuvre* qui est retenue comme variable externe pour la typologie textuelle dans le projet BFM. En cas de datation approximative, la date est fixée au milieu de la période dans laquelle le texte paraît avoir été composé. Une solution différente a été adoptée dans la base textuelle élaborée pour le *Dictionnaire de moyen français* du laboratoire ATILF <<http://atilf.atilf.fr/dmf.htm>> où c'est la date la plus ancienne possible de la composition du texte qui est retenue. Il semble toutefois que le choix de la date « médiane » soit partagé par la majorité des membres du Consortium pour les Corpus de Français Médiéval (CCFM, <<http://ccfm.ens-lsh.fr>>). Il convient de noter que selon les types de recherche sur un corpus, il faudra retenir soit l'une des « dates limites » (pour prendre en compte tous les textes qui pouvaient être composés à une période donnée), soit la date moyenne (pour, par exemple, établir des partitions diachroniques).

Plusieurs niveaux de « granularité » de datation des textes ont été définis dans la Base pour pouvoir grouper les textes appartenant à une même période plus ou moins large. Les textes peuvent ainsi être regroupés par période (ancien ou moyen français), siècle, demi-siècle ou « sous-siècle » (début, milieu ou fin d'un siècle). Cet intervalle d'un tiers de siècle a été choisi comme l'unité de regroupement chronologique par défaut dans la BFM. Le nombre de textes avec une datation imprécise ne permet en effet pas d'assurer une meilleure granularité diachronique pour l'ensemble du corpus.

Un algorithme spécifique a été élaboré pour calculer les dates limites à partir des indications « en prose libre » tirées des sources de références (éditions, catalogues, dictionnaires, etc.). Pour les textes datés approximativement de début, milieu ou fin d'un siècle, les dates limites calculées ont été respectivement -00 et -33, -34 et -66, -67 et -99.

Bien entendu, pour certains types de recherche (tels que l'étude des systèmes graphiques et de la ponctuation) c'est la (*1b*) *date de la production du manuscrit* qu'il convient de retenir. Lors du choix de textes à numériser pour un corpus, il est souhaitable de privilégier (les autres conditions étant égales) les textes pour lesquels le décalage entre la date de composition et la date du manuscrit est le moins important.

3.3.3.2. Espace

Une autre dimension importante pour la typologie des textes en français médiéval concerne la variation *diatopique* (ou régionale). Nous distinguons (2a) le *dialecte de l'auteur de l'œuvre*¹¹⁸, (2b) les *traits régionaux provenant du (/des) copiste(s)*, la (2c) *région de la production du manuscrit* et éventuellement (2d) la *ville et (2e) l'atelier*.

Les textes littéraires portent le plus souvent des traits de plusieurs dialectes (à cause notamment de copies successives effectuées dans des régions différentes). Il est bien connu également que dans toutes les sources écrites, un nombre plus ou moins important de « formes communes » sont présentes à côté de formes marquées régionalement (Picoche & Marchello-Nizia 1994, p. 20). Nous proposons donc de distinguer les traits régionaux attribuables au dialecte de l'auteur et à la région de la composition et/ou des copies successives de l'œuvre, si celle-ci est connue. D'après Anthony Lodge, les distinctions dialectales du français médiéval « semblent dans leurs grandes lignes celles qui se sont maintenues jusqu'à l'époque moderne » (Lodge 1997, p. 101). On peut ainsi distinguer pour le « françois » (langue d'*oïl*) les dialectes suivants : le picard, le normand, l'anglo-normand, le champenois, le lorrain, le gallo, l'angevin, le francien, le poitevin, le berrichon, le bourguignon et le franc-comtois (cf. la carte, *ibidem*, p. 102). Cependant, tous ces dialectes ne possédaient pas à l'époque médiévale de centres de production écrite importants (cf. *ibidem*, p. 152), ce qui explique la faible représentation de certains d'entre eux dans les textes conservés. Les dialectes qui sont susceptibles d'être plus ou moins bien représentés dans les textes médiévaux sont, d'après nos estimations, les suivants : le picard, l'anglo-normand et le normand, le lorrain, le poitevin, le champenois et le bourguignon.

Pour identifier la région de composition de nos textes, nous nous servons d'une grille plus détaillée, utilisée dans les atlas linguistiques, notamment dans celui d'Anthonij Dees (1978). La définition du lieu de la composition d'une œuvre est effectuée d'après les indications données par l'éditeur scientifique, d'après les études

118 L'usage du terme dialecte est controversé par rapport aux textes de français médiéval. Certains chercheurs préfèrent parler de la *scripta* (Stanovaïa 2003). Nous continuons néanmoins à utiliser le terme traditionnel sans prendre partie dans la polémique en cours et en lui donnant le sens de 'traits régionaux remarquables'.

Méthodologie de corpus

spécifiques ou, le cas échéant, d'après les calculs de Dees (*ibidem*) sur les variables internes des textes, les critiques qui portent sur sa méthode étant prises en compte.

La région du manuscrit peut être différente de celle de la composition du texte. Dans le cas des recherches qui portent sur les manuscrits, il est souhaitable de connaître non seulement la région de la production du manuscrit, mais aussi le scriptorium ou l'atelier précis, mais ces informations ne sont que rarement disponibles en dehors des chartes datées et localisées.

3.3.3.3. Auteur

Les données sur l'*auteur du texte* peuvent être considérées comme une variable externe de la typologie textuelle. Si les traits régionaux sont modélisés par une autre variable, l'idiolecte de l'auteur déterminé en partie par des facteurs socio-linguistiques (et notamment son appartenance à un milieu social et son âge) peuvent avoir une incidence importante sur les caractéristiques linguistiques de ses productions écrites.

A l'époque médiévale, cependant, une partie importante des textes est anonyme. On ne peut donc identifier l'auteur que pour un sous-ensemble du corpus. En revanche, même si (3a) le *nom de l'auteur* est inconnu, il peut être possible de supposer avec un bon degré de certitude qu'il appartenait à (3b) un certain *milieu social*. Si (3c) l'*âge de l'auteur* à l'époque de la composition de l'œuvre est connu, cette information peut s'avérer importante : certains chercheurs sont convaincus que sans la prise en compte de l'âge de l'auteur, la date de composition d'une œuvre n'est pas suffisante pour une analyse précise des traces de changements linguistiques dans un texte.

Par ailleurs, dans la littérature médiévale, les scribes pouvaient modifier le texte de façon considérable en fonction de leur interprétation de certains passages, de leurs goûts personnels et de ceux du destinataire ou du commanditaire du manuscrit. Ils pouvaient devenir donc dans une certaine mesure des « coauteurs » du texte (Kennedy 1970). L'un des cas les plus connus est celui du scribe Guiot, qui a « édité » les romans de Chrétien de Troyes. Pour Bernard Cerquiglini, Guiot a anticipé les pratiques des éditeurs interventionnistes de la fin du XIX^e et du début du XX^e siècle (Cerquiglini 1989). Dans notre modèle de la description du texte, les informations sur le scribe font

partie des variables spécifiques de la caractérisation des transcriptions de manuscrit (*cf.* la section 3.3.3.10).

3.3.3.4. *Type de document*

Nous passons à présent à la description d'un certain nombre de variables liées à la thématique et à la forme des textes. Elles sont dans une large mesure définies par la structure interne des textes, mais elles sont néanmoins extralinguistiques, donc externes par rapport aux phénomènes linguistiques étudiés dans un corpus. Ces variables occupent une place centrale dans la plupart des typologies textuelles et en même temps leur définition et leur dénomination sont loin d'être consensuelles (*cf.* Habert 2000 ; Lee 2001 cités plus haut).

Le projet BFM a été conduit à adopter certains choix méthodologiques et terminologiques, même si la réflexion et les discussions sur la meilleure façon de modéliser cet aspect de la typologie textuelle sont toujours en cours dans le cadre notamment du CCFM et de projets de recherche associés.

A ce jour quatre principales variables de ce type sont distinguées dans la BFM : le *type de document*, la *forme*, le *domaine* et le *genre* du texte.

La distinction de deux (4) *types de document* – document d'archive (ou acte de la pratique) ou œuvre (littéraire ou non littéraire)¹¹⁹ – est très ancienne dans la tradition d'étude et d'édition des textes médiévaux¹²⁰. Elle a même entraîné la formation de deux disciplines différentes : la diplomatique pour les documents d'archives, et la philologie pour les textes littéraires au sens large. Cependant, pour les recherches linguistiques, les deux types de sources sont aussi intéressants l'un que l'autre. Si les chartes présentent l'avantage d'être souvent datées et localisées précisément, elles ne possèdent pas la même richesse de vocabulaire ni de constructions syntaxiques que les œuvres de taille plus importante. Par ailleurs, « les scribes des chancelleries étaient aussi capricieux et imaginatifs que leurs confrères des ateliers de copistes » (Picoche & Marchello-Nizia

119 *Cf.* la distinction de « sources documentaires » (l'ensemble des documents issus des chartriers médiévaux) et « textes littéraires » (toute autre création, des plus ludiques aux plus pratiques) faite dans (Viellard & Guyotjeannin coord. 2001, p. 19).

120 La distinction entre l'écriture des chartes et celle des manuscrits littéraires est déjà observée par Jean Mabillon dans son traité *De re diplomatica* (1709).

1994, p. 19), ce qui ne permet pas de considérer les chartes comme représentant « la langue vulgaire dans toute sa vérité »¹²¹.

Actuellement, la BFM est composée presque exclusivement d'« œuvres » et ne comprend qu'un nombre très limité d'actes de la pratique¹²².

3.3.3.5. Forme

(5a) La *forme* du texte (vers ou prose) est une variable relativement facile à déterminer. Un certain problème est lié aux textes où les passages en prose alternent avec ceux en vers. Actuellement, la forme de ces textes est caractérisée dans la BFM comme « mixte ». Dans l'avenir, l'étiquetage de la forme de ces textes pourra se faire au niveau intratextuel, il sera ainsi possible de calculer pour un texte donné la proportion de passages en prose et en vers et de limiter éventuellement les requêtes à des parties homogènes du point de vue de la forme.

Une forme de texte particulière est constituée par les gloses : ces petits segments ne peuvent être assimilés avec certitude ni aux vers, ni à la prose¹²³. Actuellement, aucune glose n'est incluse dans la BFM en tant qu'unité textuelle, mais cette situation peut changer.

Il peut s'avérer utile de préciser davantage la forme du texte : (5b) le *type de vers* (octosyllabique, décasyllabique, etc.) ou (5c) la *structure interne* pour les œuvres en prose (division en livres, parties, chapitres, etc.).

121 Gaston Raynaud (1876), cité par Jacqueline Picoche et Christiane Marchello-Nizia (1994, p. 19).

122 Des projets de numérisation de chartes et de création de bases textuelles interrogeables en ligne sont en cours dans des centres de recherche partenaires du CCFM, notamment le projet d'édition en ligne de l'ensemble des cartulaires d'Ile-de-France de l'Ecole des Chartes, <<http://elec.enc.sorbonne.fr/cartulaires/>>, le projet de numérisation de chartes lorraines dirigé par Martin-Dietrich Gleßgen (Université de Zurich) et le projet de mise en ligne du corpus d'Anthonij Dees piloté par Achim Stein (Université de Stuttgart). Par ailleurs, un projet de numérisation de chartes wallonnes est en cours à l'Université de Liège (projet *Khartés* dirigé par Marie-Guy Boutier, cf. Mazziotta 2004).

123 Nous remercions l'équipe du Dictionnaire étymologique de l'ancien français (DEAF) et, en particulier, le professeur Frankwalt Möhren, de nous avoir signalé la nécessité de distinguer cette forme de texte particulière.

3.3.3.6. *Domaine*

Le rangement des textes dans des *domaines* ou des *thèmes* est assez répandu dans le monde des corpus. Pour Benoît Habert, ces classifications sont « sujettes à caution », car « trop raffinées, elles se trouvent vite battues en brèche par l'évolution des sociétés, des techniques et des mentalités ; grossières, elles sont trop floues pour être utiles » (Habert 2000). Il y voit uniquement un moyen de « débroussaillage imparfait mais commode ».

Dans la BFM, la répartition des textes par domaine joue néanmoins un rôle important : c'est en effet l'un des principaux paramètres pris en compte pour l'équilibrage du corpus. La détermination du domaine (ou des domaines, car ils peuvent être multiples) d'un texte est en effet souvent plus objective que celle du genre, et la « stabilité diachronique » de ce classement paraît plus importante.

Le domaine du texte, selon nous, est lié à la sphère de son fonctionnement ou, autrement dit, à sa destination principale (Guillot *et al.* 2007). Les textes de la BFM actuelle sont répartis en 5 domaines : *littéraire* (destination : divertissement), *didactico-scientifique* (destination : enseignement et recherche), *historico-politique* (destination : rapporter les événements historiques et contribuer à des débats politiques), *juridique* (destination : gérer les relations sociales) et *religieux* (destination : servir à la vie spirituelle). Cette grille a été élaborée en tenant compte des pratiques des principaux manuels bibliographiques de la littérature médiévale (Bossuat 1951 et suppl. ; Jauss *et al.* 1972 et suiv.). La grille des domaines permet à notre avis de catégoriser une grande majorité des textes médiévaux, mais elle n'est certainement pas exhaustive. On pourrait par exemple y ajouter le domaine de la *vie privée* (correspondance, journal intime...), mais aucun texte antérieur au XVI^e siècle relevant de ce domaine ne nous est parvenu.

Un texte peut naturellement appartenir à plusieurs domaines à la fois : un *mystère* (pièce de théâtre) religieux est en même temps une œuvre littéraire ; un traité de théologie est aussi bien religieux que didactique ; le *Roman de la rose* est une œuvre avant tout didactique, mais aussi littéraire. Cette multitude de domaines peut poser un problème méthodologique pour l'équilibrage du corpus, mais c'est une réalité, dont il faut tenir compte dans la modélisation.

Méthodologie de corpus

La solution temporaire, sûrement imparfaite, adoptée actuellement dans la BFM, consiste à choisir pour chaque texte (6a) un *domaine principal*. Un ou plusieurs autres domaines peuvent être considérés comme (6b) des *domaines secondaires*.

Pour les textes du domaine scientifico-didactique, il peut être utile de préciser (6c) le *thème* (médecine, géographie, langue, vie quotidienne, etc.).

3.3.3.7. Genre

La notion de *genre* est largement utilisée dans les études littéraires et dans la linguistique de corpus. Il existe plusieurs critères de définition et plusieurs classifications possibles pour les genres. Le problème est encore plus complexe pour un corpus de textes qui s'étale sur plusieurs siècles.

Le terme de genre remonte à la tradition rhétorique gréco-latine, la notion du genre littéraire a été développée dans les théories de poétique et dans la critique littéraire (Branca-Rosoff 1996) ; un nouveau sens (proche d'un certain point de vue de ses origines) est apparu dans la linguistique de corpus : « Les genres sont des catégories de textes distinguées spontanément par les locuteurs confirmés d'une langue ; par exemple, les genres de l'anglais incluent les romans, les articles de journaux, les éditoriaux, les articles de recherche, les discours en public, les nouvelles radiophoniques et la conversation de tous les jours » (Biber 1989, p. 5-6). Plus récemment, Douglas Biber (1998) a préféré désigner cette notion élargie par le terme de *registre*. B. Habert (2000) a adopté cet usage.

Pour l'époque médiévale, il est pourtant très difficile de déterminer ce qu'étaient ces registres relevant de la « typologie populaire ». Si pour les textes modernes, l'inventaire des registres se caractérise toujours par une « géométrie variable » (Habert 2000), nos connaissances sur la typologie des textes médiévaux dans son ensemble sont encore trop limitées pour tenter d'en dresser une liste. Le terme de registre est donc réservé dans la BFM à un usage futur, quand les progrès d'études de la typologie textuelle médiévale le permettront.

Pour l'instant, nous nous contentons d'une classification plus traditionnelle en genres, basée en premier lieu sur les propriétés sémantico-structurelles du texte. Un

système de genres particulier se forme dans chaque domaine textuel, même si certains genres sont « transversaux » (comme le genre dramatique, qui est représenté dans les domaines littéraire et religieux). Certains textes portent une caractéristique générique dans le titre (*Chanson de Roland*, *Roman de la rose*, *Cent nouvelles nouvelles*), mais celle-ci ne correspond pas nécessairement au genre attribué à l'œuvre en question par les ouvrages de référence de philologie traditionnelle ou dans les classements utilisés dans les corpus textuels modernes.

La complexité de la définition du genre nous oblige à tenir compte de plusieurs classements alternatifs, ce qui amène à la multiplication de variables « genresques ». Il s'agit du (7a) « genre auto-proclamé » (dans le titre ou dans le texte même), du (7b) « genre traditionnel » (défini dans les ouvrages de références bibliographiques ou philologiques¹²⁴) et du (7c) « genre moderne » (dans lequel l'œuvre en question pourrait être classée dans un corpus de textes modernes)¹²⁵. Selon les cas, un ou plusieurs de ces champs peuvent rester vides.

Pour faciliter les recherches pour lesquelles il n'est pas nécessaire de prendre en compte toute la complexité du classement générique, un champ (7d) « genre BFM » a été créé, dont la valeur est une sorte de compromis entre les différentes classifications (il s'agit le plus souvent d'une reprise soit du genre traditionnel, soit du genre moderne).

3.3.3.8. Séquence textuelle

Les variables présentées ci-dessus s'appliquent normalement aux textes dans leur ensemble¹²⁶. Cependant beaucoup de textes sont composés de passages plus ou moins hétérogènes. Jean-Michel Adam (1999, p. 82) insiste sur le fait que la distinction de types de discours dans les séquences textuelles homogènes est plus importante que la distinction de types de textes entiers : « L'unité " texte " est trop complexe et trop hétérogène pour présenter des régularités linguistiquement observables et codifiables, du moins à ce niveau de complexité ».

124 Notamment, dans (Hasenohr *et al.* 1994) et dans (Bossuat 1951 et suppl.), ainsi que dans la bibliographie du DEAF et dans l'introduction de l'édition scientifique correspondante.

125 C'est donc un « prototype » du *registre*.

126 Des exceptions sont toutefois possibles. Nous avons évoqué la forme « mixte ». Certains textes peuvent contenir des passages appartenant de toute à un domaine ou à un genre différent du reste de l'œuvre.

Méthodologie de corpus

Parmi les différentes *séquences textuelles* nous distinguons trois variables liées entre elles mais gardant une certaine autonomie : (8a) récit vs discours direct ; (8b) éléments d'organisation textuelle (titres, tables de matières, explicits, etc.) vs le corps du texte ; (8c) discours narratif, descriptif, argumentatif, explicatif ou dialogal (Adam 1992). Si les deux premières variables sont identifiables de façon presque automatique grâce à la ponctuation et à la mise en page des éditions, le balisage des types de discours nécessite un investissement plus important et ne pourra se faire dans un premier temps que sur une partie des textes de la BFM.

En général, dans l'étude de manuscrits, il convient en plus de traiter de façon particulière (8d) la première page (recto du 1^{er} folio) du texte, qui peut présenter une mise en page, une ponctuation et une qualité d'écriture différentes du reste du manuscrit.

3.3.3.9. Variables « conceptionnelles » et « médiales »

Une piste originale pour la caractérisation des textes a été proposée par Peter Koch et Wulf Oesterreicher (Koch 1993 ; Koch & Oesterreicher 2001). S'appuyant sur la distinction entre les aspects médial (phonique vs graphique) et « conceptionnel » (immédiat vs distant) de l'opposition entre l'oralité et la scripturalité formulée dans (Koch & Oesterreicher 1985), les linguistes allemands définissent un certain nombre de traits qui caractérisent la position du texte sur l'axe « conceptionnel » de la distance communicative.

Ces traits incluent le cadre de la communication (privée vs publique), les relations entre les interlocuteurs (intimes vs inconnus), le degré de l'émotionnalité (forte vs faible), etc. (*cf.* Fig. 9 ci-dessous).

'immédiat'		'distance'	
communicatif		communicative	
←		→	
① privée	(communication)	publique ❶	
② intime	(interlocuteur)	inconnu ❷	
③ forte	(émotionnalité)	faible ❸	
④ ancrage situationnel		détachement ~ ❹	
⑤ ancrage référentiel	(origo)	détachement ~ ❺	
⑥ coprésence	(espace/temps)	séparation ❻	
⑦ intense	(coopération communicative)	minime ❼	
⑧ dialogue		monologue ❽	
⑨ spontanée	(communication)	préparée ❾	
⑩ liberté thématique		fixation thématique ❿	

Fig. 9 : Paramètres de la distance communicative (Koch & Osterreicher 2001, p. 186)

Ces traits peuvent constituer des variables intéressantes pour l'établissement d'une typologie textuelle, il a donc été décidé de caractériser progressivement l'ensemble des textes de la BFM par rapport à certains de ces traits. Les traits retenus sont les suivants :

- (9a) émotionnalité (forte, faible ou moyenne) ;
- (9b) ancrage ou détachement actionnel et situationnel ;
- (9c) ancrage ou détachement référentiel dans la situation ;
- (9d) coprésence ou séparation spatio-temporelle ;
- (9e) dialogue ou monologue.

Ce travail n'en est qu'à ses débuts pour le moment, et il est trop tôt pour utiliser effectivement ces variables pour la composition de corpus.

3.3.3.10. Caractérisation du manuscrit

Dans un corpus de transcriptions de manuscrits, il est nécessaire de prendre en compte, en plus, un certain nombre de variables liées au type d'écriture et aux « rapports étroits et complexes qui existent entre le commanditaire, l'écrivain – souvent à la fois traducteur, copiste, éditeur – les scribes, les correcteurs et les miniaturistes » (Stiennon 1991, p. 166 ; Opsomer 2003) – et parfois les musiciens. N. Andrieux-Reix et

Méthodologie de corpus

S. Monsonégo, dans leur article consacré à la segmentation graphique et basé sur un corpus d'une quarantaine d'extraits de manuscrits, ont pris en compte (outre la date, le domaine et la forme du texte) « la pratique des écritures : calligraphique et cursive, livresque, usuelle et documentaire », ainsi que la « qualité de l'écriture » (Andrieux & Monsonégo 1997, p. 293).

Concernant cet aspect de la *caractérisation du manuscrit*, les variables suivantes sont distinguées dans le projet BFM-MSS où se situera notre corpus de textes transcrits : (10a) type d'écriture (caroline, gothique ; brisée, cursive, etc.)¹²⁷ ; (10b) « qualité » de l'écriture (plus ou moins « soignée »)¹²⁸ ; (10c) usage auquel le manuscrit est destiné (manuscrits « utilitaires » vs manuscrits « de luxe »).

Par ailleurs, en caractérisant les manuscrits, on prend en compte si possible les informations sur (10d) le destinataire (ou le dédicataire) et (10e) le commanditaire du manuscrit, ainsi que sur (10f) le(s) scribe(s). Un cas particulier, quoi que rare à l'époque médiévale est constitué par (10g) les manuscrits d'auteur (écrits ou corrigés par l'auteur du texte).

Il convient, enfin, de distinguer les manuscrits qui peuvent être décrits de façon précise (production ou correction par l'auteur du texte, date de production, nom du scribe, du scriptorium, du commanditaire, etc.) de ceux qu'on ne peut caractériser qu'approximativement.

Enfin, il peut être intéressant, dans un corpus de transcriptions de manuscrits, de contraster plusieurs versions d'une même œuvre pour mettre en évidence la variation graphique dans des contextes linguistiques proches ou identiques (Andrieux-Reix & Monsonégo 1997, p. 293). Dans ce cas, seules les variables « spécifiques » au manuscrit permettront d'opposer les unités du corpus.

127 Compte-tenu de la grande variabilité des écritures médiévales et de l'absence d'un inventaire systématique des types d'écritures, il est difficile de dresser une « liste d'autorité » pour les valeurs de cette variable. Pour les manuscrits de notre corpus, nous avons tenté de déterminer la « grande catégorie » à laquelle l'écriture semble appartenir et de préciser éventuellement des traits caractéristiques en nous appuyant sur les exemples cités dans (Stiennon 1991 ; 70 Careri *et al.* 2001)

128 Outre (Andrieux & Monsonégo 1997) ce paramètre a été retenu comme pertinent dans (Marchello-Nizia), même s'il ne figure pas habituellement dans les notices paléographiques. Il s'agit en effet d'une appréciation un peu grossière portant à la fois sur le type d'écriture et la destination du manuscrit. Son avantage consiste en sa simplicité.

3.3.4. Variables internes

Les variables internes sont définies par des indices calculés à partir de l'étiquetage de phénomènes (ou traits) linguistiques dans les textes. Elles dépendent donc de l'objet de recherche visé. L'analyse de ces variables permet de construire des « typologies inductives » des textes (Habert 2000).

Il existe probablement des corrélations entre divers phénomènes linguistiques de nature différente, dont la mise en évidence peut être en soi l'objectif d'une étude linguistique intéressante. A la lumière des études sur la typologie des langues et sur les universaux du langage, on peut supposer l'existence d'une certaine hiérarchie de phénomènes linguistiques. Autrement dit, la présence d'un phénomène linguistique peut « entraîner » la présence d'une série d'autres. On peut distinguer ainsi les traits linguistiques « dominants » des traits « impliqués ». Bien entendu, il ne s'agit pas forcément des mêmes traits dans une étude sur un corpus d'une langue et dans des recherches sur les universaux du langage¹²⁹, mais des liens sont possibles.

Dans le cas d'un corpus textuel unilingue, le « vecteur » de la dépendance entre traits peut être difficile à établir *a priori*. Par contre, il est possible de déterminer un certain nombre de traits « basiques » qui peuvent servir de repères pour une analyse approfondie de traits plus spécifiques¹³⁰. Parmi ces traits on peut citer la fréquence relative des parties du discours, des valeurs de certaines catégories morphologiques (par exemple, les temps verbaux) ou encore de certaines constructions syntaxiques¹³¹.

Les traits spécifiques sont sélectionnés en fonction de l'objectif d'une recherche linguistique concrète. Par exemple, pour l'étude des tendances de la ponctuation médiévale, le premier indice que nous utiliserons pour le classement des textes est le « taux de ponctuation » (fréquence relative des marques de ponctuation dans l'extrait analysé). Par la suite, nous calculons des indices plus fins en prenant en compte la forme de la marque de ponctuation, la « force » de la ponctuation et le type de frontière

129 Cf. chapitre 5 dans (Croft 2002).

130 Cf. les « typologies inductives “générales” » (Habert 2000).

131 Des exemples de calculs de certains traits morphosyntaxiques généraux et de leurs corrélations avec les genres textuels sont donnés dans (Bronckart 1996 ; Biber 1998, p. 55-76 ; Malrieux 2004, p. 73-85).

ponctuable sur laquelle la marque est employée (*cf* la méthodologie d'analyse linguistique présentée dans les sections 2.6 à 2.8 du chapitre précédent).

3.4. Constitution d'un corpus pour l'étude de la ponctuation médiévale

3.4.1. Grille des variables externes

Dans la pratique, il est très difficile, voire impossible, de constituer un corpus équilibré compte tenu de toutes les variables évoquées ci-dessus. D'autant que la plupart de ces variables sont en elles-mêmes assez complexes : elles peuvent se scinder en plusieurs sous-variables (date de composition de l'œuvre ou du manuscrit, etc.), et leurs valeurs ne peuvent pas toujours être déterminées avec précision (dates imprécises, textes appartenant à plusieurs domaines, etc.). Comme nous l'avons déjà remarqué, la constitution d'un corpus est toujours le résultat d'un compromis. Il faut sélectionner les variables les plus pertinentes et établir une grille des valeurs possibles. Une première proposition de grille de ce genre pour le projet BFM-MSS, est présentée dans les *Tableaux 1* et *2* donnés ci-dessous.

Le premier tableau contient les variables qui sont applicables à l'ensemble du corpus (éditions et/ou manuscrits), tandis que le deuxième présente les variables supplémentaires qui peuvent être prises en compte (en totalité ou en partie) uniquement dans le sous-corpus des textes « bien caractérisés ». La première colonne des tableaux spécifie le nom de la variable. La deuxième colonne précise l'objet pour lequel la variable en question est pertinente : corpus d'éditions critiques, de manuscrits ou les deux. La dernière colonne contient les types de valeurs utilisables pour chaque variable (entre crochets) et/ou des exemples de valeurs (entre guillemets).

Variable	Objet	Valeurs
(1a) Date de composition	Éditions, texte entier	[sous-siècle] (« début 13 ^e », « milieu 13 ^e », etc.)
(1b) Date du manuscrit	Manuscrits et éditions, texte entier	[sous-siècle]
(2a) Dialecte de l'auteur	Éditions, texte entier	[nom du dialecte] (si identifiable : « picard », « normand » ou « bourguignon »)
		« traits de [nom du dialecte] »
		« non déterminé »
(2c) Région du manuscrit	Manuscrits et éditions, texte entier	[nom de la région]
		« inconnu »
(4) Type de document	Éditions et manuscrits	« œuvre » ou « acte »
(5) Forme	Éditions et manuscrits, texte entier	« vers », « prose » ou « mixte »
	Séquences homogènes dans les textes mixtes	« vers » ou « prose »
(6) Domaine	Éditions et manuscrits, textes entiers	« littéraire », « religieux », « didactico-scientifique », « historico-politique » ou « juridique »
		« mixte » (préciser les domaines impliqués)
(8a) Type de discours	Éditions ou manuscrits, séquences textuelles « marquées »	« récit » (par défaut) ou « direct »
(8b) Élément de structure textuelle	Éditions ou manuscrits, séquences textuelles « marquées »	« corps » (par défaut), « titre », « liste », « note », « table des matières »
(8c) Type de séquence	Éditions ou manuscrits, séquences textuelles homogènes	« narratif », « descriptif », « argumentatif », « explicatif » ou « dialogal »
(8d) Première page	Manuscrits, une page physique de la source	« non » (par défaut), « oui »
(10b) « Qualité » de l'écriture	Manuscrits, textes entiers	« très soignée », « soignée » ou « peu soignée »
(10c) Destination du manuscrit	Manuscrits, textes entiers	« luxe », « ordinaire », « utilitaire »

Tableau 2: Variables applicables à l'ensemble du corpus.

Variable	Objet	Valeurs
(1a) Date de composition précise	Éditions	[année]
(1b) Date du manuscrit précise	Manuscrits	[année]
(2d) Lieu d'exécution	Manuscrits	[ville]
		[type] (« scriptorium », « chancellerie » ou « atelier »)
		[nom]
(3) Auteur	Éditions et manuscrits	[nom]
		[position sociale]
(6c) Thème ¹³²	Éditions et manuscrits, textes entiers	« médecine », « grammaire », « vie quotidienne », etc.
(10d) Destinataire	Éditions et manuscrits	[nom] et/ou [position sociale]
(10e) Commanditaire	Manuscrits	[nom] et/ou [position sociale]
(10f) Scribe(s)	Manuscrits	[nombre de mains]
		[nom(s)]
(10g) Manuscrit d'auteur	Manuscrits	« non », « oui », « corrigé par l'auteur », « inconnu »

Tableau 3: Variables applicables à un sous-corpus de textes caractérisés plus précisément.

C'est dans la deuxième partie de notre travail (chapitres 6 à 8) que nous présenterons la description de chacun des manuscrits de notre corpus. Par ailleurs, des fiches descriptives de chacun des extraits transcrits et analysés sont présentées dans l'annexe III.

3.4.2. Échantillonnage et représentativité du corpus

Comme nous l'avons vu plus haut, la représentativité et la notion même du corpus sont conditionnées par les objectifs des recherches auxquelles le corpus est destiné.

Un corpus représentatif est censé minimiser le risque des deux types d'erreurs statistiques qui menacent les généralisations à partir de lui : l'« incertitude » (*random error*) et la « déformation » (*bias error*) :

¹³² S'applique aux textes du domaine didactico-scientifique.

L'incertitude survient quand un échantillon est trop petit pour représenter avec précision la population réelle. Une déformation se produit quand les caractéristiques d'un échantillon sont systématiquement différentes de celles de la population que cet échantillon a pour objectif de refléter (Habert 2000 ; cf. Biber 1993, p. 219-220).

On peut donc distinguer deux dimensions de la représentativité du corpus : qualitative et quantitative. La dimension qualitative (liée au problème de la déformation) se traduit par la représentation de toutes les combinaisons possibles des variables externes pertinentes, et la dimension quantitative (liée au problème de l'incertitude) consiste en la sélection d'un nombre suffisant et équilibré d'unités textuelles pour chacune des combinaisons de variables et dans la « taille » suffisante de ces unités, s'il s'agit d'échantillons. Le choix des variables externes et l'équilibrage du corpus peuvent éventuellement être reconsidérés *a posteriori*, compte tenu des variables internes de la typologie « inductive » (cf. la section 3.3.4. ci-dessus).

La grille des variables externes retenues pour notre analyse a été présentée dans la section précédente, il convient à présent d'aborder la dimension quantitative. Il n'existe pas de solution universelle :

Il est en effet difficile d'envisager une norme absolue en matière de « taille » [du corpus], du fait qu'il n'existe pas de critères véritablement objectifs pour la fonder. L'essentiel demeure donc de veiller à la cohérence et à l'homogénéisation quantitatives internes du corpus (Prévost 2008, p. 37).

Les premiers corpus utilisés en linguistique étaient le plus souvent composés d'échantillons de quelques milliers de mots de textes de types variés. Par exemple, les corpus d'anglais « Brown » (Université Brown, États-Unis) et LOB (Lancaster-Oslo-Bergen) sont composés de 500 échantillons de 2000 mots chacun ; le Corpus d'Amsterdam (textes littéraires) constitué par Anthonij Dees dans les années 1980 inclut quelques textes intégraux, mais surtout des échantillons, dont un quart (74 sur 300 unités textuelles) comptent entre 1000 et 5000 mots¹³³.

L'une des raisons de cette limitation était le coût de la numérisation des textes (alors saisis à la main) et la difficulté de gérer des masses de données trop importantes.

133 D'après la bibliographie de la nouvelle version du corpus mise au point par Achim Stein et Pierre Kunstmann <<http://www.uni-stuttgart.de/lingrom/stein/corpus/>>.

Méthodologie de corpus

Les contraintes de gestion de masses de données textuelles ne sont plus d'actualité, et le nombre de textes numérisés (ou numériques « de naissance ») ne cesse de croître pour les langues modernes. Des corpus de plus en plus grands et variés peuvent être constitués aux moindres frais, et l'usage des textes intégraux est à privilégier (*cf.* Prévost 2005, p. 157) Néanmoins, lorsqu'il s'agit de transcriptions fines de manuscrits, le problème du coût et du temps nécessaire à la préparation des des textes reste d'actualité, c'est pourquoi le recours à l'échantillonnage est indispensable dans notre cas par exemple.

Par ailleurs, l'usage des textes intégraux n'est sans doute pas absolument nécessaire pour les recherches portant sur les systèmes graphiques et la ponctuation : même un échantillon de quelques centaines de mots permet d'estimer la fréquence des principales marques de ponctuation et de repérer les phénomènes graphiques caractéristiques. Il convient toutefois de rester prudent en procédant à des généralisations. Il faut notamment de tenir compte des variables caractérisant les séquences à l'intérieur des textes pour faire face à l'hétérogénéité éventuelle d'une unité textuelle (Prévost 2005, p. 157).

Les premières transcriptions du corpus BFM-MSS portaient sur des échantillons de 500 à 700 mots (1 à 2 folios du manuscrit), ce qui correspond à la taille des fragments de manuscrits cités dans des albums paléographiques¹³⁴. Pour les textes transcrits plus récemment, la taille des échantillons a été augmentée pour atteindre 2000 mots environ. Les échantillons représentent le plus souvent le début du texte (mais pas nécessairement le début du manuscrit), mais ne se limitent jamais à la première page. Certains passages appartenant à des éléments particuliers de la structure textuelle ont été exclus afin de réduire le risque de déformation. Il s'agit des tables de matières, de longues listes ou énumérations. Dans l'avenir, le corpus devra être complété par des échantillons du milieu et de la fin de chaque texte.

En ce qui concerne la dimension qualitative, Douglas Biber (1990) suggérait que chaque catégorie soit représentée par une dizaine de textes. Compte tenu du grand nombre de variables prises en compte dans un corpus diachronique (en particulier, le temps et l'espace), une telle exigence devient difficile à respecter. Le seuil minimal que

¹³⁴ Par exemple, 300 – 400 mots dans (Careri *et al.* 2001).

nous avons fixé initialement (Lavrentiev 2007e) s'élevait à 3 textes pour chaque combinaison des variables externes retenues pour l'équilibrage du corpus. L'étiquetage de certains marqueurs linguistiques dans cette première sélection de textes devrait permettre d'obtenir une typologie inductive qui pourrait ensuite servir à rééquilibrer le corpus. En réalité, certaines combinaisons de variables peuvent tout simplement ne pas être attestées dans les manuscrits conservés jusqu'à nos jours, il risque donc d'y avoir des cases qui ne seront jamais remplies. Il convient de tenir compte de ces lacunes dans l'évaluation de la généralité des résultats.

Un corpus peut être délibérément limité par rapport à certaines variables, pour mieux assurer la représentativité dans un sous-ensemble plus restreint. C'est le cas de notre travail, où le corpus est limité aux œuvres en prose du XIII^e au XV^e siècle (*cf* l'Introduction pour les raisons de ce choix)¹³⁵.

3.4.3. Caractéristiques du corpus BFM MSS

Le corpus que nous avons constitué pour cette thèse est composé de 27 échantillons de textes de 550 à 2168 mots¹³⁶, soit 28 219 occurrences-mots. L'ensemble des textes du corpus sont transcrits selon la méthode de représentation « multi-facettes » décrite dans les chapitres 4 et 5 ci dessous. Les marques de ponctuation (définies par leur « forme » et par leur « force ») et les types de « frontières ponctuables » sont annotées de façon exhaustive. Le corpus intégré dans l'outil de recherche et d'analyse

135 Trois textes du corpus se situent toutefois « en dehors » des limites ainsi posées : les *Serments de Strasbourg* et deux imprimés du XVI^e siècle. Ils constituent en quelque sorte les « repères externes » pour situer nos observations sur le corpus principal.

136 Sans compter la transcription intégrale du texte en ancien français des *Serments de Strasbourg*, un texte très court, de 115 mots seulement et la transcription du manuscrit d'Oxford (Bodleian, Laud. Misc. 587) de Villehardouin (359 occurrences-mots) qui n'était pas achevée au moment de l'annotation et de l'intégration du corpus. Si on comptabilise ces deux textes supplémentaires, on obtient le chiffre de 28 693 mots qui figure dans l'Introduction et au début de ce chapitre. Pour deux incunables, *La mort au roy artus* (1488) et *L'image du monde*, nous disposons par ailleurs de transcriptions moins « fines » sans annotation de frontières ponctuables réalisées dans le cadre de notre DEA. Ces transcriptions comportent près de 25 000 et 10 000 occurrences-mots environ respectivement. Certains indices (comme le taux de ponctuation général et la fréquence relative des différentes marques) ont pu être obtenus à partir de ces échantillons élargis pour ces textes. Nous avons également pu obtenir certaines indices à partir de l'édition intégrale de la *Queste del Saint Graal* (ms. Lyon, BM, P.A. 77, près de 120 000 occurrences-mots) par Ch. Marchello-Nizia à laquelle nous avons collaboré et qui respecte scrupuleusement certains aspects de la ponctuation scribale.

Méthodologie de corpus

textométrique Weblex¹³⁷ est interrogeable en ligne¹³⁸. Les dimensions de chaque extrait avec l'indication du domaine, du genre et de la date « médiane »¹³⁹ du manuscrit (ou de l'imprimé ancien) sont présentées dans le Tableau 4 ci-dessous. Les textes du corpus sont référencés par les identifiants utilisés dans notre base des descripteurs et dans Weblex. Les titres complets et les références des manuscrits sont présentés dans une section spéciale de notre bibliographie. Les fiches complètes comportant l'ensemble des descripteurs significatifs concernant chacun des textes sont données dans l'annexe III.

Les dimensions des extraits sont indiquées par le nombre d'occurrences-mots. Ces données sont complétées par le nombre d'unités ponctuables annotées et par le nombre total d'occurrences des marques de ponctuation dans chaque échantillon. Cela nous permet de voir très rapidement la densité de ponctuation que présente chaque texte.

Comme nous l'avons déjà mentionné, les dimensions modestes de notre corpus s'expliquent par notre choix d'investir dans la qualité des transcriptions, condition nécessaire pour pouvoir disposer de données fiables et pour assurer la pérennité, l'interopérabilité et l'extensibilité du corpus.

137 Logiciel développé par S. Heiden (UMR 5191 ICAR, Université de Lyon / CNRS), <<http://weblex.ens-lsh.fr/wlx/>>.

138 L'accès au corpus est soumis à l'acceptation des conditions générales d'utilisation de la Base de Français Médiéval. Il convient de s'adresser au gestionnaire de la BFM (bfm@ens-lsh.fr) pour obtenir de plus amples informations et les codes d'accès.

139 Voir la section 3.3.3.1. pour l'explication de la méthode de calcul de cette date.

Siècle	Dom.	Genre	Texte	Date	Mots	U.P.	M.P.
<i>Manuscrits</i>							
11	juridique	serment	strasb9768	999	115	26	15
13	didactique	encyclop.	im587	1299	644	145	21
		historique	chronique	tyr828	1279	572	126
	chronique		villehardouin12203	1282	673	135	80
	juridique	traité	beauma11652	1299	1267	250	64
	littéraire	roman	lancelot77	1249	565	122	51
		roman	qgraal77	1249	2065	406	169
roman		qgraal1119	1269	724	159	29	
14	didactique	encyclop.	im574	1316	550	121	74
		traité	dictier840	1399	2093	502	155
	historique	memoires	joinville13568	1334	1107	188	105
		chronique	villehardouin587	1374	359	66	34
		chronique	froissart869	1401	1053	190	58
	religieux	prédication	joachim768	1349	592	112	46
15	didactique	traité	chasse619	1439	751	155	75
		traité	chasse765	1449	585	118	37
		desc. geo.	egipte769	1457	613	109	31
	historique	chronique	monstre2682	1461	761	156	57
	littéraire	lyrique	quadrilogue126	1471	1655	307	125
		roman	artois11610	1435	1119	217	51
		nouvelle	qjm1052	1441	659	147	32
		r. voyage	anglure15217	1464	1962	382	152
		biographie	phares1357	1474	2168	517	177
	politique	proposition	ursins2701	1494	1578	280	64
religieux	sermon	gerson13318	1449	2102	422	115	
<i>Imprimés</i>							
15	littéraire	roman	artu1488	1488	596	122	48
	didactique	encyclop.	im1495	1495	575	140	27
16	didactique	encyclop.	im1522	1522	577	140	40
		traité	dolet1540	1540	614	119	107

Légende :

La colonne **Texte** présente l'identifiant de l'échantillon dans notre corpus (voir la Bibliographie pour les références complètes). La colonne **Date** indique la « date médiane » calculée. La colonne **Mots** donne le nombre d'occurrences-mots dans l'échantillon, la colonne **U.P.** indique le nombre d'unités ponctuables annotées et la colonne **M.P.** présente le nombre d'occurrences des marques de ponctuation relevées.

Tableau 4 : Dimensions du corpus BFM-MSS

Méthodologie de corpus

Cinq des extraits intégrés dans le corpus actuel comportent près de 2000 occurrences-mots chacun et correspondent à l'« échantillon du début », selon notre projet de corpus. Les autres échantillons, plus restreints, ont été saisis antérieurement, dans le cadre de préparation de communications à des colloques. Une « première couche » de 16 transcriptions de 500 à 700 mots environ correspond en gros au corpus utilisé pour notre intervention au séminaire du GDR « Diachronie du français » en 2001 ; une « deuxième couche » de 6 textes de 1000 à 1500 mots est liée à la préparation de la communication au colloque « Diachro 2 » (Paris, janvier 2004)¹⁴². Deux transcriptions, les *Serments de Strasbourg* (115 mots) et un extrait du manuscrit d'Oxford de Villehardouin (359 mots) ont été réalisées « en marge » de notre travail pour des raisons diverses¹⁴³ et nous en avons profité pour les ajouter au « pot commun ».

Tous ces extraits¹⁴⁴ devraient être complétés pour atteindre les dimensions prévues par le projet de corpus, mais les contraintes de temps nous ont obligé à procéder à l'annotation et à l'analyse des données du corpus dans son état inachevé.

En aucun cas, le corpus actuel ne peut être considéré comme représentatif, d'autant plus que les pratiques de la ponctuation médiévale se caractérisent par une extrême variabilité selon les manuscrits (et parfois même à l'intérieur d'un manuscrit), ce qui rend difficile toute tentative de généralisation (*cf.* Marchello-Nizia 1978, p. 42 ; Baddeley 2001, p. 139 ; Llamas Pombo 2001, p. 153). Nous avons par ailleurs présenté un exemple concret de la manière dont la prise en compte de nouveaux textes peut amener à reconsidérer les hypothèses formulées à partir d'un corpus plus restreint (Lavrentiev 2007e).

Cependant malgré la taille limitée du corpus, la description détaillée des textes et la prise en compte des différentes variables de la typologie textuelle permettent de contextualiser les phénomènes observés et d'émettre des hypothèses concernant les corrélations entre les valeurs des variables et le résultats de l'analyse linguistique. La prise en compte des résultats des recherches sur la ponctuation médiévale déjà publiés

142 Cette communication a donné lieu à une publication (Lavrentiev 2007c).

143 Nous y reviendrons, lors de l'analyse détaillée de ces textes dans les chapitres 6 et 7 respectivement.

144 A l'exception des *Serments de Strasbourg* et de la *Proposition* de Jean Juvénal des Ursins, deux petits textes déjà transcrits intégralement.

permettra de corriger ou d'appuyer nos hypothèses, ce qui constituera une avancée des connaissances dans ce domaine.

3.5. Conclusion : notre corpus comparé aux autres

Les caractéristiques du corpus que nous avons constitué diffèrent d'une façon significative de celles des corpus des autres recherches menées jusqu'à ce jour. Les corpus de la grande majorité des études ont été limités à un nombre très limité de manuscrits représentés soit par des échantillons, soit par des sondages non exhaustifs (Marchello-Nizia 1978 ; Barbance 1995 ; Lavrentiev 2000 ; Baddeley 2001 ; 71 Llamas Pombo 2001 ; Li 2007).

Parmi les rares exceptions il convient de mentionner la thèse récente de Nicolas Mazziotta (2007b), qui a procédé à une analyse systématique d'un corpus de 285 chartes liégeoises datées de 1236 à 1291 (64 252 occurrences-mots au total). C'est certainement le corpus le plus important qui ait jamais été constitué pour une étude de la ponctuation médiévale, mais il est très restreint du point de vue typologique : il ne couvre qu'un type de document (la charte), localisé à un endroit (région de Liège) sur une période de moins de 60 ans. Cette limitation volontaire a permis au chercheur de construire un corpus représentatif du type de texte choisi, mais elle ne permet pas de situer les résultats dans un contexte plus général de l'évolution des pratiques de ponctuation dans les textes en français médiéval. Par ailleurs, l'annotation de ce corpus est basée sur des critères exclusivement syntaxiques.

Un autre corpus assez grand et diversifié a été constitué par Nelly Andrieux-Reix et Simone Monsonégo pour leur étude des segments graphiques (1997). Il s'agit d'une collection de 43 manuscrits¹⁴⁵ sélectionnés en fonction des paramètres suivants :

- la diachronie : sondages dans des écrits produits du IX^e au XI^e siècle, mais surtout du XII^e à la fin du XV^e s. ;
- l'usage particulier de la langue selon la destination du texte : vers et prose, œuvres littéraires et documents d'archives, prose rythmique et prose didactique ;

¹⁴⁵ Soit 32 textes différents, car certaines œuvres sont représentées par 3 ou 5 manuscrits.

Méthodologie de corpus

- la pratique des écritures : calligraphique et cursive, livresque, usuelle et documentaire ;
- la possibilité de comparaison, entre plusieurs manuscrits de la même œuvre, des occurrences de groupes composés à l'identique en contexte identique ;
- la proximité dans le temps de l'original et des copies qui en ont transmis la teneur (*ibidem*, p. 292-293).

Ces paramètres sont en effet très proches des variables de la typologie textuelle de la BFM, mais la description des textes est toutefois moins précise chez Andrieux-Reix et Monsonégo que dans la BFM : les types de document, les domaines et les genres y sont mélangés, la variation géographique n'est pas prise en compte. L'intention de prendre en compte « un peu de tout » amène à ce que chaque catégorie de textes est représentée par un très petit nombre d'échantillons. Leur article est accompagné d'une annexe contenant des transcriptions d'extraits de certains manuscrits : 14 au total, d'une centaine de mots chacun (soit près de 1600 occurrences-mots en tout).

Par ailleurs, on ne sait presque rien sur la « cuisine interne » de ce corpus. Tous les manuscrits du corpus ont-ils été transcrits systématiquement ? Si oui, quelle est la taille des échantillons ? Quel est le format des transcriptions ?¹⁴⁶ Comment les textes ont-ils été annotés ? Autant de questions qui restent sans réponse à la lecture de l'article (par ailleurs extrêmement intéressant et novateur). On ne peut que regretter que ce corpus unique dans son genre ne soit pas rendu accessible pour des utilisations et enrichissements ultérieurs.

L'*Album de manuscrits français du XIII^e siècle* (Careri *et al.* 2001) pourrait être utilisé comme un corpus (ou plutôt un « réservoir à corpus ») pour les études sur les systèmes graphiques, si les transcriptions « diplomatiques » qu'il contient étaient accessibles sous forme électronique. Cet ouvrage est en effet composé de 52 notices paléographiques accompagnées de reproductions d'une page de chacun des manuscrits étudiés, ainsi que d'une « transcription diplomatique » et d'une « édition »¹⁴⁷ de cette même page (300 – 400 mots environ). La valeur de cette publication, qui est immédiatement devenue un outil précieux pour le travail de linguistes et de

146 Nous reviendrons sur l'aspect méthodologique des transcriptions du corpus Andrieux-Reix-Monsonégo dans la section 4.3.

147 Nous examinerons de plus près les caractéristiques de ces transcriptions dans la section 4.3.

paléographes, serait certainement encore augmentée si elle était éditée électroniquement et dotée d'un moteur de recherche et d'outils de navigation permettant de passer d'un mode de lecture à un autre.

Les sections consacrées à l'histoire de la ponctuation française médiévale dans (Mortara Garavelli éd. 2008) sont rédigées à partir d'observations sur quelques manuscrits sans indication de la taille des échantillons, il ne s'agit donc pas d'un véritable travail sur corpus.

Notre corpus a été conçu sur la base des réflexions que nous venons de mener ; il a pour but de pouvoir être réutilisé, élargi et enrichi par tous les chercheurs qui s'intéressent au fonctionnement des systèmes graphiques de manuscrits médiévaux et de livres anciens, ainsi qu'aux rapports entre le texte « normalisé » établi par les éditeurs philologues et la réalité du manuscrit médiéval. Il contient des textes assez variés s'étendant sur un empan temporel de 200 ans. N'étant pas représentatif au sens de la linguistique de corpus, il est néanmoins considérablement plus varié que les corpus de nos prédécesseurs et permet de dresser une première ébauche, même si elle est encore partiellement impressionniste, des grandes tendances de l'évolution de la ponctuation française médiévale.

Chapitre 4

Représentation des données primaires : transcription de manuscrits médiévaux

Pour une étude de la ponctuation médiévale, la nécessité de pouvoir disposer d'un corpus de transcriptions fines des sources primaires est incontestable : la grande majorité des éditions traditionnelles ne conserve aucune trace de la ponctuation des manuscrits.

Le problème de l'authenticité des données que le chercheur a à disposition se pose de façon plus générale pour toute approche de la langue médiévale. Tout linguiste qui s'intéresse aux textes de cette période doit se poser la question de savoir s'il peut travailler à partir des éditions scientifiques préparées par des philologues ou historiens médiévistes ou si la consultation directe de manuscrits est indispensable pour son étude.

Le problème de la représentation des données linguistiques dans les éditions traditionnelles et électroniques des textes médiévaux nous semble suffisamment important pour qu'on lui consacre un chapitre entier.

Nous n'évoquerons pas ici les questions du choix des variantes, de corrections, etc., mais insisterons davantage sur les questions de graphies et de ponctuation, qui concernent notre étude.

4.1. Éditions modernes comme source de données linguistiques

Lydia Stanovaïa (2003) critique vivement les linguistes qui ne font pas de distinction entre le texte comme élément de discours ('contenu') et le manuscrit ('contenant'). Les éditeurs des textes cherchent, selon elle, à « séparer la langue de l'auteur de celle des scribes, pour reconstruire tout le texte original et pour le faire éditer ensuite ». La chercheuse affirme que « toute segmentation, quelque argumentée qu'elle semble (formes archaïques vs modernes ; littéraires vs populaires ; correctes vs incorrectes ; etc.) détruit *l'unité* sémantique et linguistique du texte ». Elle en conclut qu'« un chercheur visant à obtenir les données réelles doit absolument étudier les *manuscrits* réels, existants, conservés, faits à l'époque étudiée, et non pas les *textes* irréels, hypothétiques, reconstruits par des philologues » (*ibidem*, p. 244-246).

Il est pourtant bien connu qu'une grande partie des éditions de textes médiévaux en langue vernaculaire publiées en France à partir des années 1920 sont de type « bédieriste » (Foulet & Speer 1979, p. 28 ; Bourgain & Viellard coord. 2002, p. 19), c'est à dire fidèles au « meilleur manuscrit » choisi par l'éditeur scientifique comme le manuscrit de base :

Aussi la méthode d'édition la plus recommandable est-elle peut-être, en dernière analyse, celle qui régit un esprit de défiance de soi, de prudence, d'extrême "conservatisme", un énergique vouloir, porté jusqu'au parti pris, d'ouvrir aux scribes le plus grand crédit et de ne toucher au texte d'un manuscrit que l'on imprime qu'en cas d'extrême et presque évidente nécessité : toutes les corrections conjecturales devraient être reléguées en quelque appendice (Bédier 1928, p. 356).

Il est donc possible de supposer que les « bonnes » éditions bédieristes représentent le texte écrit du manuscrit. Il faut, bien entendu, tenir compte de la perte d'une partie des données primaires due à la normalisation typographique pratiquée dans les éditions (nous y reviendrons plus bas), mais ces données ne sont utiles que pour certains types de recherches linguistiques très ciblées (comme notre étude de la ponctuation). De même que dans les recherches sur la langue orale il n'est pas toujours nécessaire de recourir à des transcriptions phonétiques fines, ainsi dans les recherches

Représentation des données primaires : transcription de manuscrits médiévaux

sur le lexique ou la morphosyntaxe du français médiéval, les traits particuliers des systèmes graphiques des manuscrits n'ont pas beaucoup d'importance, et l'effort qui serait nécessaire pour assumer leur gestion peut, au contraire, être contre-productif. Les grands corpus de textes de français médiéval (la Base de Français Médiéval¹⁴⁸ et la Base Textuelle du *Dictionnaire du Moyen Français*¹⁴⁹, par exemple) ont été constitués à partir des textes d'éditions critiques, et ces corpus sont devenus des outils pratiquement incontournables pour les recherches linguistiques sur le français médiéval.

En réalité, le degré d'interventionnisme varie beaucoup selon les éditeurs. Comme le remarquent Alfred Foulet et Mary Blakely Speer, pratiquement chaque éditeur « formule en effet sa propre philosophie éditoriale » (1979, p. XIII). Cette variation est due non seulement (et non tellement) aux goûts personnels des éditeurs, mais aussi et surtout à la nature des œuvres éditées et à la visée de l'édition. Selon Foulet et Speer,

un système d'édition purement mécanique n'est pas adapté pour les textes d'ancien français... Toute édition est une conjecture, elle constitue aussi une interprétation critique du texte. En établissant le texte, l'éditeur doit tenir compte de trois facteurs principaux : l'auteur, le scribe et le lecteur. Des circonstances de la tradition manuscrite aident souvent à déterminer si la priorité doit être attribuée au texte de l'auteur ou à la copie du scribe... Autrement dit, l'édition n'est pas une science, mais un art (*ibidem*, p. 38-39).

Pascale Bourgain et Françoise Vielliard, plus récemment, formulent la recommandation suivante :

Il faut adapter l'édition envisagée à l'intérêt du texte et à l'utilisation qui en sera faite. Si l'œuvre risque de ne pas avoir d'autre édition, autant donner le plus de renseignements possibles, donc ne pas se contenter de l'apparat léger suffisant pour une édition "générale", proche des éditions scolaires, recommandable pour les textes très souvent édités et lus par des utilisateurs qui n'ont besoin que du résultat de la démarche critique et non de sa démonstration. Et si l'œuvre est très répandue, il sera utile de garder pas mal de variantes en apparat, pour qu'on puisse voir dans quelle famille de manuscrits se sont servies les œuvres postérieures qui ont utilisé le texte.

148 Site internet : <<http://bfm.ens-lsh.fr>>. L'accès au contenu du site et à la base textuelle est à ce jour (1 mai 2009) suspendu pour des raisons juridiques. Nous espérons que le site public et la base textuelle deviendront de nouveau accessibles à la communauté des chercheurs dans les meilleurs délais.

149 Page web d'accès au DMF et à sa base textuelle : <<http://www.atilf.fr/dmf/>>.

On ne peut espérer faire une édition utile pour l'histoire des systèmes graphiques et de la langue avec 50 manuscrits, mais avec un manuscrit de base ancien ou autographe cela vaut la peine. Combiner l'agrément du lecteur et l'intérêt de la recherche, savoir à quoi son édition pourra raisonnablement servir, c'est cela aussi le jugement critique de l'éditeur (Bourgain & Viellard coord 2002, p. 25 ; cf. Viellard & Guyotjeannin coord. 2001, p. 18).

Il convient donc d'étudier soigneusement les déclarations éditoriales (ainsi que les compte-rendus publiés par la suite) avant de juger si telle ou telle édition est utilisable en tant que source de données textuelles pour tel ou tel type de recherche linguistique. Le mieux serait de disposer d'un outil permettant de quantifier précisément l'écart entre les données primaires du manuscrit et les données traitées de l'édition. La création d'un tel outil est tout à fait possible grâce à l'informatique, qui va par ailleurs changer profondément les pratiques éditoriales dans un avenir relativement proche.

Dans les années 1980 et 1990, quand les partisans de la « nouvelle philologie » parlaient des éditions informatisées comme de l'alternative aux éditions classiques permettant de représenter d'une façon plus adéquate la « mouvance » de la tradition littéraire médiévale, la majorité des philologues étaient restés sceptiques sur ce sujet, faute d'exemples concrets convaincants.

De nos jours, ce qui était un rêve aux années 1980 est sur le point de devenir une réalité banale du point de vue technologique. Les éditions électroniques de tous genres commencent à concurrencer sérieusement les livres imprimés (dont la préparation ne se passe plus de l'informatique depuis plusieurs décennies déjà). Pour l'instant, les livres traditionnels continuent d'être imprimés pour trois raisons principales : le confort de lecture qu'ils offrent, le coût du matériel informatique et l'exploitation encore faible des avantages que l'édition informatisée puisse présenter par rapport à un ouvrage imprimé. La rapidité du progrès technologique actuel fait penser que le confort de lecture sur un support informatique sera très bientôt supérieur à celui offert par un livre traditionnel. Des prototypes de « papier électronique » (*electronic paper*) (Daviss 1999) et des modèles de lecteurs de livres électroniques légers et performants existent déjà. Le prix de ce matériel reste élevé, mais il deviendra certainement de plus en plus accessible dans les années qui viennent. En ce qui concerne l'exploitation des possibilités offertes par l'informatique, celle-ci est limitée par les habitudes des auteurs et des éditeurs,

plutôt que par des contraintes technologiques. Les éditions de textes médiévaux, pour lesquels le problème d'adaptation de la forme de présentation aux intérêts des différents groupes de lecteurs (spécialistes en histoire, littérature ou linguistique, ou encore un public plus large) et d'outillage efficace (apparat critique, index, glossaires, etc.) est particulièrement important, peuvent représenter un très bon terrain d'essai pour les approches novatrices.

Bien entendu, il n'est pas question que l'informatique se substitue au travail de l'éditeur scientifique : l'interprétation des passages obscurs, la correction des erreurs, le repérage et l'évaluation des variantes sont indispensables pour la compréhension d'une œuvre littéraire ou d'un document historique. Les outils informatiques peuvent aider l'éditeur dans son travail, mais ne le remplaceront jamais. En revanche, la souplesse d'une édition hypertextuelle permettant d'adapter l'affichage du texte aux intérêts du lecteur épargne à l'éditeur de faire le choix entre la lisibilité pour les lecteurs non spécialistes et la richesse de l'apparat critique. Tout travail d'édition commence par une transcription diplomatique d'un ou de plusieurs manuscrits d'une œuvre médiévale. Chacune de ces transcriptions peut être intégrée dans une édition électronique, connectée d'une manière ou d'une autre au texte critique établi par l'éditeur¹⁵⁰. Le projet *Charrette*¹⁵¹, initié par Karl D. Uitti à la suite de l'édition « traditionnelle » du *Chevalier de la charrette* de Chrétien de Troyes (Foulet & Uitti 1989), est un excellent exemple de ce type d'édition.

Il est clair que tous les textes médiévaux ne pourront pas être réédités intégralement d'une façon semblable dans un avenir proche. En revanche, on peut envisager de compléter les éditions traditionnelles par des échantillons de transcriptions diplomatiques de quelques pages à l'occasion de leur publication sur un support numérique : toutes les grandes maisons d'édition de textes médiévaux ont actuellement des projets de numérisation de leur collections¹⁵². Ceci permettrait, entre autre, de

150 François Masai (1950, p. 182-185) a présenté les arguments, toujours valables, en faveur des éditions diplomatiques qui doivent accompagner les éditions pour la plupart des textes médiévaux. Ce n'est que grâce à l'informatique qu'une telle entreprise devient envisageable dans la pratique courante.

151 <<http://lancelot.baylor.edu/>>.

152 Une grande partie des textes édités par la maison *Champion* sont accessibles sous forme numérique dans la version électronique de la collection *Classiques Garnier* (anciennement *Champion électronique*) <<http://www.classiques-garnier.com/>> accessible sur abonnement. La librairie Droz <<http://www.droz.org/>> s'apprête quant à elle à numériser ses éditions de textes médiévaux dans les

quantifier précisément l'écart entre les données textuelles du manuscrit et celles de l'édition¹⁵³.

L'intégration de transcriptions diplomatiques dans une édition électronique, en complément du texte critique et d'une manière qui ne gêne en aucun cas pas le lecteur, ne peut donc être que bénéfique. Il convient alors de préciser ce qu'est une transcription diplomatique : quelles données primaires de l'écriture manuscrite y sont représentées et comment.

4.2. Tradition « diplomatique »

La transcription du manuscrit (ou de plusieurs manuscrits) fait partie de la préparation de toute édition. Mais ces transcriptions sortent rarement de la « cuisine interne » de l'éditeur et restent conservées, au mieux, dans son archive personnelle (*cf.* Masai 1950, p. 184). Les seules ressources publiées où l'on puisse trouver de telles transcriptions sont des éditions dites diplomatiques et des manuels de paléographie.

L'histoire des tentatives d'éditeurs pour reproduire aussi fidèlement que possible l'original manuscrit est très ancienne. On peut même dire que les incunables étaient à l'origine des sortes d'« édition imitative », car les premiers imprimeurs cherchaient à produire un objet qui ressemblait à un « beau manuscrit »¹⁵⁴. Progressivement, des normes propres aux textes imprimés se sont formées (et ont commencé à influencer, à leur tour, l'écriture manuscrite). Le système graphique des manuscrits médiévaux, avec sa multitude de variantes calligraphiques des caractères, d'éléments décoratifs et d'abréviations, est devenu peu à peu complètement étranger à la pratique courante.

Cet écart est devenu sensible dès le XVII^e siècle où est apparue *la diplomatique*, « la science qui étudie la tradition, la forme et l'élaboration des actes écrits » (CID 1997). Ses origines remontent aux premières tentatives de la critique des faux au Moyen

années qui viennent. La société savante *Anglo-Norman Text Society* (ANTS) a pour sa part autorisé le projet Anglo-Norman Online Hub <<http://www.anglo-norman.net/>> à numériser et à mettre en ligne un nombre important de ses éditions. Ces textes peuvent être librement interrogés ou « feuilletés » par n'importe quel internaute.

153 Le calcul d'un « indice de fidélité » pour les quelques textes qui font à la fois partie de notre corpus et de la Base de Français Médiéval est l'un des « produits dérivés » de notre travail (Lavrentiev 2007a).

154 Parfois il est même difficile de distinguer une page d'un incunable illuminé d'une feuille d'un manuscrit de la même époque (Saenger 1997, p. ix).

Représentation des données primaires : transcription de manuscrits médiévaux

Âge, mais le terme et la discipline sont nés avec les travaux du moine bénédictin dom Jean Mabillon (1632 – 1707). Dans les premiers traités de diplomatique, Mabillon (1709) et ses successeurs (Toussain & Tassin 1750) ont reproduit (en dessinant) des échantillons d'écriture manuscrite et les ont accompagnés de transcriptions en caractères modernes.

Par la suite, les reproductions dessinées (comme chez Mabillon ou Toussain et Tassin) ont été remplacées, d'une part, par des transcriptions typographiées imitant plus ou moins finement les marques graphiques des manuscrits (*cf.*, par exemple, de Wailly 1838, p. 454-467) et, d'autre part, à partir du XX^e siècle, par des facsimilés photographiques.

Avec le développement de la tradition des éditions critiques à partir du milieu du XIX^e siècle, les éditions reproduisant fidèlement (voire « mécaniquement ») un seul document manuscrit deviennent rares en dehors du domaine des documents d'archive et des manuels de paléographie. Si pour les éditions critiques de textes français les normes de transcription « normalisée » se sont plus ou moins stabilisées avec la publication des recommandations de la 2^e commission à la réunion des romanistes de 1925 (Roques 1926), les pratiques dans le domaine diplomatique restent plus variées.

François Masai (1950) plaide pour « une standardisation des techniques, encore trop individuelles et incontrôlables » et propose un certain nombre de principes et de signes conventionnels. Pour lui, « l'effort principal doit porter sur le relevé méticuleux des altérations subies par les textes » (p. 189), la plupart de ses signes conventionnels servent donc à noter les additions, les suppressions et les substitutions visibles dans le document transcrit. Ses recommandations ne sont pas, en revanche, très précises en ce qui concerne la représentation des traits particuliers des systèmes graphiques des manuscrits. Le chercheur parle du soin qu'il faut mettre « à copier exactement le texte avec toutes ses particularités d'orthographe et de ponctuation », sans pour autant aller « à l'encontre des intérêts supérieurs du lecteur ni des possibilités de l'imprimerie ». Ainsi, « employer le même style de lettres que le manuscrit à éditer » est souhaitable, mais pas nécessaire. Rien n'est dit sur la distinction des variantes de caractères, aucune précision n'est donnée non plus concernant la représentation de la ponctuation médiévale. Pour les manuscrits en *scriptura continua*, Masai propose d'introduire des

petits blancs entre les mots pour aider le lecteur à comprendre le texte et d'utiliser des espaces nettement plus grands pour représenter les espaces du manuscrit (espaces servant de ponctuation). Les abréviations peuvent « suivant les circonstances » être rendues soit comme dans le manuscrit (barres, ponctuation, etc.), soit par la solution indiquée entre parenthèses ou imprimée en italique ou encore soulignée.

Plus récemment, la Commission internationale de la diplomatique a émis des recommandations pour l'édition des actes médiévaux (CID & CIS 1984 ; cf. Guyotjeannin 1993 ; 84 Guyotjeannin coord. 2001). Celles-ci proposent de respecter dans une certaine mesure la mise en page et la mise en relief des fragments de texte original (usage du gras pour les caractères allongés et des petites capitales pour les majuscules), mais adoptent par ailleurs tous les traitements de normalisation des graphies et de la ponctuation pratiqués dans les éditions critiques.

Ces dernières années, sous l'influence de la « nouvelle philologie » et dans l'esprit du respect du témoin manuscrit, quelques éditions « ultra-diplomatiques » (ou « imitatives ») ont vu le jour. Des articles de linguistes s'intéressant à l'histoire des systèmes graphiques et des éditions de notices paléographiques tentent parfois de représenter encore plus finement les données primaires. Il n'existe donc pas à ce jour d'un standard d'édition diplomatique généralement accepté.

Pour comprendre quelles sont les pratiques réelles des éditeurs en matière de transcriptions diplomatiques, nous avons jugé utile d'analyser quelques cas précis.

Dans l'introduction, nous avons cité un exemple d'une transcription diplomatique dans le manuel de *Paleographie du Moyen Age* de Jacques Stiennon (1991) qui est en somme conforme aux recommandations de la Commission internationale de diplomatique. D'autres éditeurs préfèrent néanmoins rester plus proches du système graphique de leurs sources.

Ce sont les textes de grande importance historique et linguistique (et en particulier ceux qui nous sont parvenus dans un seul manuscrit) qui sont le plus susceptibles d'être édités de façon diplomatique rapprochée de la forme graphique

Représentation des données primaires : transcription de manuscrits médiévaux

originale. Tel est le cas, en premier lieu, des deux plus anciens textes en langue française, les *Serments de Strasbourg* et la *Cantilène de Sainte Eulalie*.

Les *Serments de Strasbourg* sont sans doute le texte, dont les éditions tentent le plus de se rapprocher du système graphique de la source manuscrite. Le degré de ce « rapprochement » est toutefois inégal selon les éditions. Plusieurs éditeurs accompagnent la transcription d'une image du manuscrit (mais dont la basse qualité ne permet pas le plus souvent de vérifier l'exactitude de la transcription)¹⁵⁵.

Nous reproduisons ci-dessous une photographie du début¹⁵⁶ du texte des *Serments de Strasbourg* du manuscrit Paris, BnF, lat. 9768 (*circa* 1000), ainsi que des facsimilés de transcriptions issues de plusieurs éditions¹⁵⁷.

155 Des images du fragment de la page du manuscrit Paris, BnF, lat. 9768 contenant le texte des *Serments de Strasbourg* sont accessibles sur plusieurs sites Internet. La qualité de toutes ces images est malheureusement médiocre. Il est étonnant que la Bibliothèque nationale de France, qui joue par ailleurs un rôle actif dans le développement des bibliothèques numériques, ne souhaite pas exposer en ligne une image de bonne qualité du premier texte français. La seule solution proposée par la BnF est d'acheter la reproduction à la « Banque d'images ».

156 Il s'agit du serment prononcé par Louis le Germanique. Notre transcription (*cf.* Annexe IV.6.1) inclut également le serment prononcé par les soldats de Charles le Chauve, autrement dit l'intégralité du texte en ancien français.

157 Voici la traduction ce texte donnée par M. Gally et C. Marchello-Nizia (1985) avec deux légères modifications suggérées *a posteriori* par C. Marchello-Nizia et signalées par l'italique : « Pour l'amour de Dieu et pour le salut du peuple chrétien et notre salut commun, à partir d'aujourd'hui, autant que Dieu m'en donnera savoir et pouvoir, je secourrai ce mien frère Charles en toute chose, comme *tout homme* doit légitimement secourir son frère, à condition qu'il en fasse autant pour moi. Et je ne conclurai, *selon ma volonté*, aucun pacte avec Lothaire qui pourrait nuire à mon frère Charles ».

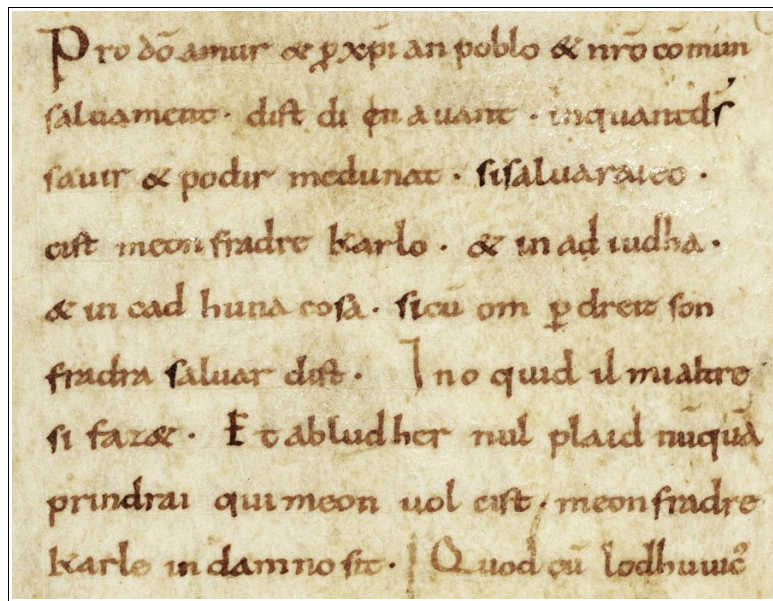


Fig. 10 : Facsimilé des Serments de Strasbourg
 (ms. Paris, BnF lat. 9768, 13r; cliché BnF)

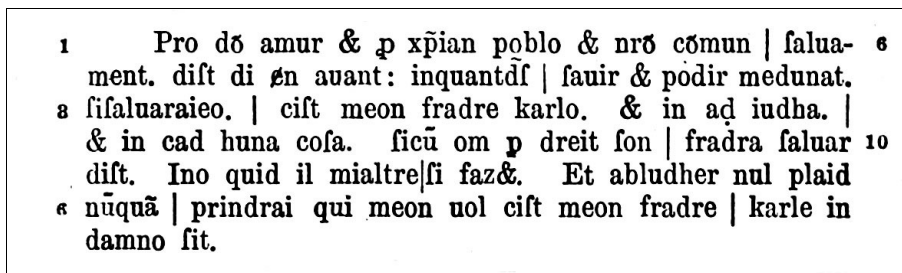


Fig. 11 : Transcription des Serments de Strasbourg (Koschwitz 1920, p. 2)

E. Koschwitz (1920) (Fig. 11) reproduit dans sa transcription les marques d'abréviation < & >, < p >, < p̄ >, etc., certaines segmentations particulières, la variante du caractère *s* « long » < f >. Il reproduit également le caractère « e barré » (ligne 2), sans doute une correction scribale *e* → *i*, mais la barre verticale du manuscrit devient oblique dans la transcription < e/ >. Est également reproduit un point souscrit au-dessous de *d* dans *ad iudha* (ligne 4), marque habituellement utilisée pour supprimer une lettre erronée. Le *v* n'est pas utilisé dans la transcription pour noter le *u* « consonantique » (*auant, saluar*)¹⁵⁸.

158 Sur ce point, la majorité des éditeurs des *Serments de Strasbourg* s'accordent. Pourtant, d'après les recherches en phonétique historique, la consonne labio-dentale [v] existait bien dès le III^e siècle (Zink 1986, p. 62-63).

Représentation des données primaires : transcription de manuscrits médiévaux

Koschwitz ne reproduit pas tels quels les sauts de ligne du manuscrit, mais les marque à l'aide d'une barre verticale. La ponctuation scribale est représentée par des marques modernes : le point en bas de ligne placé immédiatement après la dernière lettre du mot précédent, tandis que dans le manuscrit le point se situe à mi-hauteur de la ligne avec des espaces blancs plus ou moins grands des deux côtés. Il est intéressant de noter qu'à la deuxième ligne de la transcription, Koschwitz utilise un deux-points entre les mots *auant* et *inquant*. Dans la reproduction photographique du manuscrit dont nous disposons, la marque visible à cet endroit est un point « médian » tout à fait semblable aux autres. En revanche, dans le facsimilé fourni à la fin de l'édition de Koschwitz, on voit une marque semblable à un deux-points, qui peut être interprétée comme un *punctus elevatus*. Il s'agit sans doute d'un défaut matériel du facsimilé utilisé par Koschwitz (à moins que le deuxième élément de la marque de ponctuation n'ait été effacé du manuscrit entre le moment de la production du facsimilé de Koschwitz et la prise de la photographie que nous avons). Cela montre bien à quel point il est important de disposer d'une reproduction de bonne qualité pour les manuscrits dont on étudie la ponctuation.

Certaines données graphiques originales ne sont pas représentées dans la transcription. On peut noter parmi ces données la position exacte de la barre horizontale (ou tilde) d'abréviation dans *n̄ro* et *nūq̄ua* : dans le manuscrit elle est située au-dessus du milieu des groupes de lettres *ro* et *ua* respectivement, tandis que dans la transcription les tildes sont placés au-dessus d'une seule lettre à chaque fois. Une autre information perdue dans la transcription est la taille de l'espace blanc entre les mots. Le ms. 9768 semble en effet faire une distinction entre les blancs « grands » et « petits ». Par exemple, à la ligne 8, on trouve des blancs petits dans les séquences *qui meon* et *meon fradre*, tandis que les autres blancs sur cette ligne sont de taille nettement plus grande. On trouve par ailleurs des petits blancs (que nous représentons par des traits d'union) à l'intérieur de ce qu'on considère comme des mots unis : *xp̄i-an* (l. 1), *a-uant* (l. 2).

Le troisième caractère du mot *dift/dift* à la ligne 6 est ambigu : rien dans sa forme ne permet de conclure s'il s'agit d'un < f > ou d'un < f > (cf. la discussion que Koschwitz présente dans une note). Koschwitz utilise un *s* « long », tandis que les autres éditeurs mettent un < f > (en s'appuyant avant tout sur l'étymologie < lat. *debet*).

Il convient de noter que tous ces détails très fins (d'intérêt essentiellement paléographique) sont extrêmement difficiles à représenter à l'aide des techniques typographiques traditionnelles. Une perte d'informations et un ajout d'une « couche d'interprétation » sont inévitables dans toute transcription : notre but ici n'est pas de critiquer telle ou telle transcription, mais d'en montrer les limites.

Une nouvelle édition de la version allemande du manuel de Koschwitz et de Foerster a été préparée en 1932 par Alfons Hilka. La transcription des *Serments de Strasbourg* y est quasiment identique à celle que nous venons examiner, à une exception près : Hilka a introduit la distinction *u/v* voyelle/consonne. Si, du point de vue de phonétique historique, ce choix paraît explicable, il est néanmoins étonnant compte-tenu du caractère « ultra-diplomatique » du reste de la transcription.

En 1926, Philippe Lauer publie une édition de la *Chronique* de Nithard, dans laquelle apparaissent les *Serments de Strasbourg* (Fig. 12).

« Pro Deo amur et pro christian poblo et nostro com-
 « mun salvament, d'ist di in^b avant, in quant Deus savir
 « et podir² me dunat, si salvarai^c eo cist meon fradre
 « Karlo et in aiudha^d et³ in cadhuna cosa, si cum om per
 « dreit son fradra salvar dift^{e4}, in o quid il mi altresi
 « fazet et ab Ludher nul plaïd nunquam prindrai, qui,
 « meon vol, cist meon fradre Karle in damno sit. »

Fig. 12 : Transcription des *Serments de Strasbourg* (Lauer 1926, p. 104)

La transcription de Lauer est entièrement normalisée : la ponctuation et l'usage des majuscules sont modernisés, la segmentation est régularisée, les abréviations sont résolues. L'éditeur introduit même la distinction *u/v* (voyelle/consonne), contrairement aux autres éditeurs de ce plus ancien texte français. Quelques remarques sur les corrections sribales visibles dans le manuscrit et sur l'ambiguïté de la ligature < ft > sont données dans les notes.

Une chrestomathie de *Textes d'études (Ancien et moyen français)* a été publiée en 1949 par Robert-Léon Wagner (Fig. 13), ayant sans doute pour but de remplacer le manuel de Koschwitz. Wagner propose une transcription beaucoup plus « normalisée » des *Serments de Strasbourg* que celle de Koschwitz : il résout les abréviations (sans

Représentation des données primaires : transcription de manuscrits médiévaux

mettre en évidence les lettres restituées), introduit la ponctuation et la segmentation des mots moderne, met des majuscules aux noms propres et au mot *Deus* 'Dieu'. En revanche il respecte les sauts de lignes de l'original et n'introduit pas le caractère *v*.

Pro Deo amur et pro christian poblo et nostro commun
saluament, d'ist di in auant, in quant Deus
saur et podir me dunat, si saluarai eo
cist meon fradre Karlo et in adiudha
et in cadhuna cosa, si cum om per dreit son
fradra saluar dift, in o quid il mi altre
si fazet. Et ab Ludher nul plaid nunquam
prindrai qui meon uol cist meon fradre
Karle in damno sit.

Fig. 13 : Transcription des Serments de Strasbourg (Wagner 1949, p. 5)

Une autre chrestomathie de la littérature française médiévale a été publié par Albert Henry (1953).

Pro dō amur & p̄xpian poblo & nrō cōmun
saluament. dist di en auant. inquantd's
saur & podir medunat . sisaluarai eo.
cist meon fradre karlo. & in ad iudha.
& in cad huna cosa. sicū om p̄ dreit son
fradra saluar dift. Ino quid il mialtre
si faz&. Et abludher nul plaid nūquā
prindrai qui meon uol cist meon fradre
karle in damno sit. . . .

Fig. 14 : Transcription des Serments de Strasbourg (Henry 1953, p. 1)

Dans l'introduction à la transcription des *Serments de Strasbourg* (Fig. 14), Henry cite Kochwitz parmi ses sources et suit effectivement de très près la transcription de ce dernier. Il remplace toutefois les *s* « longs » par les *s* « courts »¹⁵⁹ et utilise apparemment le même caractère *p* « barré » pour le *p* « barré droit » (l. 6, abréviation de *per*) et le *p* « barré courbe » (l. 1, abréviation de *pro*) distincts dans le manuscrit et chez Kochwitz. En ce qui concerne la ponctuation, Henry utilise le « point bas », mais tente, semble-t-il, de reproduire l'espacement du manuscrit : il met un blanc avant et après le

¹⁵⁹ Il convient de noter que le manuscrit en question utilise uniquement les *s* longs. Il n'y a donc pas d'opposition des variantes de *s*, et le fait d'utiliser le *s* court dans la transcription n'amène pas à une perte de données.

point bas là où, dans le manuscrit, la distance est à peu près égale entre le point médian et les caractères qui le précèdent et le suivent (lignes 3 et 4, par exemple). Au contraire, il ne met pas de blanc avant le point si le point dans le manuscrit est placé un peu plus près du caractère précédent (lignes 6 et 7). Cependant, à la ligne 2, Henry espace ses points d'une façon strictement inverse de l'usage du manuscrit, ce qui rend sa pratique peu compréhensible.

Siegfried Becker publie en 1972 un livre entièrement consacré aux recherches sur les *Serments de Strasbourg*. Confronté sans doute à des contraintes matérielles de son imprimeur, il fournit un facsimilé de la transcription de Foerster/Koschwitz revue par Hilka (1932) (avec le *v* normalisé !) et ajoute sa propre transcription, entièrement normalisée, à l'exception de l'opposition *u/v*.

En 1995, Olivier Collet entreprend une réédition de la chrestomathie de Wagner¹⁶⁰, où il modifie très légèrement la transcription des *Serments de Strasbourg* « normalisée » de ce dernier : il supprime la lettre *d* in *ađiudha* (car le point souscrit laisse supposer une correction scribale), graphie « en un mot » *altresi* (*si* étant séparé par un saut de ligne dans le manuscrit et chez Wagner) et introduit un point médian d'enclise¹⁶¹ dans *nunqua·m* (car il considère que le *m* final est un morphème séparé : afr. *nunqua·m* < lat. **ne unquam magis*).

Pro dō amur & px̄pian poblo & nrō cōmun
saluament · dist di ꝑn auant · inquantd's
saur & podir medunat · sisaluaraiēo ·
cist meon fradre karlo · & in ađ iudha ·
& in cad huna cosa · sicū om ꝑ dreit son
fradra saluar dift · I no quid il mialtre
si faz& · Et abludher nul plaid nūquā
prindrai qui meon uol cist meon fradre
karle in damno sit ·

Fig. 15 : Transcription des *Serments de Strasbourg* (Wagner & Collet 1995, p. 9)

160 Il convient de noter que la date du manuscrit indiquée dans cette édition (fin du IX^e siècle) est erronée. La datation du manuscrit lat. 9768 généralement admise aujourd'hui est la fin du X^e siècle.

161 Cette marque est couramment utilisée dans l'édition de textes occitans.

Représentation des données primaires : transcription de manuscrits médiévaux

En plus de la transcription normalisée, Collet fournit une transcription diplomatique (Fig. 15), très proche de celles de Koschwitz et d'Henry.

Comme Henry, il remplace les *s* « longs » par des *s* « courts ». Pour représenter la ponctuation scribale, Collet utilise le point médian entouré d'espaces blancs. La qualité physique de l'impression ne permet pas de voir clairement si l'éditeur avait l'intention de distinguer le *p* « barré droit » du *p* « barré courbe »¹⁶². A la ligne 6, Collet met un blanc entre *I* et *no* (interprété 'in o' dans la transcription normalisée par tous les éditeurs). Ce choix semble difficilement justifiable, car dans le manuscrit l'espace entre *I* et *n* est à peine plus grand qu'entre *n* et *o* et nettement plus petit que les blancs qui séparent les autres mots sur cette ligne. Il s'agit donc tout au plus d'un « petit blanc ». Par ailleurs, un espace similaire sépare *E* et *t* dans la conjonction 'et' à la ligne suivante, mais Collet n'introduit aucun blanc dans ce cas.

L'une des plus récentes éditions des *Serments de Strasbourg* a été réalisée par Günter Holtus (1998) au sein d'un article publié dans les actes du colloque sur la transition du latin aux langues romanes. Dans le même esprit que Collet, Holtus fournit deux versions du texte : « édition diplomatique » et « édition textuelle ». La version « diplomatique » est très proche de celle de Koschwitz (1920), jusqu'au maintien du *s* « long ». La ponctuation médiévale est représentée par des points médians. Les abréviations *p* barrés sont résolues entre parenthèses, sans doute en raison de contraintes matérielles de l'imprimeur. Les autres abréviations (perluètes et barres horizontales) sont imitées. La transcription est accompagnée de commentaires très détaillés.

L'édition « textuelle » de Holtus est en réalité moins normalisée que la plupart des éditions dites diplomatiques. La normalisation n'a touché que la segmentation et les abréviations, qui sont résolues tacitement. En revanche, la ponctuation originale est conservée (avec l'ajout d'un saut de ligne à chaque occurrence), la distinction *u/v* n'est pas introduite. Plus surprenant est le choix de Holtus de conserver le *s* « long », dont le remplacement par un *s* « court » n'aurait aucune incidence sur l'analyse linguistique.

Le bref aperçu des différentes éditions des *Serments de Strasbourg* que nous venons de tracer permet de constater que les différences de pratiques éditoriales

162 Nous utiliserons le terme de M. Prou (1924).

apparaissent clairement même sur les quelques lignes du plus ancien monument de la langue française.

Nous n'analyserons pas de façon aussi détaillée les transcriptions du deuxième des plus anciens textes français¹⁶³, la *Séquence de Sainte Eulalie*¹⁶⁴, mais noterons simplement que les règles de transcription appliquées à ce texte varient considérablement selon les éditions et peuvent par ailleurs être différentes de celles des *Serments de Strasbourg*, même dans le cas où ces deux textes sont publiés dans une même chrestomathie.

Un exemple d'édition diplomatique d'un texte de taille plus importante peut être tiré du *Mystère d'Adam* (1968). Ce drame religieux composé dans la deuxième moitié du XII^e siècle a été conservé dans un seul manuscrit (Tours, Bibliothèque municipale, 927, 2^e quart du XIII^e siècle). Il a été édité à plusieurs reprises, de 1854 à 1963, avec des émendations éditoriales plus ou moins nombreuses : l'absence d'autres témoins de la tradition manuscrite n'a pas empêché les éditeurs de proposer des remaniements visant à faciliter la « bonne compréhension » du texte.

Leif Sletsjöe s'est donné pour objectif de « présenter telle quelle la teneur de la copie unique conservée » (p. VII) et de donner au lecteur la possibilité de juger du bien-fondé des corrections effectuées par les différents éditeurs. Son édition donne, dans les pages de gauche, une reproduction photographique du manuscrit (en noir et blanc, d'assez bonne qualité) et, en regard, dans les pages de droite, une édition diplomatique du texte. En bas des pages se trouvent les corrections et émendations des huit éditions antérieures (cf. Fig. 16).

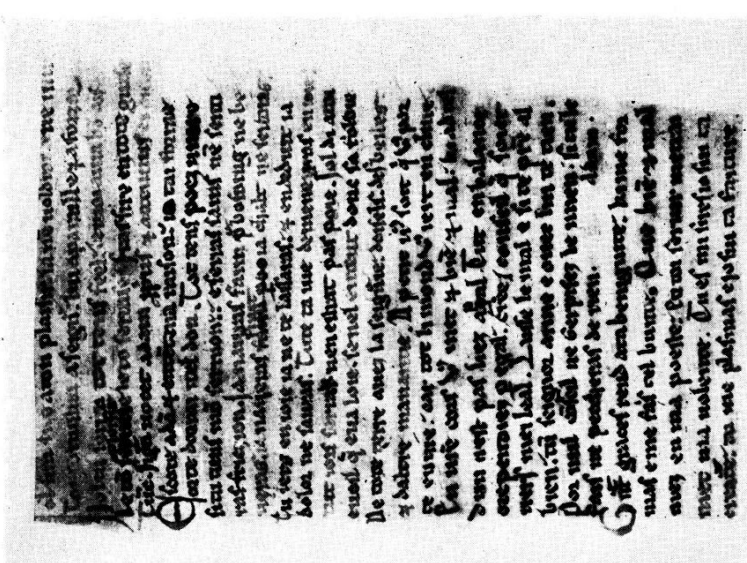
Bien qu'il affirme présenter une version diplomatique du texte, Sletsjöe n'a pas hésité à introduire certains changements, en prenant soin de les signaler dans l'introduction, car il considère une transcription « pure et simple » comme « superflue ».

Dans sa version du texte, Sletsjöe a choisi de résoudre les abréviations sans signaler les lettres restituées : à son avis, un tel signalement serait superflu en présence

163 Le manuscrit Valenciennes, BM, 150 qui contient la *Séquence* est par ailleurs plus ancien que celui des *Serments de Strasbourg*.

164 Le système graphique de la *Séquence de Sainte Eulalie* a été étudié en détail par Liselotte Biedermann-Pasques (2001). La ponctuation de ce texte a été analysée par S. Baddeley (2001)

F^o 21



7

VERS 40-77

40. E. Jol frai sire a ton plaisir
ia nen uoldrai ne issir
Toi comuatrai a seignor
lui a paraille e a forzor
Jo lui serrai tot tens feel
de moi aura bon conseil
le ton plaisir le ton seruisse
Frai sire en tote guise
(F) Escote adam e entent ma raison
io tai forme or te doerai itel don
Tot tens poez uire si tu tiens mon sermon
e serras sains nen sentiras frizion
Ja nauras fain por hosoing ne beueras
ia nauras frait ia chalt ne sentiras
tu iers en loie ia ne te lassaras
e en deduit ia dolor ne sauras
Tute la uie demeneras en ioie
tut iors serra nen estrat pas poie
Jol di a toi e uoil que eua loie
se nel entent donc safoloie
De tote terre auez la seignorie
doisels des bestes e daltre manantie
A petit uos soit qui uos porte enuie *
car tot li mond uis iert encline
En uostre cors uis met e bien e mal
ki ad tel dun nest pas liez a pal
Tut en balance ore penditez par egal
crees conseil que soiet uers mel leal
Laisse le mal e si te pren al bien
tun seignor aime e onec lui le tien
Por nul conseil ne gerpisez le mien
si tu le fais ne pecheras de rien
ki me formes e me fais tel bunte
A. Grant graces rend a ta benignite
Que bien e mal mez en ma poeste
En toi seruir metrai met ma uolente *
Tu es mi sires io sui ta creature
tu me plasmas e jo sui ta faiture

45.

50.

55.

60.

65.

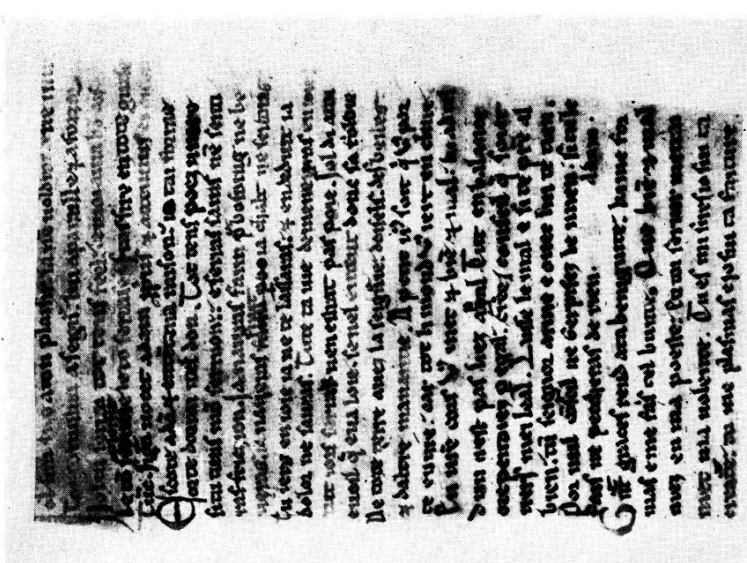
70.

75.

57. J. serras, n'en; P. Toz jors serras, nen estera pas poie; G 1 esterat;
G 2-3 serras; Sf. Tuz jors serras, ta vie n'iert pas poie
58. P. el; Sf. tel, ... que Eve l'ore
59. P. Seie nel entent, donkes safoloie; G 1 done ele; Sf. donc a folor
61. P. apote
62. P. daltre; G 2-3 de(s); Sf. de; Aeb. D'oiseiz
63. P. seit; G 1 Petit uis soit, qui uos en; Sf. A petit uis seie
64. P. a uis; G 1 Car tote terre uis iert tot tens encline; Sf. iert en vostre
65. P. el; Sf. En vostres chois
66. P. or; G 1 or pendez; G 2-3 ore pend(0)ez; Sf. or pendez; Aeb. Tot ...
67. P. paraille; G 1 conseil, seioez; G 2-3 conseil, que soiet; Sf. conseil, seiet;
Aeb. conseil; omis; Tous; Crez
68. G 1-2-3 rend; Sf. rend; 69. P. el avec; Sf. e si od lui
70. Sf. guerpisez; 72. P. grace rend; Sf. Granz
73 et 74. P. et - 75. P. et Sf. tel; Tous ont omis; met - 77. P. et

6

F^o 21



6

VERS 40-77

40. Jo l'era; G 1 et Sf. fera; G 2-3 f(e)rai
41. L. G 1, Ch et Aeb n'en; G 2-3 ne(n); Sf. ne; Tous: de rien issir
42. P. Tel ... a mon; G 1 Toi reconstrai; Sf. Tel ... a [mon] seignor
43. G 1 parail; G 2-3 parail(0); Sf. parail e a forzor
44. P. fedell; Sf. feel
45. P. De mei aura null; G 1 mult bon; Sf. De mei avra [null]
46. Sf. plaisir
47. P. Fera; sires; G 1 Fera; G 2-3 F(e)rai; Sf. Fera; siré
48. P. durrat tel; G 1 et Sf. dorrai tel; G 2-3 dorrai (0)tel; Ch et Aeb
dorrai itel
49. P. bez vivre, se l'iens
50. P. El ... ne sentiras frizon; G 1 frizon; G 2-3 ne(n); Sf. ne sentiras
frizon; G 1 sijn
51. P. boiz; G 1 et Sf. boiras; G 2-3 boiz(e)mas
52. P. et Sf. boiras frait; G 1 n'avras; G 2-3 n'av(e)mas
53. P. et Sf. boiras frait; G 1 n'avras; G 2-3 n'av(e)mas
54. P. ers; G 1-2-3 et Sf. lasseras - 55. P. El - 56. P. tot

Fig. 16 : Le mystère d'Adam (Sljetsjöe 1968, p. 6-7)

de la photocopie. Il a supprimé les lettres biffées ou exponctuées et a admis les corrections en forme d'additions visibles dans le manuscrit, même s'il n'excluait pas la possibilité qu'il s'agisse de corrections faites par d'autres mains que celle du scribe.

Par ailleurs, Sletsjøe a séparé les mots agglutinés dans les « cas évidents », mais s'en est tenu au texte du manuscrit là où « la division est rendue compliquée par les incertitudes du scribe » (p. IX), son traitement des segmentations particulières n'est donc pas complètement homogène.

En revanche, l'éditeur norvégien n'a pas introduit de ponctuation moderne ni de signes diacritiques habituellement utilisés dans les éditions de textes de français médiéval. Il est également resté fidèle au manuscrit dans la transcription des lettres *i*, *j*, *u* et *v*.

En 2001, voit le jour l'*Album de manuscrits français du XIII^e siècle* préparé par une équipe de philologues et de paléographes de l'IRHT¹⁶⁵ (Careri *et al.* 2001). L'idée à l'origine du projet ayant abouti à la publication de cet album consistait à confronter un nombre important de notices très riches de manuscrits afin de « faire apparaître, dans la présentation des textes, des paramètres, constantes et variables, qui pourraient être corrélés avec ces autres variables que sont, par exemple, la date et le lieu de copie, le format du volume ou son contenu, qu'il s'agisse du nombre ou de la typologie des textes » (*ibidem*, p. xi). Au cours du travail, les chercheurs ont été amenés à revoir leurs objectifs à la baisse : limiter à une cinquantaine le nombre de notices rédigées, accentuer la cohérence spatio-temporelle du corpus et renoncer aux théorèmes ambitieux. L'*Album* finalement publié avait pour but de « poser des prémisses fiables, aider le lecteur moderne à voir une page manuscrite, à la percevoir telle qu'en elle-même [...], fournir des points de comparaison à l'éditeur de textes souvent embarrassé pour dater et caractériser l'écriture des copies » (*ibidem*). Il s'agit donc d'une édition « technique » à visée paléographique.

Les notices, qui constituent la partie principale de l'*Album* (à côté d'une importante introduction), incluent une photographie en noir et blanc d'une page du manuscrit¹⁶⁶, sa description physique et paléographique détaillée, une transcription

165 Institut de Recherche et d'Histoire des Textes.

166 16 des 53 photographies de manuscrits sont reproduites en couleur, à la fin du volume.

Représentation des données primaires : transcription de manuscrits médiévaux

diplomatique et une « édition » d'un fragment de la page photographiée. L'explication des principes de la transcription diplomatique semble un peu laconique pour une édition de ce type :

La transcription de la page quatre conserve de l'original les agglutinations, la ponctuation et les abréviations légèrement stylisées. Elle s'efforce de respecter le jeu des formes et des modules des lettres. Les lettres de forme majuscule et les grandes onciales sont rendues par des capitales, les lettres de type minuscule et les petites onciales sont rendues par des minuscules, les lettres de forme majuscule de petit format sont rendues par de petites capitales et les minuscule de grand format par des minuscules d'un corps supérieur. Enfin, la répartition *u/v*, *i/j* est maintenue telle quelle » (*ibidem*, p. xiv).

L'analyse des transcriptions elles-mêmes montre que celles-ci sont effectivement très précises dans la représentation des marques d'abréviations, des signes de ponctuation médiévale et dans la distinction des initiales / majuscules / petites majuscules / grandes minuscules / minuscules. Par contre, l'absence de distinction entre les grands et les petits blancs diminue la fiabilité des données sur la segmentation¹⁶⁷. Rien n'est dit sur les variantes de caractères (*s* « long », *r* « rond », etc.), ni sur les ligatures, ni sur les séries de jambages. Les caractères modernes sont utilisés dans tous ces cas. Néanmoins, les transcriptions diplomatiques de l'*Album* sont à ce jour parmi les plus fines qu'on puisse trouver dans des éditions imprimées.

Certains linguistes s'intéressant à l'histoire des systèmes graphiques du français poussent plus loin la finesse des transcriptions de manuscrits qu'ils citent dans leurs articles. Par exemple, Liselotte Biedermann-Pasques (1998 ; 2001) utilise, en plus des marques d'abréviation médiévales, le *s* « long ». Nicolas Mazziotta (2007a) ajoute à cela une distinction fine des marques de ponctuation médiévales, ainsi que la distinction < *ı* >/< *ı̇* >. Elena Llamas Pombo (2001 ; 2007) emploie, en plus de tout cela, le caractère < *ÿ* >.

Des transcriptions fines de l'écriture manuscrite se trouvent dans plusieurs éditions réalisées ces dernières années par certains philologues allemands. Nous avons déjà mentionné l'édition des *Serments de Strasbourg* par Günter Holtus (1998). Un

¹⁶⁷ Les auteurs sont bien conscients du problème de la segmentation et lui consacrent un passage très intéressant dans l'Introduction (Careri *et al.*, p. xxxvi-xxxvii).

autre exemple de ce type d'édition (2003) se caractérise par le même degré de fidélité à la source primaire (usage de *s* « long », opposition < i >/< í >¹⁶⁸, segmentations particulières du manuscrit, ponctuation médiévale et résolutions d'abréviations entre parenthèses, etc.). Les auteurs de cette dernière édition vont jusqu'à utiliser les graphies « ultra-diplomatiques » dans le glossaire, ce qui a provoqué des critiques à juste titre (Möhren 2005).

Parfois, le souhait des chercheurs d'utiliser des transcriptions fines dans leurs articles publiés dans des revues ou dans des ouvrages collectifs se heurte tout simplement à des contraintes matérielles : l'absence de polices adéquates, des problèmes de conversion entre les différents formats informatiques utilisés pendant la rédaction, la relecture et la finition typographique. Ainsi, Nelly Andrieux-Reix et Simone Monsonégo (1997) ont été amenées à utiliser le *S* majuscule pour le *s* « long » et le symbole < @ > pour l'abréviation < q̄ > (= *que*). Le besoin d'un système de codage standardisé pour ces transcriptions fines mis à disposition de la communauté des chercheurs devient donc de plus en plus évident.

L'analyse de différentes transcriptions diplomatiques que nous venons d'effectuer montre que les pratiques varient considérablement selon les éditeurs, la nature des textes édités et les objectifs de l'édition. Si le but de toute édition diplomatique est de représenter fidèlement le manuscrit choisi, cette représentation peut prendre des formes bien différentes. D'une part, l'éditeur peut chercher à imiter dans la mesure du possible les données graphiques de son manuscrit (reproduire les marques d'abréviation, les variantes de caractères, la mise en page, etc.) et à se rapprocher ainsi d'un facsimilé photographique. Il peut, au contraire, analyser et interpréter les données graphiques pour les présenter ensuite dans une forme pas forcément identique à celle du manuscrit (par exemple, Koschwitz utilise les barres verticales pour les fins de lignes ; Collet utilise le *s* « court » dans sa transcription). Il peut également utiliser des marques diacritiques de désambiguïsation (l'accent aigu sur le *e*, le tréma, la cédille) à condition qu'elles ne masquent pas les données originales¹⁶⁹.

168 Le < i > est toutefois représenté par < í >.

169 La distinction des *u/v* et *i/j* (voyelle/consonne) peut en revanche masquer les données originales, car ces paires de caractères peuvent être utilisées comme des variantes dans un manuscrit.

Représentation des données primaires : transcription de manuscrits médiévaux

D'autre part, l'éditeur peut choisir de formater sa transcription d'une façon plus ou moins conforme aux règles typographiques modernes. Le « formatage » de la transcription peut se traduire par l'introduction de la ponctuation moderne, la suppression des variantes de caractères (et en même temps la désambiguïsation des lettres ambiguës), la résolution des abréviations, etc. Ce formatage facilite considérablement la lecture de la transcription, mais amène à la perte d'une partie des données primaires.

Strictement parlant, ces deux axes de caractérisation des éditions (« imitation vs interprétation » et « respect des données primaires vs formatage typographique ») ne sont pas identiques. Il est néanmoins possible d'établir une corrélation entre la tendance à l'imitation et la volonté de respecter des données primaires d'une part, et, d'autre part, entre la tendance à l'interprétation et la pratique de la normalisation typographique. On peut alors comparer les différentes éditions (transcriptions) par leur position sur un axe général de normalisation (Fig. 17).

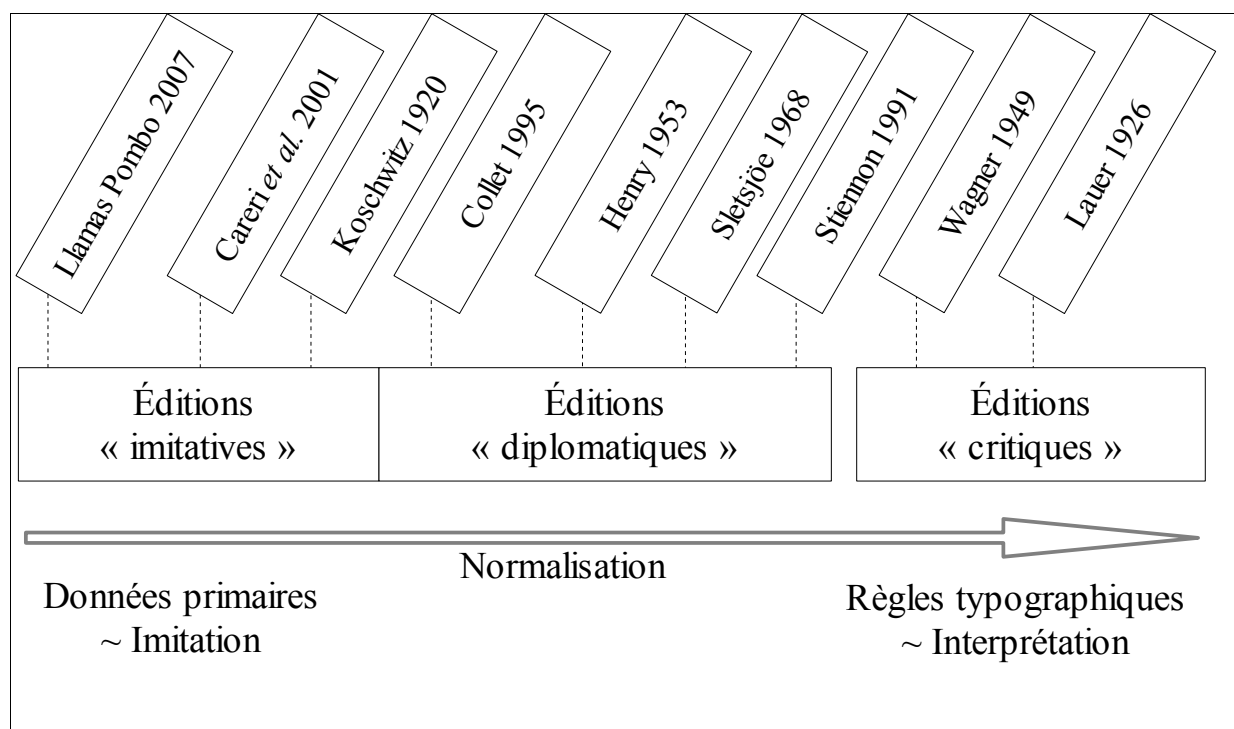


Fig. 17 : Différents types d'éditions sur un axe de normalisation

Le terme de « diplomatique » au sens strict peut être réservé aux transcriptions (et aux éditions les utilisant) dans lesquelles les graphies médiévales sont représentées

par les caractères modernes, sans introduction de distinctions ou de diacritiques désambiguïsantes (par exemple, *u/v* ou accent aigu sur un *e* final accentué).

Les transcriptions (et les éditions) utilisant des caractères ou toute autre sorte de marques supplémentaires pour noter les traits particuliers des systèmes graphiques médiévaux peuvent être qualifiées d'imitatives. Quel est l'intérêt de ce dernier type de transcriptions ?

4.3. Éditions comportant des transcriptions imitatives

Les détracteurs des éditions « imitatives » évoquent le plus souvent les arguments suivants (*cf.* Viellard & Guyotjeannin coord. 2001, p. 14) :

1. La reproduction des variantes des lettres, des marques d'abréviation, des segmentations particulières et de la ponctuation scribale rendent plus difficile la lecture du texte ;

2. Le refus de prendre parti vis-à-vis des ambiguïtés et des incohérences du manuscrit ne relève que de la paresse d'un éditeur qui ne veut pas faire le travail d'analyse philologique ;

3. L'imitation des marques sribales dans une édition n'est jamais qu'une illusion de fidélité au manuscrit : il est impossible de reproduire typographiquement toutes les nuances de l'écriture manuscrite ;

4. Le fait d'accompagner le texte critique d'une édition des images (facsimilé) de manuscrits rend inutile la transcription imitative ;

5. La transcription fine des marques sribales est beaucoup plus longue et coûteuse que la transcription en caractères modernisés.

Les deux premiers arguments sont valables uniquement pour les éditions traditionnelles sur papier. Dans une édition hypertextuelle, la conservation des données primaires **n'empêche ni une lecture simplifiée** (et même assistée) **ni la normalisation éditoriale** et l'introduction de l'apparat critique (plus d'excuse pour un éditeur

paresseux !) : grâce à la possibilité d'afficher tel ou tel type de traits seulement, le lecteur peut visualiser le texte de la façon la mieux adaptée à ses intérêts¹⁷⁰.

Le troisième argument (**impossibilité d'une fidélité absolue**) est plus important. Toute transcription, quelque fidèle ou fine qu'elle soit, est une interprétation. Mais est-ce vraiment une raison pour renoncer à toute tentative de représenter de façon adéquate les données primaires dans un format permettant leur analyse et traitement automatisé ?

Le même problème se pose pour les transcriptions de l'oral, car la finesse de la distinction des nuances sonores et articulatoires peut être extrêmement variable :

« Une notation fine est nécessaire, mais il ne faut pas s'imaginer qu'une transcription, même bien différenciée et bien adaptée, permet de décrire une prononciation dans toutes ses subtilités. Il est impossible de représenter avec exactitude la parole à l'aide des signes, quels qu'ils soient. » (Carton 1974, p. 10)

Le besoin de distinguer les traits pertinents dans la caractérisation (et donc transcription) des sons s'est fait sentir très tôt en phonétique : l'idée du phonème comme unité distinctive pertinente psychologiquement remonte aux travaux de l'École de Kazan (J. Baudouin de Courtenay et N. Kruszewski) de la fin du XIX^e siècle et a été développée pleinement par l'École de Prague (Morpurgo Davies 1998, p. 303). Une analyse phonématique implicite est par ailleurs sous-jacente à tout système d'écriture alphabétique. De nos jours, les termes de *phonème* (unité caractérisée par un ensemble de traits distinctifs), *allophone* (variante identifiable d'un phonème) et *phone* (réalisation concrète d'un phonème) sont généralement acceptés par la communauté des linguistes (cf. Dubois *et al.* 1999). Des conventions ont été élaborées permettant de distinguer les transcriptions phonématiques des transcriptions phonétiques plus ou moins fines¹⁷¹. Les transcriptions de l'oral qui ne sont pas destinées à des recherches phonétiques sont souvent effectuées en mode « orthographique » (cf. Baud *et al.* 2006, p. 29-31).

170 Les avantages des éditions électroniques sont par ailleurs bien reconnus par les défenseurs de la philologie traditionnelle (Busby 1993, p. 41-42 ; Viellard & Guyotjeannin coord. 2001, p. 14).

171 L'Association Phonétique Internationale (API) dans ses recommandations parle des transcriptions « larges » (*broad*) et « fines » (*narrow*), ces dernières pouvant être « légèrement fines » (*slightly narrow*) ou « très fines » (*very narrow*) (IPA 1999, p. 28-30).

Dans le domaine de la transcription des manuscrits médiévaux, on peut parler de transcriptions normalisées (selon les règles établies pour les éditions critiques, *cf.* transcriptions « orthographiques » de l'oral) et de transcriptions diplomatiques plus ou moins fines. Des transcriptions diplomatiques « moins fines » peuvent être comparées aux transcriptions phonématiques : on reproduit les graphèmes du manuscrit tout en neutralisant les variantes positionnelles ou purement calligraphiques. Ce niveau de représentation devrait être suffisant pour des études sur l'histoire de l'orthographe. Un niveau de finesse supérieur est nécessaire pour les recherches sur le fonctionnement des systèmes graphiques des manuscrits : il faut dans ce cas représenter toutes les variantes des caractères et des autres marques graphiques (comme par exemple les espaces de taille variable entre les mots) qui auraient potentiellement une charge fonctionnelle. Nous proposons d'utiliser le terme d'*allographe* pour nommer ces variantes potentiellement fonctionnelles (*cf.* *allophones* en phonétique)¹⁷². Une transcription « allographique » d'un texte manuscrit est donc justifiable du point de vue linguistique. Si la « fidélité absolue » d'une telle transcription à la source ne peut être qu'une illusion, elle peut néanmoins représenter d'une manière adéquate les données linguistiquement pertinentes, du moins dans l'état de nos connaissances, à condition que la transcription soit fondée sur des principes clairs et raisonnés.

Une représentation encore plus fine, allant jusqu'à la prise en compte des traits individuels de l'écriture des scribes, pourrait être nécessaire pour des recherches paléographiques ou graphologiques.

Le fait que toute transcription comporte une analyse et une interprétation des données est d'ailleurs la raison pour laquelle le **facsimilé** du manuscrit « en mode image » **ne peut pas la remplacer** (*cf.* Petrucci 1977). Là encore, on peut établir un parallèle avec la linguistique de l'oral, où les enregistrements sonores ne remplacent pas les transcriptions.

172 Dans (Dubois *et al.* 1999), l'*allographe* est défini comme une « représentation concrète du phonème ». Vu la définition d'allophone donnée dans ce même dictionnaire (variante, surtout combinatoire, d'un phonème), il nous semble préférable de définir l'*allographe* comme une variante identifiable d'un graphème et d'utiliser le terme *graphie* pour une réalisation concrète. Bien entendu, la distinction des *allographes* et des *graphes* ne fait sens que dans l'analyse de l'écriture manuscrite.

Bien entendu, la transcription ne peut à son tour remplacer une reproduction photographique du manuscrit. Il est très important pour un chercheur travaillant sur une transcription de pouvoir à tout moment vérifier son exactitude en consultant un facsimilé du manuscrit¹⁷³. Certaines éditions diplomatiques sur papier comportent en effet des images du manuscrit à côté du texte transcrit, mais la qualité de ces images est souvent loin d'être optimale. De plus, il n'est pas toujours très facile de mettre côte à côte un passage de transcription et la zone correspondante de l'image. Dans une édition électronique, l'image et la transcription peuvent en principe être alignées à une ligne près (voire à un mot près), même si dans la pratique l'alignement se limite assez souvent au niveau de pages ou de colonnes. La qualité des images numériques qu'on peut réaliser depuis quelques années déjà avec un matériel non excessivement cher est telle qu'on y voit bien les détails à peine perceptibles à l'œil nu.

Le problème **du cout des éditions** comportant des transcriptions fines n'est pas négligeable. Ce sont d'ailleurs ces contraintes matérielles qui nous ont obligé à restreindre considérablement la taille de notre corpus. Il convient de préciser toutefois que les transcriptions très fines sont nécessaires avant tout pour des recherches sur certaines propriétés des systèmes graphiques, pour lesquelles des échantillons de taille limitée à quelques pages (soit quelques milliers de mots) peuvent suffire. En revanche, les transcriptions diplomatiques (graphématiques) et normalisées sont pour les linguistes des sources de données plus fiables que les textes critiques, qui sont à leur tour plus intéressantes pour les littéraires et aux historiens, ainsi pour des lecteurs « non professionnels ». Des éditions qui permettent aux lecteurs d'accéder au texte critique aussi bien qu'aux transcriptions diplomatiques ne sont pas forcément plus difficiles à réaliser que des éditions critiques traditionnelles, car l'éditeur scientifique transcrit de toute façon ses sources avant d'établir le texte critique. S'il existait des conventions adoptées par la communauté des médiévistes pour ce type de transcriptions et si les éditeurs disposaient d'un outil permettant d'effectuer ces transcription avec le maximum

173 Il convient de rester attentif aux déformations que l'image photographique peut contenir par rapport à l'original. La qualité photographique du facsimilé conditionne en grande partie son utilité pour la recherche. Avec les images numériques de très haute résolution il est possible de voir les détails imperceptibles à l'œil nu dans l'original (par exemple, en augmentant le contraste pour rendre visibles des traits de plume presque effacés). La copie peut donc devenir dans un certain sens une meilleure source que l'original.

de confort et de les sauvegarder dans un format leur assurant une certaine pérennité, le cout des éditions « multi-facettes » pourrait considérablement diminuer. Des transcriptions « allographiques » fines peuvent être effectuées pour une partie du texte seulement, à l'exception peut-être des textes de valeur artistique, historique ou littéraire exceptionnelle. Là aussi l'adoption de normes d'encodage et le développement des outils informatiques d'aide à la transcription peuvent réduire le cout du travail et d'assurer la réutilisabilité des données et la possibilité d'extension progressive de ces transcriptions fines. C'est un pas dans cette direction que nous nous proposons de franchir.

Il est possible, comme nous l'avons démontré, de distinguer au moins quatre types de transcriptions de manuscrits, dont chacun convient à certains types de recherches, ce qui nous amène à préciser le schéma positionnant les différentes pratiques de transcription sur un axe normalisation :

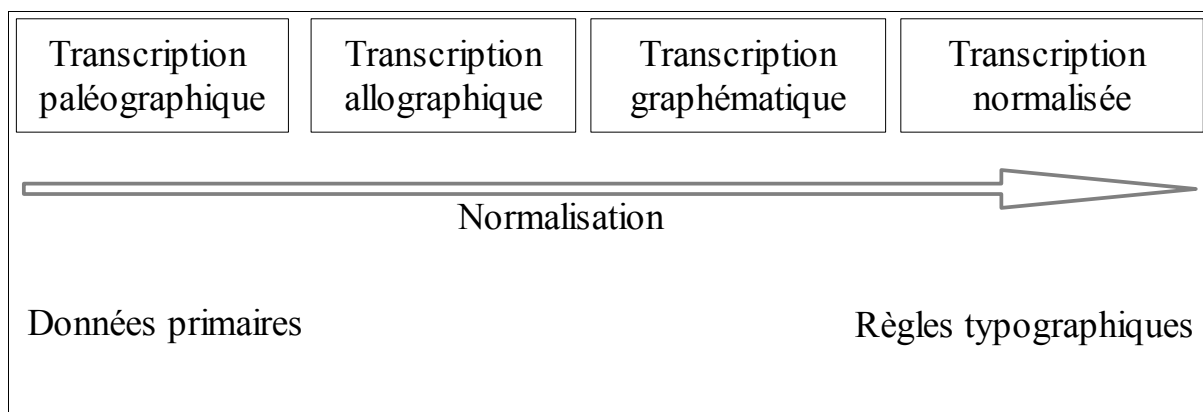


Fig. 18 : Types de transcriptions précisés

Notre recherche étant linguistique, nous nous limiterons dans notre corpus aux transcriptions de niveau allographique, tout en y ajoutant les « couches » graphématique et normalisée, ce qui doit permettre de créer des visualisations multiples correspondant aux différents types de lectures et de requêtes possibles sur notre corpus.

Le seul projet d'édition électronique ayant appliqué à ce jour la distinction systématique de trois niveaux de transcription (normalisé, diplomatique et « facsimilaire ») de manuscrits médiévaux est le *Medieval Nordic Text Archive*

(Menota)¹⁷⁴. Les protocoles d'encodage des textes et certains outils de travail de ce projet étant mis à disposition des chercheurs sur Internet, nous avons pu en profiter dans l'élaboration du format de notre corpus (cf. la section 5.3. ci-dessous). La « philosophie » d'encodage de transcriptions à plusieurs niveaux du projet Menota est exposée dans (Haugen 2004).

A la suite de Nicolas Mazziotta (2006), nous utiliserons le terme de *diffraction* pour caractériser la multiplication des « vues » et des représentations des données sous-jacentes.

4.4. Présentation des données « sensibles » dans les éditions « multi-facettes »

4.4.1. Typologie des données

Nous parlerons dans la présente section de la façon dont les données doivent apparaître au lecteur de l'édition électronique selon qu'il choisit une vue normalisée, diplomatique ou imitative.

Les caractéristiques de la vue normalisée s'appuient sur les règles suivies dans la plupart des éditions de textes médiévaux français en France (Roques 1926 ; Viellard & Guyotjeannin coord. 2001). Les caractéristiques de la vue diplomatique s'appuient sur les recommandations de F. Masai (1950) et sur les pratiques de certains éditeurs (comme, par exemple, Sletsjøe 1968), avec quelques modifications et précisions, notamment en matière de la ponctuation. Pour nous, le principe de la transcription diplomatique (= graphématique) est qu'elle doit utiliser les caractères modernes, en neutralisant les variantes calligraphiques, mais sans introduire de marques ou de distinctions désambiguïsantes. La vue imitative, enfin, s'inspire de certaines éditions « ultra-diplomatiques » sur papier et électroniques, mais il n'existe pas à ce jour de conventions stables dans ce domaine. Nous espérons que nos propositions pourront contribuer à leur élaboration.

174 <<http://www.menota.org>>.

Nous faisons bien la distinction entre la *représentation* des données et la façon dont ces données sont *présentées* au lecteur. Dans les éditions papier, seules les données typographiquement présentées sont représentées, tandis que dans l'informatique la représentation des données et la visualisation peuvent être bien dissociées. On peut par exemple encoder l'information sur la présence d'une correction scribale dans le manuscrit sans définir a priori comment elle sera visualisée : le lecteur pourra voir uniquement la forme corrigée ou les deux, avec ou sans l'usage de marques conventionnelles (crochets, barres obliques, etc.).

Puisque les traditions d'éditions existantes sont notre point de départ, nous aborderons dans la présente section l'aspect des transcriptions (autrement dit, la présentation des données) selon la perspective choisie. Nous n'oublierons pas cependant que c'est la représentation sous-jacente qui est la plus importante, c'est pourquoi nous parlerons également (et surtout) des niveaux de représentation des données qui se répercutent dans les différentes formes d'affichage. La section terminera par un exemple concret d'une transcription à vues multiples.

Une réflexion plus technique sur les possibilités d'encodage de ces données sera proposée dans le chapitre 5, et le protocole complet d'encodage de notre corpus est présenté dans l'annexe I.

La différenciation des données représentées et des formes de présentation selon la vue choisie concerne plusieurs types de données graphiques. Nous les examinerons tour à tour.

4.4.1.1. Abréviations

Une abréviation peut être définie comme une suppression volontaire d'un ou de plusieurs caractères dans une chaîne écrite alphabétique qui n'empêche pas, normalement, la lecture orale et la compréhension du texte. Le plus souvent l'abréviation est signalée par une *marque* graphique, mais ce n'est pas obligatoire. La marque de l'abréviation n'est pas par ailleurs forcément placée à l'endroit où les caractères sont omis.

Représentation des données primaires : transcription de manuscrits médiévaux

Les manuscrits médiévaux contiennent un nombre très important d'abréviations de types très variés. Selon Careri *et al.* (2001, p. xxx-xxx1) :

Lorsqu'ils furent amenés à noter les textes vernaculaires, les copistes adoptèrent naturellement le système graphique du latin auquel ils étaient habitués, c'est-à-dire simultanément un alphabet et un code d'abréviations bien rodé et efficace, dont les principes remontaient à l'Antiquité et les outils avaient été perfectionnés par les *scriptoria* monastiques, avant que la pratique n'en soit bouleversée, au XIII^e siècle, par les méthodes de la scolastique.

Même si les abréviations semblent de façon générale moins fréquentes et moins variées dans les textes français que dans les textes latins, elles jouent dans le système graphique du français médiéval un rôle tout à fait remarquable. La quasi-totalité des abréviations françaises ont été empruntées au latin, mais en français elles ne fonctionnent pas toujours exactement de la même manière qu'en latin (*cf.* Prou 1924, p. 110). Nous nous intéresserons en premier lieu aux abréviations présentes dans les manuscrits en français, mais les principes d'analyse proposés devrait être valables pour tous les manuscrits en écriture latine.

Dès les débuts de la diplomatique et de la paléographie, le problème de la résolution des abréviations occupe une place importante dans les traités et dans les manuels de ces disciplines. Au XVIII^e siècle déjà, des dictionnaires d'abréviations voient le jour (Walther 1752). Natalis de Wailly (1838), dans le chapitre consacré aux « différents systèmes d'abréviations », traite des :

- sigles ;
- notes tironiennes ;
- abréviations « proprement dites » (sans les classer davantage) ;
- lettres conjointes, monogrammatiques et enclavées.

Maurice Prou (1924) inclut toujours les notes tironiennes, les sigles et les monogrammes dans le chapitre des abréviations, mais propose un classement plus fin des abréviations « au sens propre ». Il distingue les abréviations :

- par suspension ;
- par contraction ;
- par lettres suscrites.

Il consacre des sections à part aux remarques sur quelques lettres, aux abréviations métrologiques, ainsi qu'à la cryptographie.

Jacques Stiennon (1991) reprend les distinctions de Prou, mais restreint le contenu du chapitre correspondant à ce qui était pour de Wailly les « abréviations proprement dites » (les notes tironiennes et la cryptographie sont étudiées à part).

Adriano Cappelli (1949) parle également des abréviations par lettres suscrites, par troncature et par contraction. Pour ces dernières, il propose une sous-catégorie de

- désinences contractées.

En ce qui concerne les marques spéciales, le savant italien les subdivise en :

- marques « avec une signification propre » et
- celles avec une signification « relative ».

Bernhard Bischoff (1993) propose une typologie différente :

- 1) *-m* en forme de < 3 > et *a* suscrit en forme de < ~ > ;
- 2) syllabes et lettres indiquées par un tilde ;
- 3) lettres suscrites ;
- 4) pronoms et particules fréquemment utilisés ;
- 5) contractions ;
- 6) contractions avec désinence suscrite (subdivisées en parties du discours).

A notre connaissance, les abréviations médiévales n'ont été l'objet que de très peu d'études linguistiques spécialisés. G. Hasenohr (1998) a étudié le rôle des abréviations dans la segmentation textuelle. Le système graphique du français médiéval e en effet « emprunté » les abréviations latines sans toujours tenir compte des rapports phono-graphiques propres à la langues française : leur usage pouvait en fait amener à dissocier les éléments de digrammes et de trigrammes correspondant à des diphtongues et des triptongues en voie de réduction phonétique en français médiéval. D'autre abréviations pouvaient enfreindre les règles de syllabation. En revanche, puisque « l'unité de base su système abréviatif est le mot » (*ibidem*, p. 26), les abréviations ne franchissaient jamais (à quelques exception près) les frontières d'un mot ne ne pouvaient donc « brouiller » la perception de l'unité lexicale. L'article reprend la distinction courante des abréviations en contractions d'une part et tildes et « signes

tachygraphiques » de l'autre, en mentionnant que ces derniers « ne font que se substituer à une lettre ou à un groupe de lettre parmi les plus fréquents en latin » (*ibidem*, p. 27). Cette constatation n'amène toutefois pas l'auteur à revoir sa position qui consiste à relier les abréviations uniquement aux unités-mots.

Nicolas Mazziotta (2008b) a proposé une analyse exhaustive du phénomène d'abréviation dans le cadre de son modèle de la langue écrite (*cf.* section 1.2.2). Selon lui, « la relation d'abréviation lie un caténogramme¹⁷⁵, significatif, à un logogramme synonyme... La portée des logogrammes impliqués dans un mécanisme d'abréviation consiste en un caténogramme » (*ibidem*, p. 198). La portée des marques d'abréviation peut selon les cas être *totale* (concerne un mot ou au moins un morphème entier) ou *partielle* (concerne une chaîne de caractères sans signification particulière). La portée peut également être *continue* ou *discontinue*, comme dans le cas de <chl'r> pour <chevalier>. L'article se termine par des propositions de codage des abréviations médiévales compatibles avec les recommandations de la TEI et tenant compte de l'analyse linguistique du phénomène d'abréviation.

Notre analyse des abréviations médiévales qui s'est traduite dans notre proposition de les représenter aux différents niveaux de la transcription et dans le schéma d'encodage XML-TEI a été effectuée avant la parution de l'article de N. Mazziotta. Sur certains points, elle rejoint les réflexions de N. Mazziotta, mais notre modélisation est de façon générale moins complexe.

Les différents classements proposés par les ouvrages de paléographie s'appuient sur des critères hétérogènes (position des caractères supprimés, forme de la marque, catégorie des mots abrégés, etc.) et ne semblent pas pertinents du point de vue de l'analyse linguistique.

En partant de la définition proposée au début de la présente section, nous proposons une typologie basée sur la relation entre la marque et la *portée* de l'abréviation. Cette dernière est définie comme l'ensemble des unités graphiques ou

¹⁷⁵ Rappelons que *caténogramme* dans le système terminologique de N. Mazziotta à ce qu'on pourrait appeler « chaîne de caractères ». Nous tenterons par la suite de « traduire » les termes du chercheur, tout en étant conscient qu'il ne peut pas s'agir d'une transposition adéquate. Nous en assumons les conséquences.

linguistiques nécessaires à la résolution (restitution des caractères alphabétiques omis) de l'abréviation¹⁷⁶. La portée d'une abréviation peut s'étendre à :

- une marque d'abréviation seule, en cas de correspondance univoque entre la marque et les caractères omis (comme c'est le cas de < ʹ > = *us*, < ε > = *et*, etc.) ;
- une marque d'abréviation associée à un ou plusieurs caractères alphabétiques (par exemple, une barre horizontale au dessus d'une voyelle correspond à une consonne nasale : < ā > = *an* ou *am* selon la consonne qui suit ; une marque de même forme correspond à *ue* si elle se trouve au-dessus d'un *q* : < q̄ > = *que*) ;
- un morphème entier (radical ou affixe : dans *ch'z'*, par exemple, l'abréviation porte sur la racine *chevalier* et non sur la désinence casuelle *-s*) ;
- un mot entier (comme, par exemple, *ml't* pour *moult*) ;
- un syntagme (cf. l'exemple de *Nostre Seigneur Jhesu Crist* ci-dessous).


Image	Transcription		
	allographique	diplomatique	normalisée
	n̄rēꝑ' ihūcꝑ	<i>nostre seigneur ihesu crist</i>	Nostre Seigneur Jhesu Crist

Fig. 19 : Représentation d'une abréviation dans les trois niveaux de transcription

Dans ce cas, il est possible de supposer qu'une abréviation unique est utilisée pour chacun des syntagmes : *Nostre Seigneur* et *Jhesu Crist*. On peut toutefois proposer deux autres analyses, qui permettent d'éviter la « propagation » d'une abréviation au-delà d'une unité-mot :

1. Les syntagmes *Nostre Seigneur* et *Jhesu Crist* sont tellement figés que chacun d'entre eux forme une unité-mot. L'absence d'espace blanc dans le manuscrit semble confirmer cette hypothèse (même si on peut trouver dans des manuscrits des cas d'agglutination d'unités-mots indubitablement distincts) ;

176 Notre vision de la portée est donc différente de celle de N. Mazziotta pour qui la relation d'abréviation se forme uniquement entre la marque d'abréviation et les caractères qu'elle « remplace ».

2. On peut considérer $n\bar{r}e$ et $f̄$, ainsi que $ih\bar{u}$ et $c\bar{f}$, comme des abréviations indépendantes qui sont simplement écrites sans espace blanc.

La réalité se situe sans doute entre ces deux analyses : on a affaire dans ce cas à des unités linguistiques qui occupent une position intermédiaire dans le continuum d'autonomie syntagmatique : ce sont en effet des locutions en voie de devenir des mots composés¹⁷⁷. Pour des raisons pratiques, l'encodage de ces abréviations sera basé sur la deuxième analyse, ce qui n'empêchera pas de les afficher « collées » au niveau imitatif (*cf.* plus bas la discussion sur les segmentations particulières).

Les abréviations marquées par la superscription de certains caractères sont un cas particulier : c'est la position et la taille diminuée du caractère qui constituent la marque. Dans les éditions diplomatiques où les lettres reconstituées lors de la résolution des abréviations sont mises en italique, les caractères suscrits restent en romain, car ils ne sont pas reconstitués.

Le plus souvent, le seul caractère « omis » dans ces abréviations est un *u* (après *q*) ou un *r* (après les autres lettres) qui précède immédiatement le caractère suscrit. Les limites de l'abréviation ne dépassent pas alors la combinaison du caractère normal et suscrit.

Dans certains cas, les limites de plusieurs niveaux coïncident. Ainsi, la perluète (ϵ) est une marque d'abréviation qui correspond presque toujours à deux lettres *et*, mais dans la plupart des manuscrits français, elle est utilisée uniquement pour la conjonction *et*, morphologiquement indivisible. La portée de l'abréviation s'étend donc en même temps à la marque seule, à un morphème et à un mot.

La résolution des abréviations n'est pas toujours univoque. Le *p* « barré » peut selon les mots être résolu en *par* ou en *per*. Dans l'exemple de *Nostre Seigneur Jhesu Crist*, rien n'indique en effet qu'il devait y avoir un *i* dans *Seigneur* (selon les manuscrits, ce mot pouvait également être graphié *segneur*, ou *seigneur*, sans mentionner la variation dans la notation de la voyelle de la dernière syllabe). Dans une édition électronique, il est possible d'ajouter aux résolutions des abréviations des « étiquettes de certitude », ce qui permettrait aux lecteurs de distinguer en cas de besoin les données attestées des reconstructions probabilistes.

¹⁷⁷ Voir Haspelmath (2002) pour les critères de distinction.

En cas de transcription destinée à des visualisations multiples, les abréviations peuvent être résolues tacitement au niveau normalisé, les lettres restituées doivent être balisées au niveau diplomatique (afin de pouvoir les visualiser en italique par exemple ou encore d'utiliser la couleur pour distinguer les résolutions sûres des résolutions hypothétiques), tandis qu'au niveau imitatif il convient de représenter les marques d'abréviation dans la forme la plus proche possible de l'original sans tenter de les résoudre.

4.4.1.2. Caractères ambigus

On peut regrouper dans cette catégorie des phénomènes assez variés qui se caractérisent tous par une ambiguïté de lecture d'un caractère médiéval qui est résolue dans l'écriture moderne (bien que le caractère en question existe toujours).

-x = -us ?

Dans certains manuscrits français le caractère *-x* final pouvait être utilisé en alternance avec *-us*, ce qui a poussé certains médiévistes à le considérer comme une abréviation. Cependant le *x* final peut également être utilisé après un *u* ou un *ul* (parfois dans les mêmes manuscrits où il semble avoir la valeur de *us*), ce qui complique l'analyse. Des propositions d'explication de cet usage ont été nombreuses et des discussions sur la façon de représenter ce *x* « abrégatif » dans les éditions ont duré bien longtemps (cf. Acher 1913 ; Rajna 1929). En 2001 encore, Vielliard & Guyotjeannin s'abstiennent de donner une recommandation universelle et préconisent une analyse du système graphique du manuscrit préalable à toute décision éditoriale :

« Si *< x >* n'est employé qu'avec la valeur de *< us >*, on peut le considérer comme une abréviation et le développer, à condition de justifier ce parti dans l'introduction. Dans le cas contraire, il vaut mieux respecter la graphie du manuscrit » (Vielliard & Guyotjeannin coord. 2001, p. 28).

En cas d'édition « multi-facettes », il semble évident que le caractère *x* doit être préservé dans les transcriptions imitative et diplomatique. La représentation normalisée peut être basée sur les consignes de Vielliard et Guyotjeannin.

***u/v, i/j* et des séries de jambages**

Représentation des données primaires : transcription de manuscrits médiévaux

Les caractères ambigus sans doute les plus connus sont les deux paires *u/v* et *i/j*. La distinction actuelle basée sur le principe consonne/voyelle a été proposée par Pierre de la Ramée au XVI^e siècle. Dans les manuscrits médiévaux, ces paires pouvaient être utilisées soit comme variantes libres (avec une fréquence supérieure de *u* et de *i*), soit comme variantes positionnelles (*v* initial / *u* médian et final ; *j* initial et final / *i* médian). La grande majorité des éditeurs français introduisent la distinction phonétique, mais il existe cependant des éditions diplomatiques où les distinctions originales sont conservées (cf. les cas analysés dans la section 2.2 ci-dessus). Dans certains cas, l'interprétation phonétique de l'*u* peut s'avérer difficile. Vielliard et Guyotjeannin recommandent alors de chercher des indications comme l'aboutissement phonétique des noms géographiques ou encore des alternances graphiques dans le manuscrit même : si le futur du verbe *avoir* est graphié tantôt *aura*, tantôt *auera*, il faut transcrire *avra* ; si, au contraire, l'alternance est avec *ora*, la transcription correcte serait *aura*¹⁷⁸.

Dans une édition comportant des couches de transcription multiples, on peut adopter les principes exposés ci-dessus pour le niveau normalisé. Au niveau imitatif, il convient bien entendu de préserver la distinction présente dans le manuscrit. Au niveau diplomatique, si on suivait « à la lettre » le principe de transcription graphématique, il faudrait neutraliser les oppositions *i/j* et *u/v*, qui relèvent de la variation libre ou positionnelle, et tout transcrire par *i* et *u* respectivement. Toutefois, dans ce cas, puisque ces variantes manuscrites correspondent à des lettres distinctes dans le système graphique moderne, leur maintien dans une transcription diplomatique n'augmentera pas le temps de sa saisie, ni ne gênera pas sa lecture pour un lecteur averti. Il semble de façon générale préférable de conserver une information dans la transcription si celle-là ne complique pas son exécution et son interprétation, il n'y a donc pas de raison pour « forcer » la neutralisation des variantes *i/j* et *u/v*. Ces caractères seront donc transcrits « tels quels » au niveau diplomatique.

Il convient de noter que *i* et *j* ne portent pas de point suscrit dans la plupart des manuscrits médiévaux datant d'avant le XV^e siècle, et son usage ne devient

178 Pour le futur des verbes *avoir* et *savoir*, on transcrit généralement avec un *v* les textes d'ancien français et avec *u*, les textes de moyen français, en s'appuyant sur les conclusions de A.T. Baker (1937). Il est cependant évident que la transformation phonétique sous-jacente ne s'était pas produite d'un jour au lendemain et en même temps dans toutes les régions.

systematique que dans les incunables (Beaulieux 1927, p. 6)¹⁷⁹. Ces caractères peuvent en revanche être surmontés d'un léger trait oblique (ressemblant à un accent aigu)¹⁸⁰. Le point suscrit fait partie du tracé de la lettre moderne, il est donc raisonnable de l'utiliser au niveau normalisé et diplomatique. Au niveau imitatif, on cherche à préserver les distinctions des variantes (allographes) perceptibles, mais on n'imite pas forcément le tracé du caractère manuscrit. Puisqu'il n'y a pas d'opposition *ι/i* dans les manuscrits médiévaux français, il n'aurait pas été gênant, en principe, d'utiliser le caractère moderne (avec le point) dans la transcription. L'opposition *ι/í* du manuscrit pourrait ainsi être représentée par l'opposition *ι/í*.

Cependant, en l'absence d'une marque diacritique, le *ι* pouvait parfois se confondre dans le manuscrit avec des jambages faisant parties des *n*, *m* et *u*. Même si souvent des petits traits de liaison en haut ou en bas permettent de faire la distinction, cette ambiguïté est un trait important du système graphique des manuscrits médiévaux. Prenons, par exemple, la graphie *luire* :



Fig. 20 : Série de jambages dans le mot 'livre' (im574, 172r).

Celle-ci peut être transcrite (en mode normalisé) de trois façons différentes, si on ne prend que des formes réelles de mots français : *livre*, *livré* ou *luire*. Bien sûr, le contexte permet le plus souvent d'interpréter la forme graphique de façon univoque, mais des cas d'ambiguïté effective et de confusion existent. C'est dans ces cas rares seulement qu'il nous semble nécessaire de représenter l'ambiguïté des séries de jambages dans la transcription imitative. En revanche, l'emploi du caractère *ι* (sans point) peut s'avérer utile dans la visualisation imitative, car cela accentuerait l'opposition visuelle avec *í* et rapprocherait l'aspect de la transcription de sa source sans compliquer la lecture.

179 Des exceptions ont toutefois été constatées par certains paléographes (Delisle 1852).

180 D'après Charles Beaulieux (1927, p. 6), « les accents sur l'*i* apparaissent au IX^e siècle, mais ne deviennent fréquents qu'à partir du XIII^e siècle, époque à laquelle la minuscule caroline s'est transformée en écriture anguleuse improprement appelée gothique ». L'usage de l'accent aigu sur le *i* et sur certaines autres lettres est particulièrement remarquable certains manuscrits anglo-normands (dont ceux de la *Chanson de Roland* ou de *Tristan et Yseut*). Nous reviendrons sur cette question dans la section 4.4.1.3. consacrée aux marques diacritiques médiévales.

En effet, comme nous l'avons déjà vu, il y a une part d'interprétation dans toute transcription. Selon le niveau de la transcription, il faut veiller à ce que l'interprétation ne masque pas les données importantes pour ce niveau. Il convient aussi de tenir compte des habitudes de lecture (formées par des traditions éditoriales) et des facilités techniques d'encodage (même si ces dernières ne doivent pas être déterminantes). Du point de vue linguistique, l'ambiguïté des séries de jambages ne constitue une donnée importante que lorsqu'elle entraîne une ambiguïté réelle des formes de mots. Du point de vue technique, l'encodage systématique de cette ambiguïté risque de compliquer considérablement la saisie et le traitement des données. Tout ceci justifie, à notre avis, notre choix qui peut sembler paradoxal à une première vue.

ci/ti et ct/tt

Dans ces groupes de caractères, il peut être difficile, selon le type d'écriture, de déterminer si le premier d'entre eux est un *c* ou un *t*. Voici deux images extraites d'une même page du ms. Paris, BnF, fr. 840, f° 395r (XVe s., *L'art de dictier* d'Eustache Deschamps) avec leur transcription par l'éditeur¹⁸¹ :

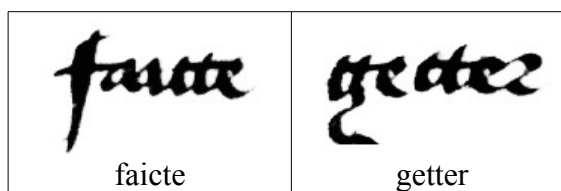


Fig. 21 : Ambiguïté du groupe de lettres *ct/tt* (*dictier* 840, 395r).

Ici la difficulté de distinction graphique correspond à l'absence de distinction phonétique, l'opposition *c/t* n'a aucune importance pour la compréhension du mot.

Vielliard et Guyotjeannin recommandent de généraliser en cas de doute la graphie la plus fréquente dans le manuscrit édité (d'après les occurrences qu'on peut lire clairement) ou, le cas échéant, de « choisir le parti le plus simple pour le lecteur, en raisonnant au cas par cas et en tenant compte, si besoin est, de l'étymologie » (2001, p. 27).

Cette solution semble la plus raisonnable pour les transcriptions diplomatique et normalisée. Dans la transcription imitative, il convient d'utiliser une même transcription

¹⁸¹ Éd. Gaston Raynaud, Paris, F. Didot, 1891

dans tous les cas de doute, sans tenir compte de l'étymologie. Une ligature < ct > peut éventuellement être utilisée pour le groupe *ct*.

diacritiques modernes (ç, é, ï, etc.)

Il est d'usage d'introduire dans les éditions certains signes diacritiques qui permettent de faciliter la lecture en désambiguïsant la « valeur phonétique » de certains groupes de lettres. Les recommandations relatives à l'édition des textes français médiévaux limitent cette pratique à :

- l'usage d'accent aigu sur le *e* accentué [e] en fin de mot ou devant un *s* final ;
- l'usage de tréma sur le *i* dans les cas où il faut distinguer un *i* syllabique d'un élément d'une diphtongue (*païs, meïsmes, oiï*) ;
- l'usage de la cédille pour signaler la lecture fricative (et non occlusive) d'un *c* devant les consonnes *a, o, u* (*ço, ainçois, sçay*).

Certains éditeurs introduisent davantage de diacritiques (par exemple, un accent grave dans la préposition *à*), dont la fonction n'est pas de faciliter l'interprétation phonétique, mais de distinguer graphiquement des homophones. Cette pratique n'est pas recommandée, mais elle a l'avantage de faciliter les procédures d'étiquetage morphosyntaxique et de lemmatisation automatiques des textes. Elle a tendance à se répandre dans les éditions de textes des XIV^e et XV^e siècles (*cf.* par exemple de Crecy 1997).

Dans tous ces cas, il s'agit d'enrichir la transcription de marques de désambiguïsation sans pour autant perdre les données originales. La seule chose qui soit « masquée » par ces diacritiques est l'ambiguïté des graphies originales, qui est de toute façon résolue, à quelques exceptions près, au niveau du mot ou du contexte immédiat. L'usage des diacritiques est donc justifié au niveau de la représentation normalisée. La généralisation de l'usage de l'accent grave sur la préposition *à* peut même être envisagée : elle serait utile pour le traitement automatique, et il est très facile de masquer ces accents au moment de la visualisation. En revanche, aux niveaux imitatif et diplomatique, l'usage de ces marques de désambiguïsation ne semble pas approprié : si une ambiguïté existe dans le système graphique des manuscrits, il convient de la préserver dans les transcriptions graphématique et allographique.

4.4.1.3. Diacritiques médiévaux

Dans les manuscrits français qui constituent notre corpus, l'usage des marques diacritiques autres que les marques d'abréviation se limite à deux cas :

- un léger trait oblique (une sorte d'accent aigu) sur le *i* (*í*) ;
- un point (ou parfois une autre marque, comme un trait oblique ou une sorte de crochet) au-dessus de *y* (*ȳ*, *ȝ* ou *ȝ̅*).

Nous avons déjà mentionné le rôle du trait oblique sur le *i* dans les séries de jambages ambiguës. Le point sur le *y* n'a pas de fonction particulière (tout comme le point sur le *i* dans la typographie moderne). Il constitue toutefois un élément bien visible du caractère, ce qui lui donne un potentiel fonctionnel. Il convient donc de représenter ce point (ou trait oblique) suscrit dans la transcription imitative, mais pas dans les transcriptions diplomatique et normalisée.

Certains manuscrits médiévaux, en particulier anglo-normands, utilisent des accents (simples ou doubles) non seulement sur des *i*, mais aussi sur des *o*, *a* et d'autres lettres (Monfrin 1958, p. 395-396 ; Marchello-Nizia 1995, p. 1240 ; Short 2005, p. I/16). Ces accents pouvaient marquer certains mots monosyllabiques (par exemple, la préposition *á*), la diérèse, etc. Selon Charles Samaran (Samaran 1929, p. 164), l'accent simple ou double sur le *i* pouvait servir indiquer la prononciation conconantique [dʒ]. Certains éditeurs ont choisi de conserver ces accents dans les éditions, tout en procédant par ailleurs à la normalisation graphique (par exemple Curtius 1911). Il est vrai que l'accentuation originale constitue une donnée intéressante, et il est souhaitable de la préserver dans une édition. Cependant, les éditions papier traditionnelles qui emploient en même temps les diacritiques médiévaux et modernes sont difficiles à lire et à traiter à l'aide des outils d'analyse automatique du texte. La distinction des niveaux de représentation permet de résoudre ce problème en conservant les accents médiévaux au niveau imitatif, en les supprimant au niveau diplomatique et en introduisant si nécessaire des diacritiques modernes au niveau normalisé.

4.4.1.4. Variantes calligraphiques

L'écriture manuscrite est toujours riche en variation des tracés de caractères. Une analyse linguistique permet toutefois de distinguer les variantes aléatoires, dues à l'individualité du scripteur et aux conditions matérielles de son travail, des variantes faisant partie du système graphique et pouvant dans certaines conditions prendre une charge fonctionnelle. Bien sûr, la limite entre ces deux catégories n'est pas infranchissable, et on peut avoir affaire à des cas transitoires. Nous sommes néanmoins convaincu que pour une époque et un système d'écriture donnés, il est possible d'établir une liste précise des variantes linguistiquement pertinentes. Ce sont ces variantes qu'il convient de représenter dans la transcription imitative (allographique).

Le développement de l'imprimerie conduit progressivement à la réduction du nombre de variantes calligraphiques : leur nombre était encore important dans les incunables et a baissé progressivement au cours du XVI^e et du XVII^e siècle. La variante ayant résisté le plus longtemps est le *s* « long » < f > (on le trouve encore dans des imprimés du XVIII^e siècle).

Dans les éditions scientifiques des textes médiévaux, les variantes calligraphiques ont presque toujours été neutralisées, à l'exception de quelques éditions « hyper-diplomatiques » (*cf.* la section 2.2 plus haut).

Il est possible de distinguer trois types de variantes calligraphiques :

- variantes conditionnées par la position dans le mot : initiale, médiane ou finale ;
- variantes conditionnées par l'entourage graphique ;
- variantes libres.

Assez souvent, une variante donnée ne se range pas complètement dans un seul type, mais on peut juste constater une tendance à l'usage positionnel ou graphiquement conditionné. Le *n* et le *m* avec le dernier jambage prolongé (< n > et < η >), ainsi que le *s* « court » < ε >, sont des variantes finales, le *v* devient au XV^e siècle la variante initiale de *u*. Le *r* « rond » < z > est une variante utilisée habituellement après des caractères ayant un tracé plus ou moins rond (par exemple, *o*, *p* ou *ð*).

Les variantes positionnelles jouent un rôle dans le fonctionnement du système graphique : elles renforcent et parfois remplacent les espaces blancs en tant que séparateurs des unités-mots. Les autres types de variantes ont un potentiel de fonctionnalité dans la mesure où ils peuvent devenir des variantes positionnelles ou des caractères distincts (*u* et *v*, *i* et *j* sont à l'origine des variantes).

La liste complète des variantes de caractères que nous représentons dans les transcriptions imitatives est présentée dans l'annexe II.1. Les problèmes d'encodage de ces variantes (ainsi que des autres éléments graphiques « non-standard ») sont analysés dans le chapitre 5 ci-dessous.

4.4.1.5. Ligatures

Les ligatures sont assez nombreuses dans l'écriture manuscrite. Comme la plupart des variantes calligraphiques, elles ne sont jamais représentées dans les éditions critiques, ni diplomatiques. Dans les transcriptions imitatives, il convient de représenter les ligatures, dont la forme se distingue nettement de celle qu'aurait donnée la combinaison de deux caractères indépendants « accolés ». Nous avons déjà mentionné le groupe *ct* (souvent difficile à distinguer de *tt*). Une autre ligature caractéristique est le « double *p* » < *pp* >.

Le « potentiel fonctionnel » des ligatures est plutôt faible. Elles peuvent être considérées comme des marques qui renforcent l'unité du mot : il est peu probable qu'une ligature se trouve entre la dernière lettre d'un mot et la première d'un autre. Si tel était le cas, il serait raisonnable de supposer que les deux « mots » forme en fait une seule unité segmentale.

4.4.1.6. Grandes initiales (lettrines)

Les grandes lettrines initiales réalisées en couleur et souvent dotées d'éléments décoratifs jouent un rôle important dans la structuration du texte dans les manuscrits médiévaux. Ces lettres décoratives étaient ajoutées par un rubricateur ou par un enlumineur quand le scribe avait fini l'écriture du « corps » du texte. C'était le scribe qui établissait la hiérarchie des initiales en laissant un alinéa plus ou moins grand

(en nombre de lignes). Il existe d'ailleurs des manuscrits et des incunables où les initiales n'ont jamais été réalisées, mais les alinéas permettent d'appréhender la hiérarchie des unités textuelles. Le rubricateur pouvait, pour sa part, dessiner des initiales dépassant le cadre initialement prévu.

De toute façon, une initiale marque toujours le début d'une unité textuelle de haut niveau. Les plus petites initiales (une ou deux lignes d'alinéa selon les manuscrits) correspondent à des paragraphes (ou des groupes de vers) dans les textes modernes. Des initiales plus grandes correspondent à des unités supérieures (épisodes, sections, chapitres, parties, etc. selon la complexité de la structure de l'œuvre).

A l'exception de quelques éditions particulièrement attentives à l'aspect décoratif du manuscrit, les initiales ne sont pas reproduites dans les éditions modernes, qu'elles soient critiques ou diplomatiques. Certains éditeurs choisissent de respecter la division en chapitres et paragraphes de leur manuscrit de base, d'autres la modifient plus ou moins profondément. Vielliard & Guyotjeannin (2001) énoncent le principe de modernisation inconditionnelle : « ne pas tenir compte des usages du scribe » en ce qui concerne l'emploi des majuscules. Ils ajoutent toutefois que « l'emploi des lettrines et des majuscules d'un manuscrit bien ponctué aide néanmoins à percevoir les intentions stylistiques du scribe, sinon celles de l'auteur ; il doit être étudié en tant que tel et à part » (Vielliard & Guyotjeannin coord. 2001, p. 55). A notre avis, il faut bien distinguer les majuscules ordinaires (faisant partie de la ponctuation « lexicale » ou « syntaxique »¹⁸²) et les initiales (qui participent à la mise en page et à la structuration « globale » du texte).

Les grandes divisions représentées par la mise en page du manuscrit ne sont jamais dues au hasard¹⁸³, elles résultent de l'analyse du texte par le scripteur médiéval, il nous semble donc raisonnable d'en tenir compte à tous les niveaux de transcription.

Au niveau normalisé, il est toutefois possible d'introduire des divisions de paragraphes dans de longs passages continus dans le manuscrit. Au niveau diplomatique, les paragraphes doivent correspondre aux unités structurelles du manuscrit. Autrement dit, il ne doit pas y avoir de « sauts de paragraphe » là où il n'y en a pas dans le manuscrit.

182 Voir la section 2.2.1. pour la distinction des trois niveaux de ponctuation que nous adoptons.

183 Cf. l'article de D. Poirion (1978) sur les paragraphes dans les manuscrits de Villehardouin.

Au niveau imitatif, il est utile de représenter des informations plus détaillées sur l'initiale : sa couleur, sa taille prévue (nombre de lignes avec alinéa) et sa taille réelle. La présence d'un élément décoratif peut être signalée, mais la description complète de la décoration relève plutôt du niveau paléographique.

4.4.1.7. Marques de ponctuation

Comme nous l'avons déjà mentionné à plusieurs reprises, la ponctuation des manuscrits n'a presque jamais été prise en compte dans les éditions critiques ou diplomatiques.

Nous avons déjà présenté l'ensemble des marques de ponctuation médiévales attestées dans notre corpus (section 2.7.1) et nous reviendrons sur les problèmes particuliers qui se sont posés lors de la transcription de certains textes dans les sections consacrées à l'analyse de la ponctuation dans chacun des textes de notre corpus (chapitres 6 à 8). Ci-dessous, nous n'évoquerons que les éléments d'analyse nécessaires pour justifier la façon dont il convient de représenter ces marques dans les différents types de transcription.

Une ponctuation moderne correspondant à l'interprétation de la structuration du texte par l'éditeur scientifique est appropriée à la représentation normalisée. Bourgain et Vielliard recommandent dans tous les cas de « s'inspirer le plus possible de la ponctuation médiévale, surtout lorsqu'on a un manuscrit proche de l'époque de l'auteur et soigneux, car il rend alors au mieux les intentions de l'auteur, notamment l'analyse de la phrase telle qu'on la percevait à l'époque » (Bourgain & Vielliard coord. 2002, p. 66).

Puisque la transcription diplomatique est fondée sur le principe graphématique, il conviendrait d'y représenter les graphèmes ponctuants (ou *punctogrammes*, si on adopte le terme de Nicolas Mazziotta (2007b)) en neutralisant les variantes combinatoires. Dans la majorité des manuscrits, on peut distinguer, du point de vue fonctionnel, deux grandes catégories des ponctuants : des ponctuants forts (une marque de ponctuation quelconque suivie d'une majuscule) et des ponctuants faibles (une marque de ponctuation quelconque suivie d'une minuscule). La forme de la marque de ponctuation est moins importante que le type de caractère qui la suit : majuscule ou

minuscule. Le seul fait d'employer une majuscule (même non précédée d'un signe de ponctuation proprement dit) peut constituer une ponctuation forte dans certaines conditions. Nous proposons donc d'utiliser deux marques de ponctuation dans les transcriptions diplomatiques :

- un point suivi d'une majuscule pour toutes les ponctuations fortes ;
- une virgule suivie d'une minuscule pour toutes les ponctuations faibles.

Dans certains manuscrits et incunables, une distinction fonctionnelle liée à la forme de la marque de ponctuation semble se profiler. Par exemple, dès le XIII^e siècle, on peut rencontrer parfois un *punctus interrogativus* < ? > dans des phrases selon toute vraisemblance interrogatives ou exclamatives. Cette marque de ponctuation peut être représentée par un point d'interrogation dans la transcription diplomatique.

Au niveau imitatif, il convient de représenter toutes les marques de ponctuation médiévales qu'on peut identifier dans un manuscrit donné.

Il s'agit le plus souvent du point seul (placé habituellement à la mi-hauteur de la ligne) < · >, du *comma* (ou *punctus elevatus*) < ; > et de la barre oblique < / >. Dans certains manuscrits, une distinction de variantes de commas (< ; > vs < / >) ou de barres obliques (< / > vs < / >) semble se profiler, il convient d'en tenir compte. Plus rarement, on trouve des occurrences du *punctus interrogativus* et d'autres marques décrites dans les traités de ponctuation médiévaux.

Le « pied-de-mouche » (ou *paraph*) < ¶ > est un cas particulier : cette marque (issue sans doute d'un grand *C* du mot latin *Capitulum*) peut être placée au début de divisions de texte : elle est souvent utilisée dans des tables de matières, devant les titres de chapitres, mais peut aussi apparaître au milieu du texte (Parkes 1992, p. 43). Les pied-de-mouches étaient souvent écrits en couleur (bleue ou rouge), par un rubricateur¹⁸⁴. L'origine de cette marque, le mode de son écriture et son fonctionnement la rapprochent des lettrines initiales plutôt que des marques de ponctuation.

Dans la présentation normalisée, où la ponctuation est modernisée, les pied-de-mouches peuvent complètement disparaître. Il est toutefois souhaitable de mettre des « sauts de paragraphe » à tous les endroits où les pied-de-mouches sont placés, à

184 Artisan chargé d'écrire toutes les lettres, les mots et les phrases à l'encre rouge dans le manuscrit (Stiennon 1991, p. 349)

l'exception des cas où leur usage semble aberrant et risque de désorienter le lecteur, et dans ce cas on signalera cette modification dans une note. Dans la présentation diplomatique, il convient d'utiliser systématiquement un saut de paragraphe à l'endroit où un pied-de-mouche est placé dans le manuscrit. Éventuellement, un symbole conventionnel (§) peut être utilisé en complément du saut de paragraphe. Dans la présentation imitative, il convient d'utiliser un symbole de forme plus proche de celle d'un pied-de-mouche médiéval et de conserver l'information sur sa couleur originale.

4.4.1.8. Segmentations particulières

Dans les manuscrits latins écrits en France, la séparation des mots par des espaces blancs commence à être pratiquée systématiquement à partir du XI^e siècle (Saenger 1997, p. 23), mais les « mots » graphiques des manuscrits médiévaux ne correspondent pas toujours à ce que nous considérons comme des mots du point de vue de l'analyse linguistique moderne. Des phénomènes de ce genre peuvent être observés dans les manuscrits français comme dans les manuscrits latins. Les articles, les prépositions et autres clitiques y sont assez souvent « collés » aux mots « pleins », mais les syntagmes agglutinés peuvent également contenir plusieurs mots lexicaux. Dans d'autres cas, un espace blanc peut couper en deux un mot unique. Cette coupure peut correspondre ou ne pas correspondre à une division en morphèmes. En revanche, le plus souvent, au moins une des parties coupées est graphiquement identique à un mot existant¹⁸⁵.

Un problème supplémentaire est lié à la taille de l'espace blanc. Dans un nombre de manuscrits certains espaces blancs entre les mots sont nettement plus courts que les autres (cf. le facsimilé des *Serments de Strasbourg* ci-dessus). L'opposition entre les grands et les petits blancs est-elle significative ? Il est difficile de répondre à cette question faute d'étude approfondie à ce sujet, aucune n'ayant jamais été réalisée jusqu'à présent. Il convient de préciser en tout cas que la différence entre les blancs « longs » et « courts » (ainsi qu'entre les blancs « courts » et l'absence de blanc) est parfois très

185 Voir (Andrieux-Reix & Monsonégo 1997), étude la plus profonde des segmentations particulières dans les manuscrits français médiévaux publié à ce jour. L'histoire générale de l'introduction de la séparation des mots dans les manuscrits d'écriture latine est étudiée dans (Saenger 1997).

difficile à établir. En cas de doute, nous nous tenons à l'opposition binaire (blanc / absence de blanc).

En ce qui concerne la présentation des segmentations particulières, la tradition veut que celles-ci soient neutralisées (autrement dit, que les mots soient séparés « à la moderne ») dans les vues normalisée et diplomatique. En revanche, la segmentation réelle du manuscrit doit être conservée dans la version imitative. Les blancs courts doivent être encodés pour permettre leur étude ultérieure, mais il n'est pas nécessaire de les visualiser d'une façon différente des blancs « normaux ».

4.4.1.9. Marques de corrections sribales

Les manuscrits médiévaux peuvent porter des traces de corrections effectuées soit par le scribe, soit par une autre personne à une époque proche de la réalisation du manuscrit (même si elles sont beaucoup moins nombreuses que les corrections dans les manuscrits d'auteurs de l'époque moderne). Pour la bonne représentation des données linguistiques, il est important de distinguer ces corrections témoins de l'époque des marques ajoutées plus tardivement. Pour un linguiste, ces dernières n'ont pas plus d'intérêt que les émendations de l'éditeur scientifique moderne, tandis que les premières font partie des données primaires.

En cas de correction contemporaine de l'écriture du manuscrit, on peut faire l'hypothèse que l'état du texte corrigé représente la langue de l'époque étudiée. C'est donc le texte corrigé qui intéressera la grande majorité des chercheurs. Il est inutile de préserver les traces du texte initial et de l'opération de correction dans les transcriptions normalisée et diplomatique.

En revanche, la façon dont les corrections sont effectuées dans le manuscrit peut intéresser un paléographe ou un linguiste travaillant sur les systèmes graphiques. L'erreur initiale et le fait qu'elle a été rapidement corrigée peut dans certains cas servir d'indice de variation ou d'évolution de la langue. Les données sur la correction (la forme initiale, si encore accessible, la forme corrigée et la marque de correction) peuvent donc être représentées dans la transcription imitative pour satisfaire ce genre d'intérêt.

Représentation des données primaires : transcription de manuscrits médiévaux

Il est plus important dans ce cas de représenter clairement les trois éléments de la correction que de tenter de l'imiter telle qu'elle se présente dans le manuscrit. François Masai (1950) a proposé un système de notation des corrections sribales (additions, suppressions et substitutions) que nous avons décidé d'adopter : des barres obliques signalent des ajouts, tandis que les crochets marquent les suppressions. Des marques supplémentaires indiquent la méthode de la suppression : par exemple un point au-dessous d'une lettre signifie que cette dernière est à supprimer. La liste complète des marques utilisées est présentée dans l'annexe II.2.

4.4.2. Synthèse

Le traitement des différents types de données graphiques selon les différents modes de présentation correspondant aux niveaux de transcription du manuscrit français médiéval est résumé dans le tableau ci-dessous :

Type de données	Transcription normalisée	Transcription diplomatique (graphématique)	Transcription imitative (allographique)
Abréviations médiévales	résolution tacite	résolution avec la mise en italique des caractères ajoutés	reproduction de la marque d'abréviation
Caractères ambigus <i>x</i> final pouvant correspondre à <i>us</i> <i>u/v</i> ; <i>i/j</i> (voyelle/consonne) séries de jambages distinction <i>ci/ti</i> , <i>ct/tt</i> <i>e/é/ē</i> , <i>c/ç</i> (diacritiques modernes)	maintien de <i>x</i> <i>v, j</i> = consonnes <i>u, i</i> = voyelles interprétation interprétation usage de diacritiques dans des cas précis	maintien de <i>x</i> identique au manuscrit (+ point suscrit) interprétation interprétation absence de totale diacritiques	maintien de <i>x</i> identique au manuscrit (pas de point suscrit) interprétation interprétation absence de diacritiques modernes
Diacritiques médiévaux <i>í</i> , <i>ȳ</i>	suppression	suppression	maintien
Variantes calligraphiques (allographes) : <i>s</i> long ; <i>r</i> rond ; <i>n</i> et <i>m</i> finaux, etc.	neutralisation	neutralisation	maintien
Ligatures	séparation	séparation	maintien partiel
Grandes initiales	neutralisation (la structuration du texte est effectuée par l'éditeur, avec la prise en compte des données du manuscrit)	neutralisation, mais placement au début d'un paragraphe	représentation de la taille et de la couleur de l'initiale
Ponctuation	application des règles modernes	point + majuscule pour une ponctuation forte virgule + minuscule pour une ponctuation faible	maintien des marques médiévales
Segmentations particulières	normalisation	normalisation	maintien
Corrections sribales	texte corrigé	texte corrigé	maintien des marques de correction

Tableau 5 : Synthèse de la représentation des « données sensibles » aux différents niveaux de la transcription de manuscrits médiévaux

4.4.3. Exemples d'application

4.4.3.1. Transcription multi-facettes des Serments de Strasbourg

Pour donner un exemple concret d'application des principes formulés ci-dessus, revenons aux *Serments de Strasbourg* et plaçons l'une après l'autre les trois transcriptions-types¹⁸⁶ :

Pro Deo amur et pro christian poblo et nostro commun salvament, d'ist di in avant, in quant Deus savir et podir me dunat, si salvarai eo cist meon fradre Karlo et in aiudha et in cadhuna cosa, si cum om per dreit son fradra salvar dift, in o quid il mi altresi fazet. Et ab Ludher nul plaid nunqua·m prindrai qui, meon vol, cist meon fradre Karle in damno sit.

Fig. 22 : Transcription normalisée (« orthographique ») des Serments de Strasbourg.

Pro deo amur *et* pro christian poblo *et* nostro commun saluament, d'ist di In auant, in quant *deus* sauir *et* podir me dunat, si saluarai eo, cist meon fradre karlo, *et* in aiudha, *et* in cadhuna cosa, si *cum* om *per* dreit son fradra saluar dift. In o quid il mi altresi fazet. Et ab ludher nul plaid nunqua·m prindrai qui meon uol cist meon fradre karle in damno sit.

Fig. 23 : Transcription diplomatique (graphématique) des Serments de Strasbourg.

186 Voir Fig. 10 (p. 172) et la note 157 (p. 171) pour la traduction.

P ro dō amur & p xp̄i an poblo & nrō cōmun
 faluament · dift dī [e + i] n auant · inquantdf
 faur & podir medunat · fīfaluaraieo ·
 cist meon fradre karlo · & in a[d] iudha ·
 & in cad huna cofa · ficū om p dret fon
 fradra faluar dift · In o quid il mialtre
 fī faz & · Et ab ludher nul plaid nūquā
 prindraī qui meon uol cist meon fradre
 karle in damno fit ·

Fig. 24 : Transcription imitative (allographique) des Serments de Strasbourg.

Dans la transcription normalisée, nous avons appliqué toutes les règles recommandées pour l'édition des textes de français médiéval, y compris l'introduction de la distinction *u/v* voyelle et consonne, ce que les autres éditeurs (à l'exception de Lauer) n'ont pas fait¹⁸⁷. Les majuscules ont été ajoutés aux noms propres (*Karlo* et *Ludher*) et au mot *Deus* 'Dieu'. Nous avons également utilisé le « point médian » pour signaler l'enclise dans *nunqua-m*, en suivant O. Collet (Wagner & Collet 1995).

En ce qui concerne la ponctuation, où l'éditeur scientifique garde en principe toute sa liberté, nous avons suivi les choix de Lauer et Wagner là où ils concordaient (les autres éditeurs ayant respecté la ponctuation de manuscrit).

Dans quatre cas, la ponctuation des éditeurs correspond à celle du manuscrit : il s'agit de séparer des compléments circonstanciels introductifs et de marquer le début d'une proposition circonstancielle comparative à l'aide d'une ponctuation faible (l. 2, 3 et 5).

Dans un cas, les éditeurs ont remplacé la ponctuation forte du manuscrit par une ponctuation faible au début d'une proposition conditionnelle (*in o quid...* 'à condition que', l. 6).

Les éditeurs ont été unanimes à ne pas préserver trois marques de ponctuation entre le verbe avec un sujet pronominal postposé et son complément d'objet direct (*saluarai eo _ cist meon fradre*, l. 3) et devant deux compléments circonstanciels avec la répétition de la conjonction *et* (*_ et in aiudha _ et in cadhuna cosa*, l. 4). Ces usages, aussi étonnants qu'ils puissent paraître à un lecteur moderne, ont été attestés dans bien

187 Cf. la note 158.

d'autres manuscrits médiévaux¹⁸⁸. Nous avons également relevé un certain nombre de cas semblables dans notre corpus.

Dans deux cas, la ponctuation des éditeurs diverge. Premièrement, à la ligne 7, Lauer ne met aucune ponctuation au début d'une proposition indépendante introduite par une conjonction *et* (*et ab ludher nul plaid...*), tandis qu'on trouve une ponctuation forte à cet endroit dans le manuscrit et chez Wagner. Ce choix est d'autant plus surprenant que Lauer met bien une virgule dans le passage correspondant de sa traduction. Nous avons donc choisi de suivre le manuscrit et Wagner dans ce cas.

Deuxièmement, Lauer met des virgules avant et après le complément circonstanciel (de type « commentaire énonciatif ») *meon vol* 'selon ma volonté' à la différence du manuscrit et de Wagner. Cette ponctuation nous semble utile à la bonne compréhension du texte, nous avons donc suivi Lauer dans la transcription normalisée.

Dans la transcription diplomatique, nous avons respecté le principe graphématique (un caractère moderne correspond à un graphème du manuscrit), avec la mise en italique des lettres « restituées » dans les résolutions des abréviations. La ponctuation faible du manuscrit (point + minuscule) est représentée par une virgule suivie d'une minuscule, et la ponctuation forte, par un point suivi d'une majuscule. La segmentation est en revanche normalisée.

Dans la transcription imitative, nous avons reproduit les allographes du manuscrit (variantes des lettres, marques d'abréviation et de ponctuation) et la segmentation effective. Nous avons utilisé le surlignage en couleur pour mettre en évidence les « petits blancs » (en bleu clair entre des mots séparés et en mauve à l'intérieur d'un mot). La correction sribale (barre verticale écrite par-dessus un < e > et interprétée comme un < I >) est présentée conformément aux recommandations de F. Masai.

Les mêmes principes ont été respectés dans la transcription des fragments de manuscrits qui ont constitué notre corpus pour l'étude de la ponctuation médiévale. Toutes ces transcriptions peuvent être consultées dans les trois modes de présentation dans l'annexe IV sur le CD ROM accompagnant la présente thèse. Un format spécifique

¹⁸⁸ A commencer par les romans de Chrétien de Troyes dans l'« édition » du scribe Guiot (Roques 1952).

a été élaboré pour la présentation des citations du corpus dans ce mémoire. La section suivante est consacrée à sa description.

4.4.3.2. Format des citations du corpus

Dans le cadre de citations d'occurrences d'un corpus dans une publication scientifique, il est impossible d'afficher systématiquement côte à côte les trois niveaux de représentation des données de transcriptions. Selon les objectifs de recherche visés par la publication, on peut privilégier tel ou tel niveau, mais il peut également être nécessaire de compléter un niveau avec des traits appartenant à un autre.

L'objet de notre recherche est le système de ponctuation, nous prenons donc soin de représenter précisément toutes les données qui concernent la ponctuation et « allégeons » l'aspect des éléments qui n'y sont pas directement liés.

1. Le niveau de représentation graphématique semble le plus adapté pour **les graphies des mots**. Les **allographes des lettres** et les **marques d'abréviation** ne jouent pas de rôle important dans la structuration de texte au niveau « syntaxique ». Leur reproduction systématique dans les exemples risquerait de gêner la lecture et de « brouiller » les éléments qui font l'objet même de la citation (structures syntaxiques et marques de ponctuation). Nous neutralisons donc les variantes des lettres et résolvons les abréviations en indiquant les lettres « reconstituées » par les italiques. Il est en même temps indispensable de rester proche du système graphique original sans le masquer par une couche de normalisation, nous n'ajoutons donc pas de signes diacritiques et n'introduisons pas les distinctions *i/j* et *u/v*.

2. La **segmentation** du texte fait partie du système de ponctuation au sens large (ponctuation « lexical »). L'agglutination de deux mots peut d'ailleurs être utilisée comme indice d'absence de « frontière ponctuable » entre eux. Il est important de conserver une trace des segmentations particulières, tout en marquant les limites de mots « normalisées » pour faciliter la lecture. Nous indiquons donc les agglutinations par un < + > et les agglutinations, par un < _ >. L'apostrophe est utilisée en cas d'élision.

3. Les marques de **ponctuation** proprement dites doivent être représentées avec le maximum de précision, au niveau allographique donc. Par ailleurs, il peut être utile

Représentation des données primaires : transcription de manuscrits médiévaux

dans certains cas de contraster la ponctuation originale du manuscrit et celle utilisée par un éditeur scientifique qui applique les règles modernes.

4. La reproduction pure et simple de la **disposition du texte sur les lignes** peut s'avérer peu commode en cas de citations courtes commençant au milieu d'une ligne et finissant au milieu d'une autre, il est en revanche nécessaire de garder une trace de cette disposition.

5. En ce qui concerne les **corrections scribales**, on peut se contenter de présenter la forme finale du texte après correction, à moins que la correction n'intervienne dans le fonctionnement même de la ponctuation. Les corrections éditoriales indispensables à la compréhension du texte (ajout de lettres ou de mots omis, suppression de mots répétés) peuvent être introduites avec une mise en évidence typographique (entre crochets).

De façon générale, les crochets sont utilisés pour introduire dans les citations toute sorte d'annotation nécessaire (par exemple, l'indication du type de frontière ponctuable).

Les conventions de présentation des citations du corpus adoptées sont résumées dans le tableau ci-dessous :

Élément du système graphique	Règle de présentation	Exemple
Graphies des mots, abréviations	présentation graphématique	... <i>deus</i> sauir <i>et</i> podir ...
Corrections sribales	forme corrigée	
Corrections éditoriales		
- ajouts	entre crochets	li v[i]ns
- suppressions	barré	jusque ts
Segmentations particulières		
- agglutinations (simples)	symbole < + >	in+quant+ <i>deus</i>
- agglutinations (élision)	apostrophe	qu'il
- déglutinations	symbole < _ >	<i>christi</i> _an
Marques de ponctuation		
- cas général	présentation allographique	dunat · si+saluarai+eo ·
- comparaison original / édition	entre crochets : forme allographique, symbole < > , marque de l'éditeur (normalisée)	dunat [·>] si+saluarai+eo [·>]
Saut de ligne	double barre verticale	<i>commun</i> saluament

Tableau 6 : Conventions de la présentation des citations du corpus.

Voici le texte des *Serments de Strasbourg* mis en forme en mode « citation »¹⁸⁹ (avec la mise en contraste de la ponctuation originale et normalisée) :

- (4.1) **Pro deo**+amur *et*+**pro**+*christi*_an poblo *et* **nostro** *commun* || saluament [·>] d'ist di In auant [·>] in+quant+*deus* || sauir *et* podir me+dunat [·>] si+saluarai+eo [·>] || cist meon fradre karlo [·>] *et* in a iudha [·>] || *et* in cad_huna cosa [·>] si+cum om per dreit son || fradra saluar dift [·>] I_n+o quid il mi+altresi || fazet [·>] E_t+ab+ludher nul plaid *nunqua*·m || prindrai qui+[>] meon uol [·>] cist meon+fradre || karle in+damno+sit [·>].

Dans la pratique, les règles de présentation des transcriptions dans les citations que nous adoptons sont assez proches de celles appliquées par M. B. Parkes (1992) dans les transcriptions des plaques à la fin de son ouvrage (avec une différence notable dans le traitement des segmentations particulières).

¹⁸⁹ Toutes les citations du corpus sont numérotées avec l'indication du numéro de chapitre (la numérotation recommence dans chaque chapitre).

4.4.4. Conclusion

La représentation précise des données primaires a pris une part importante dans notre travail. Nous considérons que cet aspect est important compte tenu des changements profonds que les technologies numériques apportent dans le domaine de l'édition numérique et de la recherche linguistique et philologique.

Une réflexion méthodologique profonde nous paraît indispensable afin de produire des transcriptions et des éditions qui pourront « survivre » et contribuer au progrès de la science dans ce nouveau « monde numérique ». Nous espérons que les principes méthodologiques de la représentation des données primaires des manuscrits médiévaux élaborés au cours de notre travail pourront être développés et appliqués dans bien d'autres projets.

Dans le chapitre suivant, nous exposerons le côté « technique » de notre travail de transcription de manuscrits, ou autrement dit, l'implémentation technologique des principes méthodologiques que nous venons d'exposer.

Chapitre 5

Aspects techniques

Nous avons jusqu'à présent décrit l'aspect des trois principaux niveaux de transcriptions électroniques de manuscrits médiévaux sans entrer dans les détails techniques de leur réalisation. Dans ce chapitre, nous aborderons en premier lieu le problème conceptuel de la granularité d'« éclatement » de ces niveaux dans une édition électronique qui vise à les mettre en rapport et à permettre au l'utilisateur de « naviguer » plus ou moins librement entre eux. En suivant la proposition de Nicolas Mazziotta (2006), nous utiliserons le terme de *diffraction* pour désigner cet éclatement de représentation et de visualisation du texte transcrit de la source primaire.

Nous expliquerons ensuite les raisons de notre choix technologique (XML-TEI¹⁹⁰) pour l'encodage des transcriptions et de leur annotation avant parler des éléments d'extension que nous avons été amené à apporter au jeu de balises « standard » recommandé par la TEI. Nous nous arrêterons ensuite sur notre façon de traiter les « caractères spéciaux » présents dans les manuscrits médiévaux et dans les incunables. A la fin du chapitre, nous présenterons brièvement la « syntaxe compacte » que nous utilisons lors de la saisie et de la correction de nos transcriptions.

190 XML, *eXtensible Markup Language* <<http://www.w3.org/XML/>> ; TEI, *Text Encoding Initiative* <<http://www.tei-c.org>>.

5.1. Niveau de diffraction : mot ou caractère ?

Indépendamment de la solution technologique choisie, il convient avant tout de prendre une décision méthodologique : à quel niveau de hiérarchie structurelle du texte cette diffraction doit-elle avoir lieu ? La diffraction est en effet inévitable dès lors qu'on décide de produire pour un passage de texte plusieurs vues divergentes. En revanche, la finesse de l'alignement des différentes vues et les solutions technologiques concrètes peuvent varier considérablement.

Théoriquement, on peut réaliser trois transcriptions indépendantes de chaque texte ou extrait de texte inclus dans le corpus, mais cela n'a pas beaucoup d'intérêt : en l'absence d'un alignement plus fin entre les différentes transcriptions, il risque d'être difficile d'assurer efficacement leur exploitation et les mises à jour. Par exemple, un alignement mot-à-mot est nécessaire pour pouvoir effectuer des requêtes de type « rechercher toutes les graphies imitatives correspondant à une forme normalisée ». S'il est nécessaire de corriger une faute de frappe dans une transcription, il est désirable de pouvoir le faire simultanément à tous les niveaux de la transcription.

Le choix réel est donc entre la diffraction (ou alignement des différentes vues) au niveau du mot, ou bien au niveau du caractère.

Il existe par ailleurs une solution « **opportuniste** » qui consiste à « éclater » la transcription uniquement pour les segments où les vues normalisée, diplomatique et imitative sont effectivement différentes. En réalité, ces trois vues sont assez souvent identiques : il suffit qu'un mot ne soit pas abrégé et ne contienne ni variantes calligraphiques, ni caractères ambigus. L'avantage de cette solution est qu'elle est économique : ne sont encodées que les informations utiles. Elle peut en revanche rendre la transcription plus difficile à traiter informatiquement. De toute façon, la question du niveau de diffraction reste d'actualité pour les cas où les différentes vues divergent. Si le choix n'est pas constant dans l'ensemble de la transcription (selon les cas, la diffraction a lieu au niveau d'un caractère, d'un groupe de caractères ou d'un mot), la difficulté du traitement et le risque d'erreur deviennent très importants.

Aspects techniques

La diffraction **au niveau du caractère** est la solution qui paraît la plus logique *a priori*. Le caractère est la plus petite unité distinctive de la chaîne graphique signifiante linguistiquement. L'alignement des différentes formes de représentation à ce niveau permet d'atteindre la plus grande souplesse dans la personnalisation de la visualisation et dans la formulation des requêtes croisées.

Le problème principal lié à l'alignement des transcriptions au niveau du caractère est d'ordre technique. Cet alignement est assez difficile à implémenter et nécessite forcément le recours à des logiciels propriétaires pour la saisie et la visualisation des transcriptions. Or, pour assurer la pérennité des données et l'interopérabilité des corpus, il est préférable d'utiliser les formats les plus ouverts possible, indépendants des logiciels propriétaires, voire lisibles « à l'œil nu » en mode « texte brut ».

Il existe également des problèmes d'ordre méthodologique. On peut par exemple se poser la question de savoir si les marques diacritiques doivent être dissociées des caractères de base ou s'il faut traiter la lettre et sa marque diacritique comme un ensemble. La correspondance « un à un » entre les caractères des différents niveaux de représentation n'est pas toujours évidente. Cela concerne en premier lieu les abréviations, où la position de la marque ne correspond pas toujours à celle des lettres « omises »¹⁹¹.

L'alignement des transcriptions **au mot par mot** donne moins de souplesse dans les choix de la visualisation : il est difficile dans ce cas de combiner les éléments des différents niveaux de représentation (par exemple, à la fois afficher les marques d'abréviations et distinguer les caractères *u* et *v*). S'il est nécessaire de corriger un caractère d'un mot sur un niveau de transcription, la mise à jour des deux autres niveaux pourra difficilement être réalisée en mode automatique. Si la diffraction systématique était appliquée dès le moment de la saisie des transcriptions, il serait nécessaire de taper trois fois chaque mot du texte.

En revanche, les transcriptions avec la diffraction au niveau du mot ont l'avantage d'être relativement faciles à lire dans leur forme brute. Les feuilles de style

191 Cf. la section 4.4.1.1 du chapitre précédent et (Mazziotta 2008b).

pour leur visualisation ou préparation pour des traitements ultérieurs sont faciles à écrire et à modifier.

Du point de vue méthodologique, la diffraction au niveau des mots peut être justifiée par la présence d'éléments de logographie dans le système graphique du français, encore plus forte dans l'écriture manuscrite médiévale qu'en français moderne grâce notamment à la fréquence élevée des abréviations.

En cas de « segmentations particulières » (où les limites des mots graphiques du manuscrit ne correspondent pas), on peut baliser les unités lexicales conformément à la pratique normalisée tout en assignant un trait « agglutiné » à chaque mot qui n'est pas séparé du mot suivant par un espace blanc. Cela permettrait d'une part de visualiser dans la transcription imitative la segmentation réelle du manuscrit et, d'autre part, d'effectuer des requêtes ciblées sur les occurrences comportant des segmentations particulières.

Compte tenu des avantages et des défauts des deux options de diffraction, nous avons fait le choix de la **diffraction systématique au niveau des mots** pour la version définitive de nos transcriptions. En revanche, pour la saisie de celles-ci, nous avons adopté un « style de saisie » (ou « syntaxe compacte ») avec un marquage spécial réservé aux caractères ou groupes de caractères pour lesquels il y a une divergence effective des trois niveaux de transcription.

5.2. Choix technologique : XML-TEI

Depuis 2001, l'équipe de la Base de Français Médiéval a fait le choix d'utiliser le métalangage XML (*eXtensible Markup Language*) et de suivre les recommandations de la TEI (*Text Encoding Initiative*). Les raisons et les enjeux de cette décision sont présentés dans (Heiden & Guillot 2003), certains aspects méthodologiques d'utilisation de la norme XML dans les corpus de français médiéval sont par ailleurs discutés dans (Heiden & Lavrentiev 2004). Nous n'y reviendrons pas ici, mais présenterons très brièvement le principe de codage (ou plutôt de « balisage ») de textes en XML et évoquerons les objectifs et le mode de fonctionnement du consortium TEI.

Aspects techniques

Le standard XML, développé par le consortium W3C¹⁹² depuis 1996 sur la base de l'ancienne norme SGML (ISO 8879:1986), semble à ce jour le mieux adapté aux besoins d'éditions électroniques et de corpus textuels indépendants d'une plateforme logicielle propriétaire particulière. Le nombre de projets utilisant la norme XML et respectant les recommandations de la TEI¹⁹³, la « transparence » du code, la facilité que ce format offre pour le traitement et l'échange des données sont des gages de pérennité du corpus.

Le langage XML est décrit en détail dans un grand nombre d'ouvrages imprimés et de ressources accessibles en ligne¹⁹⁴, les recommandations complètes de la TEI sont également publiées sur Internet¹⁹⁵. Nous n'évoquerons donc ici que quelques principes de base indispensables pour la compréhension des exemples du corpus que nous citerons plus bas.

Le principe à la base de la norme XML (et de beaucoup d'autres langages de balisage) est de séparer le texte (caractères imprimables et espaces blancs entre les mots, ou CDATA, en termes informatiques) des annotations (ou méta-données) de toutes sortes, qui sont placées dans des *balises* (*tags*, en anglais). Toutes les balises XML sont délimitées par des chevrons (par exemple, `<p>` pour la balise de paragraphe la plus répandue). Les balises peuvent « entourer » un segment de texte annoté ou d'autres balises. Dans ce cas, on distingue des balises de début et de fin : par exemple, `<p>` et `</p>` pour le début et la fin d'un paragraphe. Une balise peut également être placée toute seule, comme borne kilométrique sur une route, à un certain endroit dans le texte. Ces balises « vides », qu'on appelle *milestone* en anglais, ont une forme particulière : une barre oblique avant le chevron fermant. Par exemple, `<pb/>` pour un saut de page.

L'ensemble formé par les balises de début et de fin et le segment qu'elles entourent (ou bien la balise « vide » toute seule) est appelé un *élément* XML. Un élément peut être doté d'un nombre en principe illimité d'*attributs* destinés à porter des informations supplémentaires (par exemple, le numéro pour un saut de page).

192 <http://www.w3c.org>.

193 Ce choix a notamment été fait par le Consortium pour les corpus de français médiéval (CCFM), <http://ccfm.ens-lsh.fr>. A la date du 2 mai 2009, plus de 120 projets utilisant la TEI sont référencés sur le site web <http://www.tei-c.org>.

194 La spécification officielle la plus récente est accessible sur <http://www.w3.org/xml/>.

195 <http://www.tei-c.org/Guidelines/>.

La norme XML définit la morphologie des balises et pose un certain nombre de contraintes (par exemple, l'interdiction du chevauchement d'éléments), mais les noms des éléments (« vocabulaire ») et leurs propriétés (« grammaire ») peuvent être définis en fonction des besoins d'un projet donné.

Le consortium TEI vise, depuis sa fondation en 1987, à proposer un standard de balisage (« vocabulaire » et « grammaire ») XML¹⁹⁶ utilisable pour l'édition et la description de tous types de ressources textuelles (textes écrits, transcriptions de l'oral, dictionnaires, corpus textuels et autres). L'enjeu de ce travail est évident : la pérennité des données encodées et l'interopérabilité des outils.

A ce jour, la dernière version des recommandations de la TEI¹⁹⁷ comporte 23 chapitres et définit près de 500 éléments XML répartis dans 20 modules. Un module spécifique est consacré aux transcriptions des sources primaires, un autre module permet de gérer l'encodage des caractères « non-standards », un autre encore est destiné à la description des manuscrits.

En réalité, il est extrêmement difficile d'élaborer des normes de balisage de textes qui satisfassent tous les projets d'édition et de recherche possibles et imaginables. Les balises nécessaires à l'encodage fin des transcriptions de sources primaires n'ont été introduites que dans la dernière version des recommandations de la TEI et tous les problèmes ne sont pas encore réglés.

Dans le cadre d'un projet précis, il est souvent nécessaire de trouver un compromis entre les intérêts immédiats du projet et le respect des normes communautaires. Nous avons toujours cherché à suivre les recommandations de la TEI, mais dans certains cas celles-ci nous ont paru insuffisantes pour les intérêts de notre recherche. Par exemple, la TEI ne propose pas de balise spécifique pour les marques de ponctuation, elle ne permet pas non plus d'implémenter la diffraction des transcriptions au niveau des mots.

La TEI propose par ailleurs elle-même des mécanismes de « personnalisation » (choix des options permettant de réduire la plage des possibilités d'encodage) et d'« extension » (ajout de nouveaux éléments et/ou de nouveaux attributs aux éléments

196 SGML jusqu'en 2002, car XML n'existait pas aux débuts de la TEI (1987).

197 <<http://www.tei-c.org/Guidelines/P5/>>.

Aspects techniques

existants). Ce sont ces mécanismes que nous avons utilisés pour adapter les schémas de balisage de la TEI aux besoins de notre projet.

5.3. Extension de la TEI

La personnalisation et l'extension du schéma d'encodage recommandé par la TEI que nous avons effectuées pour nos transcriptions s'appuient, bien entendu, sur l'expérience d'autres projets ayant réalisé des transcriptions fines de manuscrits, en particulier, des projets *Charrette* et *Menota* (déjà mentionnés).

Le projet *Charrette*, initié en 1989 par Karl D. Uitti de l'Université de Princeton, est la première édition savante en ligne comprenant des transcriptions allographiques dont l'encodage se basait sur les recommandations de la TEI. Nous avons participé à ce projet en 2001/2002 en mettant au point une conversion des anciennes transcriptions SGML vers le format XML en concertation avec K.D. Uitti, R. Alvarado et S.-J. Murray. Nous continuons par ailleurs toujours la collaboration avec l'équipe du projet en qualité d'éditeur associé.

Inspiré du principe fondamental d'XML qui consiste en la dissociation de la représentation des données de la forme de présentation, nous avons conçu trois formes de présentation pour les transcriptions du projet *Charrette* destinées à différentes catégories de lecteurs : « manuscrit », « riche » et « édition » (Lavrentiev 2005). Dans nos travaux ultérieurs, nous avons cherché à rapprocher les modes de présentation des traditions d'éditions existantes.

Les pratiques adoptées dans le projet *Menota* et documentées dans *Menota Handbook* (Haugen éd. 2008) nous ont semblé bien adaptées à nos besoins. La diffraction systématique de trois niveaux de représentation du texte au niveau des mots pratiquée dans les transcriptions *Menota* correspondait à notre choix méthodologique (cf. la section 5.1. ci-dessus). Les outils développés dans le cadre du projet *Menota*, et en particulier la spécification de la personnalisation de la TEI (*menota.odd*), ainsi que les normes de codage des « caractères spéciaux » au niveau imitatif (cf. la section 5.4. ci-dessous) sont extrêmement précieux pour l'optimisation de la saisie et du traitement des transcriptions.

Nous avons donc décidé d'aligner dans la mesure du possible nos pratiques sur les recommandations de *Menota Handbook*, en documentant tous les écarts et les extensions qui nous ont paru nécessaires.

Le balisage d'éléments de structure textuelle supérieurs à un mot (grandes divisions, paragraphes, pages, etc.) a pu se faire sans s'écarter des recommandations de la TEI.

En revanche, l'introduction de la diffraction systématique des transcriptions au niveau des mots, qui n'est pas prévue par les recommandations de la TEI, a nécessité l'extension du schéma d'encodage de cette dernière. Le principal changement par rapport au jeu de balises « standard » de la TEI consiste donc dans l'ajout d'éléments correspondants aux trois niveaux de transcription : `<me: norm>`¹⁹⁸ pour le niveau normalisé (orthographique), `<me: dipl>` pour le niveau diplomatique (graphématique) et `<me: facts>` pour le niveau imitatif (allographique). Tous ces éléments peuvent être « emballés » par un élément TEI `<choice>` destiné à baliser des groupes d'éléments présentant un même segment de texte sous formes alternatives. Le `<choice>` est à son tour « emballé » dans un élément `<w>` correspondant à une unité-mot :

```

<w>
  <choice>
    <me: norm>ſçavoyr</me: norm>
    <me: dipl>ſcauoyr</me: dipl>
    <me: facts>fcauoyr</me: facts>
  </choice>
</w>

```

Fig. 25 : Exemple de balisage d'un mot « diffracté ».

Un autre ajout important est l'élément `<me: punct>` utilisé pour baliser les marques de ponctuation. Selon les recLa TEI propose en effet d'utiliser l'élément `<c>` (caractère) en cas de besoin pour baliser les marques de ponctuation, or une marque de ponctuation est un signe linguistique de nature différente des caractères composant la

198 Le préfixe d'« espace de nommage » (*namespace*) `me:` rajouté au nom d'une balise permet d'indiquer que celle-ci appartient à un schéma différent que celui utilisé par défaut (TEI). Ce mécanisme d'identification de jeux de balises est prévu par la norme XML et son utilisation est requise pour les extensions reconnues « conformes » à la TEI.

Aspects techniques

graphie d'un mot. Du point de vue de la diffraction des niveaux de transcription, les propriétés d'une marque de ponctuation sont plus proches de celles d'un mot que de celles d'un caractère. Nous jugeons donc l'introduction d'une balise spécialisée totalement justifiée dans ce cas.

Nous avons exposé nos arguments à la réunion annuelle de la TEI en novembre 2008 (Lavrentiev 2008), et le conseil de la TEI, lors de sa réunion à Lyon le 3 avril 2009, a pris la décision d'accepter notre proposition d'inclure cette nouvelle balise lors de la prochaine mise à jour des *Recommandations*.

En cas d'agglutination de plusieurs mots, *Menota Handbook* propose d'emballer la séquence agglutinée (plusieurs éléments `<w>`) dans un élément `<seg>` portant un attribut `type="nb"`¹⁹⁹ :

```
<seg type="nb">
  <w>en</w>
  <w>egipte</w>
</seg>
```

Fig. 26 : Balisage d'une agglutination selon les recommandations du projet Menota.

Cette solution est conforme aux recommandations de la TEI, mais elle amène à la complication de la structure du document et ne permet pas de gérer facilement les cas de doutes (en présence notamment des « petits blancs »).

Nous avons préféré associer l'information sur la segmentation à la balise même de l'unité-mot en y ajoutant des attributs `bfm:aggl` et `bfm:agglCert` :

```
<w bfm:aggl="simple"
bfm:agglCert="no">en</w>
<w>egipte</w>
```

Fig. 27 : Balisage d'une agglutination dans notre corpus.

Cette solution permet d'encoder sans difficulté n'importe quelle combinaison de mots agglutinés avec ou sans élision (l'attribut `bfm:aggl` pouvant prendre les valeurs

¹⁹⁹ Pour faciliter la lecture des exemples, nous supprimons la diffraction des niveaux de transcription dans les cas où celle-ci n'a pas d'importance dans le contexte de la citation.

"**simple**" ou "**elision**"), avec ou sans un petit blanc (l'attribut **bfm:agglCert** prenant la valeur "**no**" en cas de présence d'un petit blanc).

En cas de déglutination, *Menota Handbook* recommande d'insérer un simple espace blanc à l'intérieur de l'élément **<w>**. Du point de vue de l'analyse linguistique, ce blanc interne à un mot a un statut différent d'un blanc « normal ». Nous avons par conséquent choisi d'introduire une balise spéciale **<bfm:sb/>** pour l'encoder. Son attribut **type** permet de préciser s'il s'agit d'un petit blanc. Cette balise sera utilisée uniquement aux niveaux de représentation allographique et graphématique.

```
<me:fac>por<bfm:sb/>tees</me:fac>
```

Fig. 28 : Balisage d'une déglutination dans notre corpus.

5.4. « Caractères spéciaux » dans les manuscrits médiévaux

De façon générale, l'encodage des transcriptions aux niveaux normalisé et diplomatique ne pose pas de problème particulier : les caractères utilisés sont les mêmes que dans les textes modernes, ils sont donc pris en charge par les normes en vigueur. Pour annoter les éléments nécessitant un traitement particulier (comme par exemple les caractères « reconstitués » des abréviations au niveau diplomatique), les balises XML-TEI conventionnelles sont largement suffisantes.

Le niveau « imitatif » est plus problématique, car l'encodage des « caractères spéciaux » médiévaux (tels que les variantes calligraphiques, les marques d'abréviation et de ponctuation, etc.) n'est pas pris en charge par la norme actuelle la plus complète (Unicode, version 5.0)²⁰⁰. Il existe cependant une initiative internationale MUFI²⁰¹, qui vise à établir la liste exhaustive des caractères utilisés dans les manuscrits médiévaux

200 Il y a toutefois quelques exceptions. Le *s* « long » **<ſ>**, par exemple, fait partie du standard Unicode.

201 *Medieval Unicode Font Initiative*, <http://www.mufl.info/>. L'activité de MUFI est actuellement coordonnée par Odd Einar Haugen, de l'Université de Bergen. Cette initiative s'est développée dans le cadre du projet Menota afin de répondre aux besoins d'encodage d'éditions « imitatives » comportant des caractères médiévaux qui ne sont pas pris en charge par les standards en vigueur.

Aspects techniques

d'Europe occidentale et à la soumettre au consortium Unicode afin d'intégrer dans le standard les caractères manquants. A ce jour, les tables de MUFI (version 2.0) incluent 1326 caractères, dont près de 750 dans la « zone privée » (caractères absents dans la norme actuelle). La proposition MUFI inclut pour chaque caractère son dessin, son code hexadécimal, sa description selon les règles Unicode et un nom d'« entité SGML » permettant d'éviter des problèmes d'affichage et de conversion à l'aide d'outils ne prenant pas en charge les caractères Unicode.

Des polices de caractères sont développées par des partenaires de la MUFI. Pour visualiser les transcriptions imitatives par des navigateurs Web à l'écran et à l'impression, ainsi que dans les exemples cités dans la présente thèse, nous avons profité en particulier de la police « Andron Scriptor Web » développée par Andreas Stötzner à Leipzig et mise à disposition gratuitement dans le cadre du projet MUFI.

A notre avis, certaines propositions de MUFI vont au-delà des distinctions de variantes allographiques potentiellement significatives du point de vue linguistique et atteignent ainsi un niveau de finesse où l'intérêt et la possibilité même de standardisation ne sont pas évidents. Cela ne remet pas pour autant en question l'intérêt et l'utilité de l'initiative MUFI en général. Le simple fait de décrire formellement et de répertorier tous les caractères distincts utilisés dans les manuscrits occidentaux au Moyen Âge serait une avancée scientifique considérable.

Nous avons donc décidé d'utiliser les codes MUFI pour encoder les caractères « spéciaux » médiévaux au niveau imitatif de nos transcriptions : les variantes de lettres, les ligatures, les marques d'abréviation et de ponctuation médiévales. A quelques exceptions près, nous avons trouvé dans les tables MUFI tous les caractères nécessaires à la transcription allographique des manuscrits et des incunables de notre corpus.

Parmi les caractères manquants, à notre avis, dans les tables MUFI, nous pouvons citer le « pied-de-mouche » (Fig. 29, *cf.* la section 2.7.1.) et la marque de coupure de mot en fin de ligne en forme d'une double barre oblique < ≧ > (Fig. 30). Ces deux caractères faisant partie de l'inventaire des marques de ponctuation, nous nous arrêterons plus en détail sur le problème de leur codage.

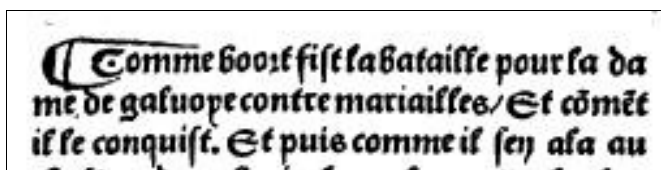


Fig. 29 : Un pied-de-mouche dans la *Mort au roy artus* (1488)

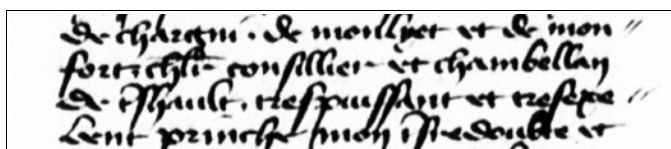


Fig. 30 : Doubles barres obliques dans le ms. Paris, BnF fr. 2862 (XVe s.)

Le deuxième de ces caractères est en effet déjà présent dans Unicode sous le nom de « double trait d'union oblique » (code 2E17), dont l'usage est signalé en linguistique de langues anciennes du Proche Orient. Il suffit donc d'intégrer ce caractère dans les tables MUFI, afin qu'il soit pris en charge par les polices compatible MUFI.

En ce qui concerne le pied-de-mouche, MUFI inclut déjà deux caractères dont la forme et la fonction sont assez proches de celui-ci, < ¶ > (code 00B6) et < ꝛ > (code F1E1). Le premier d'entre eux est utilisé pour visualiser les fins de paragraphes dans plusieurs logiciels de traitement de texte. Il est en effet le résultat de la transformation progressive du pied-de-mouche dans la typographie. Les deux tracés représentent certainement le même graphème, mais la transformation est suffisamment importante pour les considérer comme des allographes : le pied de mouche médiéval garde un lien fort avec le caractère < C > (du mot latin *capitulum*) et, en général, avec les lettres initiales : réalisé, comme ces dernières, en couleur par un rubricateur, il marquait le début et non la fin d'une unité textuelle. La barre verticale du pied de mouche provient d'ailleurs des barres en couleur utilisées dans certains manuscrits pour marquer toutes les majuscules (Parkes 1992, p. 43).

Le crochet adlinéaire < ꝛ > placé habituellement dans les marges provient à son tour du caractère grec < Γ > (du mot *graphos*). Sa fonction est similaire à celle du pied de mouche et, selon les écritures, son tracé peut se rapprocher de celui du pied-de-mouche. Il s'agit donc d'un autre allographe du même graphème.

Aspects techniques

Le pied-de-mouche médiéval en forme d'un grand < C > avec une barre verticale interne mérite donc à notre avis de disposer de son propre code, distinct de la marque de paragraphe typographique et du crochet adlinéaire.

Nous avons soumis à la MUFI une proposition pour inclure les deux caractères (le pied-de-mouche et la double barre oblique) dans la prochaine version de ses tables de codes. Cette proposition a été prise en compte, notre argumentaire pour l'intégration de ces deux caractères est publié sur la page web de propositions de nouveaux caractères MUFI²⁰². En attendant, nous utilisons le code 00B6 (entité standard `¶`) pour le pied de mouche et le code 2E17 (entité proposée `&dbloblhyphen;`) pour la double barre oblique.

D'autres cas (très peu nombreux) dans lesquels il nous semble utile d'ajouter les caractères aux tables MUFI sont présentés dans la documentation technique relative à nos protocoles d'encodage (annexe II).

5.5. Syntaxe « compacte » : saisie et corrections

Parmi les inconvénients de la diffraction systématique des transcriptions, nous avons mentionné le fait que chaque mot du texte est répété trois fois, ce qui est loin d'être pratique au moment de la saisie.

Pour optimiser la procédure de saisie, nous avons mis au point un format compact qui peut être transformé automatiquement en format diffracté complet.

Le format compact utilise certaines balises XML, mais aussi des caractères spéciaux (ou « raccourcis typographiques ») pour encoder une partie des données pertinentes. Notre transcription compacte est un document XML bien formé, mais il contient des annotations non XML qui ne peuvent pas être directement prises en charge par les outils standard de traitement d'XML (schémas de validation, feuilles de style, éditeurs XML, etc.). La conversion au format « 100% XML » est donc obligatoire à un moment donné.

Des solutions similaires sont assez répandues dans le monde de l'édition électronique. Pour ne citer qu'un exemple, l'outil d'édition collaborative en ligne

202 <http://www.mufi.info/pipeline/>.

MediaWiki²⁰³ utilise des raccourcis typographiques pour la saisie des hyperliens, des variantes de polices (gras et italique), des tables, etc.²⁰⁴

Le principe général de notre format compact consiste à saisir pour chaque élément « diffractable » de la chaîne graphique (caractère ou groupe de caractères) la forme de représentation la plus riche en information : les allographes du niveau imitatif d'une part, mais aussi les diacritiques désambiguïsants du niveau normalisé.

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
p&rrrot;i&slong;e ²⁰⁵	pɹɪfe	prise	prise
ob&slong;tiné	obftine	obstine	obstiné

Des caractères « modificateurs » sont rajoutés en cas de transformation d'un caractère au niveau normalisé. Par exemple, l'ajout d'un astérisque < * > devant un *u* signifie que celui-ci doit être transformé en *v* au niveau normalisé. L'ajout d'un dièse < # > devant une minuscule signale que celle-ci doit être normalisée en majuscule et *vice versa*.

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
roy de #na*uarre	roy de nauarre	roy de nauarre	roy de Navarre

L'apostrophe signale une agglutination avec élision (comme dans l'orthographe française contemporaine). Une agglutination simple (sans élision) est notée par un < + >, une agglutination incertaine (petit blanc) est signalée par < +? >.

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
d'outre+ mer	doutremer	d'outre mer	d'outre mer

203 <<http://www.mediawiki.org/>>.

204 Cf. <<http://en.wikipedia.org/wiki/Wikipedia:Cheatsheet>>.

205 Les exemples cités dans la présente section proviennent des transcriptions **joinville13568** et **froissart869**.

Aspects techniques

Les déglutinations sont représentées par une marque de soulignement < _ > éventuellement suivie d'un point d'interrogation en cas de doute ou de présence d'une « petit blanc ».

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
bien_viengnans	bien viengnans	bienviengnans	bienviengnans

Les grandes initiales (letrines) sont marquées par de doubles accolades ; les informations sur la taille, la couleur et une décoration éventuelle y figurent séparées par des deux-points :

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
{{P:4:4:blue:doré}}#Our	P _{4/4} Our	P Our	Pour

Les limites d'une abréviation sont marquées par des doubles parenthèses. Pour les abréviations fréquentes et régulières (résolution « automatique » à partir de la marque ou d'association de la marque à quelques caractères), seule la représentation imitative est notée. Dans ce cas les représentations normalisée et diplomatique peuvent être générées automatiquement à partir d'une liste de correspondance préétablie. Dans les autres cas, la forme abrégée et la résolution sont notées côte à côte, séparées par une marque < _ >. Les lettres restituées dans la résolution sont placées entre crochets :

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
qu((a&bar;))t	quāt	quant	quant
#en((gl&dblbar;z_gl[oi]z))	englz	engloiz	Engloiz

Les marques de ponctuation sont balisées <c type="punct_strong"> ou <c type="punct_weak">, les trois niveaux de représentation sont systématiquement notés, le caractère < % > est utilisé comme séparateur :

Transcription compacte	Présentation imitative	Présentation diplomatique	Présentation normalisée
<punct>.%,%&virgmin;</punct>	/	,	.
<punct>,%%</punct>			,

La transformation des transcriptions compactes au format diffracté complet s'effectue à l'aide d'un script Perl que nous avons mis au point à cet effet (cf. le dossier <Corpus/scripts> du CD-ROM accompagnant la présente thèse).

5.6. Conclusion

Après avoir constaté l'existence et l'importance du problème de l'authenticité des données textuelles que les linguistes utilisent pour l'étude de langues anciennes et d'états historiques de langues vivantes, nous avons analysé la diversité des pratiques de représentation du texte en prenant l'exemple d'un certain nombre d'éditions du plus ancien monument de la langue française, les *Serments de Strasbourg*. Certains éditeurs vont très loin dans la représentation des moindres détails graphiques de leurs sources, d'autres préfèrent normaliser une partie des données pour se conformer aux règles typographiques et aux habitudes de lecture modernes. Chaque solution répond aux intérêts de lecteurs différents. L'édition électronique permet de concilier des intérêts souvent contradictoires en proposant pour un même document plusieurs « vues » (ou niveaux de représentation des données).

Des prototypes très intéressants de ce genre d'éditions existent déjà, mais les pratiques sont loin de se stabiliser. Or la pérennité et l'interopérabilité des données numériques dépend fortement de l'application des normes des normes internationales et de la transparence des choix d'encodage. Il est également important que ces choix soient appuyés par une analyse scientifique des données traitées.

En ce qui concerne les manuscrits médiévaux, nous distinguons trois niveaux de représentation des données correspondant d'une part à des niveaux d'analyse linguistique et, d'autre part, à une certaine tradition éditoriale. Un certain parallélisme

Aspects techniques

peut être établi entre la représentation des données primaires dans les transcriptions de manuscrits et d'enregistrements de la langue orale.

Au niveau *normalisé* (ou orthographique) les données prennent une forme régularisée (élimination de variantes peu significatives, désambiguïsation de certaines marques, résolution tacite des abréviations, application des normes modernes dans la segmentation des mots et dans la ponctuation). Cela facilite, pour un lecteur moderne, la compréhension du texte. Une telle représentation est pratiquée dans la grande majorité d'éditions critiques de textes français médiévaux.

Au niveau *graphématique* (*cf.* les transcriptions phonématiques de l'oral), la transcription représente les graphèmes de la source primaire : les variantes positionnelles et calligraphiques sont neutralisées, tandis que les ambiguïtés propres au système graphique des manuscrits médiévaux ne sont pas levées (telles que la non distinction de *u* voyelle et consonne ou l'absence d'une marque diacritique sur un *e* final accentué). Le même principe est appliqué aux marques de ponctuation : les ponctuations faibles du manuscrit sont généralement représentées par une virgule et les ponctuations fortes, par un point. Ce choix est fondé sur notre modèle du fonctionnement de la ponctuation médiévale. La segmentation des unités-mots reste normalisée à ce niveau.

La représentation des manuscrits dans certaines éditions diplomatiques est assez proche du niveau graphématique, même si le plus souvent les éditeurs ont tendance soit à normaliser davantage leurs transcriptions, soit à distinguer certaines variantes de graphèmes.

Au niveau *allographique* (*cf.* les transcriptions phonétiques avec la distinction des allophones), la transcription représente toutes les variantes des graphèmes (ou allographe) nettement identifiables. Ces variantes peuvent potentiellement assumer une charge fonctionnelle : ainsi les variantes utilisées uniquement en position initiale ou finale dans un mot peuvent servir de marque de séparation des mots. Certains éditeurs ont tenté d'imiter dans leurs transcriptions les variantes de caractères, les marques d'abréviations et de ponctuation médiévales, mais faute d'arguments scientifiques pour appuyer leurs choix, ils étaient souvent accusés d'inconsistance et même de paresse. Les transcriptions allographiques sont pourtant indispensables pour des recherches sur des systèmes graphiques et peuvent être utiles aux paléographes. Si celles-ci ne remplacent

pas, mais complètent les transcriptions normalisée et graphématique, les principaux arguments contre les éditions « imitatives » sont réfutables. C'est donc le principe de représentation multi-facettes que nous avons adopté pour notre corpus de transcriptions de manuscrits.

Sur le plan technique, après avoir analysé les avantages et les défauts de toutes les options possibles, nous avons choisi d'opérer systématiquement la diffraction des trois niveaux de représentation au niveau des mots. Pour assurer la pérennité et l'interopérabilité des données numériques, nous avons suivi dans la mesure du possible les recommandations de la TEI. Dans les cas où les solutions proposées par la TEI ont été insuffisantes pour les besoins de notre projet, nous avons adopté les normes élaborées dans le cadre d'un projet similaire (Menota) ou, le cas échéant, procédé à l'extension du schéma de balisage TEI en respectant le protocole prévu à cet effet.

Sur cette conclusion, nous terminons la partie de notre mémoire consacrée à la présentation du cadre historique, théorique, méthodologique et technique de notre recherche et passons à l'analyse des données concernant l'usage de la ponctuation dans les textes de notre corpus.

Partie II

Analyse des données texte par texte et synthèse

Chapitre 6

Manuscrits antérieurs à 1400

6.1. Paris, BnF, lat. 9768 (*Serments de Strasbourg*)

Les *Serments de Strasbourg* sont considérés comme le premier document en langue française. Il s'agit de quelques phrases prononcées en 842 par deux des petits-fils de Charlemagne, Louis le Pieux et Charles le Chauve, devant leurs armées, pour s'engager à ne pas s'allier séparément à leur frère, Lothaire.

Cet engagement solennel devait être compris par les armées respectives, ce qui a imposé l'usage de langues vernaculaires (français et allemand) et non du latin. Le texte en français et en allemand est cité dans la chronique latine de Nithard, conservée à ce jour dans un seul manuscrit, BnF, lat. 9768. Dans notre corpus, ce texte porte l'identifiant **strasb9768**, que nous utiliserons désormais.

Le manuscrit est daté de fin du X^e – début du XI^e siècle (d'après la majorité des chercheurs s'étant prononcés sur le sujet). Il a probablement été écrit à Soissons.

Ce petit texte de haute importance symbolique a été publié maintes fois, et peut ainsi servir d'excellent exemple de différentes traditions éditoriales.

Même si **strasb9768** se situe en dehors des limites temporelles adoptées pour notre corpus, nous avons choisi de l'y intégrer pour avoir un point de comparaison

« extérieur », une préhistoire de la ponctuation française, en quelque sorte²⁰⁶. Comparaison d'autant plus intéressante qu'il s'agit d'un texte en prose, tandis que la grande majorité des premiers textes français étaient en vers.

La position particulière que ce texte occupe dans le corpus (et dans l'histoire du français !) fait que ses caractéristiques typologiques n'ont pas beaucoup d'importance : nous ne disposons pas de textes d'autres types de la même période pour pouvoir isoler l'impact de tel ou tel trait. Notons néanmoins que le texte appartient au domaine juridique, et le terme de *serment* peut être retenu en tant que spécification du genre. En même temps, **strasb9768**²⁰⁷ n'est qu'une citation insérée dans une chronique latine, c'est-à-dire un texte du domaine historique. Dans la grille de types de séquence textuelles de J.-M. Adam, les *Serments* peuvent être classés sous la catégorie « description », car il s'agit ici de présenter les engagements de chaque partie concluant le pacte. Le texte est certainement trop court pour pouvoir identifier des traits dialectaux.

Les deux parties en français des *Serments de Strasbourg* ne comptent que 115 mots et 16 occurrences de marques de ponctuation. La seule marque utilisée est le point médian. Le taux de ponctuation²⁰⁸ y atteint près de 14%, une valeur largement supérieure à la moyenne de notre corpus. Dans 5 cas le point médian dans le manuscrit est suivi d'une majuscule, ce qui donne un taux de ponctuation forte de 4,3%.

En ponctuant ce texte en conformité avec les règles modernes (comme l'a fait, par exemple Lauer), on utilise 18 marques, la différence quantitative avec la ponctuation originale est donc très faible.

Dans 11 cas (8 occurrences de ponctuation faible et 3 occurrences de ponctuation forte), la ponctuation du manuscrit et la ponctuation modernisée sont identiques²⁰⁹. Dans 10 cas, il s'agit de limites de propositions (indépendantes ou subordonnées) et dans un cas, d'un complément circonstanciel détaché. Les « ajouts » dans la ponctuation modernisée se limitent au marquage d'un groupe nominal employé

206 De la même façon, le traité de ponctuation d'Étienne Dolet (**dolet1540**) constitue un point de comparaison « tardif ».

207 Cf. le sigle DEAF : **Serments** (nous enlèverons la dernière lettre du sigle DEAF chaque fois qu'il s'agira d'une œuvre médiévale indépendamment de ses éditions modernes).

208 Nombre moyen de marques de ponctuation pour un extrait de 100 mots du texte.

209 Voir la section 2.8. pour les critères de comparaison.

comme modificateur modal détaché (6.1), d'une apposition du sujet (6.2) et du début d'une proposition principale après une subordonnée (6.3)²¹⁰ :

- (6.1) nul plaid *nunqua·m* || prindrai qui+> meon uol > cist meon+fradre || karle in+damno+sit (*strasb9768, 13b:12*) ;
- (6.2) Et karlus > meos sendra > || de+suo part+non los+tanit (*strasb9768, 13b:29*) ;
- (6.3) Si+lodhuuigs || sagrament , que son fradre karlo || iurat > conseruat (*strasb9768, 13b:29*).

Dans trois cas, l'emploi d'une marque de ponctuation dans le manuscrit ne correspond pas aux règles modernes. Toutes ces occurrences se succèdent immédiatement sur deux lignes du manuscrit :

- (6.4) si+saluarai+eo > || cist meon fradre karlo > et in adiudha > || et in cadhuna cosa (*strasb9768, 13b:7-9*).

Dans le premier cas, un point sépare le sujet pronominal postposé du complément direct d'un verbe, dans les deux autres, la ponctuation est utilisée au début de compléments circonstanciels coordonnées, avec une répétition de la conjonction *et*. Une autre analyse est également possible : les deux premiers points dans cet exemple servent à encadrer le syntagme nominal complexe *cist meon fradre karlo*. L'usage de la ponctuation peut dans tous les cas s'expliquer par le besoin de séparer des groupes syntaxiques distincts à l'intérieur d'une proposition étendue.

Nous pouvons donc conclure que la ponctuation du premier texte français est bien organisée et étonnamment proche du système moderne.

6.2. Lyon, BM, Palais des Arts 77 (*Lancelot dou Lac et la Queste del saint Graal*)

Ce manuscrit lyonnais est représenté dans notre corpus par deux extraits d'œuvres différentes appartenant au « cycle du Lancelot - Graal » : cinq grands romans en prose composés par un auteur inconnu vers 1220-1240. Ces romans racontent l'intégralité de l'histoire de la Table Ronde et du roi Arthur, de sa conception à sa mort.

210 Nous rappelons que toutes les citations du corpus sont présentées selon les conventions décrites dans 4.4.3.2.

Ce cycle de romans en prose a été très populaire au Moyen Âge en France, ce dont témoigne le grand nombre de manuscrits et d'incunables parvenus jusqu'à nos jours (une cinquantaine de manuscrits pour la *Queste*, par exemple). La caractérisation typologique ne pose pas de problème pour ces textes : il s'agit d'œuvres littéraires appartenant au genre du roman. Le texte ne comporte pas de traits dialectaux remarquables.

Le manuscrit lyonnais (connu dans la tradition éditoriale sous les sigles *K* pour la *Queste du Graal* et *Ly* pour *Lancelot*) n'est daté que très approximativement du XIII^e siècle. Il est écrit sur parchemin et comporte 264 feuillets de 322×222 mm. Il est considéré comme un « bon » manuscrit et a servi de base à l'édition de la *Queste* par A. Pauphilet (1923)²¹¹, ainsi que pour l'édition électronique en cours de préparation sous la direction de Ch. Marchello-Nizia et à laquelle nous collaborons. Malheureusement, la version du *Lancelot* présentée par ce manuscrit est incomplète, et il manque également quelques fragments au début et à la fin de la *Queste*.

Le fragment du *Lancelot* qui fait partie de notre corpus (folios 8v – 9r, 565 mots) a été transcrit en 2001 et entièrement révisé en 2007. Cette transcription porte l'identifiant **lancelot77**²¹² que nous utiliserons désormais.

Notre transcription d'un fragment de la *Queste* fait partie du projet éditorial de Ch. Marchello-Nizia qui vise à proposer une version de texte « très fidèle » au manuscrit et en même temps dotée d'un appareil critique et d'outils d'analyse linguistique, accessible aux différentes catégories de lecteurs et utilisant pleinement les possibilités offertes par les technologies numériques.

L'ensemble du texte a été transcrit par Ch. Marchello-Nizia selon des règles précises permettant à la fois de préserver des informations précieuses concernant la langue du scribe (y compris la ponctuation), et de faciliter la lecture du texte en appliquant avec précision certaines règles de normalisation adoptées dans les éditions « critiques ». Les principes éditoriaux de Ch. Marchello-Nizia sont exposés dans l'introduction à l'édition électronique en ligne (Marchello-Nizia 2008). Il s'agit en quelque sorte d'un nouveau compromis (par rapport à la tradition éditoriale dominante) entre la représentation fidèle de la source et l'accessibilité pour le lecteur.

211 La motivation du choix du manuscrit est détaillée dans (Pauphilet 1921).

212 Cf. l'identifiant DEAF de ce roman : **SGraalIVLanc**.

Manuscrits antérieurs à 1400

Pour notre part, nous avons appliqué à un fragment de 2000 mots environ (folios 160v – 164r) notre principe de transcription « multi-couches », où le texte établi par Ch. Marchello-Nizia a constitué le niveau « normalisé ». Cette transcription est entrée dans notre corpus sous l'identifiant **qgraal77**²¹³.

Nous avons donc pour un même manuscrit deux échantillons tirés d'œuvres différentes appartenant toutefois à un même cycle. Cela nous a permis de vérifier si les pratiques de ponctuation sont stables dans des parties différentes du codex.

Le taux général de ponctuation paraît effectivement constant dans les deux fragments (8,7% dans **lancelot77** et 8,2% dans **qgraal77**). En revanche, il semblerait que le taux de ponctuation forte soit légèrement plus élevé dans **qgrraal77** (2,7% contre 1,6% dans **lancelot77**), malgré un taux général moins élevé.

La marque de ponctuation principale utilisée dans ce manuscrit est toujours le « point médian », mais on trouve également des occurrences, rares mais pas isolées, du *comma* (ou *punctus elevatus* < ʒ >) et du *punctus interrogativus* < ʔ >, sur lesquels nous reviendrons plus bas.

Une analyse plus détaillée permet d'apercevoir quelques différences, qui restent toutefois limitées. Dans les deux textes, la position la plus favorable à la ponctuation est la frontière entre propositions indépendantes : cette articulation est ponctuée dans près de 90% des cas (80 occurrences sur 90 dans **qgraal77** et même 24 occurrences sur 25 dans **lancelot77**). La ponctuation forte est toutefois nettement plus fréquente dans cette position dans **qgraal77** (28% contre 8% dans **lancelot77**). La même tendance est observable aux frontières du discours direct et en cas de changement de locuteur. La frontière entre propositions ayant un élément commun (sujet, circonstant ou encore une subordonnée) est marquée dans près de 50% des cas dans les deux textes. La frontière entre propositions principale et subordonnée n'est ponctuée que très rarement, quelle que soit la position de la subordonnée par rapport à la principale et quel que soit le type de subordonnée : 2 occurrences sur 38 dans **lancelot77** et 4 occurrences sur 131 dans **qgraal77**. Les syntagmes coordonnés à l'intérieur de propositions sont également peu ponctués de façon générale (entre 20 et 25% dans les deux textes).

213 Cf. l'identifiant DEAF de la *Queste* : **SGraalIVQueste**.

Il est difficile de trouver une explication sûre pour les cas où le scribe met une ponctuation à une frontière syntaxique généralement peu ponctuée. On peut toutefois remarquer que le plus souvent ces ponctuations permettent de diviser des segments assez longs contenant des structures syntaxiques assez complexes.

(6.5) Mes+tant que ce est cil qui a este hui fet *cheualier* || nouel ¶ que mes sires *lancelot* fist *cheualier* de+sa+main (*qgraal77, 162a:11*) ;

Ainsi, dans (6.5) une relative en *que* est placée à la suite d'une autre relative, et il n'est même pas évident de déterminer à quel antécédent elle est « accrochée » (*cil* ou *chevalier novel*).

Un cas particulier est constitué par des constructions (qu'on analyserait certainement comme des phrases complexes en français moderne) en deux parties avec un lien logique fort qu'on peut qualifier de corrélation. Il s'agit typiquement de proposition subordonnées circonstancielles suivies d'une principale, comme dans l'exemple (6.6). Ch. Marchello-Nizia (2007) a constaté que cette position est particulièrement favorable, dans le manuscrit qui nous intéresse, à l'emploi d'une marque de ponctuation spéciale, le *comma* (ou *punctus elevatus*, < ʒ >)²¹⁴. Cette même marque est par ailleurs fréquemment utilisée après des interjections au début du discours direct (6.7) :

(6.6) Et+qui que || il soit ou cil que uos dites ou autres ¶ ie uoldroie || que bien li+uenist (*qgraal77, 161c:27*) ;

(6.7) ha ¶ sire *cheualiers* fait li || nains tant mal auez exploitie · ha ¶ sire *cheualiers* fait || la+dame tant m'aeuz traie (*lancelot77, 8c:6-8*).

L'hypothèse la plus plausible, selon Ch. Marchello-Nizia, est que cette marque de ponctuation relativement rare (à peine 200 occurrences dans toute la *Queste* contre plusieurs milliers de points) est que sa présence « indique... la présence d'une intonation montante – intonation... destinée à indiquer que la phrase n'est pas finie, et que l'élément que clôt le comma n'est que le premier d'un énoncé en deux pans » (Marchello-Nizia 2007, p. 304). Les arguments proposés par la chercheuse sont bien

214 Cf. la section 1.3. où les travaux de Ch. Marchello-Nizia sur la ponctuation médiévale sont présentés d'une manière plus détaillée.

convaincants, même s'il n'existe aucune « preuve directe » concernant l'intonation médiévale.

Les extraits intégrés dans notre corpus ne contiennent malheureusement que très peu d'occurrences du *comma* : elles se trouvent toutes dans les exemples (6.6) et (6.7), ce qui montre une limitation importante liée à l'usage d'un corpus composé de textes non intégraux : ils ne conviennent que pour l'étude des grandes tendances et des phénomènes très fréquents. Les phénomènes rares risquent de ne pas y apparaître.

C'est exactement ce qui est arrivé au *punctus interrogativus*, dont nous n'avons relevé aucune occurrence dans les extraits de **lancelot77** et de **qgraal77**. Cependant, 72 occurrences de cette marque ont pu être retrouvées dans le texte intégral de la *Queste* grâce à l'édition de Ch. Marchello-Nizia. Elles sont toutes situées à la fin de phrases interrogatives à l'intérieur d'un discours direct, mais une centaine d'autres phrases interrogatives ne portent pas cette marque. On trouve à leur fin soit un point médian (6.8), soit aucune ponctuation (6.9) :

- (6.8) biau frere sauez uos qui cist cheualiers || est qui siet ou siege perilleux ¶ Je+nel
sai+mie+bien+fet || lion ¶ (*qgraal77*, 162a:9) ;
- (6.9) Coment auez uous || non ¶ fait ele ¶ (*lancelot77*, 8c:39).

Il convient de noter toutefois que la ponctuation n'est absente que devant une incise finale, et il y a toujours une ponctuation à la fin de cette incise (6.9). Il s'agit en effet d'une régularité dans la ponctuation scribale : ne pas séparer l'incise du discours direct. Cette tendance se confirme par 2 occurrences dans la *Queste* (à l'extérieur de notre corpus) où le *punctus interrogativus* est placé à la fin de l'incise :

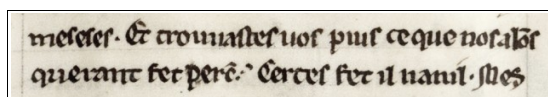


Fig. 31 : *Punctus interrogativus* à la fin d'une incise (*qgraal77*, 3b).

- (6.10) Et trouuastes uos puis ce que nos+alors || querant fet Perceval ¶ (*qgraal77*, 223b:21).

Si l'on compare la ponctuation du manuscrit avec celle des éditions, la différence la plus marquée concerne précisément les incises et les apostrophes, qui ne sont jamais ponctuées dans notre fragment (une quarantaine d'occurrences dans

qgraal77 et une quinzaine dans **lancelot77**). Les autres cas d'ajout de ponctuation par les éditeurs sont plutôt isolés : une ou deux occurrences dans chaque catégorie. Il convient de rappeler que l'édition de Ch. Marchello-Nizia, qui sert de base à notre représentation normalisée de **qgraal77**, vise à préserver la ponctuation scribale, à l'exception notable des frontières des passages en discours direct et des incises. A la différence de l'édition de Ch. Marchello-Nizia, nous avons choisi d'introduire dans la transcription normalisée des virgules pour isoler les apostrophes : il s'agit d'une règle précise de la ponctuation moderne (du même ordre que le marquage des incises) qui mérite d'être respectée au niveau normalisé, d'autant que la ponctuation scribale reste représentée aux niveaux facsimilaire et diplomatique.

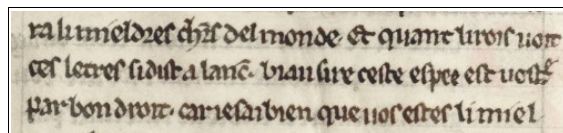


Fig. 32 : Ponctuation du discours direct (*qgraal77*, 161a).

- (6.11) Et quant li+rois uoit || ces lettres si+dist a *lancelot* [>:] [>«] biau sire [>], ceste espee est uostre || par bon droit (*qgraal77*, 161a:22-24).

Dans l'exemple (6.11) ci-dessus, le deux-points et le guillemet ouvrant après *lancelot* sont présents dans l'édition Marchello-Nizia, la virgule dans la représentation normalisée après *biau sire* a été ajoutée par nous. Remarquons que le nom *lancelot* est représenté dans le manuscrit par une abréviation *lanc.* avec un point qui peut également être interprété comme une marque de ponctuation syntaxique, si on tient compte de la « loi de neutralisation » (Tournier 1980, p. 37). Une telle analyse ne semble toutefois pas être fondée dans le passage en question, car ici le discours direct est précédé du verbe de parole avec un complément. Dans ces conditions, où il s'agit en quelque sorte d'une incise antéposée, le début du discours direct est rarement ponctué dans **qgraal77** (8 occurrences sans ponctuation contre 3 avec dans des contextes similaires), cf. les exemples (6.12) et (6.13) :

- (6.12) si+dist au || pseudome [>:] [>«] sire [>], bien soiez uos uenez (*qgraal77*, 161c:22) ;
- (6.13) Quant il l'a uestu *et* apareillie [>], si li dist [.>:] [>«] siuez || [37] moi [>], sire *cheualiers* (*qgraal77*, 161c:36-37).

Dans **lancelot77**, l'éditeur (Micha 1979)²¹⁵ a parfois ajouté une ponctuation faible à des frontières entre les propositions principale et subordonnée, même si dans la majorité des cas la ponctuation est absente dans l'édition aussi bien que dans le manuscrit. Prenons l'exemple des relatives introduites par le pronom *qui*. Sur 8 occurrences présentes dans notre extrait, aucune n'est précédée d'une ponctuation dans le manuscrit. A. Micha a ajouté une virgule dans 3 cas, où la relative ne suit pas immédiatement l'antécédent du pronom relatif, comme dans (6.14) :

(6.14) la dame deuant lui [**>**,] qui mout li sembloit || bele (*lancelot77*, 8d:13).

Les occurrences où la ponctuation du manuscrit a été supprimée par l'éditeur se trouvent uniquement dans **lancelot77**²¹⁶. Dans les 7 cas, il s'agit de propositions complètes ou de prédicats coordonnés, comme dans les exemples (6.15) et (6.16) :

(6.15) lors monte la dame [**.**>] || *et*+li+nains destent les paueillons (*lancelot77*, 8d:11) ;

(6.16) lor+en+uindrent || a+l'entree d'une forest [**.**>] *et*+trouuerent desoz un home || un+cheualier+arme (*lancelot77*, 8d:16).

Il est donc possible de conclure que la ponctuation dans le manuscrit Lyon BM PA 77 est assez conséquente et parfois plus rigoureuse du point de vue syntaxique que dans les éditions. Ainsi, les incises n'y sont jamais ponctuées, tout comme les relatives en *qui* (en tout cas, dans les extraits dépouillés). En revanche, aux frontières de propositions indépendantes la ponctuation peut apparaître dans certaines conditions où les normes modernes n'en prévoient pas (par exemple, des propositions courtes ou simplement des prédicats coordonnés avec la conjonction *et*).

215 Le manuscrit lyonnais n'est pas son manuscrit de base (Oxford, Bodleian, Rawlinson 899), mais les deux manuscrits sont très proches, ce qui nous permet d'utiliser la ponctuation de Micha au niveau normalisé de notre transcription.

216 Cela n'a rien de surprenant, puisque le respect des marques de ponctuation sribales est un principe éditorial de Ch. Marchello-Nizia pour la *Queste du Graal*.

6.3. Paris, BnF, n.a. fr. 1119 (*Queste del saint Graal*)

Comme le manuscrit précédent, ce codex de la BnF est daté d'une façon très approximative : XIII^e – début du XIV^e siècle, il est sans doute légèrement plus tardif que le manuscrit lyonnais. C'est un codex de parchemin, de 246 feuillets, 340×230 mm, il contient, outre la *Queste*, les romans *Agravain*, *Lancelot* et la *Mort Artu* (cf. Micha 1960, p. 179). On ne sait rien sur la région de sa production, ni sur la personnalité du scribe. La version de la *Queste* proposée par ce manuscrit (connu sous le sigle Z) est assez proche de celui de Lyon, et c'est ce manuscrit que Ch. Marchello-Nizia a utilisé dans leurs éditions pour fournir le texte des passages manquants du manuscrit K²¹⁷.

L'extrait qui fait partie de notre corpus correspond au folio 138r, où commence le texte de la *Queste*, et compte 715 occurrences-mots. Son identifiant est **qgraal1119**.

Le taux de ponctuation dans ce manuscrit est à peu près 2 fois moins élevé que dans **qgraal77** et n'atteint que 4,2%. La seule marque de ponctuation attestée dans notre extrait est le point. Même les propositions indépendantes ne sont ponctuées que dans 7 cas sur 34. Les frontières du discours direct (à l'exception des incisives) sont ponctuées dans 8 cas sur 20. En revanche, ce manuscrit présente un nombre relativement élevé d'occurrences « atypiques », où la ponctuation apparaît à des frontières a priori peu susceptibles d'être ponctuées. Par exemple, la fréquence relative de la ponctuation à la frontière entre les propositions principale et subordonnée y est 4 fois plus élevée que dans **qgraal77**. Ce texte présente également 6 occurrences de ponctuation entre syntagmes n'ayant pas la même fonction (donc, non coordonnés, ni juxtaposés) à l'intérieur d'une proposition. Là aussi, la fréquence relative est nettement supérieure à celle du manuscrit lyonnais (0,84% contre 0,28% dans **qgraal77** et 0 occurrences dans **lancelot77**).

La ponctuation au niveau normalisé de notre transcription est basée sur celle de l'édition Pauphilet. L'écart entre le manuscrit et l'édition sous cet aspect est très important : 75 marques de ponctuation ajoutées dans l'édition, dont 17 fortes. On trouve

217 Quant à A. Pauphilet, il a préféré le manuscrit V (BnF, fr. 751), appartenant à une autre famille, pour la fin du texte, et il n'indique pas sa source pour la lacune au début.

Manuscrits antérieurs à 1400

par ailleurs 8 occurrences où une marque de ponctuation du manuscrit a disparu dans l'édition. Dans 21 cas seulement, on trouve des marques de ponctuation au même endroit dans le manuscrit et dans l'édition. Parmi ces occurrences, on compte 7 « renforcements » (ponctuation faible du manuscrit remplacée par une ponctuation forte dans l'édition) et 1 cas d'« affaiblissement ».

Toutes ces observations sommaires nous amènent à croire que le manuscrit BnF n.a. fr. 1119 présente un système de ponctuation bien différent du manuscrit lyonnais étudié dans la section précédente, et aussi plus éloigné du système moderne.

Certaines tendances sont néanmoins communes aux deux manuscrits : l'absence de ponctuation aux incises, par exemple (17 occurrences sur 17) ; la ponctuation aux limites du discours direct et de propositions indépendantes est moins régulière, mais toutefois plus fréquente que dans les autres contextes.

Dans le domaine de l'hypotaxe, on trouve 3 occurrences où la ponctuation du manuscrit a été supprimée dans l'édition. Dans 2 cas, il s'agit d'introduire un discours indirect interrogatif (6.17), (6.18) et dans un cas, de marquer la principale après une relative attachée à un circonstant de cause au début d'une phrase (6.19) :

(6.17) Sire fet ele || por dieu dites moi [.>] se *lancelot* est caienz (*qgraal1119, 138a:18*) ;

(6.18) *et* il+li demande [.>] a qui elle est • (*qgraal1119, 138a:22*) ;

(6.19) *et* por sa+simplece qu'il || i+uoit [.>] i+espoire il tant de+bien (*qgraal1119, 138b:23*).

L'exemple (6.19) nous semble par ailleurs assez proche des constructions corrélatives (avec peut-être un modèle intonatif particulier), pour lesquelles Ch. Marchello-Nizia a mis en évidence la tendance de **qgraal77** à utiliser le *comma*.

Dans 10 cas, l'éditeur a jugé nécessaire d'introduire une ponctuation (toujours faible) à la limite des propositions principale et subordonnée qui n'est pas marquée dans le manuscrit. On en trouve, par exemple, plusieurs occurrences à la fin de propositions circonstancielle, devant une principale introduite par l'adverbe *si* ou *lors*, ce qui fait double emploi :

(6.20) *et quant* il le+uoient [>.] si l'acolent *et* besent (*qgraal1119, 138b:3*).

L'absence de ponctuation dans le manuscrit n'est pas étonnante ici, car l'adverbe *si* en ancien français est plutôt un marqueur de lien que de séparation (Marchello-Nizia 1985). Nous avons toutefois relevé un contreexemple à cette « règle » :

(6.21) *et quant li+rois et li altres qui el+pales || sont uoient ce [.>] si lor en+poise moult ·*
(*qgraal1119, 138a:28*).

Il convient de noter que dans ce cas la subordonnée circonstancielle gouverne à son tour une subordonnée relative et que la principale est immédiatement précédée d'un complément d'objet direct exprimé par un pronom démonstratif *ce*. Dans ce contexte, la nécessité de marquer la frontière syntaxique (et sans doute rythmique) semble plus importante que dans des constructions moins complexes.

La seule occurrence où la ponctuation forte du manuscrit n'a pas été conservée dans l'édition provient d'un contexte syntaxique assez complexe :

(6.22) *Sire ie uos amein nostre || norricon [.>] Itant de+ioie com+nos auons [.>] nostre confort || et nostre espoir [.>] que uos en faciez [.>] cheualier* (*qgraal1119, 118b:15-17*).

Voici la traduction de ce passage proposée par I. Vedrenne : 'Seigneur, je vous amène ce jeune homme que nous avons élevé, et qui est toute notre joie, notre réconfort et notre espoir, afin que vous le fassiez chevalier'. On peut analyser *itant de ioie* comme un complément d'objet direct du verbe *amein*, juxtaposé au précédent (*nostre norricon*). Viennent ensuite deux autres compléments (*nostre confort* et *nostre espoir*) et, enfin, une subordonnée circonstancielle de but introduite par un simple *que*. Il faut dire que le manuscrit K (*qgraal77, 160b:33*), où il n'y a pas de ponctuation devant *itant* et où un point précède *nostre confort*, concorde avec une telle analyse. Cela ne nous empêche pas pour autant de supposer que le scribe du manuscrit Z ait pu interpréter cette séquence différemment : 'Seigneur, je vous amène ce jeune homme que nous avons élevé. [Il y a] tellement de joie, puisque nous avons notre réconfort et notre espoir que vous le ferez chevalier'. Cette interprétation paraît peu probable après la lecture attentive du passage, mais elle pouvait être provoquée, au cours de la copie rapide, par l'hétérogénéité des éléments juxtaposés. Dans ce cas, *itant* se trouverait au début d'une proposition indépendante. Il n'est pas exclu par ailleurs que ce mot soit rapproché par le

scribe de l'adverbe *item* 'de même' (emprunté au latin) utilisé parfois comme un « ponctuant lexical » (cf. l'analyse du texte **joachim768** la section 6.10.).

Quelques occurrences de marques de ponctuation échappent apparemment à l'explication en termes d'unités syntaxiques « ponctuables ». Il s'agit du point devant le mot *chevalier* dans (6.22) et des trois occurrences suivantes :

- (6.23) lors entra en la || sale a+cheual [.>] une || *moult* bele damoisele||le (*qgraal1119*, 138a:10) ;
- (6.24) *et* || *quant* il orent mene en [.>] une chambre (*qgraal1119*, 138a:49) ;
- (6.25) *et* il lor conte *comment* [.>] une damoisele || l'a laienz amene (*qgraal1119*, 138b:7).

On voit clairement que dans les trois cas, la marque de ponctuation est placée devant un déterminant indéfini *une*. Ce déterminant n'est sans doute pas toujours bien distingué de l'adjectif numéral '1' en français médiéval, et il n'est pas rare qu'il soit noté par le chiffre romain *.i.* (entouré de deux points). Nous avons d'ailleurs un exemple dans le même texte, où le chiffre est utilisé pour noter le mot *un* en emploi pronominal :

- (6.26) la ioie que li+cosin firent li *.i.* de+l'autre (*qgraal1119*, 138a:4).

Faute de meilleure hypothèse, nous pouvons supposer que le point devant *une* provient de cette pratique de marquage de chiffres. La présence de l'abréviation *ch'r* peut expliquer l'usage du point devant celle-ci dans l'exemple (6.22) : certaines abréviations, comme les chiffres, pouvaient être entourées de points (même si ce n'est généralement pas le cas de *ch'r*). Un échantillon de texte beaucoup plus grand que le nôtre serait nécessaire pour confirmer ou infirmer ces explications.

La ponctuation de **qgraal1119** présente donc à la fois des traits communs et des différences avec celle de **qgraal77**, mais reste de toute façon, à quelques exceptions près, explicable en termes de frontières d'unités syntaxiques ponctuables.

6.4. Lyon, BM, 828 (*Livre d'Eracles de Guillaume de Tyr*)

Ce manuscrit lyonnais a l'avantage d'être assez précisément daté et localisé : ces informations sont rassemblées dans l'introduction à l'édition de M. R. Morgan (1982), qui a publié d'après ce manuscrit la *Continuation de Guillaume de Tyr* (située à la suite du *Livre d'Eracles*). Le manuscrit a été écrit vers 1280 à Acre (territoire d'Israël à notre époque) dans un scriptorium des croisés²¹⁸.

Le codex est en parchemin, sa taille est de 303×225 mm et il comporte 381 feuillets. L'écriture à l'encre noire sur la première page du *Livre d'Eracles* est partiellement effacée, mais grâce aux photos numériques de haute résolution²¹⁹ le texte a pu être transcrit avec un bon degré de précision. L'identifiant de la transcription dans notre corpus est **tyr828** (cf. le sigle DEAF **GuillTyr**).

Le *Livre d'Eracles* est une chronique de croisades originellement écrite en latin à la fin du XII^e siècle²²⁰ et traduite en français dans le premier tiers du XIII^e siècle. Ce texte appartient donc au domaine historique.

La transcription intégrée dans notre corpus est composée de deux fragments situés aux folios 1r et 5r et compte au total 569 mots. Le choix de deux fragments distincts a été dicté par la qualité insuffisante de la reproduction dont nous disposions au moment de la transcription primaire, mais il présente par ailleurs l'avantage de pouvoir contraster les données du premier recto du manuscrit (qui devait être réalisé avec plus de soin que le reste du texte) à celles d'une page intermédiaire.

La première colonne de la première page **tyr828** a la particularité de présenter apparemment deux « couches » de ponctuation : des points médians sans doute posés par le scribe pendant l'écriture du texte, et des barres obliques (et peut-être même une sorte de parenthèse) ajoutées postérieurement²²¹. Nous reproduisons dans la Fig. 33 la première colonne de cette page et nous indiquons à l'aide de flèches et d'étiquettes

218 Des catalogues plus anciens datent ce manuscrit du XIV^e siècle.

219 Nous tenons à remercier la Bibliothèque municipale de Lyon qui a effectué la prise de ces images à notre demande et nous les offertes gracieusement.

220 La chronique est conduite jusqu'en 1184, son auteur est mort deux ans plus tard.

221 Une occurrence d'une telle marque tardive se trouve à la première ligne de la deuxième colonne. Elle semble être la dernière.

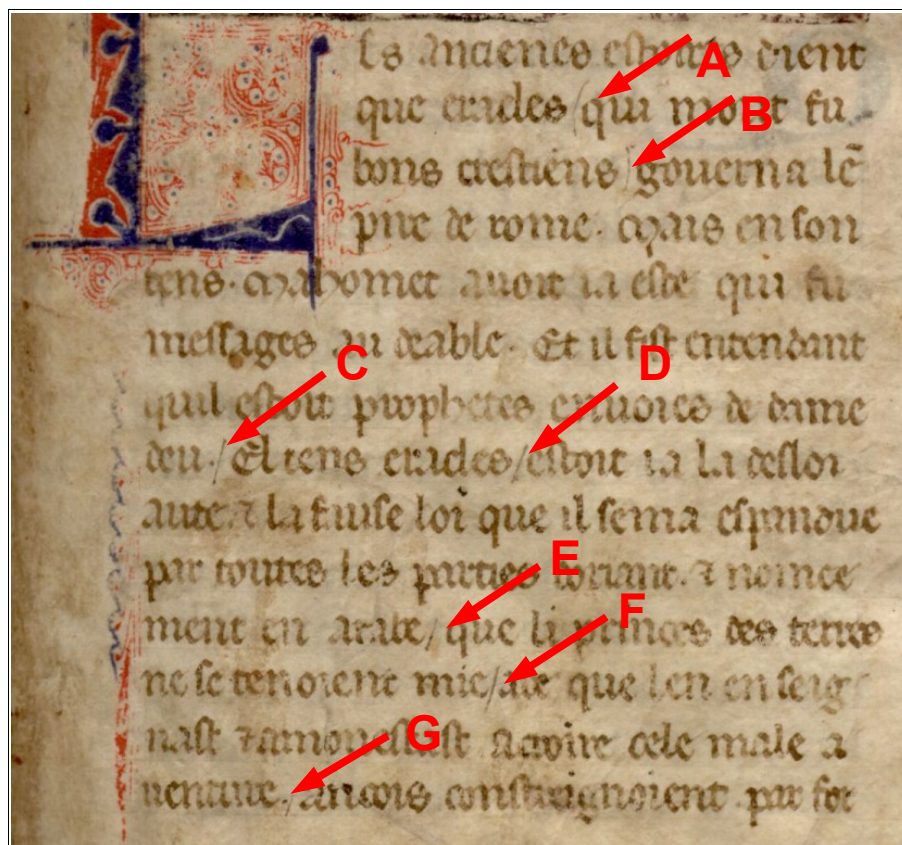


Fig. 33 : Manuscrit Lyon, BM, 828, f° 1r (fragment)

alphabétiques les occurrences de ces marques rajoutées. Certaines d'entre elles « se greffent » sur des marques initiales (C, G)²²², d'autres semblent être placées là où il n'y avait aucune ponctuation initialement (A, B, D, F). Parfois, il est impossible de déterminer avec certitude s'il y avait déjà une marque de ponctuation à l'endroit où une nouvelle marque a été ajoutée (E).

Les deux premières ponctuations « tardives » (A et B) ressemblent à une paire de parenthèses et entourent une relative insérée entre le sujet et le prédicat de sa principale, un cas où l'on pourrait bien utiliser des parenthèses dans un texte moderne. S'il s'agit vraiment de parenthèses, l'ajout ne doit pas être antérieur au XVI^e siècle, où nous rencontrons les premières occurrences de ce signe dans des textes français.

Deux autres occurrences (E et F) marquent également des propositions subordonnées. Le cas F est intéressant, car la ponctuation sépare du reste de la phrase le pronom antécédant sa relative :

²²² Une marque de ponctuation formée par la combinaison d'un point et d'une barre oblique est attestée dans d'autres manuscrits (cf. Marchello-Nizia 1978, p. 35 ; Mazziotta 2007b, p. 57). Dans notre cas, il s'agit néanmoins très probablement de deux marques qui se sont superposées.

(6.27) li princes des terres || ne se tenoient mie / (F) a+ce que l'en enseignast || et+amonestast
a+croire cele male aenture (*tyr828, 1a:12*).

Cette façon de segmenter sert sans doute à souligner le lien fort qui existe entre l'antécédent *ce* et la conjonction *que*, qui amène dans d'autres cas à la formation de conjonctions composites comme *parce que*.

Une autre occurrence de ponctuation ajoutée (D) semble marquer la fin d'un complément circonstanciel « de phrase » en position initiale (un usage bien moderne !) :

(6.28) El tens eracles / (D) estoit ia la desloiaute || et la fause loi (*tyr828, 1a:8*).

Mais il se peut qu'elle soit utilisée pour marquer une limite plutôt informationnelle que grammaticale : on peut supposer qu'ici se situe la limite entre le thème et le rhème. D'autres occurrences du même manuscrit, où une ponctuation est placée entre le sujet et le prédicat, peuvent appuyer cette hypothèse.

D'après (Parkes 1992), l'usage de la barre oblique (*virgula suspensiva*) en tant que marque de ponctuation faible est caractéristique de certains manuscrits des XIV^e, XV^e et XVI^e siècles. Dans notre corpus de textes français, les premières occurrences de la barre oblique se rencontrent au XV^e siècle. Dans les imprimés, elle est remplacée progressivement par la virgule « ronde » à partir du XVI^e siècle. Les parenthèses, quant à elles, ont été inventées à la fin du XIV^e siècle, mais les seules occurrences dans notre corpus se trouvent dans le traité de ponctuation d'Estienne Dolet (1540). Ce traité est également le seul texte à ponctuer systématiquement les propositions subordonnées. Tout cela nous amène à croire que la « seconde couche » de la ponctuation pourrait être datée de la fin du XV^e ou du XVI^e siècle par un lecteur savant, familier des pratiques des humanistes et des imprimeurs novateurs.

Dans le calcul des indices concernant la ponctuation du **tyr828**, nous ne prendrons en compte, bien entendu, que la ponctuation initiale. Le taux de ponctuation général de ce texte s'élève à 9,6% (5,4% pour la ponctuation faible et 4,2% pour la ponctuation forte).

L'extrait que nous étudions ne contient aucun passage en discours direct, nous ne pouvons donc rien dire sur la façon qu'avait le copiste de ponctuer ce type de structures.

Manuscrits antérieurs à 1400

On voit dans ce texte une nette tendance à la ponctuation de propositions indépendantes : 23 ponctuations (dont 19 fortes) sur 25 frontières de propositions autonomes et 5 ponctuations (dont 1 forte) sur 13 frontières de propositions ayant des éléments communs. Dans les deux cas d'absence de ponctuation apparente à la limite entre des propositions autonomes, on peut apercevoir dans le manuscrit une toute petite tache sombre qui peut être une trace d'un point effacé ou écrit légèrement :

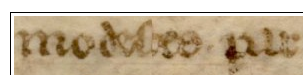
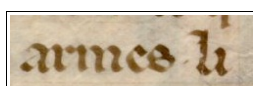


Fig. 34 : Ponctuations « incertaines » (tyr828, 5b:15 et 1b:9).

Il n'est tout de même pas exclu que ces taches soient de simples défauts du support ou bien des ajouts postérieurs à l'écriture du manuscrit. Quoi qu'il en soit, la frontière entre propositions autonomes semble une position très favorable à l'emploi d'une ponctuation dans ce texte.

Les limites entre les propositions principales et subordonnées sont ponctuées dans 8 cas sur 41 (près de 20%), et nous n'avons pas réussi à trouver une régularité dans l'alternance des occurrences ponctuées et non ponctuées. Les deux cas que nous avons classés comme des occurrences de ponctuation forte demandent un examen plus attentif :

- (6.29) Qant eracles ot conquise perse • || *et* ocis cosdroe qui estoit si puissant roi || **I** en raporta le ueraie crois • en Iherusalem • || que cil auoient porte en perse • (tyr828, 1b:4) ;
- (6.30) Mes quant li message reuindrent || Il aprist certainement • qu'il n'auoit pas || gent a+combatre a eaus (tyr828, 1b:37).

On voit que les deux occurrences sont très similaires : il s'agit d'une fin de subordonnée temporelle en *quant*, suivie de sa principale qui commence par le pronom *Il* avec une majuscule, au début d'une nouvelle ligne. Il n'y a pas de marque de ponctuation proprement dite, mais nous considérons de façon générale que l'emploi d'une majuscule seule équivaut à une ponctuation forte (cf. 2.7.2.). Dans ce cas précis, le principe général, qui nous semble justifié par ailleurs, peut être remis en cause : l'usage de la majuscule *I* en début de ligne peut être une simple variante positionnelle

(sans fonction « ponctuationnelle ») comme c'est le cas pour la lettre *R* dans un certain nombre de manuscrits. Si notre nouvelle hypothèse est vraie, nous pouvons conclure que dans l'extrait étudié la ponctuation forte n'est jamais utilisée entre une subordonnée et sa principale.

Une ponctuation est assez souvent utilisée entre des syntagmes coordonnés à l'intérieur d'une proposition (7 occurrences sur 26), surtout si le nombre de constituants coordonnés est supérieur à 2. Il y a même une ponctuation forte dans une longue énumération de malheurs qui ont frappé l'humanité ayant oublié les valeurs chrétiennes :

- (6.31) Car pestilences *et* || famines • estoient grans sur terre • espoentemens || dou ciel •
Crolles de la terre par || mains leus *et* plusors autres choses qui || deussent espoenter
les cuers des homes (tyr828, 5b:9).

Nous interprétons ici *pestilences*, *famines*, *espoentemens*, *crolles* et *autres choses* comme des sujets coordonnés, mais on peut également considérer les trois derniers comme des propositions elliptiques sans verbe.

Le manuscrit **tyr828** présente un nombre relativement élevé d'occurrences de ponctuation à l'intérieur d'une proposition en dehors des cas de coordination (10 au total).

Il y a par exemple 3 occurrences d'une ponctuation faible entre le sujet et le prédicat (toujours avec le verbe *estre*) :

- (6.32) li pueples des || crestiens • estoit mene en cele terre || d'orient (tyr828, 5a:35) ;
(6.33) il sembloit que la fins dou mon||de • fust pres par les signes que nostre sires || en dist
en l'euangile (tyr828, 5b:6)²²³.

Des occurrences de ce type ne semblent pas isolées, même si elles sont très rares par rapport au nombre total de propositions contenant un sujet et un prédicat. Comme nous l'avons déjà signalé, elles peuvent être mises en relation avec la structure informationnelle de l'énoncé : la limite entre le sujet et le verbe correspond souvent à une frontière entre le topique et le reste de l'énoncé, et rien ne contredit une telle interprétation pour les occurrences relevées. On peut aussi essayer de déterminer d'autres facteurs favorables à l'apparition d'une marque de ponctuation dans cette

²²³ Cf. aussi la première marque de ponctuation dans (6.31).

Manuscrits antérieurs à 1400

position. Ainsi, le sujet dans toutes les occurrences est exprimé par un groupe nominal avec modificateurs (et non par un simple nom ou prénom). On constate aussi que dans les 3 cas le prédicat est exprimé par le verbe *estre*, mais le nombre d'occurrences est trop faible pour exclure une simple coïncidence.

La plupart des autres occurrences de ces ponctuations « aberrantes » semblent être liées au marquage de noms propres. Ainsi, dans 2 cas, un point apparaît devant et après le syntagme prépositionnel *en Ierusalem* (cf. l'exemple (6.29)) et, dans 2 autres cas elles se situent à proximité du nom propre *Mahomet* :

(6.34) Mais en son || tens . Mahomet auoit ia este (*tyr828, 1a:5*) ;

(6.35) homar || le fis Cathap qui estoit prince d'arabe . || tiers . apres Mahomet rois
et+enseignieres || a ses comandemens uint en cele || *terre* (*tyr828, 1b:18*).

Le marquage des noms propres n'était en effet pas stabilisé à l'époque de la rédaction des manuscrits que nous étudions. Souvent, ils ne portaient aucune distinction graphique par rapport aux noms communs, mais on voit de temps en temps apparaître l'usage de majuscules, mais aussi de marques de ponctuation devant et/ou après le nom propre (éventuellement précédé d'une préposition, comme c'est le cas dans (6.35)). Assez souvent, le marquage de noms propres coïncide avec des frontières d'unités ponctuables, ce qui ouvre la voie à plusieurs interprétations parallèles. Par exemple, dans (6.34), on peut interpréter le point comme servant à séparer un circonstant initial du reste de la phrase. Les différentes interprétations ne sont pas forcément exclusives l'une de l'autre : le langage naturel est riche en éléments « multi-fonctions » ou ambigus.

La dernière édition scientifique de la chronique Guillaume de Tyr en ancien français date du XIX^e siècle (Paris 1879), le manuscrit lyonnais n'est pas son manuscrit de base, mais le texte est suffisamment proche de notre manuscrit pour qu'on puisse projeter la ponctuation de l'éditeur au niveau normalisé de notre transcription. Ce texte présente l'un des rares cas où la ponctuation de l'édition est moins dense que celle du manuscrit : nous avons relevé 11 « ajouts »²²⁴ et 24 « suppressions » de marques de

224 Dont les 2 cas évoqués précédemment, où l'état du manuscrit ne permet pas de juger s'il y avait ou non une marque de ponctuation.

ponctuation dans l'édition par rapport au manuscrit. Les suppressions concernent tous les types de frontières ponctuelles, mais le plus souvent il s'agit de structures coordonnées avec la conjonction *et*, des subordonnées en *que* et, enfin, de syntagmes à l'intérieur de proposition que nous venons d'examiner. Les « ajouts » n'obéissent pas non plus à une stricte règle syntaxique : il s'agit le plus souvent d'un usage « pragmatique » permettant de faciliter l'interprétation de structures complexes, comme dans l'exemple suivant :

- (6.36) Endementre || qu'il s'entendoit a+ce [> ,] estesuos *que* homar || le fis Cathap qui estoit prince d'arabe · || tiers · apres Mahomet [> ,] rois *et*+enseignieres || a ses comandemens [> ,] uint en cele || *terre* qui a+a+nom palestine (*tyr828, 1b:15-20*).

Notre traduction : 'Pendant qu'il [Héraclius] s'occupait de cela [> ,] voici que Omar, fils de Catab, qui était prince d'Arabie, troisième après Mahomet [> ,] roi et enseignant de ses commandements [> ,] est venu dans ce pays qui s'appelle la Palestine'.

Notre analyse de **tyr828** permet de conclure qu'il présente un type de ponctuation différent de **qgraal77** et de **qgraal1119**. Il est proche de **qgraal77** par son taux de ponctuation élevé et par le marquage régulier des frontières de propositions autonomes, mais il ponctue de façon nettement plus fréquente les propositions subordonnées et de certains syntagmes à l'intérieur de propositions. La particularité de ce manuscrit peut s'expliquer, entre autres, par sa provenance d'un scriptorium situé loin de la France métropolitaine.

6.5. Paris, BnF, fr. 12203 (*Conquête de Constantinople de Geoffroi de Villehardouin*)

Le manuscrit 12203 du fonds français de la BnF est daté approximativement de la fin du XIII^e siècle, ses 185 feuillets de 315 × 230 mm sont en parchemin, l'écriture est plutôt soignée. Outre le texte de Villehardouin (qui y occupe les folios de 69 à 112), il contient 4 autres œuvres historiques (*Histoire d'outremer et du roi Saladin*, *Histoire des comtes de Flandre*, la continuation d'Henri de Valenciennes et *Histoire des ducs de Normandie*). Son origine géographique se situe probablement en Picardie.

Manuscrits antérieurs à 1400

La *Conquête de Constantinople* de Geoffroy de Villehardouin (né vers 1150 et mort peu après 1213) est l'une des plus anciennes chroniques écrites en français qui décrit les événements de la quatrième croisade, à laquelle l'auteur a participé à partir de 1198. D'après E. Faral (1961), Villehardouin a rédigé sa chronique après les événements, pendant les dernières années de sa vie.

L'extrait que nous avons transcrit pour notre corpus se situe aux folios 69r – 71r et contient 658 mots. Il porte l'identifiant **villehardouin12203**²²⁵.

Le taux de ponctuation général dans notre extrait est assez élevé (12,3 %), la ponctuation faible est légèrement plus fréquente que la forte (47 occurrences contre 34). Le point est toujours la marque de ponctuation principale, mais on trouve également 3 occurrences du *comma*.

Notre extrait contient deux passages en discours direct dont les limites sont dans les deux cas marquées par une ponctuation forte :

(6.37) li dus *et* tout li autre [...] disent • Nos l'otrions • || Nous l'otrions • (*villehardouin12203, 71b:10*).

Nous n'avons aucune occurrence d'incise, mais il y a en revanche 3 apostrophes, jamais ponctuées :

(6.38) Segneur vees+la || miracle *que* dex nous a faite (*villehardouin12203, 71b:18*).

Toutes les frontières de propositions autonomes sont ponctuées (26 occurrences, dont 20 de ponctuation forte), ce qui est un indice significatif. Les propositions avec un élément commun sont séparées par une ponctuation (toujours faible) dans 13 cas sur 18. En effet, l'absence de ponctuation est constatée dans les propositions avec un sujet commun (coordination de prédicats), dont la deuxième est relativement courte :

(6.39) Et il estoit priestres *et* tenoit || la parroche de la uille • (*villehardouin12203, 69b:12*).

La ponctuation apparaît quand la structure syntaxique devient un peu plus complexe :

(6.40) Et apres i+enuoia || .i. sien cardonnal Maistre Pieron || de Capes croisie • *et* manda par lui || le pardon tel *com* ie vous dirai • (*villehardouin12203, 69b:26*) ;

225 Cf. le sigle DEAF **Villeh.**

- (6.41) Joifrois de ville Harduin li mareschus || de champaigne moustra la || parole **·** *et* dist *par* la volente *et par* l'acort || as autres *et* lor dist (*villehardouin12203, 71a:29, 30*).

Dans (6.40), l'objet direct de la première proposition comporte deux appositions, et la deuxième proposition régit une subordonnée comparative. Dans (6.41), nous avons aussi une apposition qui amplifie la première proposition ; la deuxième contient des compléments circonstanciels coordonnés et est suivie, à son tour, d'une troisième proposition coordonnée.

Les limites entre principales et subordonnées sont ponctuées (toujours faiblement) dans un tiers des cas (13 occurrences sur 38), ce qui donne une fréquence relative nettement plus élevée par rapport aux autres textes de notre corpus. Aucune règle stricte ne peut expliquer la variation constatée. La probabilité de trouver une ponctuation semble plus élevée devant une relative (que devant une complétive) et dans les cas où la principale vient après sa subordonnée :

- (6.42) Hanquis [...] *qui moult iert sages et* || preus **·** monta ou letrin [(*villehardouin12203, 71b:16*) ;
- (6.43) nule gent ki sour mer soient || n'ont si *grant* pooir *comme vous* aues · (*villehardouin12203, 71a:39*).

Les exemples (6.42) et (6.43) sont syntaxiquement proches, mais on trouve la ponctuation à la fin d'une relative uniquement dans (6.42). Il n'est pas exclu que la position finale du verbe dans la relative en (6.43), juste avant le verbe de la principale, et le fait que la frontière coïncide avec un saut de ligne ait joué son rôle.

Il convient de remarquer que les trois occurrences du *comma* se trouvent toutes dans cette position entre subordonnée et principale :

- (6.44) **EN** l'autre an apries que chil || preudom Fouques parla de || diu **·** ot .i. tournoi *en* champaigne (*villehardouin12203, 69c:11*) ;
- (6.45) *Moult* fu *grans* || la renommee par les *terres* **·** quant cil || doi se croisierent · (*villehardouin12203, 69c:29*).

Cet usage ne peut que conforter l'hypothèse selon laquelle le *comma* tend à être utilisé dans les constructions à deux parties « corrélées », à l'endroit où pendant la lecture on devait faire une pause avec une intonation montante (Marchello-Nizia 2007, p. 304 ; cf. notre section 6.2.).

Manuscrits antérieurs à 1400

Le manuscrit **villehardouin12203** présente un nombre assez important de marques de ponctuation entre syntagmes coordonnés à l'intérieur d'une proposition (17, dont 9 fortes). Cela s'explique par la présence d'une longue liste de noms de personnes participant à un tournoi :

(6.46) **EN** la terre le conte de champaigne || se croisa Gerniers li euesques || de Troies **et** li cuens Gautiers || de Braine **;** Joffrois de Joinuile **;** || qui estoit senescaus de la tiere **;** Robiers ses freres **;** Gautiers de || voignori **;** Gautiers de Monbelyart **;** || Eustases d'escouflans **;** Guis || dou plaissie **et** ses freres Henris || D'ardillieres **;** Ogiers de saint chienon **;** (*villehardouin12203, 69c:31-40*).

Cette liste continue en effet encore longuement, et nous avons été amené à couper la transcription à cet endroit et la reprendre plus loin afin de ne pas biaiser les données par une pratique de ponctuation très particulière propre à ce passage.

Les énumérations (ou séquences de syntagmes nominaux de même fonction), et particulièrement celles de noms de personnes ou de noms géographiques, constituent des séquences textuelles où des règles de ponctuation spéciales s'appliquent : chaque terme tend à être ponctué (qu'il soit ou non précédé de la conjonction *et*) et les majuscules servent à marquer les noms propres. Il convient de noter que **villehardouin12203** présente une pratique très régulière en matière de marquage de noms propres. Dans d'autres manuscrits, l'usage de la majuscule est beaucoup moins systématique, et la majuscule « passe » souvent du nom propre à la conjonction qui le précède. On peut expliquer cette tendance par l'importance attachée dans cette chronique aux noms de personnages historiques. En dehors de cette liste de noms, la ponctuation de syntagmes coordonnés ou juxtaposés est rare.

Les occurrences de la ponctuation autour de syntagmes particuliers à l'intérieur de la proposition sont peu fréquentes, mais pas exceptionnelles. Dans 2 cas, un point intervient après des appositions et devant le verbe, on peut supposer ici que la présence de noms propres dans un contexte proche favorise cet emploi, mais l'emploi de la ponctuation autour des appositions est loin d'être systématique :

(6.47) **Joifrois** de ville Harduin li mareschaus || de champaigne **;** moustra la || parole (*villehardouin12203, 71a:28*) ;

(6.48) Thiebaus quens de champaigne || **et** de Brie **;** prist la crois (*villehardouin12203, 69c:15*).

Deux autres occurrences surviennent dans une proposition autour du verbe *s'agenoillier* que le scribe a sans doute voulu mettre en évidence pour une raison qui n'est pas claire :

- (6.49) **M**aintenant tout li .vi. message s'a||genoillierent a+leur pies *moult* || plourant
(*villehardouin12203, 71b:6-7*).

Chacune des trois occurrences restantes peut être expliquée par une raison ou une autre, mais elles restent de toute façon des cas très isolés.

Si l'on compare la ponctuation du manuscrit à celle du niveau normalisé, on peut constater que les cas où les règles modernes imposeraient l'usage d'une marque là où il n'y en a pas dans le manuscrit se situent presque tous dans une catégorie précise. Sur 17 occurrences de ce type, 13 concernent les appositions et les apostrophes, autrement dit, des groupes syntaxiques de niveau inférieur à la proposition. Les cas où une marque de ponctuation du manuscrit disparaît au niveau normalisé sont plus divers : sur 25 occurrences, on trouve 6 propositions avec un élément commun, 6 frontières subordonnée / principale, 6 syntagmes coordonnés et 6 syntagmes particuliers. Un cas est particulièrement intéressant, car il s'agit de la situation où la ponctuation change le sens du texte :

- (6.50) Chil || doi conte ierent neueu le roi de || france [.>] *et* cousin germain *et* neueu || le roi d'engleterre [.>] De l'autre || part [>.] auoec ces .ij. contes se croisierent || doi *moult* haut baron de || france [.>.] (*villehardouin12203, 69c:21-27*).

Le manuscrit « accroche » le complément circonstanciel *de l'autre part* à la proposition suivante, tandis que l'éditeur le place en fin de phrase précédente et commence même un nouveau paragraphe. L'interprétation de l'éditeur correspond aux faits historiques, tandis que l'interprétation du scribe est plus naturelle du point de vue de la structure syntaxique.

On peut noter en conclusion que le manuscrit BnF fr. 12203 présente un système de ponctuation assez riche, se rapprochant d'une part de celui du manuscrit Lyon BM P.A. 77 (usage du comma) et d'autre part de **tyr828** (ponctuation de subordonnées et de syntagmes particuliers).

6.6. Paris, Bibliothèque Sainte-Geneviève, 587

(Image du monde)

Ce manuscrit daté de la fin du XIII^e siècle – début du XIV^e siècle (Prior 1913)²²⁶ ne présente pas une écriture très soignée. Le manuscrit est composé de 191 feuillets, en parchemin, de 370 × 262 mm²²⁷. Outre l'*Image du monde* en prose (f. 172 – 191), ce manuscrit contient les textes suivants : *Vie des sains en romans* (f. 2 – 137), *Vie de saint Gerard* (f. 135 – 145 et 147 – 171) et *Mireors d'Amistié* (f. 146).

L'*Image du monde* est une sorte d'encyclopédie médiévale populaire couvrant les principaux domaines du savoir de l'époque : la théologie, l'astronomie, la géographie, etc. Selon la typologie de la BFM, ce texte appartient au domaine didactico-scientifique et au genre de l'encyclopédie.

Son auteur, Gossouin de Metz, est un clerc d'origine messine, probablement un franciscain (Connochie-Bourgne 1999), et son œuvre présente de nombreux traits du dialecte lorrain. Le texte français s'appuie sur de nombreuses sources antérieures, notamment latines, mais sa composition est originale²²⁸. Il était originellement composé en vers, mais assez rapidement une version en prose est apparue. La composition de la première rédaction en vers est généralement datée de 1246. D'après O.H. Prior, la rédaction en prose a pu être rédigée par Gossouin la même année et certainement avant la deuxième rédaction complète en vers (1248). D'autres chercheurs émettent toutefois des doutes sur cette hypothèse.

L'œuvre de Gossouin a rencontré un grand succès au Moyen Âge, ce dont témoigne le grand nombre de manuscrits qui nous sont parvenus (69, dont 8 en prose). Elle a également connu plusieurs éditions imprimées à la fin du XV^e et au XVI^e siècle (dont 2 sont représentées dans notre corpus : **im1495** et **im1522**), des traductions (anglaises et hébraïques) et des plagiat.

Le manuscrit de la Bibliothèque Sainte-Geneviève est l'un des plus anciens, mais il présente une version très abrégée, c'est donc un autre manuscrit (également

226 Dans la bibliographie du DEAF, on trouve la date de *circa* 1300, obtenue d'après des recherches sur les autres textes du même codex.

227 Le feuillet 146 est plus étroit (370×250 mm) et contient une œuvre différente.

228 Et cela contrairement à l'indication du texte même qui prétend être une traduction du latin.

représenté dans notre corpus : **im574**) que Prior a choisi comme manuscrit de base pour son édition.

La transcription du manuscrit Sainte-Geneviève 587 dans notre corpus porte l'identifiant **im587**²²⁹ et contient 641 occurrences-mots. Au niveau normalisé de la transcription, nous avons introduit notre propre ponctuation, en tâchant de respecter les mêmes règles que celles appliquées dans l'édition de Prior.

Le taux de la ponctuation enregistré dans ce manuscrit est particulièrement faible (3,1% au total), et il s'agit essentiellement de la ponctuation forte (18 occurrences contre 2 seulement de la ponctuation faible). Une telle domination de la ponctuation forte est un cas plutôt rare, mais on en trouvera d'autres dans notre corpus.

L'extrait représenté dans notre corpus ne contient pas de discours direct, mais c'est le cas du texte tout ensemble, ce qui n'est pas étonnant pour un texte appartenant à ce genre.

La seule marque de ponctuation incontestablement utilisée dans **im587** est le point. Nous avons par ailleurs une occurrence d'une barre oblique très légère à la ligne 20 du folio 172v, 1ère colonne :

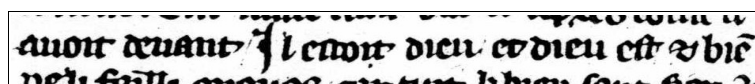


Fig. 35 : Barre oblique légère (*im587*, 172v).

Il se peut qu'il s'agisse dans ce cas d'un ajout tardif, mais de toute façon nous interprétons l'emploi d'une majuscule comme une ponctuation forte. Ce manuscrit présente par ailleurs une particularité d'écriture qui consiste à tracer de petites barres obliques à droite de certaines lettres en finale d'un mot, surtout du *t* :

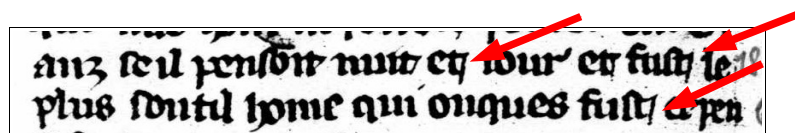


Fig. 36 : « Fausses » barres obliques (*im587*, 172d).

229 Cf. le sigle DEAF **ImMondePr**.

Ces barres peuvent ressembler parfois à des marques de ponctuation, mais ce n'en sont pas. En revanche, on peut considérer les lettres finales dotées de ces barres comme des allographes positionnels servant à renforcer la notation des limites de mots et jouant ainsi le rôle de marques de segmentation.

Une autre « pseudo » marque de ponctuation utilisée dans ce texte est un point surmonté d'une barre verticale (elle ressemble donc à un point d'exclamation moderne) : nous en avons 2 occurrences en fin de deux lignes successives :

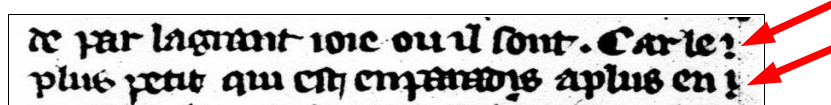


Fig. 37 : Marques de fin de ligne (im587, 172d)

Ces occurrences sont situées en dehors de frontières « ponctuables », et elles jouent un rôle essentiellement décoratif (justification de fins de ligne). Il n'est pas exclu que ces marques servent aussi à relier des éléments de syntagmes soudés mais séparés par des fins de ligne (d'une façon analogue aux traits-d'union en cas de coupure de mots).

La grande majorité des occurrences de la ponctuation dans **im587** se trouve aux frontières de propositions indépendantes (16 occurrences sur 20). Ce type de frontière syntaxique est ponctué dans près de 50% des cas. Il semblerait que la présence d'un sujet nominal (et non pronominal) dans une nouvelle proposition augmente les chances de trouver une ponctuation à son début, mais le nombre d'occurrences dont nous disposons est insuffisant pour l'affirmer avec certitude.

L'usage de la ponctuation ne paraît pas stable même dans des conditions syntaxiques très similaires. Ainsi, la conjonction de coordination de propositions à valeur causale *car* est précédée d'une ponctuation dans 4 cas sur 13, sans qu'on puisse donner une explication claire à cette variation, comme dans les deux exemples suivants :

(6.51) il fait orendroit et come il peust iamais || ualoir **!** Car autrement ne ne²³⁰ seroit il pas || dieus (*im587, 172c:31*) ;

(6.52) mais || n'est pas de+telle nature **!** car il est || dieu enterinement sanz commencement et || sanz fin (*im587, 172c:35*).

230 Adverbe de négation répété, il s'agit sans doute d'une erreur de copie.

L'emploi de la ponctuation aux autres types de frontières est très rare : 2 occurrences (sur 24) entre des propositions avec un élément commun (6.53), 2 occurrences (sur 51) entre une subordonnée (6.54) et sa principale et 1 occurrence (sur 28) dans la coordination de syntagmes (6.55) :

- (6.53) il les tient tuit en sa+main por+tens qui passe . || ne qui viengne il ne pert nule chose de sa bonte (*im587, 172c:22*) ;
- (6.54) car tuit li bien qui onques || furent faiz et sont et seront . Sont tuit || portrait deuant lui (*im587, 172d:33*) ;
- (6.55) il est || dieu enterinement sanz commencement et || sanz fin deuant cui nule chose²³¹ or uielle or || nouvelle . Mais toz iours en vne maniere || bele et freche (*im587, 172c:38*).

Dans tous ces cas, nous avons affaire à des constructions syntaxiques assez complexes, et l'emploi de la ponctuation joue effectivement un rôle d'aide à la lecture. Dans (6.53), l'une des deux occurrences de la ponctuation faible, le point est placé en fin de ligne entre deux propositions relatives coordonnées. Une frontière syntaxique plus forte devant la proposition suivante (*il ne pert...*) n'est pourtant pas ponctuée. On peut soit supposer que le scribe n'a pas bien compris ce passage complexe, soit expliquer l'usage du point par des raisons de justification en fin de ligne. Dans (6.54), le scribe a sans doute voulu séparer deux formes du verbe *estre* appartenant à des propositions distinctes. Dans (6.55), enfin, l'interprétation du passage est difficile. Apparemment, il manque le verbe dans la proposition *deuant cui nule chose [est?] or uielle or || nouvelle*. La suite peut être interprétée comme un attribut coordonné, mais on peut aussi supposer l'ellipse du verbe et classer cette occurrence dans la catégorie de propositions avec un élément commun.

Si l'on compare la ponctuation du manuscrit à celle qu'on utiliserait dans une édition moderne (niveau normalisé de la transcription), on constatera un grand nombre d'ajouts de signes de ponctuation : 19 entre propositions indépendantes, le même nombre entre des subordonnées et principales et 7 dans les autres cas.

Dans un seul cas, la ponctuation du manuscrit n'aurait pas été maintenue dans une édition, c'est l'usage du point dans l'exemple (6.53) que nous venons d'analyser.

231 Le scribe a sans doute omis un verbe (*estre ?*) à cet endroit.

Nous pouvons conclure que le manuscrit **im587** présente un système de ponctuation peu développé, avec une basse fréquence des marques, une quasi-absence de distinction entre ponctuation forte et faible (seule la ponctuation forte est utilisée sauf cas exceptionnels) et la présence de ces marques presque exclusivement aux frontières de propositions indépendantes.

6.7. Paris, BnF, fr. 11652 (*Coutumes de Beauvaisis* de Philippe de Beaumanoir)

Le manuscrit fr. 11652 de la BnF a été exécuté « à l'extrême fin du XIII^e siècle » ou « dans les premières années du XIV^e siècle » (Salmon 1899), il est chronologiquement très proche du manuscrit Sainte-Geneviève 587. L'explicit du manuscrit porte une signature abrégée : *Col d. Fl. cl. s. li. i.* A. Salmon l'interprète ainsi :

Collibertus [ou *Colinus*] de Floriaco, clericus, scripsit librum istum (Salmon 1899, p. XVIII).

Si cette hypothèse est correcte, on peut situer la rédaction du manuscrit au village de Fleury (dans l'Oise).

Le manuscrit est composé de 246 feuillets en parchemin, de 298 × 210 mm. Il a été exécuté avec beaucoup de soin, collationné et corrigé. En revanche, les miniatures, pour lesquelles une place avait été réservée au début de chaque chapitre, n'ont jamais été réalisées. L'écriture du manuscrit peut être définie comme une gothique régulière. Am. Salmon reconnaît ce manuscrit comme l'un des plus rapprochés de l'original de Beaumanoir, mais il a néanmoins préféré baser son édition sur la méthode lachmanienne, en tâchant de reconstruire l'original perdu. Il n'a donc pas hésité à corriger ce qu'il considérait comme des fautes du manuscrit. Paradoxalement, il s'est montré moins interventionniste dans le domaine de la ponctuation que certains éditeurs bédieristes dans un passé plus récent.

Les *Coutumes de Beauvaisis* de Philippe de Beaumanoir n'est pas un simple coutumier, mais plutôt un traité où les cas concrets sont accompagnés de réflexions

générales sur le droit. Le texte appartient donc au domaine juridique, mais il est proche également du domaine didactique. Le texte a été composé vers 1283, le manuscrit que nous étudions n'est donc séparé de cette date que d'une quinzaine ou une vingtaine d'années. La langue de Beaumanoir est caractérisée par un certain archaïsme : il respecte, par exemple, de façon générale la déclinaison bicasuelle.

L'extrait que nous avons transcrit se situe aux folios 4r-5v et contient 1261 mots. Il porte l'identifiant **beuma11652**.²³² Le taux de ponctuation y est plutôt bas : 5,6%. Les occurrences de la ponctuation faible y sont presque aussi fréquentes que celles de la ponctuation forte (34 contre 36 respectivement). A côté du point, largement majoritaire, on trouve 3 occurrences du *comma*, toujours utilisé comme une ponctuation faible.

L'échantillon intégré dans notre corpus ne contient pas de discours direct : il s'agit en effet d'une partie initiale de l'œuvre, où il est surtout question des qualités dont doit faire preuve un bon bailli.

Sans surprise, la position la plus favorable à l'emploi de la ponctuation est la frontière entre deux propositions autonomes : elles sont ponctuées dans 47 cas sur 52 (dont 29 ponctuations fortes). Dans tous les cas d'absence de ponctuation à ce type de frontière on peut constater que la deuxième proposition commence par une conjonction de coordination et a un sujet pronominal ou n'a pas de sujet exprimé, comme c'est le cas de la proposition impersonnelle (*et souvent avient que...*) dans l'exemple suivant :

(6.56) l'en ne doit pas tenir le || bailli pour sage qui uers touz est fel *et* || cruex *et* souuent auient *que* les simplez || gens qui ont bones *quereles et*+loiax || lessent perdre leur quereles... (*beuma11652, 4v:31*).

On voit par ailleurs que la proposition en question sert à introduire un argument en faveur de l'affirmation qui précède. Il y a donc un lien logique fort entre tous les éléments de ces deux groupes de propositions.

Dans deux autres cas, la conjonction de coordination est accompagnée de l'adverbe *autrement* pour former un connecteur argumentatif. Le lien logique est aussi très fort dans ce cas :

(6.57) Donques doit il estre hardiz || sanz couardise *et* sanz riens douter || ou autrement il ne fet pas ce || qui a+lui+apartient *et*+a+son estat (*beuma11652, 5b:34*).

232 Cf. le sigle DEAF **BeaumCout**.

Il convient de noter que l'adverbe *ausint* ('aussi'), qui joue un rôle d'organisateur textuel, semble attirer la ponctuation forte au début des propositions le contenant : dans les 5 occurrences de cet adverbe dans notre extrait il se situe à proximité d'une ponctuation forte. C'est notamment le cas de constructions corrélatives, comme dans l'exemple suivant :

- (6.58) Car tout aussint || *comme* li+mires qui pour pitie de sa maladie || lesse a+ataindre la plaie de || la_ quele il le doit garir le met en+peril || de+mort . Tout aussint li+baillis qui *est* || debonnaire uers les mesfesans de || sa baillie met ceuz qui ueulent uiure || em+pes en peril de+mort . (*beauma11652, 4d:11-15*).

Les propositions avec un élément commun ne sont séparées par une ponctuation que dans 1 occurrence sur 10 :

- (6.59) Car se il estoit || couars il n'ozeroit couroucier le || riche *homme* qui auroit a+fere *contre* || le+poure . ou+il+n'ozeroit celui qui auroit || mort deseruie fere iusticier || pour pour de son lignage (*beauma11652, 5b:26*).

Dans cet exemple, les deux propositions qui commencent par *il n'ozeroit...* ont une subordonnée commune *se il estoit couars*, mais chacune d'entre elles régit par ailleurs une relative qui ne dépend que d'elle. Toute la construction, qu'on a toutes les raisons d'appeler une phrase complexe, est effectivement longue et complexe, et l'usage d'une ponctuation entre ses deux parties principales paraît tout-à-fait justifié.

La ponctuation apparaît rarement entre les principales et subordonnées : nous avons relevé une seule occurrence de ponctuation forte et 6 occurrences de ponctuation faible sur 126 frontières syntaxiques de ce type. La seule occurrence de la ponctuation forte est la suivante :

- (6.60) Donques || se tors ou despis est fet a+son seigneur || ou a+li . Il le doit uengier hastiement || *et* sagement (*beauma11652, 5a:31*).

Il s'agit ici d'une principale qui suit sa subordonnée dans une construction hypothétique. Dans cinq autres occurrences similaires il n'y a pas de ponctuation à cet endroit. On peut toutefois remarquer une particularité de l'occurrence avec la ponctuation forte : le pronom sujet *il* de la principale y suit immédiatement le pronom

objet indirect *li* de la subordonnée. La ponctuation a pu être utilisée pour mieux séparer ces formes courtes.

Une autre occurrence présente l'usage du *comma* à ce type de frontière :

(6.61) qui mesfet au || bailli ; il mesfet au seigneur (*beauma11652, 5a:40*).

Cet usage est très similaire aux cas relevés dans **lancelot77** et **villehardouin12203** et correspond parfaitement à l'hypothèse de Ch. Marchello-Nizia.

Les deux autres occurrences du *comma* sont tout aussi intéressantes :

(6.62) La+terce uertu que li+baillis doit auoir || si+est ; que il doit estre douz *et* debonnaire
|| sanz felonnie *et* sanz cruaute (*beauma11652, 4d:4*) ;

(6.63) Car autant uaut || estre sapiens *comme* sages ; donques || dison nous que cil qui
s'entremet || de+baillie garder *et* de iustice fere doit || estre sages (*beauma11652, 4c:16*).

Dans (6.62), le *comma* est employé devant une complétive jouant le rôle d'attribut du sujet qui est un nom abstrait. En français moderne, une ponctuation à cet endroit est impossible (ou en tout cas non conforme aux règles). Du point de vue de la structure logique, cependant, il s'agit d'une frontière entre le thème et le rhème particulièrement importante dans ce passage du texte où sont énumérées et décrites les dix vertus nécessaires à un bon bailli. Une pause avec une intonation montante pourrait tout-à-fait être observée à cet endroit lors de la lecture. La lecture de la suite du texte (au delà de l'échantillon transcrit) permet de relever encore une occurrence du *comma* dans la même construction au folio 6r (lorsqu'il s'agit de la huitième vertu). En même temps, le *comma* n'est pas utilisé dans des contextes similaires, comme dans l'exemple suivant :

(6.64) La+quarte uertu qui doit estre en || baillieu ; si est *que* il soit soufrans || *et* escoutans
(*beauma11652, 4d:38*).

Cela confirme une autre caractéristique importante du *comma* observée dans les autres textes : le caractère facultatif de son emploi :

La présence du *comma* n'apparaît jamais comme obligatoire : il s'agit de constructions relativement fréquentes, dont seul un petit nombre présente cette ponctuation (Marchello-Nizia 2007, p. 303).

Revenons à l'exemple (6.63). Ici le *comma* est placé devant une proposition introduite par l'adverbe *donques* et qui sert de conclusion à l'argumentation qui précède. Toutes les autres occurrences de cet adverbe-connecteur sont accompagnées d'une ponctuation forte. L'hypothèse de l'intonation montante semble difficilement applicable dans ce cas. On peut supposer en revanche que le segment entre la dernière occurrence de la ponctuation forte (devant *car*) et *donques* a semblé trop court au scribe pour qu'il en ajoute une nouvelle : 7 mots, tandis que la distance moyenne entre les ponctuations fortes dans notre extrait est de 35 mots ! Peut-être a-t-il voulu renforcer le point simple devant ce connecteur sans utiliser la majuscule.

Les quatre occurrences de la ponctuation entre les subordonnées et les principales qu'il reste à examiner surviennent toutes dans des constructions complexes avec des successions de relatives et autres subordonnées imbriquées.

La ponctuation entre des syntagmes coordonnés ou juxtaposés dans une proposition n'est pas très fréquente : on en compte 3 occurrences (toujours faibles) sur 44 frontières de ce type identifiées. Deux de ces occurrences figurent à l'intérieur d'une énumération de sujets :

(6.65) *et* de ce puet || nestre au bailli qui est pereceuz || uilanie **;** *et* disfame **;** *et* damage **;**
(*beauma11652, 5b:18*).

Ici la ponctuation sert sans doute à souligner la gravité des conséquences de la paresse du bailli.

Le dernier cas concerne une coordination de compléments du prédicat avec des régimes différents : deux syntagmes avec la préposition *uers* (dont une avec une relative subordonnée) et un avec la préposition *es* :

(6.66) nous l'entendo[n]s || uers ceuz *qui* bien welent *et* uers le commun || pueple **;** *et* es cas
qui auient plus || par mescheance *que* par malice • (*beauma11652, 4d:25*).

Aucune occurrence de ponctuation n'a été relevée autour de syntagmes particuliers à l'intérieur de propositions.

Notre première analyse du système de la ponctuation dans **beauma11652** montre que malgré le taux de ponctuation relativement bas, ce texte présente un usage

de ponctuation bien cohérent : les propositions indépendantes y sont ponctuées presque régulièrement, dans les autres cas, la ponctuation est utilisée dans des constructions complexes, pour segmenter des séquences trop longues. L'usage du *comma* le rapproche du **lancelot77** et **villehardouin12203**.

6.8. Paris, BnF, fr. 574 (*Image du monde*)

Selon O.H. Prior (1913), ce manuscrit est l'un des plus beaux de la Bibliothèque Nationale. Il mesure 387 × 265 mm et contient 139 feuillets en parchemin. L'écriture en 4 colonnes est typique du XIV^e siècle. La banque d'images de la BnF contient plusieurs miniatures de ce manuscrit, attribuées au *Maitre du roman de Fauvel*. Le catalogue de cette banque d'images propose aussi une datation relativement précise : années 1315 – 1320.

L'*Image du monde* en tant qu'œuvre a déjà été présentée dans la section 6.6. O.H. Prior considère le manuscrit BnF, fr. 574 (auquel il attribue le sigle A) comme le meilleur témoin du texte et le choisit comme le manuscrit de base de son édition. L'extrait que nous avons transcrit et annoté correspond à la même partie de l'œuvre que **im587**, mais il est légèrement plus court et contient 549 occurrences-mots. Dans notre corpus, cet extrait porte l'identifiant **im574**.

Le taux de ponctuation dans **im574** est plutôt élevé : 13,3%. Les ponctuations fortes sont très légèrement majoritaires (36 occurrences contre 32). On compte 60 occurrences du point et 12 occurrences du *comma* (une fréquence exceptionnellement élevée pour ce dernier !).

Notre extrait ne contient pas de discours direct, ce qui est normal pour ce type de texte. Les frontières entre propositions autonomes sont ponctuées régulièrement : il y a toujours une marque de ponctuation dans cette position, et la ponctuation forte domine très nettement (33 occurrences contre 3 pour la ponctuation faible).

Toutes les occurrences de ponctuation faible concernent des propositions à sujet pronominal ou absent, ayant un lien logique fort avec la proposition précédente marqué par un connecteur correspondant :

Manuscrits antérieurs à 1400

- (6.67) Car ainsi feust il defaillanz || *et* nonpoissanz d'aucune || chose ; *et* de tant fust || il hons mostels · (im574, 5a:10) ;
- (6.68) Car li || mal li sont contraire ; *et* || pource couuient il qu'i || se traient en sus de+lui · (im574, 5b:11) ;
- (6.69) Car ce || n'est fors que fiens et || ordure ; si le couuient || descendre en *par*font · (im574, 5b:16).

Dans les trois cas la première proposition commence par la conjonction *car* associée à une ponctuation forte²³³. On peut supposer que la relation exprimée par cette conjonction porte sur les deux propositions suivantes et nous avons donc affaire à des propositions qui partagent un élément commun. Une autre hypothèse serait que *car* « attirant » la ponctuation forte (c'est le cas des 13 occurrences dans notre extrait), la proposition suivante aurait plutôt tendance à être ponctuée faiblement. Cela se confirme dans 9 occurrences sur 13. Parmi les contrexemples, la deuxième proposition commence elle-même par *car* une fois, une autre fois la proposition avec *car* se trouve en fin de paragraphe et le cas suivant la deuxième proposition est assez longue pour « distancer » une nouvelle ponctuation forte. La dernière exception est tout aussi intéressante :

- (6.70) Car ce couuient || il *par* nature ;
Ausi *comme* || l'en uoit de l'ordure
du || vin qui est mis el vaissel ; ||
que li laiz se de _part || du bel ;
Si que li bons || demeure en haut (im574, 5c:8).

C'est en effet l'un des exemples où on aperçoit clairement les traces de la version originale en vers : nous avons disposé l'exemple de façon à mettre en évidence les anciennes rimes (rappelons que les sauts de ligne réels du manuscrit sont représentés par des barres verticales doubles). On voit très bien que la ponctuation apparait ici aux anciennes fins de vers qui correspondent à une limite syntaxique d'un certain niveau. Ainsi, la complétive *que li laiz...* est ponctuée, contrairement à la majorité des propositions de ce type. Les subordinées circonstancielles (*Aussi comme...* et *Si que li bons...*) portent une ponctuation forte tandis qu'on attend plutôt une ponctuation faible dans cette position (cf. plus bas). En revanche, il n'y a pas de ponctuation entre *l'ordure* et *du vin*, car la frontière syntaxique entre un nom et son modificateur n'est pas, de

233 Notons au passage le *comma* dans (6.69). Nous reviendrons plus bas sur l'usage de cette marque.

façon générale, « ponctuable ». Il est donc possible que le facteur des anciennes limites de vers favorise le renforcement de la ponctuation.

L'usage de la ponctuation est variable à la frontière des propositions ayant un élément commun : on compte 3 occurrences de ponctuation forte, 8 occurrences de ponctuation faible (dont 4 avec un *comma*)²³⁴ et 4 occurrences d'absence de ponctuation. La ponctuation faible est apparemment la plus typique dans cette position.

Tous les cas d'absence de ponctuation entrent dans une catégorie précise : il s'agit de propositions courtes avec un sujet commun (« coordination de prédicats ») :

(6.71) Car diex fu || deuant *et* sera touz iourz (*im574: 4c:16*).

S'il y a plus de deux propositions avec un sujet commun, la situation est plus complexe :

(6.72) Et touz biens ||[*col. b*] sont siens a droiture . || *et* par nature s'en vount || a+lui . Car de lui viennent || touz *et* mueuent . || *et* a+lui tiennent leur droit || chemin . (*im574, 5a19:-b:4*).

Dans cet exemple, nous avons 5 propositions avec un sujet commun *touz biens*. La deuxième (*et par nature...*) et la cinquième (*et a lui tiennent...*) commencent par une ponctuation faible, comme la majorité des occurrences de ce type. La troisième proposition contient un « quantifieur flottant » *touz* (Riegel *et al.* 1997, p. 212) qui n'est en effet qu'une reprise du déterminant du sujet dans la première proposition. La présence de ce quantifieur nous amène à croire qu'il s'agit dans ce cas non pas d'une proposition, dont le sujet se trouve dans une proposition précédente, mais d'une proposition avec un sujet nul, qui doit être classée parmi les propositions autonomes, et l'usage de la ponctuation forte est donc tout-à-fait normal. Par ailleurs, comme nous l'avons déjà remarqué, le connecteur *car* « attire » incontestablement la ponctuation forte dans **im574**. La quatrième proposition (*et mueuent*), qui n'est pas ponctuée, est très étroitement liée à la troisième et présente un cas typique de coordination de prédicats.

Les frontières entre les propositions principale et subordonnée sont ponctuées dans près de la moitié des cas : sur un total de 30 occurrences, on trouve 6 ponctuations fortes (dont 2 avec un *comma*) et 8 ponctuations faibles (dont 4 *comma*).

²³⁴ On peut y ajouter les 3 occurrences avec *car* en première proposition (6.67), (6.68) et (6.69).

Nous reviendrons plus bas sur les occurrences du comma, mais analyserons d'abord les autres cas. Les trois occurrences du point suivi d'une majuscule concernent toutes les propositions circonstancielles. Nous avons déjà analysé deux de ces occurrences dans (6.70) : ce sont les anciennes limites de vers qui semblent favoriser la ponctuation forte. Le même facteur peut être retenu pour une troisième occurrence (nous avons souligné les mots qui riment) :

- (6.73) Tant a li maus |[fol. 6] plus d'oscurte *et* de douleur || en enfer : ou il est || touz dis : *et* sera tant *comme* || diex sera en paradis : || Ou diex a touz biens || deuant soi · (im574, 6a:4).

La dernière occurrence est plus difficile à expliquer :

- (6.74) Qvant || diex fist le mon||de au commencement : || Il ne+li en estoit nul mestier · (im574, 4c:11).

Il s'agit ici d'une proposition principale qui suit sa subordonnée circonstancielle. La principale commence sur une nouvelle ligne et son premier mot est le pronom *il* (dans un emploi impersonnel). Notons que les deux autres occurrences de *il* au début d'une proposition sont ponctuées de la même façon, mais il s'agit dans les deux cas de propositions autonomes. L'hypothèse d'une erreur du scribe ne semble pas improbable dans ce cas, mais de toute façon l'emploi d'une ponctuation à une frontière entre propositions n'a rien d'exceptionnel.

Sur les 7 occurrences de ponctuation faible à la frontière entre une principale et une subordonnée, 4 sont avec un *comma*. Une occurrence de cette catégorie a déjà été analysée en (6.70), une autre est une relative circonstancielle en *ou* et encore une autre est une comparative. Il est difficile de déterminer quel facteur a joué le rôle principal dans ces cas : il faudrait disposer d'un nombre plus important d'occurrences pour tester les hypothèses possibles. Une chose est sûre : dans **im574** la ponctuation aux frontières entre les propositions principale et subordonnée n'est pas rare, mais nettement moins fréquente qu'entre des propositions indépendantes l'une de l'autre.

En ce qui concerne les syntagmes coordonnés à l'intérieur d'une proposition, leur nombre est assez élevé : 30 au total (pour un échantillon de 549 mots). 12 de ces occurrences sont associées à une ponctuation, toujours faible.

Une bonne partie de ces occurrences correspond à une seule proposition avec 5 objets coordonnés :

(6.75) Il n'ot onques ne faim · || ne soif · ne tans · ne ²³⁵ || mois · ne heure · (im574, 4d:4-6).

C'est un cas typique d'énumération que Mario Roques (1952, p. 194) avait identifié comme un lieu favorable à la ponctuation.

Dans d'autres cas, la ponctuation apparaît entre certains des syntagmes coordonnés seulement :

(6.76) Car sa || bontez est toute pure et || saintisme · *et* saine *et* nete || sanz nul mal · (im574, 5b:9).

Une certaine régularité semble se dégager : s'il y a trois ou quatre éléments coordonnés, la ponctuation apparaît devant le troisième (c'est-à-dire après la première paire), mais cette règle n'est pas sans exception. Parfois, la ponctuation est différente dans des conditions très similaires, comme dans les deux paires de syntagmes prépositionnels avec *sanz* dans l'exemple suivant :

(6.77) *et* ades || les aura sanz painne · || *et* sanz annui · Touz les || a · *et* touz les enlumine || sanz nulle defaute · et || sanz nul termine · (im574, 6a:7-9).

L'extrait qui fait partie de notre corpus ne contient qu'une énumération plus ou moins longue. Et pourtant certaines parties de l'œuvre de Gossouin contiennent de longues listes, notamment de noms géographiques. Il est intéressant de regarder comment la ponctuation fonctionne à ces endroits du texte :

(6.78) Puis est danemar||che · Et hongrie · Osteri||che · Et puis germanie|| qui a maint regne vers || occident · Si i est soabe · || Et alemaingne ou u||ne eaue sourt qui a *nom* || dunoe · qui sespant et || court par .vii. flueues · || Si i est yllande · Escoce · ||[fol. 68r] Et angleterre · Et tou||te france · (im574, 67d:10-68a:2).

La densité de la ponctuation forte dans cet extrait est remarquable. On peut sans doute interpréter chaque syntagme contenant un nom de pays comme une proposition indépendante, avec une omission possible du prédicat, mais la fréquence exceptionnelle de la ponctuation forte a ici une autre explication : chaque nom de pays dans cette liste

235 Le « point d'exclamation » ici n'est pas une marque de ponctuation syntaxique mais tout simplement un « outil » de justification de fins de ligne et peut-être de liaison entre les éléments d'un syntagme soudé (cf. les occurrences du même type dans im587).

porte une information importante et doit être mis en évidence et clairement séparé des autres, quelle que soit la structure syntaxique. Dans deux cas, la majuscule apparaît au début d'un nom propre (*Osteriche* et *Escose*), mais puisque tous les autres noms géographiques, qui ne se trouvent pas au début d'une unité ponctuable, ne prennent pas de majuscule, nous avons toutes les raisons de les interpréter comme des majuscules de ponctuation syntaxique plutôt que comme des majuscules de noms propres (ponctuation de niveau lexical)²³⁶.

Notons enfin que dans **im587** le passage correspondant ne contient que deux majuscules, purement « syntaxiques » :

- (6.79) Apres vient danemarche / hongrie / osteriche · gernemie || soaue ou vn flum que l'on apele la+dunoe sort qui se || depart en .vii. flums / Apres vient · engleterre · irlande · esco||ce · france . (*im587, 182a:8-11*).

En revanche, la ponctuation faible (point ou barre oblique) est exceptionnellement dense dans ce texte où le taux de ponctuation général est plutôt faible.

Tout ceci confirme notre hypothèse que les listes (ou énumérations) constituent un type de séquence textuelle avec des règles de ponctuation très particulières.

Revenons maintenant à **im574**. Le dernier type d'unités ponctuables, les syntagmes particuliers à l'intérieur d'une proposition, ne présente dans notre extrait qu'une seule occurrence :

- (6.80) [tout li] est touz || iourz deuant les ieulz · || autresi bien li loing *comme* || li pres · (*im574, 4d:12*).

Il s'agit d'une construction comparative à deux parties, séparée du début de la proposition. Même si elle ne constitue pas une proposition à part entière, cette construction est suffisamment autonome sur le plan syntaxique et sans doute intonatif pour expliquer l'emploi d'une marque de ponctuation.

Comme nous l'avons déjà remarqué, **im574** présente un nombre relativement élevé d'occurrences du *comma* : près de 17% des marques de ponctuation relevées sont en effet des *comma*. Leur emploi n'est pas limité à un type de frontière syntaxique, mais

²³⁶ Cela fait la différence avec **villehardouin12203**, où dans un passage similaire (longue énumération de noms de personnes, exemple (6.46)), des majuscules de noms propres pouvaient apparaître en dehors de frontières ponctuables.

la majorité des occurrences se regroupent autour de deux pôles : propositions avec un sujet commun (4 occurrences) et limites entre les propositions principale et subordonnée (7 occurrences). Une seule occurrence échappe à cette règle, celle que nous avons citée dans (6.69) : proposition autonome en *si* qui suit une proposition courte en *car*. Nous avons conclu que cette construction était proche des propositions ayant des éléments communs.

Dans deux cas, le *comma* est suivi d'une majuscule, il s'agit donc d'une ponctuation forte :

- (6.81) Et de tant || *comme* li bons deüent || plus clers ; De tant retient || la lie plus d'ordure ·
(*im574, 5d:4*) ;
- (6.82) Et *con* plus || esclairsist li biens de||uant dieu ; *et* plus s'es||ioist ; Tant a li maus ||
[*fol. 6r*] plus d'oscurte *et* de dou||leur en enfer ; ou il est || touz dis ; *et* sera tant *com*||
me diex sera en para||dis · (*im574, 5d:17-6a:3*).

La ressemblance des deux occurrences est remarquable : il s'agit d'une deuxième partie dans une construction corrélatrice (*de tant*) *comme...* (*de*) *tant*. Ce sont donc les conditions « classiques » pour l'emploi d'un *comma*. Mais pourquoi utilise-t-on la majuscule ? On aperçoit que les deux parties corrélées sont assez longues. Dans (6.82), chaque partie est composée de plusieurs propositions et contient des ponctuations faibles. On peut supposer que le scribe était conscient que ce type de construction corrélatrice était une structure syntaxique de niveau supérieur à une simple proposition, et a utilisé une majuscule pour représenter cette hiérarchie. Le *comma* a servi pour sa part à souligner le lien logique (voire intonatif) entre les deux parties. On peut peut-être parler dans ce cas d'une ponctuation d'un niveau intermédiaire, entre le fort et le faible.

Un autre phénomène est saillant dans (6.82) : une succession de 4 occurrences du *comma* sur 5 lignes. Les raisons de cette accumulation d'occurrences d'une marque généralement peu fréquente sont sans doute extralinguistiques : c'est un choix subjectif du scribe déterminé par des facteurs externes au texte et donc inaccessible à notre analyse.

Manuscrits antérieurs à 1400

En dehors de ce passage extraordinaire, on trouve deux occurrences du comma entre des propositions courtes avec la répétition d'un élément (objet direct) utilisée comme figure stylistique :

(6.83) Tout uoit ¶ tout || tient en sa+main · (im574, 4d:2) ;

(6.84) Touz les || a ¶ et touz les enlumine (im574, 6a:9).

Les deux occurrences sont très similaires, même si leurs structures syntaxiques ne sont pas identiques : dans (6.83) les propositions sont juxtaposées et l'objet est exprimé par le pronom *tout* ; dans (6.84), les propositions sont coordonnées et l'objet est exprimé par un syntagme nominal composé d'un pronom personnel (forme du cas régime pluriel) et du « quantifieur flottant » *touz*. Il est possible que la ressemblance apparente joue un rôle plus important dans le choix de la ponctuation qu'une analyse syntaxique fine.

Les autres occurrences du *comma* ne semblent pas liées à des conditions syntaxiques particulières qui puissent les différencier des occurrences du point. Aucune occurrence ne semble contredire l'hypothèse logico-intonative de Ch. Marchello-Nizia et le caractère facultatif de l'usage de cette marque se confirme.

O.H. Prior dans son édition de *l'Image du mode* respecte assez souvent la ponctuation du manuscrit BnF fr. 574 (son manuscrit de base). Sur 121 unités « ponctuables », une marque de ponctuation dans l'édition correspond à une marque du manuscrit dans 56 cas et dans 40 cas les unités ponctuables ne sont ponctuées ni dans le manuscrit, ni dans l'édition. On compte dans l'édition 9 marques de ponctuation ajoutées et 16 supprimées par rapport au manuscrit.

4 occurrences d'ajouts surviennent devant un complément circonstanciel introduit par la préposition *sanz* ('sans') :

(6.85) Car il est diex entierement [> ,] || sanz commencement || et sanz fin [. > .] (im574, 5a:13).

Ces compléments sont facultatifs dans la structure syntaxique de la proposition et apportent une précision à l'information principale véhiculée par celle-ci. C'est cette valeur pragmatique de précision qui explique l'utilisation d'une ponctuation dans ce

type de contexte en français moderne. Le scribe, pour sa part, est resté conséquent dans le choix de ne pas ponctuer les syntagmes particuliers à l'intérieur d'une proposition.

Les autres occurrences d'ajout concernent les propositions relatives que l'éditeur a jugées suffisamment disjointes pour insérer une ponctuation. Là aussi, le manuscrit ne fait qu'appliquer la règle générale de ne pas mettre de ponctuation autour des relatives.

Les cas de suppression de la ponctuation scribale dans l'édition se répartissent en deux catégories : les frontières de propositions coordonnées avec un élément commun (6.86) et la coordination de syntagmes à l'intérieur d'une phrase (6.87). Dans les deux cas, toutes les occurrences relevées (sauf une) présentent l'usage de la conjonction *et* :

- (6.86) Si que li bons || demeure en haut [.>] Et la || lie demeure au fonz [>] *qui* || est mauuaise [.>] (*im574, 5c:13*) ;
- (6.87) **E**t li maus [>] qui est || obscurs *et* laiz [.>] *et* tenebreus || seur toute rien [>] || laist le bien *et* descent aual [.>] (*im574, 5c:4*).

Notons que les deux exemples donnés ci-dessus présentent à la fois l'ajout d'une marque de ponctuation par l'éditeur autour de propositions relatives et de la suppression d'une marque entre propositions ou mots coordonnés.

L'ensemble des observations présentées ci-dessus nous permet de conclure que **im574** présente un système de ponctuation assez logique et cohérent. L'ensemble des règles qui se dégagent sont proches de celles observées dans **villehardouin12203** et **tyr828**, avec toutefois une distinction plus nette des niveaux syntaxiques (pas de ponctuation de syntagmes particuliers).

6.9. Paris, BnF, fr. 13568 (*Vie de saint Louis de Jean de Joinville*)

Le manuscrit BnF fr. 13568 est un volume qui compte 391 pages²³⁷ en parchemin de 220 × 150 mm, avec le texte disposé sur deux colonnes. D'après son écriture, certains chercheurs (V.L. Delisle, Ch. Samaran) le dataient du troisième quart du XIV^e

²³⁷ Ce manuscrit a la particularité d'être numéroté en pages et non en feuillets.

Manuscrits antérieurs à 1400

siècle. Un examen attentif de ses peintures a toutefois amené J. Monfrin (1998, p. XCIV) à faire remonter la datation aux années 1330 – 1340. L'écriture du manuscrit est soignée et régulière. Le travail du scribe a été revu par un réviseur qui inscrivait en marge les corrections introduites ensuite dans le texte, après grattage. L'exécution du manuscrit peut être approximativement localisée dans le Nord-Est de la France.

L'histoire de la composition de la *Vie de saint Louis* (aussi connue sous les titres de *Mémoires* ou d'*Histoire de saint Louis*) par Jean Joinville (1225 – 1317) est assez complexe. Le noyau du livre est constitué par le récit de la croisade qu'il avait rédigé, sans intention de le publier, vers 1272. Mais c'est vers 1305 que Joinville a entrepris, sur commande de la reine Jeanne de Navarre, la rédaction de l'ouvrage. Celui-ci a été achevé et dédié à l'héritier du trône (le futur roi Louis X) en 1309. Puisqu'il s'agit d'un récit concernant des événements historiques, on peut classer l'ouvrage dans le domaine historique, mais l'auteur avait bien en vue l'objectif « d'édifier le jeune prince, en lui proposant de suivre l'exemple de son bisaïeul, plutôt que de prendre modèle sur son père, Philippe le Bel » (Hasenor *et al.* 1994, p. 792). L'œuvre de Joinville n'est donc pas étrangère au domaine didactique.

La transcription intégrée dans notre corpus s'étend de la première à la septième page et contient 1098 occurrences-mots. Nous nous y référerons désormais sous l'identifiant **joinville13568**²³⁸.

Le taux de ponctuation général dans notre extrait atteint les 9,3%. La ponctuation faible est légèrement plus fréquente que la ponctuation forte (58 contre 44 occurrences respectivement). La marque de ponctuation largement majoritaire est le point, mais on trouve également 3 occurrences du *comma*.

La macro-structure du texte de Joinville est particulièrement bien mise en forme dans le manuscrit 13568 qui présente tout un système d'initiales associées à des alinéas de 1 à 4 lignes. L'usage de ces initiales est lié à la structure logique du texte. Tout au début, l'auteur s'adresse à son destinataire et présente une sorte de résumé de son ouvrage composé de deux livres, dont le premier se divise en deux parties. La présentation de chaque partie commence par une initiale (une ligne de hauteur pour les parties et deux pour le livre). Le corps de la narration commence par le rappel des

238 Cf. le sigle DEAF **Joinv.**

événements au cours desquels saint Louis a risqué sa vie pour la bonne cause. Les événements sont énumérés (*le premier fait...*, *la seconde foiz...*, etc.), et une initiale de deux lignes ouvre la description de chacun d'eux.

De temps en temps, de petites initiales en couleur apparaissent au milieu d'une ligne. Celles-ci ne correspondent pas forcément à des unités de la structure narrative, et leur emploi peut être motivé par des raisons diverses. Dans (6.88), par exemple, le même syntagme *a la fin de sa vie* se répète à la fin d'une proposition et au début d'une autre :



Fig. 38 : Initiale en couleur au milieu d'une ligne (*joinville13568, 2d*).

(6.88) on||ques *h*omme lay de nos||tre temps ne vesqui si || saintement de tout || son temps des le com||mencement de son reg||ne . jusques a la fin || de sa uie . *A* la fin de || sa uie ne fus ie mie . (*joinville13568, 2d:20*).

Ici l'initiale sert sans doute à souligner le fait que la répétition est utilisée par l'auteur à des fins stylistiques et ne relève pas d'une erreur du copiste²³⁹. Notons par ailleurs que la première lettre du mot *jusques* est légèrement plus grande que les minuscules autour. C'est l'un des cas où il est difficile de dire s'il s'agit d'une ponctuation forte ou faible. Nous reviendrons sur les conditions syntaxiques de cette occurrence plus bas.

L'extrait que nous étudions ne contient pas de passage citant le discours direct d'un personnage, mais au début du texte l'auteur s'adresse directement au destinataire en utilisant l'appellatif *sire* :

(6.89) Chier sire || ie vous foiz a sauoir || que ma dame la roy||ne uostre mere (*joinville13568, 1b:2*).

Comme dans tous les manuscrits que nous avons étudiés précédemment, le mot en apostrophe n'est pas séparé par une ponctuation du reste de la proposition, mais une

239 Il convient de noter que Michèle Perret (1981) a choisi cette phrase pour titre de son article qui porte sur le témoignage dans cette œuvre de Joinville.

Manuscrits antérieurs à 1400

ponctuation forte le sépare bien de la proposition précédente. Le passage contient par ailleurs deux citations, dont l'une (6.90) provient de la Sainte Écriture, tandis que l'origine de l'autre reste non identifiée :

- (6.90) Sire pour+ce || qu'il est escript . Fai pre||mier ce qu'il a _fiert a || dieu . et il te adredera . || toutes ces autres be||soignes ai ie fait es||crire . ce qui a fiert aus || [col. d] troiz choses desus dites . (joinville13568, 2c:16) ;
- (6.91) Et de ce est escript . || Se tu creins dieu si te || creindront toutes les || riens qui te verront . || Et ceste demouree fist || il tout contre son con||seil . (joinville13568, 6c:3).

Dans les deux cas, le début de la citation est marqué par une ponctuation forte. La fin de la citation porte une ponctuation forte dans (6.91), mais il faut remarquer que dans ce passage la citation s'insère dans une série de propositions indépendantes et la ponctuation peut être déterminée davantage par la présence de ce type de frontière que par le changement de plan énonciatif. L'exemple (6.90) est plus complexe. Ici la citation est « accrochée » à une proposition subordonnée circonstancielle (*pour ce que...*), dont la principale (*ai je fait escrire...*) suit immédiatement la citation. Aucune ponctuation ne sépare le début de la principale de la fin de la citation. En revanche, il y a bien une ponctuation faible entre le verbe et l'objet direct à l'intérieur de la citation et entre le verbe de la principale et son objet exprimé par le pronom *ce* avec une relative *qui a fiert*. Dans ce cas, on peut soit admettre que la ponctuation est aberrante (contraire au principe de la hiérarchie des unités syntaxiques), soit supposer que le scribe a interprété la structure du passage différemment : la citation s'arrête au mot *adredera*, la proposition suivante va de *toutes* à *escrire*, et la dernière proposition (*ce qui a fiert...*) est une sorte d'addition. Cette interprétation est certes moins logique du point de vue sémantique, mais elle n'est pas agrammaticale.

Les frontières entre des propositions indépendantes sont presque toujours ponctuées : 29 occurrences de ponctuation forte, 9 occurrences de ponctuation faible et une occurrence d'absence de ponctuation.

La seule occurrence enregistrée d'absence de ponctuation se trouve dans le contexte suivant (devant *et il ne uoult...*) :

- (6.92) Et || se il demouroit en sa || neif . par son cors pe||ust il recouurer a re||conquerre la terre de || egypte . et il ne uoult || nullui croire . ains sail||li en la mer tout ar||me . (joinville13568, 4c:15).

On peut supposer que pour le scribe cette nouvelle proposition était coordonnée avec la précédente et partageait avec elle une subordonnée conditionnelle commune (*se il demouroit...*). Une telle structuration rend le sens du passage obscur, mais en cas de lecture rapide, qui ne vise pas à la compréhension totale de structures complexes, cette incohérence peut passer inaperçue.

En ce qui concerne le choix entre la ponctuation forte ou faible à la frontière de propositions autonomes, nous n'avons pas réussi à identifier de facteurs qui le conditionnent. La ponctuation forte est dominante, mais l'usage de la ponctuation faible ne semble pas être se caractériser par des conditions particulières. 8 occurrences sur 9 de la ponctuation faible sont suivies d'un connecteur de coordination (*et, car* ou *ains*), mais ces mêmes connecteurs sont aussi bien utilisés avec la ponctuation forte. Notre hypothèse est que l'opposition entre la ponctuation forte et faible n'était pas considérée comme très importante par le scribe de **joinville13568**, et qu'il faisait son choix plus ou moins librement. Il avait néanmoins une tendance sans doute inconsciente à privilégier la ponctuation forte à des frontières syntaxiques entre des propositions indépendantes.

Les propositions avec des éléments communs sont peu nombreuses dans notre extrait. Sur un total de 7 occurrences, 6 portent une ponctuation faible et l'une d'entre elles, une ponctuation forte. La ponctuation forte est utilisée entre deux propositions complétives dans un discours indirect, coordonnées à l'aide de la conjonction *mez* ('mais') :

(6.93) [II] dist que son peuple ne || lairoit il ia ! Mez feroit || tele fin comme il feroi||ent !
(*joinville13568, 5a:8*).

Ici la ponctuation forte peut servir à souligner la fermeté avec laquelle saint Louis s'engageait à partager le sort de son peuple. Cette façon de ponctuer évoque des phrases courtes, espacées de pauses bien marquées, prononcées par une personne exprimant une décision irrévocable. D'autre part, comme nous l'avons déjà indiqué, l'opposition entre la ponctuation forte et faible ne semble pas jouer un rôle fonctionnel très important dans **joinville13568**.

Manuscrits antérieurs à 1400

La ponctuation entre les propositions subordonnées et leurs principales apparaît très variable : elle est absente dans 59 occurrences (soit près de 60%) , la ponctuation faible est utilisée dans 28 cas ,et la ponctuation forte dans 9 cas.

La ponctuation forte, la moins fréquente dans cette position, ne semble pas pour autant être associée à un type de frontière particulier : on trouve des occurrences qui sont associées à des circonstancielles, relatives et complétives, qu'elles se trouvent avant ou après la proposition principale. Il n'y a aucune occurrence de ponctuation forte aux frontières de propositions infinitives et participiales, plus étroitement liées à la principale, mais le nombre peu élevé d'occurrences de ces propositions ne permet pas d'affirmer que la ponctuation forte y est exclue.

Certains passages semblent toutefois propices à l'usage d'une ponctuation forte :

- (6.94) Le secont liure || nous parlera || de ses granz cheualeri||es . et de ses granz har||demens .
Les quiex sont || tiex . Que ie li vi qua||tre foiz mettre son cors || en aenture de mort . ||
Aussi comme vous || orrez ci apres pour es||pargnier le doumage || de son peuple .
(*joinville13568, 3b:2-13*).

Dans cet exemple, nous avons une construction syntaxique assez complexe, où une principale contient deux compléments de prédicat coordonnés (dont chacun est un syntagme de 4 mots), suivie d'une relative en *lequel*, suivie à son tour d'une complétive. Vient ensuite une subordonnée comparative (*Aussi comme vous aurez...*) jouant le rôle d'un organisateur textuel (elle renvoie à un passage situé plus loin dans le texte où l'épisode mentionné sera présenté plus en détail) qui est suivie d'une proposition infinitive servant de complément circonstanciel à la proposition d'avant. La relative, la comparative et, ce qui est plus étonnant, la complétive sont marquées par une ponctuation forte. Il se peut que dans ce passage l'usage de la ponctuation forte soit favorisé par le fait qu'il s'agit d'un résumé du second livre, où chaque proposition est sensée représenter un bloc de texte important dans la suite de l'ouvrage. Même si le passage en question n'a pas la mise en forme typique pour une table de matières, les règles de ponctuation peuvent y être différentes du corps du texte, comme c'est le cas dans les longues listes auxquelles nous avons été confronté dans **villehardouin12203** et **im574**.

Le pronom *ce* antécédent des relatives semble se rattacher plutôt à la subordonnée qu'à la principale :

- (6.95) ai ie fait es||crite ¶ || ce qui afiert aus ||[col. d] troiz choses desus dites ¶ || c'est a sauoir ¶
ce qui af||fiert au profit des ames || et des cors ¶ et ce qui af||fiert au gouuernement ¶ du
peuple ¶ (*joinville13568, 2c:20-d:4*).

Dans cet exemple, nous avons une triple répétition de la formule *ce qui afiert* ('ce qui convient'), précédée à chaque fois d'une ponctuation (soit directement, soit espacée par la conjonction *et*). Ici le facteur de répétition joue sans doute un rôle : la ponctuation met en valeur chaque élément dans l'énumération : dans le système graphique moderne on a souvent recours à des listes « à puces » dans des cas similaires. Néanmoins, le positionnement du point devant *ce* est un indice clair de son rattachement à la subordonnée.

Les syntagmes coordonnés à l'intérieur d'une proposition ne sont pas très fréquents dans notre extrait : 18 occurrences au total, dont 6 avec une ponctuation faible et une avec une ponctuation forte.

La ponctuation forte survient devant un troisième complément d'objet direct coordonné représenté par un syntagme nominal, qui régit à son tour deux groupes prépositionnels :

- (6.96) Car il auoit dou||[p. 5]ble tierceinne et me||noison moult fort ¶ || Et la maladie de l'ost ||
en la bouche et es iam||bes ¶ (*joinville13568, 5a:2*).

C'est donc encore une fois une construction complexe où la ponctuation sert à indiquer des blocs de syntagmes.

Notre extrait présente une occurrence d'un complément circonstanciel représenté par une construction corrélatrice en *que... que* ('aussi bien... que') :

- (6.97) le roy enuoia || querre .xiiij. mestres || nothonniers ¶ que de || celle neif que d'autres ||
qui estoient en sa com||paignie ¶ pour li con||seiller que il feroit ¶ (*joinville13568, 6d:8-9*).

Une ponctuation faible est présente à gauche et à droite de cette construction, et il n'y a pas de ponctuation à l'intérieur. La ponctuation correspond donc à la hiérarchie des unités syntaxiques.

Manuscrits antérieurs à 1400

Tous les cas d'absence de ponctuation entre syntagmes coordonnés concernent la frontière entre le premier et le deuxième terme coordonnés à l'aide de la conjonction *et*. Si un troisième terme est présent, il est séparé par une ponctuation :

(6.98) Salut et amour ¶ et [[col. b] honneur ¶ et son serui||se appareille ¶ (*joinville13568, 1a:11*).

Bien entendu, il ne s'agit ici que d'une hypothèse à vérifier dans l'avenir sur un échantillon beaucoup plus important de ce manuscrit.

La ponctuation autour de syntagmes particuliers est assez rare dans **joinville13568** : on a relevé dans cette catégorie 5 occurrences de ponctuation faible et deux ou trois occurrences de ponctuation forte.

Le nombre d'occurrences de la ponctuation forte est incertain à cause de la difficulté d'interprétation de la « casse » de la lettre *j* déjà citée dans l'exemple (6.88). Syntactiquement, il s'agit ici de la deuxième moitié d'un complément circonstanciel exprimant une étendue temporelle ('depuis le commencement de son règne jusqu'à la fin de sa vie'). Du point de vue purement syntaxique la ponctuation paraît peu probable ici. Mais il se peut que le scribe ait utilisé la ponctuation pour souligner la distance entre les deux événements.

Une autre occurrence de la ponctuation forte qui sert à marquer un syntagme particulier survient dans le passage suivant :

(6.99) ains sail||li || en la mer tout ar||me ¶ L'escu au col ¶ le || glaiue ou poing ¶ et || fu des premiers a ter||re ¶ (*joinville13568, 4c:18*).

Il s'agit ici d'un complément circonstanciel rattaché à un attribut du sujet. Dans ce passage le rôle de la ponctuation est avant tout stylistique : le découpage des unités permet d'accentuer l'expressivité de la scène, car chaque détail est ainsi mis en valeur.

Plusieurs occurrences de ponctuation intrapositionnelle se trouvent dans la longue formule de dédicace initiale :

(6.100) **A** S_on bon || seigneur || looys filz || du Roy de || france ¶ par la grace de || dieu roy de nauarre ¶ || de champaigne et de bri||e conte palazin ¶ J_ehan || sire de ioinuille son se||neschal || de champaigne ¶ || Salut et amour ¶ et [[col. b] honneur ¶ et son serui||se appareille ¶ (*joinville13568, 1a:1-1b2*).

Ici la ponctuation apparaît entre des syntagmes apposés et coordonnés. L'emploi d'une majuscule dans *Jehan*, qui constitue formellement une ponctuation forte, est motivé avant tout par le marquage du nom propre de l'auteur.

Les trois occurrences du *comma* dans notre extrait sont utilisées dans des conditions similaires, même si chacune a ses particularités :

- (6.101) uostre mere qui || moult m'amoit a cui || dieu bone merci face ; || me pria si a certes con||me elle pot que ie li fe||isse faire .i. liure (*joinville13568, 1b:7*) ;
- (6.102) **L**a seconde foiz |[col. d] qu'il mist son cors || en auanture de mort || si fu tele . que au par||tir qu'il fist de l'auma||sourre pour uenir a da||miete ; Son conseil || li loa si comme l'en || me donna a entendre || que il s'en uenist a da||miete en galies . (*joinville13568, 4d:5*) ;
- (6.103) Car quant le roy fu de||moure en acre ; pour || vn home a armes que || il auoit en sa compaign||nie . Ceulz d'acre en a||uoient bien .xxx. quant la uille fu prise . (*joinville13568, 5b:8*).

Dans (6.101) le *comma* est employé devant la reprise d'une principale après une relative insérée, un contexte « classique » pour cette marque de ponctuation. Dans (6.102), le *comma* suivi d'une majuscule est aussi utilisé devant une reprise de la principale, mais cela se passe dans le cadre d'une construction complexe avec plusieurs niveaux de subordonnées et de nombreux circonstants. La combinaison *comma*-majuscule peut servir à marquer une frontière importante dans cet ensemble complexe tout en signalant que la fin n'est pas atteinte. Dans (6.103), le *comma* se trouve devant un circonstant de cause d'une subordonnée temporelle. Ce circonstant régit pour sa part une relative, et c'est après cette relative que commence la principale de la temporelle. Bref, une autre construction complexe, mais ici le *comma* est utilisé au niveau hiérarchique le plus bas, tandis qu'une frontière plus importante est marquée par une ponctuation forte « classique ».

J. Monfrin dans son édition (1998) suit assez souvent la ponctuation du manuscrit : dans 70 cas une marque de ponctuation de l'édition correspond à une marque présente dans le manuscrit et dans 56 cas, la ponctuation est absente à une frontière ponctuable aussi bien dans le manuscrit que dans l'édition.

Manuscrits antérieurs à 1400

Dans 28 cas, la ponctuation du manuscrit est supprimée dans l'édition. Il s'agit le plus souvent des frontières entre subordonnées et principales, mais aussi de propositions et de syntagmes coordonnés. L'éditeur a ajouté une marque de ponctuation dans 25 cas. Il s'agit là aussi quelquefois de frontières de propositions, mais la différence la plus notable est l'ajout de la ponctuation pour marquer les noms en apostrophe et les appositions.

De façon générale, **joinville13568** présente un système de ponctuation assez cohérent, proche de celui de **im574**.

6.10. Lyon, BM, 768 (*Prophéties du maitre Joachim*).

La datation dont nous disposons pour ce manuscrit lyonnais est très approximative : le XIV^e siècle ; la région de son exécution est elle aussi inconnue. Le codex contient 179 feuillets en parchemin et mesure 233 × 165 mm. Outre les *Prophéties* de Joachim (f. 97v – 100), il contient un fragment en prose des prophéties des Sybilles (f. 95v - 97) et le *Pèlerinage de la vie humaine et de l'âme* de Guillaume de Digulleville (f. 1 – 95 et 100r – 179). L'écriture du manuscrit est assez soignée, mais il n'est pas très richement décoré : les décorations se limitent à des initiales et titres à l'encre rouge ou bleue.

Du point de vue typologique, ce texte appartient incontestablement au domaine religieux et son genre peut sans doute être défini comme une prédication. Nous n'avons pas relevé de formes marquées dialectalement dans l'extrait étudié.

A notre connaissance, les *Prophéties* de Joachim n'ont pas été éditées à l'époque moderne, nous avons donc élaboré nous-même le niveau normalisé de la transcription en appliquant les règles en vigueur.

Le fragment que nous avons transcrit correspond au début de la prophétie de Joachim situé aux folios 97v – 98r et contient 578 occurrences-mots. Son identifiant dans notre corpus est **joachim768**.

Avec 46 occurrences de marques de ponctuation (22 pour la ponctuation forte et 24 pour la ponctuation faible) le taux de ponctuation y atteint les 8,0%.

Au début du texte de Joachim se trouve un titre inscrit à l'encre rouge, et le corps du texte commence par une initiale de 3 lignes. Plus loin, des initiales en couleur apparaissent parfois au milieu d'une ligne, et quand elles se trouvent au début d'une ligne, le texte occupe entièrement la ligne précédente, ce qui nous amène à la conclusion qu'il s'agit dans les deux cas d'un même type d'initiales qui ne sont pas associées à un saut de ligne.

Dans la grande majorité des occurrences les passages marqués par ces initiales commencent par un même mot, *item* :

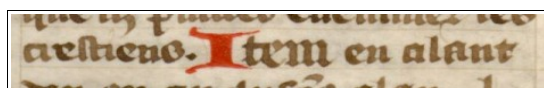


Fig. 39 : « Item » avec une initiale en couleur (joachim768, 97d).

Ce mot est souvent écrit en caractères nettement plus grands que le reste du texte. Il s'agit donc d'un mot (marque lexicale) associé à une mise en forme particulière (marque graphique) ayant pour fonction principale la structuration du texte. C'est un cas dans lequel le terme de 'marque de ponctuation lexicale' paraît entièrement approprié.

Le passage dépouillé contient par ailleurs une occurrence d'un *pied-de-mouche* entre la fin d'une courte introduction et le début de la première prophétie. Cette marque de ponctuation écrite à l'encre rouge est suivie de la forme *Premierement*, avec une grande majuscule marquée d'un trait rouge.

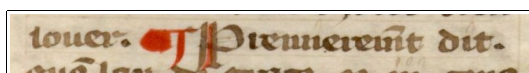


Fig. 40 : Pied-de-mouche (joachim768, 97c).

C'est sans doute un cas un cas intermédiaire entre une majuscule ordinaire et une initiale.

Toutes ces observations nous permettent de tirer une première conclusion selon laquelle **joachim768** présente un système de ponctuation « extra-forte » particulier conciliant l'économie de place (absence de sauts de ligne) et la facilité de repérage des macro-unités textuelles. Avec un total de 8 occurrences, la fréquence relative de la ponctuation extra-forte dans ce texte est la plus élevée de notre corpus.

Manuscrits antérieurs à 1400

En ce qui concerne les autres types d'unités ponctuables, la frontière entre des propositions indépendantes reste la position la plus favorable : 21 occurrences de ponctuation (dont 13 fortes) contre seulement 2 occurrences d'absence de ponctuation.

Dans un cas, l'absence de ponctuation est sans doute liée à une erreur (ou plutôt à une « hyper-correction ») du scribe :

(6.104) **I**tem en || celui temps *enuiron* l'an .l. iij. || aura *moult grant* guerre en grece || entre les turs *et les grecs* seront || *desconfiz par* .iij. foiz . (joachim768, 98a:12).

Ici, le syntagme *et les grecs* joue à la fois le rôle d'un complément dans la proposition [*il y*] *ara moult grant guerre* et celui du sujet dans la proposition suivante. Le scribe a sans doute supprimé le syntagme répété de sa source en croyant, à tort, que c'était une erreur. Après cette suppression, on ne sait plus où passe la frontière entre les propositions, l'absence de ponctuation est donc compréhensible.

L'autre occurrence d'absence de ponctuation est plus difficile à analyser. Un point apparaît en effet non pas avant mais après la conjonction *et* qui ouvre une nouvelle proposition :

(6.105) **I**tem *sera* || vne *grant* bataille *apres* laquelle sera || a+la *confusion* du noble sanc de || france . *et y*+ara pris vn roy et . || n'*entendez pas que* pour la+*prinse* de celui || ne pour la *desconfiture que* les guer||res ne le batailles *doient finer* . (joachim768, 98a:17).

La proposition précédente est courte, elle commence aussi par la conjonction *et* et est précédée d'une ponctuation. La frontière logique semble pourtant plus forte devant la deuxième proposition (*et n'entendez pas...*), car il y a un changement du temps verbal et du mode énonciatif : on passe de la narration à l'interpellation directe du lecteur. Pourquoi le point est-il placé après *et* ? Cela peut être une simple erreur. On peut supposer aussi que le point sert ici à remplir la fin de ligne (comme c'était le cas de marques spécifiques dans **im587**, **im574**, etc.). On peut enfin avancer l'hypothèse selon laquelle ce point marque effectivement une frontière (voire une pause à observer pendant la lecture) devant un changement de plan énonciatif.

L'hypothèse d'une marque servant à remplir la fin de ligne nous semble la plus probable, et l'absence de la ponctuation devant *et* s'explique sans doute par la distance trop courte de la ponctuation précédente.

Aucune condition formelle ne semble déterminer le choix entre la ponctuation forte ou faible à ce type de frontière. Sur 13 occurrences de la ponctuation forte 8 sont suivies de la conjonction *et*, tandis que pour la ponctuation faible la fréquence de cette conjonction est de 6 occurrences sur 8. La fréquence relative de la cooccurrence de cette conjonction avec la ponctuation faible à ce type de frontière syntaxique est donc plus élevée, mais le nombre total des occurrences dont nous disposons est trop bas pour juger si cette tendance se confirme dans l'ensemble du texte.

Notre extrait ne présente que 5 occurrences de frontières entre des propositions coordonnées ayant un élément commun, ce qui rend très hypothétique toute tentative d'en tirer des conclusions sur les tendances d'usage. Il est néanmoins intéressant de constater que 4 des 5 occurrences ne sont pas ponctuées et que la seule occurrence ponctuée est associée à une avancée temporelle dans la narration (exprimée par l'adverbe *après*) :

(6.106) les grecs seront || desconfiz par .iij. foiz **.** et apres aront || vitoire par la venue d'un prince (*joachim768, 98a:13*)²⁴⁰.

La ponctuation est rare à la frontières entre des propositions principales et subordonnées : sur 31 frontières 5 seulement portent une ponctuation, dont 1 forte. Cette dernière s'explique en effet par des raisons extralinguistiques : il s'agit de la première de la première ligne d'une colonne : une position dans laquelle le scribe de *joachim768* met presque régulièrement une majuscule, quelles que soient les conditions syntaxiques. Le scribe utilise par ailleurs souvent des *litterae oblongatae* sur les premières lignes de colonne :

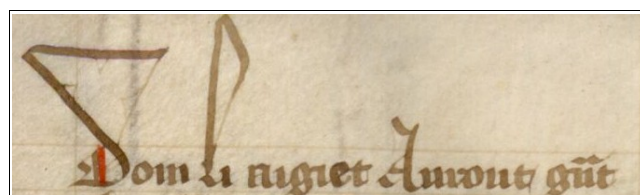


Fig. 41: *Litterae oblongatae* (*joachim768, 97d*).

La rencontre d'une majuscule et du début d'une relative est donc une pure coïncidence ici.

²⁴⁰ Notons que cet exemple se croise partiellement avec (6.105).

Toutes les autres occurrences de ponctuation entre principale et subordonnée surviennent également dans une position physique particulière, à savoir la fin d'une ligne. Il est donc possible qu'elles n'aient rien à voir avec la ponctuation « syntaxique ». On peut toutefois analyser les facteurs linguistiques qui favorisent (sinon déterminent) ces emplois.

- (6.107) Premièrement dit ¶ || quen l'an de de grace .M. iij. cens || xlv. commencerent les
orribles a||uentures (joachim768, 97c:18) ;
- (6.108) Et || dedans ce temps sera vne com||munite et vn parlement tenu || des ours ¶ Lyons ¶ et
lepars ¶ || pour auoir nouvelles uesteures ¶ (joachim768, 97d:30).

Dans (6.107), la proposition principale remplit avant tout la fonction d'un organisateur textuel (la même que les mots *item* dont nous avons déjà parlé). Le fait de marquer la limite entre cet élément et le contenu de la prédication n'a rien d'étonnant.

Dans (6.108), le début d'une proposition infinitive coïncide avec la fin d'une énumération : l'emploi d'une ponctuation permet de montrer que la subordonnée dépend non pas du dernier élément de l'énumération, mais de l'ensemble de la proposition principale.

Dans l'exemple suivant, l'hypothèse d'une marque décorative semble en revanche la plus probable :

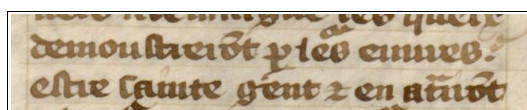


Fig. 42 : Ponctuation « décorative » (joachim768, 97d).

- (6.109) les quelx || demoustreront par leurs euures ¶ || estre sainte gent et en atrairont ||
plusours a eulx (joachim768, 97d:11).

Ici un point surmonté de deux traits en forme de *breve*, d'une couleur moins foncée que le reste du texte²⁴¹, est placé en fin de ligne et suivi, sur une nouvelle ligne, par une proposition infinitive jouant le rôle du complément d'objet direct de la principale. Du point de vue syntaxique et sémantique, la frontière syntaxique suivante (devant *et en atrairont*) paraît beaucoup plus importante, et pourtant elle n'est pas

²⁴¹ On peut donc se demander s'il ne s'agit pas d'un ajout postérieur à la rédaction initiale.

ponctuée. Nous avons donc toutes les raisons de croire qu'il s'agit là d'une marque de « remplissage » de fin de ligne.

Les syntagmes coordonnés à l'intérieur d'une proposition sont ponctués assez rarement : on compte 5 occurrences de ponctuation contre 18 absences. Dans un cas, l'état du manuscrit ne permet pas de déterminer avec certitude si le scribe avait l'intention d'utiliser une marque de ponctuation :

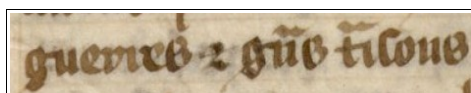


Fig. 43 : Ponctuation incertaine (ou trace de grattage) (joachim768, 98a).

Il s'agit en effet, selon toute vraisemblance, d'une correction sur grattage, et on ne peut pas dire si la petite tache qu'on voit est une marque de ponctuation ou la trace d'un caractère effacé.

Toutes les occurrences où la ponctuation est indubitablement présente concernent des énumérations avec au moins trois éléments coordonnés, comme celle qui est citée dans (6.108). Les autres occurrences de ce type se trouvent toutes dans la longue formule introductive :

(6.110) **C**i commence la prophecie || Mestre Joachim l'astro||nomien ¶ Selom son enten|| dement en vraies parolles declae||rees lequel prie a+touz que nul || n'y prengne erreur ¶ mes gou||uernance ¶ dotrine ¶ et science ¶ de || vraie jntroduction ¶ a+l'encontre || des degrez de+l'euure de nature || sans diffamer en rien leur chouses || qui ci apres s'ensieuent ¶ (joachim768, 97c:3-13).

Ici le fait qu'il s'agit d'une formule initiale s'ajoute au facteur d'énumération, ce qui crée des conditions particulièrement favorables à l'apparition d'une ponctuation. On voit dans ce même passage un point qui sépare un nom (*science*) de son modificateur (*de vraie introduction*), sans doute mal interprété comme un autre terme coordonné ou comme une fin d'énumération. La vraie fin de l'énumération (après *introduction*) est d'ailleurs également ponctuée.

En ce qui concerne des syntagmes particuliers à l'intérieur d'une proposition, ceux-ci sont rarement ponctués. Nous avons une occurrence de ponctuation devant un complément du comparatif (Riegel *et al.*, p. 365) introduit par la conjonction *que* :

Manuscrits antérieurs à 1400

- (6.111) Ains en seront *plus grans et plus horribles* . que deuant *et croistront* de || plus en plus . (joachim768, 98a:22).

Ce type de complément de comparatif ressemble beaucoup à une proposition complétive (même s'il n'y a pas de verbe exprimé), ce qui peut avoir favorisé l'apparition d'une ponctuation.

Dans un cas, un point est placé entre un circonstant « à fonction scénique » (Riegel *et al.*, p. 144) initial et le prédicat :

- (6.112) Item apres la+prise || du *grant lyon et du petit lioncel* . || auendra .j. si *grant meschief* ou roy||aume francois *que ce sera merueilles* (joachim768, 98a:27).

Cet emploi correspond bien à une règle de ponctuation moderne, mais notre occurrence peut s'expliquer par des raisons purement décoratives (*cf.* les autres occurrences de ponctuation en fin de ligne). Il convient toutefois de noter que dans ce cas précis l'espace entre la dernière lettre du mot *lioncel* et la ligne de réglure de la marge droite ne semble pas suffisamment important pour justifier l'emploi d'un point de « remplissage ».

Une seule occurrence se prête difficilement à une explication syntaxique ou graphique :

- (6.113) *quat* ilz *perdront* || la .x^e. *partie* . de lour *consail* . (joachim768, 97d:38).

Ici un point sépare un nom de son modificateur (comme dans (6.110)), la proposition ne contient pas d'éléments coordonnés et l'occurrence ne se trouve pas en fin de ligne. En revanche, le nom en question (*partie*) est précédé d'un adjectif numéral 'dixième' graphié par un chiffre romain suivi d'un *e* suscrit et entouré de deux points : .x^e. On peut supposer que l'emploi du point après *partie* est lié au marquage graphique d'un chiffre qui peut s'étendre à l'ensemble du syntagme auquel le chiffre appartient.

La comparaison de la ponctuation du manuscrit avec celle du niveau de transcription normalisé ne présente pas d'intérêt pour ce texte, puisque c'est nous-même qui l'avons réalisée. Notre principe est en effet de respecter, même au niveau normalisé, la ponctuation scribale, à l'exception des cas où celle-ci est incompatible avec les règles modernes ou qu'un ajout d'une marque facilite la compréhension d'un passage. Nous

avons ainsi supprimé 10 marques scribales devant certains syntagmes particuliers et termes coordonnés par la conjonction *et*. Nous avons ajouté 14 marques, essentiellement devant des circonstants « de cadre » et devant des précisions ou des reformulations, comme dans l'exemple suivant :

(6.114) Item en alant || d'an en an jusques a+l'an .l. [> ,] || sera grant guerre [> ,]
amenuise||ment et mortalite de peuple [> ,] || tant par espee que autrement
(*joachim768, 97d:25*).

La ponctuation scribale a été maintenue dans la version normalisée dans 35 cas, tandis que 47 frontières « ponctuables » ne sont marquées à aucun niveau de transcription.

Le manuscrit **joachim768** présente donc un système de ponctuation tout-à-fait conforme aux tendances générales constatées dans la grande majorité des textes de notre corpus avec toutefois deux particularités : une fréquence élevée de la ponctuation « extra-forte » et l'usage possible de la ponctuation à des fins de justification en fin de ligne.

6.11. Oxford, Bodleian, Laud Misc 587 (*Conquête de Constantinople de Geoffroi de Villehardouin*)

Ce manuscrit de la Bibliothèque Bodléienne est daté de la deuxième moitié du XIV^e siècle. Les feuillets sont en parchemin, l'écriture est régulière, les lettres ornées et les miniatures sont réalisées avec beaucoup de soin. C'est ce manuscrit que E. Faral a choisi comme base pour son édition (1961).

Ce manuscrit ne figurait en effet pas dans notre projet de corpus initial, mais nous avons profité d'un séjour à Oxford pour en transcrire un petit fragment de 352 mots que nous avons intégré dans notre corpus sous l'identifiant **villehardouin587**. Les dimensions de ce fragment étant donc considérablement plus restreintes que celles des autres échantillons de textes intégrés dans notre corpus (à l'exception des *Serments de Strasbourg*), nous ne pourrions y observer que les tendances les plus marquées.

L'extrait transcrit compte 19 occurrences de ponctuation faible et 13 occurrences de ponctuation forte, ce qui porte le taux général de la ponctuation à 9,4%.

Dans 2 cas, la première lettre d'un mot apparaît légèrement plus grande que les autres, il se peut qu'il s'agisse ici d'un type de lettres particulier occupant une position intermédiaire entre les minuscules et les majuscules et permettant ainsi de parler de trois (et non de deux) « forces » de ponctuation distinctes (sans parler de la ponctuation « extra-forte » liée à l'usage des initiales et des sauts de ligne). L'usage de « grandes minuscules » est connu dans les manuscrits médiévaux (Haugen éd. 2006b, p. 82). Elles sont traditionnellement interprétées comme des variantes de majuscules, mais certains chercheurs les considèrent comme un type de caractères indépendant (Leeuw van Weenen 2000). Le très faible nombre d'occurrences de ce type de lettres dans notre extrait ne permet pas de vérifier l'hypothèse de la « ponctuation moyenne », mais elle ouvre sûrement une piste qui mérite d'être explorée dans l'avenir.

Une autre particularité intéressante de ce manuscrit est la présence de deux formes de marques de ponctuation (point et barre oblique) employées avec des fréquences comparables (21 contre 13 occurrences respectivement) : dans les autres manuscrits que nous avons examinés jusqu'à présent la dominance du point était beaucoup plus marquée. Rien dans l'aspect de ces barres obliques (la couleur d'encre, la disposition du texte autour) ne laisse supposer qu'il s'agisse d'ajouts tardifs par rapport à la rédaction du manuscrit, même si une telle hypothèse ne peut être entièrement exclue.

Les barres obliques sont apparemment « spécialisées » dans le marquage de la ponctuation faible (où elles sont employées aussi fréquemment que les points), mais nous avons toutefois une occurrence de barre oblique suivie d'une majuscule et une autre, d'une « grande minuscule ».

Le fragment étudié ne contient aucun passage au discours direct. La tendance à l'emploi de la ponctuation aux frontières de propositions autonomes est indubitable : toutes les occurrences de ce type de frontière (19 au total) sont ponctuées il s'agit dans 13 cas d'une ponctuation forte. Toutes les occurrences de ponctuation faible dans cette position (6 au total) sont associées à l'emploi de la conjonction de coordination *et*. Dans

deux occurrences seulement cette conjonction est précédée d'une ponctuation forte, et il s'agit dans les deux cas d'une frontière narrative importante :

- (6.115) **EN** l'autre an apres **que** cil preudon || Folques parla ensi de diu / ot || [col. b] un tornoi en+la campaingne a+un cha||stel qui ot nom Aicris . Et par la grace || de dieu si auint que tibauz quens de || campaingne et de brie prist la croiz . (villehardouin587, 1b:2).

Dans cet exemple, il est question de deux événements distincts que l'auteur met en relation tout en marquant la frontière entre les deux à l'aide de la formule *par la grace de dieu* et de la construction impersonnelle *si auint que* pour introduire le deuxième événement.

Les propositions ayant un élément commun sont trop peu nombreuses (6 au total) dans notre fragment pour en tirer quelque conclusion que ce soit. 4 occurrences portent une ponctuation faible et dans 2 cas la ponctuation est absente. Un échantillon de texte plus important sera nécessaire pour parvenir à dégager les facteurs qui déterminent la présence ou l'absence de la ponctuation dans cette position.

Les limites entre les principales et les subordonnées sont ponctuées dans un tiers des occurrences (7 sur 21), et nous n'avons relevé aucune occurrence de la ponctuation forte à ce type de frontière. Il est intéressant de noter que dans 6 occurrences sur 7 la marque de ponctuation utilisée dans ce cas est la barre oblique, mais les données sont insuffisantes pour pouvoir parler d'une tendance à la spécialisation syntaxique de cette marque. L'absence de ponctuation semble plus probable devant des relatives et des complétives qui suivent immédiatement le terme qui leur est lié (antécédent ou verbe régissant), tandis que l'interposition d'un autre élément peut favoriser l'apparition d'une marque :

- (6.116) Or || sachiez que cil+quens Thibauz ere || iones hom . (villehardouin587, 1b:7) ;

- (6.117) l'a _postoilles enuoia en fran||ce / et manda as prodome / **que** il pen||chast²⁴² des+croiz par s'autorite . (villehardouin587, 1a:21).

Dans (6.116) la complétive qui suit immédiatement le verbe antécédent n'est pas ponctuée²⁴³, tandis que dans (6.117) la complétive qui est séparée de son verbe régissant par un complément d'objet indirect est précédée d'une barre oblique. Dans l'exemple

242 Sans doute une erreur de copie : E. Faral corrige en *preechast* d'après un autre manuscrit.

243 Cf. aussi la construction *auint que* dans (6.115).

Manuscrits antérieurs à 1400

suivant, trois relatives sont attachées à une même principale : l'une au sujet et les deux autres au complément du verbe :

- (6.118) Tuit cil qui se croissero||ient et feroient le+seruice deu un an || en+l'ost seroient quites de toz les pe||chiez que il auoient faiz / dont il se||roient confes . (villehardouin587, 1a:26-29).

La seule marque de ponctuation employée à l'intérieur de cette construction est une barre oblique séparant la deuxième relative de la troisième : selon notre analyse, c'est une frontière « horizontale » entre deux propositions subordonnées avec une principale commune. La présence d'une ponctuation ici corrobore notre hypothèse, selon laquelle les frontières « horizontales » ont une plus grande tendance à être ponctuées que les frontières verticales.

Les syntagmes coordonnés à l'intérieur d'une proposition sont ponctués dans moins de la moitié des occurrences (5 occurrences sur 13). Les données dont nous disposons ne permettent pas de dégager de facteurs plus spécifiques. Il est toutefois intéressant d'analyser les quelques propositions au début du texte qui formeraient une phrase complexe pour un lecteur moderne :

- (6.119) SAchiez que .M. et || C. et quatre+uinz anz || apres l'incarnation || nostre sengnor iesucrist / || As+tens Innocent A||postoil de Rome / et || phelippe Roy de France . et Richart || Roy d'engleterre et [corr. > ot] un saint home || en france / qui ot nom folques de || nuilli . Cil nuillis si est entre li||gni sor marne e paris . et il ere pre||stres et tenoit la parroche de+la+uil||le . (villehardouin587, 1a:1-12).

Le passage commence par la présentation du cadre historique des événements : l'auteur précise d'abord l'année, puis nomme le Pape et les rois de France et d'Angleterre de l'époque – tous ces circonstants de temps sont séparés les uns des autres par des marques de ponctuation, il y a même une ponctuation forte devant *as temps*. Celle-ci est sans doute liée aux conditions particulières créées par une énumération de noms propres (cf. notre analyse de **villehardouin12203** et de **im574**), certains noms propres étant d'ailleurs notés avec des majuscules²⁴⁴. Dans le manuscrit examiné, un autre personnage, Foulques de Neuilly, est ajouté à cette série à l'aide de la

244 Il convient toutefois d'interpréter avec prudence la valeur de la majuscule *R* : la majuscule peut être employée comme une simple variante positionnelle (au début d'un mot) dans certains manuscrits.

conjonction *et*. Il s'agit très probablement d'une erreur scribale, car dans ce cas il n'y aurait aucun centre prédicatif (sujet + verbe) dans cette proposition. Pour cette raison, l'éditeur du texte a proposé une correction (appuyée par d'autres manuscrits) qui consiste à remplacer la conjonction *et* par une forme verbale *ot* 'il y eut'. Cette correction rend la phrase tout-à-fait compréhensible. L'absence de ponctuation devant le verbe est plus prévisible que devant une troisième répétition de la conjonction *et* dans une longue énumération. On peut imaginer alors que le scribe avait devant lui un manuscrit où cette erreur ne figurait pas et qui était ponctué d'une façon conséquente. En commettant l'erreur de confondre une forme verbale et une conjonction, le scribe a néanmoins respecté la ponctuation « correcte ». Le nom du saint homme est donné dans une relative qui est séparée de son antécédent, et il y a bien une barre oblique devant. Vient ensuite une précision concernant la position géographique de la ville de Neuilly. Il s'agit d'un type de frontière syntaxique particulier qui est marqué, dans le système moderne, par des parenthèses (c'est cette marque que l'éditeur a utilisée à cet endroit). Dans le manuscrit, cette précision est précédée d'une ponctuation forte et suivie d'une ponctuation faible : la nécessité de marquer les frontières de cette unité était donc bien ressentie par le scribe, mais il n'avait pas à sa disposition de signe spécialisé pour ce cas de figure.

Aucune marque de ponctuation à une frontière d'un syntagme particulier n'a été relevée dans notre extrait.

Même si l'extrait dont nous disposons pour **villehardouin587** est trop bref pour définir avec précision les facteurs qui déterminent l'emploi des marques de ponctuation, les grandes tendances constatées dans les autres textes du corpus semblent se confirmer : la ponctuation apparaît le plus souvent aux frontières « horizontales » de propositions ; des règles particulières s'appliquent à des énumérations de noms propres.

6.12. Paris, BnF, fr. 840 (*Art de dictier* d'Eustache Deschamps)

Le manuscrit 840 du fonds français de la BnF est remarquable par son volume physique : 1162 pages en parchemin. Le manuscrit réunit l'ensemble des œuvres d'Eustache Deschamps, en majorité de la poésie lyrique (ballades, rondeaux, virelais, etc.), mais aussi quelques traités, lettres et épîtres. Trois œuvres seulement sont en prose, dont le traité intitulé *l'Art de dictier* (folios 394v – 400r), que nous avons sélectionné pour notre corpus. Ce texte appartient donc au domaine didactique.

D'après la bibliographie du DEAF, ce manuscrit peut être daté de *circa* 1400, il se situe donc à la limite entre le XIV^e et le XV^e siècle. La date de composition de *l'Art de dictier* est connue précisément, c'est la fin de l'année 1392. Le manuscrit est donc chronologiquement proche de l'original de l'auteur.

L'écriture de ce volume n'est pas particulièrement soignée et présente des tracés de lettres nettement plus cursifs que les manuscrits étudiés précédemment.

L'extrait que nous avons transcrit se situe au début du traité qui nous intéresse (folios 394r à 396v) et comporte 2082 occurrences-mots. Dans notre corpus, il porte l'identifiant **dictier840**²⁴⁵.

Une caractéristique du système de ponctuation de ce manuscrit qui frappe dès la première approche est la fréquence d'usage de majuscules qui ne sont pas précédées d'un signe de ponctuation, mais qui jouent sans aucun doute un rôle de ponctuations syntaxiques. Dans notre système, ces occurrences sont considérées comme des marques de ponctuation forte, et nous les comptabilisons lors du calcul du taux de ponctuation général, qui atteint les 7,5%, avec une très nette domination de la ponctuation forte (128 occurrences contre 29 ponctuations faibles). Si on ne prenait en compte que les occurrences de signes de ponctuation, le taux général ne dépasserait pas les 4%, la ponctuation forte serait toujours majoritaire (53 occurrences contre 29), mais d'une façon moins marquée. Il convient de noter que le taux de ponctuation général dans notre

²⁴⁵ *L'Art de dictier* n'a pas d'identifiant spécifique dans la bibliographie du DEAF. Les œuvres complètes d'Eustache Deschamps sont rassemblées sous le sigle **Desch**.

extrait est fortement influencé par des passages contenant des listes de notes et de lettres :

(6.120) ¶ Item les liquides sont B C D F G || K L M N P Q R S T X V Z !
(dictier840, 396a:31-32).

En effet, la majorité des occurrences du point suivi d'une majuscule sont enregistrées dans les quelques énumérations de ce type. Dans le reste du texte, le taux de ponctuation n'atteint que 5% en moyenne, et la marque dominante est la majuscule seule. Le point est le signe de ponctuation le plus fréquent (68 occurrences), mais on trouve également 8 occurrences de la barre oblique (sous forme d'un léger trait) et une occurrence d'une marque en forme de trois points disposés en triangle avec un sommet dirigé vers le bas :

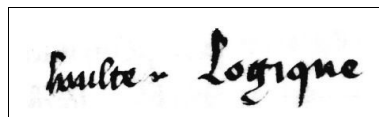


Fig. 44 : Ponctuation en forme de trois points (*dictier840*).

Nous avons enfin relevé 4 occurrences du pied-de-mouche placées sûrement par un rubricateur en marge gauche, et qui servaient à marquer le début d'une nouvelle section (en l'occurrence la description d'un nouvel art). La nouvelle section peut commencer au milieu d'une ligne, mais le pied-de-mouche se situe toujours dans la marge :

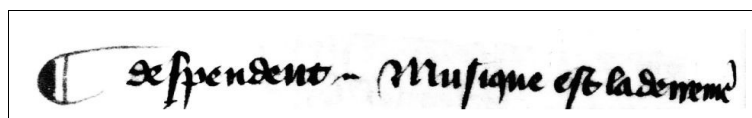


Fig. 45 : Pied-de-mouche en marge (*dictier840*).

Notre extrait contient quelques occurrences qu'on peut interpréter comme un discours direct. Il s'agit de l'explication de la façon dont il faut prononcer certains mots qui commencent par un *h* « muet » :

(6.121) Et n'est pas H ! proprement *letre* mais || n'est que vne aspiracion sonnante se||*[col. b]*lon la+maniere des noms ainsi *comme* || se on vouloit dire *hannequin*²⁴⁶ ou || *hannote*²⁴⁷ qui

246 'Une sorte de jeu de cartes', d'après le dictionnaire de Godefroy et le DMF. G. Roques (1982, p. 112) donne un autre sens : 'enfant turbulent'.

247 Le sens de ce mot reste obscur. La seule occurrence mentionnée dans le DMF est celle que nous citons. On peut juste supposer que ce mot appartient au même groupe thématique que *hannequin*.

sanz ladict h n'aroit || pas son plain son Aincois droit || on **Annequin** *et* **Annote**
(*dictier840, 396a:33-b:5*).

Il ne s'agit donc pas d'un discours direct à part entière, mais plutôt de citations de mots ou de phrases dans une publication linguistique. La présence du verbe de parole dans les deux cas justifie néanmoins le rapprochement avec le discours direct. Il est intéressant de noter que dans le premier cas aucune marque graphique ne distingue les mots cités comme exemples du texte environnant, tandis que dans le deuxième cas chacun de ces mots prend une majuscule. La différence est qu'on cite d'abord la forme écrite (avec *h*) et qu'ensuite on indique la prononciation (sans *h*). L'emploi de la majuscule a pu servir à mettre en évidence le fait que [a] est le premier son devant être prononcé. Il est possible que la différence de traitement entre les deux cas soit due à des raisons purement graphiques : pour la lettre *h* la différence entre la majuscule et la minuscule n'est pas très marquée dans **dictier840** (cf. l'illustration présentée ci-dessous avec une majuscule à gauche et une minuscule à droite). Il se peut que le scribe ait eu l'intention d'utiliser une *littera notabilior* à cet endroit, mais que le résultat ne soit pas suffisamment perceptible.

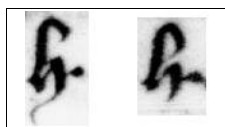


Fig. 46 : Lettre H majuscule et minuscule (*dictier 840*).

Il est enfin possible que les majuscules *A* soient utilisées ici comme de simples variantes sans valeur « ponctuationnelle », comme dans l'exemple suivant :

(6.122) Et **A** || venir par getter *et*+compter (*dictier840, 394c:18*).

La dernière hypothèse nous semble toutefois peu probable pour l'exemple (6.121), car ici les occurrences ne se trouvent pas en fin de ligne et la mise en relief intentionnelle des premières lettres des mots cités est très utile pour la compréhension du texte.

Concernant l'exemple (6.121), il convient de noter enfin l'usage du point situé immédiatement après la citation d'une lettre. Le rôle de cette marque consiste certainement en une mise en relief de l'emploi d'une lettre en tant qu'élément autonome

dans le contexte d'une explication métalinguistique et non comme une partie de la graphie d'un mot.

Le manuscrit **dictier840** n'échappe pas à la tendance générale qui consiste à ponctuer les frontières de propositions autonomes : 39 occurrences sur 43 portent une ponctuation forte (il n'y a aucune occurrence de ponctuation faible dans cette position).

L'un des cas d'absence de ponctuation survient dans une longue succession de propositions décrivant les conditions dans lesquelles on peut écouter la musique « naturelle » (devant *ou la musique...*) :

(6.123) Et+aussis les diz || des chancons se puent souente||foiz recorder en pluseurs lieux ou || ilz sont moult volentiers ois ou || Le+chant de+la+musique artificiele || n'aroit pas tousiours lieu comme || entre seigneurs et+dames estans a || Leur priue et+secretement ou la || musique naturele se puet dire || et recorder par vn homme seul de || bouche ou lire aucun liure de ces || choses plaisans deuant vn ma||lade / et autres cas semblables ou || le+chant musicant n'aroit point || lieu pour la+haulteur d'icellui (*dictier840, 395d:8-22*).

Dans cet exemple, on peut hésiter, au début de la ligne 4, entre l'interprétation de *ou* comme conjonction de coordination ou bien comme un pronom adverbial relatif ('où'). Dans le premier cas, la proposition en question serait coordonnée avec la première proposition de l'extrait cité. Il s'agirait donc d'une frontière entre deux groupes de propositions complexes et autonomes :

Et aussis les diz des chançons se puent souventefoiz recorder... ou la musique naturele se puet dire et recorder par un homme seul...

Dans le deuxième cas, on peut interpréter la proposition *ou la musique...* comme une relative coordonnée avec les propositions *ou ilz sont...* et *ou le chant...* La deuxième interprétation nous semble moins logique, mais elle est la première à venir à l'esprit en cas de lecture rapide²⁴⁸. L'absence de la ponctuation dans le manuscrit peut signifier que le scribe, lui aussi, a dû prendre cette proposition pour une relative coordonnée.

L'exemple cité permet de constater un autre phénomène intéressant : l'usage de la majuscule *L* en début de ligne (*Le chant...* et *Leur privé...*). Ces majuscules ne se

248 G. Raynaud, dans son édition (1891) a mis une virgule devant la proposition qui nous intéresse. Cela signifie sans doute qu'il se penchait pour la deuxième interprétation, car il aurait utilisé une ponctuation plus forte dans le cas contraire.

trouvent pas aux frontières d'unités ponctuables, nous les considérons donc comme de simples variantes positionnelles. Cela n'empêche pas que ces caractères créent des frontières visuelles qui donnent l'impression d'un texte structuré et qui font que l'absence de la ponctuation syntaxique dans un passage relativement long ne « saute pas aux yeux ». Il convient de rappeler également que la grande majorité des textes contenus dans le manuscrit étudiés sont en vers. L'usage « abusif » des majuscules peut en effet provenir de la mise en forme des vers.

Les trois autres cas d'absence de ponctuation à la frontière entre propositions autonomes concernent tous la coordination (avec la conjonction *et*) de propositions relativement simples.

Il convient de noter que la fréquence relative des frontières entre des propositions autonomes est inhabituellement basse dans **dictier840** : il y en a en moyenne 2,1 occurrences pour 100 mots du texte (ce taux 3,3 à 6,6 dans les autres textes). Cette caractéristique est sans doute liée au style personnel de Deschamps qui utilise un grand nombre de constructions complexes avec des chaînes de propositions subordonnées (dont il n'y a pas moins de 115 occurrences, soit 5,5 pour 100 mots).

Les propositions coordonnées avec un élément commun sont séparées par une ponctuation dans un quart des cas environ (10 occurrences sur 42). À l'exception d'une occurrence, où il y a une barre oblique devant une majuscule, il s'agit toujours de l'emploi d'une majuscule seule.

Les propositions avec des éléments communs sont ponctuées lorsqu'elles font partie de constructions syntaxiques complexes, comme par exemple dans le paragraphe consacré à l'arithmétique qui commence par une principale *Arismetique est science...* suivie de 16 propositions infinitives (que nous avons surlignées et indexées dans l'exemple ci-dessous) :

- (6.124) ¶ Arismetique est science ^ade+getter || ^bet compter par le+nombre de augo||risme || *et* autre nombre commun ^cet || de mesurer ^det+arpenter les terres . ||[fol. 394v] les boys et choses semblables ^epour scauoir la+haulteur des choses en alant || vers le+ciel . la+largeur des euaes et || des riuieres la+parfondeur des puis || et des concaues de la+terre ^fDe scauoir || les heures les temps les minutes || *et* les momens ^gpour scauoir le+com||mencement || des jours *et* des nuis des || sepmaines / des moys *et* des ans || ^hpour venir au+grant miliaire ⁱet || scauoir par ce nombre en quercu||lant la reuolucion des temps ^jet || congnoistre le+cours du soleil et de || la+lune Et du zodiaque ^kscauoir la || maniere du poys *et*+de+la+loy des || monnoyes tant en or comme en || argent les

dragmes caras demi || dragmes *et*+les empirances *Et A* || venir^m par *getter* *et*+compter
 en mon||tant *et* multipliant son nombre || de la+plus petite somme jusques ||
 a+la+plusgrande *et* haulte / *Et pour* || *congnoistre* selon les espaces des ||
 charpenteries *a*+veoir les cours des || toiz par vn des cours seulement || quans milliers
 de+clou *et*+de+late || *et* de+tieulle il+aura sur vn+toit || *et* ainsi des autres choses en ce
 cas . (*dictier*840, 394b:31-394c:28).

Les relations entre ces propositions et les groupes qu'elles forment sont assez claires au début et un peu moins évidentes vers la fin du fragment. Les propositions *a* et *b*, ainsi que *c* et *d* forment des groupes soudés coordonnés entre eux qui jouent tous le rôle de complément du nom (*science*). La proposition *e* est une subordonnée circonstancielle rattachée à l'ensemble du groupe *a+b+c+d*. Dans tout ce passage, il n'y a aucune ponctuation entre les propositions (mais il y a deux points entre des termes énumérés sans conjonction de coordination). La proposition *f* est un nouveau complément du nom *science* (dont elle est assez éloignée). Elle commence par une majuscule qui sert à marquer une rupture avec le groupe de 5 propositions qui précèdent. Les propositions *g*, *h*, *i* et *j* sont des subordonnées circonstancielle de *f* (ou plutôt de l'ensemble de la principale initiale et de la proposition *f*). La seule ponctuation dans le segment allant de *g* à *j* est une barre oblique qui apparaît à l'intérieur de la proposition *g* entre des compléments d'objet direct énumérés (5 au total).

Les choses deviennent plus complexes ensuite. L'interprétation la plus logique est de considérer *du souleil*, *et de* || *la+lune* et *Et du zodiaque* comme des compléments coordonnés du nom *cours* et de placer *sçavoir* dans une nouvelle proposition (*k*) rattachée de nouveau à la principale et coordonnée avec les groupes *a+b+c+d* et *e+f+g+h+i+j*. La ponctuation du manuscrit suggère une autre analyse : la proposition *k* commencerait par *Et du zodiaque*. On peut supposer que la répétition de la conjonction *et* devant le deuxième et le troisième complément, l'absence de connecteur devant *sçavoir* et la complexité générale de la construction ont induit le scribe en erreur.

La proposition *l* est sans doute coordonnée avec *k* (et les deux grands groupes qui précèdent). Le changement de préposition (*a* au lieu de *de*) témoigne de l'affaiblissement du lien de coordination. La ponctuation forte est tout-à-fait légitime à son début. La majuscule de la préposition *A* (qui suit immédiatement *Et* initial) peut s'expliquer par sa position en fin de ligne et, donc, par l'influence de la justification.

Les propositions *m* et *n* forment un groupe qui sert de complément circonstanciel à *l*. Notons qu'il s'agit des mêmes verbes que dans *a* et *b*.

La proposition *o* présente encore un aspect de l'art de l'arithmétique, elle est donc coordonnée avec *k*, *l*, etc. La majuscule de la ponctuation forte est accompagnée cette fois par une barre oblique. La proposition *p* joue selon nous le rôle d'un complément du verbe (*congnoistre ... a veoir...*), l'absence de ponctuation y est normale.

Notre analyse de cette construction complexe permet de conclure que la ponctuation reflète dans une certaine mesure la hiérarchie des structures logiques et syntaxiques, mais, comme dans l'exemple (6.123), elle est peut être basée sur une segmentation spontanée au cours d'une lecture rapide, plutôt que sur une analyse profonde du sens de la construction.

Les limites entre propositions principales et subordonnées sont rarement ponctuées : 10 occurrences de ponctuation forte et 5 occurrences de ponctuation faible contre 100 occurrences d'absence de ponctuation.

Cinq de ces 15 occurrences (et 3 sur 5 pour la ponctuation faible) concernent les cas où la frontière d'une subordonnée coïncide avec la fin d'une énumération (après *A*. *B*. *C*. dans l'exemple suivant) :

(6.125) Lesquelz ars trouua du tiers aige du || monde et au+temps de habraham ¶ || zozoastres qui regnoit en baterie²⁴⁹ Et || pour ce est le+premier *et* principal art || gramaire ¶ par+lequel l'en vient et a||prant tous les autres ars par+les figures || des lettres de A . B . C ¶ que les enfans || aprannent *premierement* (*dictier840, 394a:21*).

La ponctuation dans ces cas est liée au marquage des éléments énumérés plutôt qu'à la délimitation des frontières de propositions.

L'exemple ci-dessus présente une autre occurrence d'une ponctuation devant une proposition relative (*par lequel l'en vient...*). Ici aussi, la ponctuation sert avant tout à mettre en évidence le mot *grammaire* qui porte l'information la plus importante dans ce contexte.

Le même fragment du texte présente enfin un nouvel exemple de ponctuation « superficielle », résultant d'une erreur d'interprétation due à une lecture rapide. La

249 Il s'agit de la Bactrie, une région à cheval sur les états actuels d'Afghanistan, d'Ouzbékistan et du Tadjikistan, dont Zoroastre était originaire selon certaines traditions.

connaissance des faits historiques permet de comprendre le sens du début de ce fragment : ‘lesquels arts furent découverts, au Troisième Âge du monde et à l’époque d’Abraham, par Zoroastre, qui régnait en Bactérie et en Perse’. Le nom géographique de *Perse* était sans doute inconnu du scribe, qui l’a pris pour un adverbe *pour ce* et a placé une ponctuation forte devant.

Parmi les autres occurrences de frontières ponctuées entre principales et subordonnées, on peut constater l’absence de relatives et de complétives introduites par *qui* ou par *que*²⁵⁰ : les connecteurs sont toujours plus spécifiques (*lequel, afin que, pour ce que, etc.*).

Comme nous l’avons déjà indiqué, l’extrait de **dictier840** que nous étudions est caractérisé par une très haute fréquence d’énumérations et autres syntagmes coordonnés à l’intérieur d’une proposition : 264 occurrences au total. Dans 184 cas, la ponctuation est absente, dans 22 cas il y a une ponctuation faible et dans 58 cas la ponctuation est forte.

Dans (6.120) nous avons cité un exemple d’énumération de lettres (ou plutôt de phonèmes dans la terminologie moderne), les exemples de ce genre sont nombreux. Les cas où la ponctuation est absente concernent des coordinations simples (avec 2 éléments), les frontières entre le premier et le deuxième élément d’une énumération (comme c’est le cas dans (6.120)) et certaines énumérations relativement courtes et intégrées dans des constructions syntaxiques complexes :

(6.126) Ceuls qui auoient et ont acoustume || de faire en ceste musique naturele || Seruentois de *nostre* dame | chancons || Royaulx pastourelles balades et ron||deaulx portoient chacun ce *que* fait || auoit deuant le+prince du puy (*dictier840, 395c:8-13*).

Dans cet exemple, 5 objets directs sont énumérés dans une relative avec un antécédent démonstratif qui joue le rôle de sujet dans la proposition qui suit la relative. Le premier des éléments énumérés, *Serventois*²⁵¹, porte une majuscule (qui peut s’expliquer par sa position initiale sur la ligne, cf. *Royaux* sur la ligne suivante). Il est séparé par une ponctuation faible du deuxième élément, tandis que les trois éléments qui

250 A l’exception des fins d’énumérations déjà mentionnées.

251 Selon Godefroy, *serventois* est « à l’origine, [une] pièce de vers composée par ou pour des *servants* ; plus tard pièce composée en l’honneur de la Vierge ».

restent ne sont pas ponctués. La présence d'une ponctuation entre les deux premiers éléments peut être favorisée par le fait qu'à la différence des trois derniers ceux-ci sont composés de plusieurs mots.

Un autre exemple nous permet de supposer qu'il existe une différence dans le traitement des énumérations basée sur la catégorie grammaticale du terme coordonné :

- (6.127) Car par ses vj notes qui sont appel||lees ¶ vs ¶ re ¶ my ¶ fa ¶ sol ¶ la ¶ l'en puet || aprendre a+chanter acorder doubler || quitoier tiercoier tenir deschanter || par figure de+notes par clefs *et* par || lignes le+plus rude homme du monde (*dictier840, 395a:9-14*).

Ici on trouve des points entre tous les substantifs qui désignent des notes musicales, ainsi qu'au début et à la fin de l'énumération. En revanche, aucune ponctuation n'est présente entre les 7 infinitifs juxtaposés dans la même construction.

En dehors des cas d'énumération la ponctuation est extrêmement rare à l'intérieur de la proposition. Nous avons relevé 6 occurrences de constructions exprimant une précision et introduites par *c'est assavoir* ou *c'est a dire* :

- (6.128) nul se il n'estoit liberal ¶ C'est || a+dire fils de noble homme *et* a _strait de || noble lignie n'osoit aprendre aucun || d'iceuls ars ¶ C'est+assavoir gramaire || logique . Rethorique . Geometrie . Aris||metique . || Musique Et astronomie . (*dictier840, 394a:9-14*).

Formellement, il s'agit de propositions incidentes, mais sémantiquement elles sont très proches de syntagmes apposés à l'intérieur d'une proposition (on peut considérer les formules *c'est à* + infinitif comme de simples connecteurs ayant perdu le statut d'un noyau prédicatif). Quoiqu'il en soit, 5 des 6 occurrences de ce types portent une ponctuation forte, c'est donc une construction particulière bien identifiée par le scribe de **dictier840**. La seule occurrence où la ponctuation est absente est la suivante :

- (6.129) Et *entre* ces cinq voyeux en y+a || deux ¶ c'est assavoir E²⁵² / *et* v ¶ qui se me||ttent bien ensemble (*dictier840, 396a:28*).

Ici la ponctuation est utilisée pour marquer les lettres citées, et il se peut que le scribe n'ait pas voulu surcharger le passage d'un trop grand nombre de marques.

252 Sans doute, une erreur du copiste. Logiquement, il s'agit de *I*.

D'autres occurrences de la ponctuation à l'intérieur d'une proposition sont liées au marquage de lettres citées (comme dans l'exemple (6.121)) ou à la mise en valeur de termes importants (comme *Grammaire* dans (6.124)). Seules trois occurrences échappent à cette tendance :

- (6.130) Geometrie est science [...] de faire *et* || edifier les chasteaulx salles et maisons || pour habiter / les clochiers *et+autres* || edifices en ront en triangle et en || quarreure et les mener droit sanz || boce **J**usques a+leur perfection faire || tonneaulx *et autres* vaisseaulx de || certaines pieces longueur *et+grosseur* (*dictier840, 394b:12*) ;
- (6.131) pour+ce *que* neant+plus *que* l'en || pourroit proferer **L**e+chant de mu||sique sanz la+bouche ouurir Neant || plus pourroit l'en proferer ceste mu||sique naturele sanz voix (*dictier840, 395c:18*) ;
- (6.132) Et aussi ces deux musiques sont si || consonans **L**une auecques l'*autre* que || chascune puet *bien* estre appellee mu||sique (*dictier840, 395c:24*).

Dans les trois cas, il s'agit de lettres n'ayant pas de différence de tracé très nette entre la majuscule et la minuscule. Il se peut que notre interprétation basée sur un jugement impressionniste soit erronée et qu'il s'agisse de minuscules simplement un peu plus grandes que les autres. Si on maintient l'interprétation de ces lettres comme majuscules et donc comme marques de ponctuation forte, on peut trouver une explication propre à chaque cas. Dans (6.130), il s'agit d'une longue énumération des activités associées à l'art de géométrie. Cette énumération n'est pratiquement pas ponctuée, et le scribe a pu croire que le circonstant *jusques a leur perfection* se rattache à la proposition suivante et non à la proposition précédente, ce qui aurait pourtant été plus logique.

Dans (6.131), il s'agit d'une construction complexe avec l'établissement d'un parallèle entre deux phénomènes à l'aide de la répétition d'un connecteur *neant plus que* ('pas plus que'). Le scribe a pu vouloir utiliser la ponctuation pour mettre en évidence les deux parties contrastées, mais dans la première, il a préféré attirer l'attention du lecteur sur le premier nom de la proposition et non sur la longue succession de connecteurs.

Dans (6.132), nous avons également affaire à une sorte de construction contrastive (*l'une et l'autre*), qui joue le rôle de modificateur détaché du sujet. Il se peut que cette construction attire la ponctuation de la même façon que la formule *c'est assavoir*.

Manuscrits antérieurs à 1400

Compte tenu du faible taux général de la ponctuation dans le manuscrit, il n'est pas étonnant que l'éditeur (Raynaud 1891) ait souvent procédé à l'ajout d'une marque là où il n'y en avait pas dans le manuscrit. Dans l'extrait étudié, nous avons relevé 151 occurrences de ce type. En revanche, il n'a supprimé que 18 marques présentes dans le manuscrit (il s'agit essentiellement de points autour de lettres citées et de termes coordonnés avec *et*). La ponctuation est présente à une frontière syntaxique à la fois dans le manuscrit et dans l'édition dans 138 cas et est absente dans les deux dans 186 cas.

Le manuscrit **dictier840** se différencie à plusieurs égards de toutes les autres pièces de notre corpus étudiées jusqu'à présent. Il emploie massivement les majuscules sans signes de ponctuation devant, il contient un grand nombre de longues énumérations associées à une ponctuation spécifique et relativement peu de frontières de propositions autonomes (à cause de la fréquence des énumérations et des propositions subordonnées). Les grandes tendances de l'usage de la ponctuation restent cependant les mêmes que dans les autres manuscrits : la ponctuation apparaît avant tout à la frontières de propositions autonomes, sa fréquence baisse en cas de présence d'un élément commun et aux frontières entre les principales et les subordonnées. Les occurrences de la ponctuation aux frontières de syntagmes particuliers sont très rares, ce qui témoigne sans doute de la conscience que le scribe avait de la hiérarchie syntaxique. Certaines occurrences (peu nombreuses de façon générale) relèvent sans doute d'erreurs du scribe dans l'interprétation de la structure logique du texte, ces erreurs étant dues à la lecture rapide de constructions complexes.

Chapitre 7

Manuscrits du XV^e siècle

7.1. Vatican, Reg. Lat. 869 (*Chroniques* de Jean Froissart)

Ce manuscrit de la bibliothèque du Vatican se compose de 151 feuillets en papier de 290 × 218 mm. Le filigrane du papier permet une datation assez précise. D'après G.T. Diller (1972), il correspond à la marque d'un papier fabriqué en Hollande septentrionale en 1399, selon le catalogue de V. Mošin et S. Traljić (1957). Cela permet de situer la date de la rédaction du manuscrit à 1410 au plus tard, et plus probablement entre 1399 et 1405, car « il est improbable qu'on ait gardé le papier plus d'une dizaine d'années » (Diller 1972, p. 13). Toujours d'après G. T. Diller, la graphie de ce manuscrit « appartient clairement à la scripta franco-picarde » (*idem*).

Le manuscrit qui nous intéresse est entièrement occupé par la troisième rédaction du premier tiers du premier livre des *Chroniques* de Jean Froissart, dont il est le seul témoin. Le manuscrit « a dû suivre de très près un premier brouillon, sans doute dicté par le chroniqueur, comme le prouvent à la fois la langue du texte et les dates des derniers évènements consignés dans le récit » (*idem*).

Selon la typologie textuelle de la BFM, le texte de Froissart appartient au domaine historique et représente le genre de la chronique.

L'écriture du manuscrit qui nous intéresse n'est pas très soignée. Le texte est disposé en une seule colonne sur la page, les initiales pour lesquelles le scribe avait laissé des alinéas de deux lignes au début des chapitres n'ont jamais été réalisées. Le texte de ce manuscrit porte des traces d'annotations certainement postérieures à la rédaction initiale : certains mots ou syntagmes sont soulignés, des inscriptions indiquent le numéro du chapitre (par exemple *ch. 2* à la ligne 37 du folio 1v) ou une sorte de « titre courant » (par exemple *Naturel des Angles* aux lignes 5 et 6 du folio 2r). Ces inscriptions situées dans la marge droite (ou à la fin d'une ligne si celle-ci n'est pas remplie) se distinguent du corps du texte par la couleur de l'encre (plus foncée) et une forme d'écriture différente. Compte tenu de leur caractère de toute évidence tardif, nous avons décidé de ne pas tenir compte de ces annotations dans notre transcription. En revanche, toutes les marques de ponctuation semblent avoir été posées par le scribe même : la couleur de l'encre est identique à celle des mots environnants.

L'extrait que nous avons transcrit se trouve juste après le prologue, au début du corps de la chronique de Froissart (f. 1v – 2r) et compte 1051 occurrences-mots. Son identifiant dans notre corpus est **froissart869**²⁵³.

Le taux de ponctuation dans ce manuscrit est assez bas : 56 occurrences en tout, soit 5,2%. La ponctuation forte est quasiment absente : l'extrait transcrit contient seulement deux occurrences de majuscules sur des frontières ponctuables (dont une qui ressemble plutôt à une minuscule élargie), en plus de deux initiales de 2 lignes non réalisées situées au début et à la fin de l'extrait choisi.

La seule marque de ponctuation relevée dans notre extrait est la barre oblique, souvent assez courte et située dans la partie supérieure de la ligne :

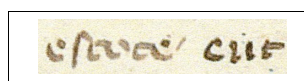


Fig. 47 : Petite barre oblique en haut de ligne (froissart869, 2v).

Le fragment transcrit ne contient pas de discours direct.

²⁵³ Cf. le sigle DEAF de cette troisième rédaction des *Chroniques* de Froissart : **FroissChron3**.

Manuscrits du XV^e siècle

Le type de frontière syntaxique le plus souvent ponctué se situe entre deux propositions autonomes : notre échantillon présente 32 occurrences ponctuées contre 23 sans ponctuation. Notre tentative de dégager des facteurs qui puissent expliquer l’alternance entre la présence ou l’absence de ponctuation dans cette position n’a pas apporté de résultat certain. Le facteur de la forme du sujet ne semble pas jouer un rôle significatif, même si une légère différence est perceptible selon que le sujet est nominal ou pronominal (dans la proposition qui suit la frontière ponctuable) :

	Ponctuation présente	Ponctuation absente
Sujet nominal	14	14
Sujet pronominal	16	8

Les chiffres laissent supposer que le sujet pronominal favorise la présence d’une ponctuation sans pour autant l’exiger.

Le second facteur examiné concerne la longueur des segments sans ponctuation qui précèdent ce type de frontière. Celle-ci est en effet très variée dans les deux cas (de 3 à plus de 30 mot), et aucune différence significative n’a pu être observée.

Le type de connecteur joue peut-être un certain rôle :

	Ponctuation présente	Ponctuation absente
Absence de connecteur	7	2
<i>et</i>	19	11
<i>car</i>	1	4
<i>mais</i>	0	3
autre connecteur	1	3

La ponctuation est présente dans 7 cas sur 9 en cas d’absence de connecteur et dans 19 cas sur 30 en présence de la conjonction *et*. En revanche, 4 occurrences sur 5 de la conjonction *car* et 3 occurrences sur 3 de la conjonction *mais* ne sont pas ponctuées.

Il convient de noter que la haute fréquence de la ponctuation en cas d’absence de connecteur peut être provoquée en partie par d’autres facteurs : dans 3 cas le premier mot de la proposition est soit *engles* (‘Anglais’), soit *engleterre* (‘Angleterre’) :

- (7.1) et en cela sont il ne et obstine ne nuls ne || lor poroit faire entendant le contraire / engles sont de merueilleuses conditions chaut || et boullant tos esmeu en ire tart apaisie ne amode en doucour (*froissart869, 2r:16*).

Il se peut que les noms propres de ce peuple et de ce pays, particulièrement importants dans le cadre de la chronique de Froissart, attirent en quelque sorte la ponctuation aux unités ponctuables les contenant. Outre les 3 occurrences où ce terme se trouve en tête d'une proposition, nous avons relevé 1 occurrence dans laquelle une ponctuation précède le syntagme prépositionnel dont il fait partie et 3 occurrences dans lesquelles ce syntagme est suivi d'une marque de ponctuation :

- (7.2) li homme de mestier et li laboureur / parmi engleterre viuent || de ce que il seuent faire / (*froissart869, 2r:28*) ;
- (7.3) et essil||la moult dou roiaulme d'engleterre / iusques a la+riuiere dou thin (*froissart869, 2v:5*).

Notons que dans les derniers exemples les frontières ponctuables sont intrapropositionnelles, donc *a priori* peu favorables à l'apparition de la ponctuation. Dans les 6 autres occurrences des mots *engles* ou *engleterre*, une marque de ponctuation n'est jamais très loin : à 6 mots de distance au maximum.

Bien entendu, notre hypothèse sur l'incidence de l'emploi de certains noms propres sur la ponctuation des unités syntaxiques qui les contiennent devra être vérifiée scrupuleusement sur un corpus beaucoup plus important que le nôtre.

Les propositions coordonnées ou juxtaposées avec un élément commun sont séparées par une ponctuation dans 5 cas sur 30. Il y a donc une nette tendance à l'absence de ponctuation dans cette position. Les occurrences ponctuées ne semblent pas comporter de traits spécifiques qui les différencient de la majorité des occurrences.

En ce qui concerne les frontières entre les propositions principale et subordonnée, 13 occurrences sur 49 sont ponctuées (soit près d'un quart). Le type de la subordonnée joue peut-être un certain rôle : près de la moitié des relatives sont ponctuées (8 occurrences sur 17), tandis que dans les autres cas la ponctuation reste exceptionnelle (par exemple, 1 occurrence sur 9 pour les complétives). Il est possible que ce soit plutôt le type de connecteur que le type de la subordonnée qui ait une

Manuscrits du XVe siècle

incidence sur la ponctuation : 4 occurrences sur 9 de *qui* sont précédées d'une ponctuation, tandis que 10 occurrences sur 11 de *que* ne le sont pas. Le fait que l'antécédent de la relative soit un nom propre peut également être un facteur favorable à l'emploi d'une ponctuation :

(7.4) car asses tos apries ce / que il fu couronnes li rois d'escoce / qui se nonma Robers de brus || desfia ce roi edouwart / (froissart869, 2v:1) ;

(7.5) et conuient bien que || vns rois qui est lors+sires se ordonne apries euls et s'encline a+moult de lors || volentes (froissart869, 2r:35).

Dans (7.4), le syntagme *li rois d'escoce* qui contient un nom propre de pays est suivi d'une ponctuation, tandis que dans (7.5) il n'y a pas de ponctuation entre *un rois* (utilisé au sens générique) et sa relative. De effet, toutes les occurrences ponctuées de *qui* contiennent un nom propre dans un contexte proche.

Les syntagmes coordonnés à l'intérieur d'une proposition ne sont que très rarement ponctués : 3 occurrences sur 45.

L'extrait dépouillé ne présente qu'une occurrence d'une énumération de plus de deux éléments, et elle ne contient aucune ponctuation :

(7.6) et li communs peuples || est de fele perilleuse orgueilleuse et desloiale condition / (froissart869, 2r:24).

L'une des occurrences ponctuées suit immédiatement un nom propre, et la construction syntaxique est par ailleurs complexifiée par la présence d'une apposition qui sert à préciser le référent du sujet :

(7.7) Adont s'esmurent chils || rois edouwars fils au bon roi edouwart / et toute la cheualerie d'engleterre (froissart869, 2v:7).

Les deux autres occurrences sont plus difficiles à expliquer :

(7.8) [engles] ne+se+pueent conioindre parfaitement / ne naturellement en l'amour ne alianca || de nation estragne / (froissart869, 2r:19) ;

(7.9) et le tiennent a en_dormi+et+a || pesant qant il ne voelt ensieuir les oeures de son pere / ou de son predicessour bon roy || qui a resgne en deuant de li / (froissart869, 2r:13).

Dans (7.8), l'occurrence intervient dans une longue succession de propositions coordonnées, où sont énumérées les différentes caractéristiques des Anglais. Toutes ces propositions commencent par la conjonction *et*, contiennent un verbe à la 6^e personne et sont précédées d'une ponctuation. Il est possible que le scribe ait confondu l'adverbe *naturelment* avec une forme verbale en *-ent* (comme par exemple *pueent* sur la même ligne) et ait jugé nécessaire d'insérer une ponctuation.

Dans (7.9) il s'agit d'une coordination de deux syntagmes prépositionnels compléments d'un nom. Si on retire le deuxième élément coordonné, la proposition reste tout-à-fait correcte. Il n'est pas exclu que le scribe ait mis la ponctuation en croyant qu'il s'agissait d'une fin de proposition.

Notre extrait présente 5 occurrences de marques de ponctuation aux frontières de syntagmes particuliers à l'intérieur d'une proposition. Deux d'entre elles peuvent être liées à la présence du nom propre *engleterre* dans un contexte très proche (7.7). Deux autres occurrences se succèdent dans une proposition où il est question d'un mariage d'une fille royale avec un autre roi :

(7.10) li biaux rois pheli||ppes de france ¶ eut mariee sa fille izabel ¶ au roi edouwart
d'engleterre / (*froissart869, 1v:42*).

Nous voyons très clairement que la marque de ponctuation est placée à chaque fois immédiatement après le syntagme contenant le nom propre. La proposition entière fait partie d'une sorte d'introduction à la chronique qui pose un cadre général. On peut supposer que les noms propres de membres de familles royales étant particulièrement importants dans ce contexte, les marques de ponctuation servent à attirer sur eux l'attention du lecteur.

La dernière occurrence survient en fin de ligne, juste après l'adverbe *or*, qui est à son tour précédé de l'une des rares occurrences de ponctuation forte :

(7.11) li peuples vodroit || monstrier sa felonnie et poissance li noble n'aueroient point de
duree a euls ¶ Or ¶ || sont il et ont este vn lonch temps moult bien d'acort ensamble
(*froissart869, 2r:25*).

Il se peut que la marque après *or* serve simplement à remplir la fin de ligne, et on peut même supposer que l'usage d'une majuscule, plus large qu'une minuscule, n'est

Manuscrits du XVe siècle

pas sans rapport avec le souci de la justification. Il est aussi possible que l'adverbe *or* souvent utilisé en français médiéval comme marqueur d'une frontière importante dans la chaîne narrative, ait attiré des marques de ponctuation à sa droite comme à sa gauche. Faute d'occurrences similaires dans l'extrait dépouillé toutes ces explications restent purement hypothétiques.

Le décalage entre la ponctuation du manuscrit et celle de l'édition de G.T. Diller (1972) est assez important : dans 52 cas l'éditeur a jugé nécessaire d'ajouter une marque de ponctuation là où il n'y en avait pas dans le manuscrit, dans 15, il n'a pas tenu compte d'une marque scribale et dans 42 cas seulement une marque de ponctuation est utilisée en même temps dans l'édition et dans le manuscrit. Par ailleurs, 82 frontières ponctuables ne sont ponctuées ni dans le manuscrit ni dans l'édition.

La suppression des marques concerne en premier lieu les contextes dans lesquels l'emploi de la ponctuation semble motivé par des raisons autres que syntaxiques, et notamment par la mise en valeur des noms propres, que nous avons évoquée ci-dessus. Les ajouts se produisent le plus souvent aux frontières de propositions autonomes et autour de subordonnées circonstancielles.

Malgré un taux général de ponctuation peu élevé, **froissart869** présente des usages qui s'inscrivent parfaitement dans les grandes tendances de la ponctuation médiévale telles qu'elles se dessinent dans les autres textes de notre corpus. La spécificité de ce manuscrit consiste dans l'extrême rareté des majuscules (et donc de la ponctuation forte) et dans la tendance à utiliser des marques de ponctuation à proximité de certains noms propres.

7.2. Paris, BnF, fr. 126 (*Quadrilogue invectif* d'Alain Chartier)

Le manuscrit fr. 126 de la Bibliothèque nationale de France est un grand in-folio contenant 263 feuillets en vélin. La différence d'écriture permet de distinguer deux parties reliées dans un même codex, mais le style des miniatures est unique, ce qui laisse supposer que le volume a été assemblé dès le XV^e siècle (Droz 1923, p. X).

Le *Quadrilogue invectif* d'Alain Chartier ouvre la deuxième partie du codex et occupe les folios 191 à 209. L'écriture de cette partie peut être datée de la première moitié du XV^e siècle et la date de composition du *Quadrilogue* (1422) permet de situer la date du manuscrit entre 1422 et 1450.

Selon E. Droz, il s'agit de l'un des plus anciens manuscrits de cette œuvre, offrant un « bon » texte dans une copie « très soignée », ce qui l'a amené à choisir ce manuscrit comme base de son édition. Il convient d'ajouter que le manuscrit est richement décoré (*cf.* par exemple la première page du texte avec une miniature et un ornement disposé autour du texte).

Le *Quadrilogue invectif* est sans doute la création la plus connue d'Alain Chartier, un chef-d'œuvre qui a eu une longue postérité. Cette « prosopopée politique, morale et satirique » (Hasenohr *et al.* 1994, p. 29) est composée de quatre très beaux discours remplis de plaintes et de récriminations et vise à réclamer l'union des trois ordres sociaux (le peuple, la chevalerie et le clergé) contre l'ennemi de la France et pour la défense du roi. C'est donc un texte caractérisé par un haut degré d'émotion et d'expressivité. Dans la typologie textuelle de la BFM cette œuvre est placée dans les domaines littéraire et politique, et son genre est défini comme lyrique.

Le texte commence par un prologue où l'auteur explique son point de vue personnel sur la situation politique en France en recourant à de nombreux parallèles historiques et en posant plusieurs questions rhétoriques. Le corps du texte est organisé comme une représentation théâtrale avec de longues prises de parole successives des personnages « ponctuées » par les répliques de l'auteur (*Acteur*) et des personnages ayant déjà prononcé leurs discours.

L'extrait que nous avons transcrit s'étend du folio 191r au folio 192v et contient 1652 occurrences mots. Son identifiant dans notre corpus est **quadrilogue126**²⁵⁴. Le taux de ponctuation dans l'extrait dépouillé atteint les 7,6%, avec une dominance de la ponctuation faible (87 occurrences contre 37 pour la ponctuation forte). Quant aux signes de ponctuation, le point est quasiment la seule marque de ponctuation syntaxique utilisée (114 occurrences). La majuscule seule est toutefois employée dans 3 cas, toujours au début d'une ligne. L'absence d'un point s'explique donc sûrement par des

254 Ce texte relativement tardif n'est pas référencé dans la base bibliographique du DEAF.

contraintes liées à la justification : chaque fois la dernière lettre du mot précédent se trouve immédiatement devant la réglure droite :

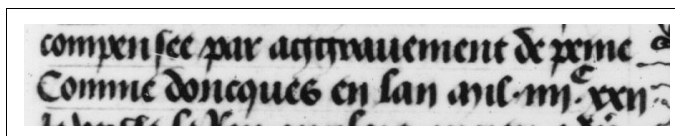


Fig. 48 : Absence de ponctuation en fin de ligne (quadrilogue126, 192a).

Le scribe n'hésite pas à dépasser la réglure pour terminer un mot, mais il évite de mettre un signe supplémentaire une fois que cette limite est atteinte.

L'extrait que nous avons transcrit contient 4 occurrences du pied-de-mouche, une marque de ponctuation destinée généralement à la structuration globale du texte. Trois de ces occurrences sont localisées dans la première page du texte, très richement décorée. Elles sont toutes situées au milieu d'une ligne et précédées d'un point. Les pied-de-mouches jouent certainement un rôle dans ce dispositif de décoration, d'autant qu'ils sont eux-mêmes entourés d'un ornement décoratif :



Fig. 49 : Pied-de-mouche (quadrilogue126, 191a).

L'usage des pied-de mouches n'est toutefois pas sans rapport avec la structure logique et syntaxique du texte : le premier d'entre eux apparaît à la fin d'une longue formule de dédicace initiale. Cette formule contient elle-même plusieurs ponctuations faibles et une ponctuation forte devant le prénom de l'auteur²⁵⁵ :

(7.12) **A** la treshaulte *et* excel||lente maieste des prin||ces . a+la treshonnouree || magnificence des || nobles . circonspection || des clers . et bonne industrie du peuple || francois . Alain Charretier humble secretaire || du roy *nostre+sire* . *et* de mon tresredoubte+seigneur monseigneur || le Regent . lointaing immitateur des orateurs || salut . en crainte de dieu . humiliacion soubz || sa iustice cognoissance de ses

255 Même si la majuscule pourrait s'expliquer ici par la volonté de mettre en relief le nom propre, sa position à une frontière ponctuable importante dans le contexte de la dédicace initiale justifie à nos yeux qu'on la considère comme une ponctuation forte.

Jugemens .et || retourner a+sa misericorde soubz la+poin||ture de sa punicion .¶
 Comme les haul||tes dignitez [...] (*quadrilogue126, 191a:1-14*).

L'usage du pied-de-mouche permet de mettre en relief la formule « protocolaire » ouvrant le texte et de représenter la hiérarchie des relations entre les éléments dont elle est composée. Il convient de noter que l'éditeur scientifique du texte (Droz 1923) a choisi d'insérer un saut de paragraphe à cet endroit.

Les deux autres occurrences du pied-de-mouche coïncident avec les seules deux frontières de propositions autonomes de la page initiale, car toutes les autres propositions sont liées par des relations de subordination ou ont des éléments communs. Il est donc possible d'affirmer que le pied-de-mouche sert là aussi à représenter la hiérarchie des relations syntaxiques.

Aucun pied-de-mouche n'apparaît au verso du premier feuillet du *Quadrilogue invectif*, même si l'organisation formelle et la visée pragmatique de la séquence textuelle restent les mêmes. Cela nous permet de formuler l'hypothèse selon laquelle la force de la ponctuation « monte d'un grade » à la page initiale, au moins dans ce texte.

La dernière occurrence du pied-de-mouche dans notre extrait est située au début du corps du *Quadrilogue*, juste après l'incipit en latin, devant une sorte de didascalie indiquant la prise de parole de l'auteur :



Fig. 50 : Lettrine et pied-de-mouche (*quadrilogue126, 192b*).

Manuscrits du XV^e siècle

Le pied-de-mouche permet ici de mettre en évidence un élément particulier de la structure textuelle et de le séparer nettement de la grande initiale qui le précède et qui est liée au texte de la ligne suivante.

Le fragment que nous avons transcrit ne contient pas de discours direct cité par les personnages qui prennent la parole.

Toutes les frontières entre les propositions autonomes sont ponctuées dans l'extrait, avec le plus souvent une ponctuation forte (32 occurrences contre 8 pour la ponctuation faible).

Les données dont nous disposons ne permettent pas d'isoler les facteurs qui déterminent le choix de la ponctuation forte ou faible dans cette position. Il semblerait même que la pratique du scribe change légèrement d'une page à l'autre. Dans les deux extraits ci-dessous, nous avons une proposition autonome introduite par la conjonction *car*.

- (7.13) Et se memoire nous puet || aucune chose ramenteuoir ¶ et les anciens || liures de noz peres apprendre a cognoistre || nos faiz par les leurs ¶ toutes anciennes || escriptures sont pleines de mutacions || subuersions et changemens des Royau||mes et des principautez ¶ car comme les || enfans naissent et croissent en hommes || parfaiz et puis declinent a viellesce et a || mort ¶ ainsi ont seigneuries leur commen||cement ¶ leur accroissement ¶ et leur declin ¶ (*quadrilogue126, 191c:30-40*) ;
- (7.14) Et combien que ces choses soient assez || euidentes a cognoistre ¶ si y errent les || pluseurs ¶ Car en racontant les faiz || qu'ilz cognoissent a l'oeil ¶ ilz demeurent || en descognoissance de la cause ¶ Et || pour+ce que les Jugemens de dieu [...]
(*quadrilogue126, 192a:9-14*).

Dans (7.13) situé au verso du folio 191, la ponctuation faible est employée devant *car* à 38 mots de la ponctuation forte et à 15 mots de la ponctuation faible qui précèdent, tandis que dans (7.14), au recto du folio suivant, la ponctuation forte est utilisée devant *car* à 15 mots de la ponctuation forte et à 5 mots seulement de la ponctuation faible qui précède. La densité de la ponctuation dans (7.14) est nettement plus élevée que dans (7.13), bien que le type de séquence textuelle et les types de constructions syntaxiques ne changent pas considérablement. Il n'y a pas non plus de changement de main apparent. La seule différence visible entre les deux pages est le décor, beaucoup plus riche au folio 192r qu'au folio 191v :

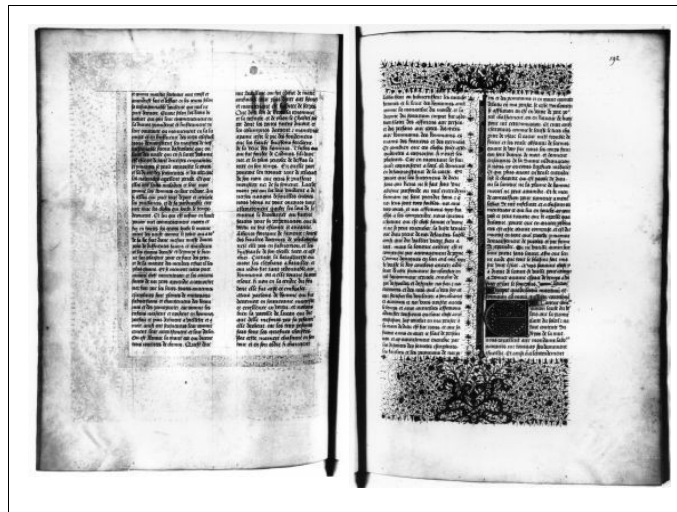


Fig. 51 : Aperçu de la décoration (quadrilogue 126, 191v et 192r).

On peut supposer que nous sommes confrontés ici au même phénomène qu'à la page initiale du texte, où le soin porté à l'aspect décoratif de la page semble se répercuter dans la densité et la force accrue de la ponctuation.

Les frontières entre les propositions partageant des éléments communs sont ponctuées dans la moitié des occurrences relevées dans notre extrait (19 sur 38). Une seule occurrence comporte une ponctuation forte :

- (7.15) Encore selon les drois de || nature qui ont leur commencement en || la diuine prouidence et l'instrument de || leur ouraige ou mouuement en la lu||miere ||et en l'influence des corps celestielz || nous demonstrent les maistres de tres||inestimable || science d'astrologie que ou || liure des cieulx qui en si large volume || est escript de+tant diuerses empreintes || et ymaiges se peult cognoistre le cours || et la duree des seigneuries et des citez que || les naturiens appellent periode . Et que || elles ont leurs maladies et leur mort || comme les hommes en leur endroit . (quadrilogue 126, 191c:15).

Cette occurrence survient vers la fin d'une longue construction comprenant plusieurs relatives et complétives contenant elles-mêmes des membres coordonnés. En effet, la complétive marquée par une ponctuation forte est coordonnée non pas avec une autre proposition en *que* (qui pourtant la précède immédiatement) mais avec un sujet nominal d'une forme réflexive du verbe 'connaître' : *se peult cognoistre le cours... et que elles ont...* Le sujet de la complétive est le pronom anaphorique *elles* qui renvoie aux modificateurs du sujet de la proposition principale (*seigneuries* et *citez*).

Manuscrits du XV^e siècle

La construction est donc loin d'être banale et peut poser des problèmes pour la compréhension lors d'une lecture rapide. L'emploi d'une ponctuation permet d'éviter ici le rapprochement avec la proposition en *que* la plus proche, et l'absence de toute ponctuation sur les 10 lignes précédentes a sans doute poussé le scribe à utiliser une marque forte.

Une différence assez nette de ponctuation semble se profiler entre les propositions avec un sujet commun et les subordinées coordonnées : les premières sont séparées par une ponctuation dans 5 cas sur 19 seulement, tandis que le taux de ponctuation est de 12 occurrences sur 13 pour les subordinées coordonnées. Le nombre d'occurrences de propositions ayant un argument commun est trop bas (4 au total) pour qu'on puisse tenter de dégager une tendance. Il est toutefois intéressant de noter que la ponctuation est présente dans deux cas sur quatre.

Le seul cas d'absence de ponctuation devant une subordinée coordonnée concerne une proposition infinitive *et auecques ce noz vices croistre...* située dans le contexte suivant :

- (7.16) Comme doncques en l'an mil .iiii^c. xxii || Je veisse le Roy anglois ancien aduer||saire de ceste seigneurie soy glorifier en || nostre Ignominieux reproche . enrichir de || noz despoilles et despriser noz faiz *et* noz || couraiges . et des *nostres* qu'il a vers soy at||raiz . fortifier les voutentez a son aliançe || *et* auecques ce noz vices croistre auecques || le temps . et a+noz aueuglees affections || adiouster tousiours quelque chose a *nostre* || confusion . (*quadrilogue*126, 192c:27-37).

Ici l'absence de ponctuation s'explique sans doute par des contraintes de justification en fin de ligne : la dernière lettre du mot précédent la frontière ponctuable touche la réglure droite. Les autres propositions infinitives de cette construction sont au contraire toutes ponctuées.

Sur 5 occurrences de frontières ponctuées entre propositions à sujet commun, deux coïncident avec des fins de propositions subordinées, comme dans l'exemple suivant :

- (7.17) Mais parlons de Romme qui fut || derreniere en souueraine mageste || et excellente en vertu || et notons || bien la parolle de Lucan (*quadrilogue*126, 191d:36).

Une des occurrences restantes est située après le dernier mot de la première page, qui est, comme nous l'avons déjà remarqué, caractérisée par un décor et une ponctuation plus riches qu'ailleurs dans le texte.

Une autre paire de propositions avec un sujet commun sont formées par des relatives coordonnées, ce qui a peut-être favorisé l'apparition d'une ponctuation :

(7.18) Je voy sur nous les coups feruz || qui sont dignes de mort **!** et donnent || ensaignes de la diuine indignacion . (*quadriologue126, 192b:10*).

La dernière occurrence d'une ponctuation à ce type de frontière syntaxique est également explicable :

(7.19) Ain||si celui qui puet tout depart et retranche || les puissances **!** et de sa pardurable eter||nite mue les choses qui soubz le temps || decourent . (*quadriologue126, 191c:19*).

Il s'agit ici d'une troisième proposition ayant un même sujet pronominal, dont la structure est plus complexe que celle des deux premières (qui ne sont composées que d'un verbe chacune et d'un objet commun), ce qui amène à la formation de deux blocs distincts. La ponctuation sert ici à marquer la hiérarchie des relations syntaxiques.

L'extrait dépouillé contient un grand nombre de propositions subordonnées (107 frontières entre principales et subordonnées relevées), trait qui caractérise le style de cette œuvre d'Alain Chartier. La majorité des occurrences (près de 70%) ne sont pas ponctuées.

Pour deux occurrences, il s'agit sans doute d'une ponctuation forte, même si dans les deux cas la lettre initiale *I/J* peut être interprétée comme une simple variante positionnelle sans valeur de ponctuation forte (ce qui est certainement le cas de *Je* et de *Ignomieux* dans l'exemple (7.16) ci-dessus et du premier *Je* dans (7.21) ci-dessous). Si toutefois on maintient l'hypothèse de la ponctuation forte, on peut remarquer que dans les deux cas, il s'agit d'une principale postposée, précédée d'une longue subordonnée contenant elle-même d'autres subordonnées et plusieurs membres coordonnés :

(7.20) Comme les haul||tes dignitez des+seigneuries soient establi||es soubz la diuine *et* *in*finie puissance qui || les eslieue en florissant en prosperite *et* en || glorieuse renomnee **!** Il est a croire *et* tenir fer||mement || *que* ainsi *que* leurs commencemens *et* leurs || accroissances *sont* maintenues et adrecees || par la diuine prouidence . (*quadriologue126, 191b:17*) ;

Manuscrits du XVe siècle

- (7.21) Et comme Je recueil||lisse en ma souenance la puissance et || diligence des ennemis .
la desloialte || de plus[i]eurs subgiez . et la perte des *princes* || et cheuallerie dont dieu
par maleureu||se bataille a laissie ce Royaume desgar||ny . qui me fait durement
ressongnier || l'issue de ceste infortune . Je contrepensoye || et pensoye a+l'encontre la
grandeur et || distance des parties de ce dit Royaume || dont les ennemis ne suffiroient
garder || le quart . (*quadrilogue*126, 192c, 15).

La frontière entre une subordonnée circonstancielle et sa principale postposée semble de façon générale assez favorable à l'emploi d'une ponctuation : 11 occurrences sur 15 sont en effet ponctuées. Les circonstancielles postposées sont quant à elles ponctuées dans 6 cas sur 15. Pour les autres types de subordonnées, soit l'usage de la ponctuation est très rare, soit le faible nombre total de ces subordonnées ne permet pas de distinguer une tendance quelconque.

Les syntagmes coordonnés à l'intérieur de propositions sont assez fréquents dans **quadrilogue126** : 86 occurrences, dont 25 ponctuées (toujours avec une ponctuation faible). Le plus souvent, il s'agit d'énumération ou de coordination d'objets directs (plus de 50 occurrences). La majorité des occurrences ponctuées concerne des énumérations composées d'au moins trois éléments, mais ce n'est pas une règle strictement respectée. La présence ou l'absence de la conjonction *et* ne semble pas avoir d'incidence sur l'emploi d'une marque de ponctuation.

Les syntagmes particuliers à l'intérieur d'une proposition ne sont que rarement ponctués dans **quadrilogue126** : nous avons relevé 9 occurrences de ce type dans l'extrait dépouillé. Dans 3 cas, il s'agit d'un point placé en fin de ligne pour des raisons de justification. Deux de ces occurrences sont visibles sur la Fig. 50 ci-dessus (entre *et* et *comicum* et entre *nuit* et *nous*).

Trois autres occurrences se trouvent dans la longue formule initiale citée dans l'exemple (7.12) devant *Alain Charretier, lointaing immitateur* et *en crainte*. A chaque fois, il s'agit d'une frontière potentiellement ponctuable (fin d'un complément circonstanciel placé devant le sujet, apposition, groupe adjectival étendu...) qui reçoit effectivement une marque dans ce contexte particulier.

Une paire de points est placée autour d'une apposition à un nom propre :

(7.22) Thebes || qui fut fondee de Cadmus filz d'age||nor et la plus peulee de dessus la || terre en son temps . (*quadrilogue126, 191d:11-12*).

Ici la ponctuation sert sans doute de moyen alternatif de mise en relief de noms propres, à côté des majuscules, déjà utilisées dans certains cas.

La dernière occurrence que nous avons classée parmi les syntagmes particuliers (en l'assimilant à des constructions comparatives) peut en effet être interprétée comme une proposition subordonnée dans laquelle le mot *non* remplace le groupe sujet-prédicat (*non* = 'elle ne tourne pas') :

(7.23) ou a elle tourne sa+grant || gloire si non en la cendre du feu || dont elle fut arse . (*quadrilogue126, 191d:32*).

En étudiant la ponctuation de l'édition d'E. Droz (1923), nous avons constaté que la ponctuation forte ou faible du manuscrit a été maintenue dans 82 cas (27% des frontières ponctuables). On trouve 6 occurrences où la ponctuation forte du manuscrit a été « affaiblie » dans l'édition, mais il n'y a en revanche aucun cas de « renforcement ». Dans 112 cas la ponctuation est absente à une frontière ponctuable à la fois dans l'édition et dans le manuscrit. Cela porte à 200 (65%) le nombre de frontières ponctuables dont le traitement est similaire dans le manuscrit et dans l'édition.

L'intervention la plus fréquente est l'ajout d'une ponctuation faible (73 occurrences). Il n'y a cependant aucun cas d'ajout d'une ponctuation forte. Ce sont les frontières de subordonnées relatives qui présentent le plus fort taux d'ajout de ponctuation.

Dans 37 cas, une ponctuation présente dans le manuscrit a été supprimée dans l'édition. Ce phénomène concerne généralement la frontière entre subordonnées coordonnées (qui ont, comme nous l'avons remarqué, une nette tendance à être ponctuées dans **quadrilogue126**).

Dans un cas, c'est une ponctuation forte du manuscrit qui a été supprimée (devant *et ainsi que* ci-dessous) :

(7.24) Enuiron l'aube du Jour [>] || lors que la premiere || clarte du soleil et na||ture contente du || Repos de la nuit [.>] || nous rappellent aux mondains labours [.>] || n'a+gaires me trouuay soudainement || esveillie [.>] Et [.>] ainsi que a l'entendement |[192c] apres repos se presente ce que l'en a plus || a cuer [.>] me vint

Manuscrits du XV^e siècle

en ymaginacion la dou||loureuse fortune et le piteux estat de la || haulte seigneurie et glorieuse maison || de france [.>] (*quadrilogue126, 192c:35-d:5*).

Il s'agit en l'occurrence d'une frontière entre propositions autonomes, où l'usage de la ponctuation forte est tout-à-fait légitime. C'est donc plutôt l'éditeur qui s'éloigne dans ce cas de la représentation de la structure syntaxique du texte.

De façon générale, la ponctuation dans **quadrilogue126** s'inscrit dans les grandes tendances mises au jour dans les manuscrits étudiés précédemment. Ce beau manuscrit nous a permis par ailleurs de constater que le soin particulier porté à la décoration des pages initiales peut également avoir une incidence sur la densité et la force de la ponctuation.

7.3. Paris, BnF, fr. 619 (*Livre de chasse de Gaston Phébus*)

Ce manuscrit en vélin de 113 folios est entièrement occupé par le *Livre de chasse* de Gaston Phébus. Le catalogue de la BnF date ce codex du XIV^e siècle. Une date différente (et sûrement plus réaliste) est avancée par Gunnar Tilander (1971, p. 24) : 1440.

L'écriture du manuscrit est assez soignée, les intertitres sont réalisés à l'encre rouge, les majuscules sont surlignées en ocre, en revanche les miniatures ne sont pas colorées.

Le *Livre de chasse*, composé par Gaston Phébus à la fin des années 1380, a connu un grand succès à la fin du Moyen Âge : G. Tilander a recensé 44 manuscrits du XV^e et du début du XVI^e siècle contenant cette œuvre, ainsi que trois éditions imprimées au cours du XVI^e siècle. Le chercheur a identifié 4 manuscrits primaires dont les autres dérivent. Le manuscrit BnF, fr. 619 (manuscrit **D** dans la nomenclature de Tilander) appartient à ce groupe, mais le texte qu'il offre diffère considérablement de la version que Tilander a préférée pour son édition (celle du manuscrit BnF, fr. 616) : il présente de nombreux changements dans l'ordre des mots, dans les graphies, ainsi que des lacunes.

Dans la grille de description typologique des textes de la BFM cette œuvre est classée dans le domaine didactique et son genre est caractérisé comme « traité ».

La transcription intégrée dans notre corpus est composée de deux sous-extraits, l'un au recto et verso du folio 1, l'autre au recto et verso du folio 4. Les folios 2 et 3 contiennent quant à eux une table des matières, un type particulier de séquence textuelle que nous avons choisi de ne pas inclure dans notre corpus pour éviter un biais trop important dans les données. L'ensemble des deux extraits comporte 750 occurrences-mots et porte l'identifiant **chasse619** dans notre corpus²⁵⁶.

Le taux de ponctuation général dans **chasse619** dépasse légèrement les 10%. La ponctuation faible est plus fréquente que la forte, mais sa prédominance n'est pas écrasante (45 occurrences contre 31). La seule marque de ponctuation utilisée régulièrement est le point (74 occurrences). Comme nous l'avons déjà signalé, les majuscules sont surlignées à l'encre ocre, ce qui permet de renforcer la mise en relief visuelle des ponctuations fortes.

Dans quelques endroits, on peut être tenté d'identifier des occurrences d'autres marques. Par exemple, l'image suivante présente un cas où l'on voit un ou deux points placés en haut de la ligne, au-dessus d'un point ordinaire. En réalité, ces points hauts font partie du tracé du caractère *L* majuscule dans le type d'écriture de ce manuscrit. Un point est également utilisé ici dans le tracé de la majuscule *I*.

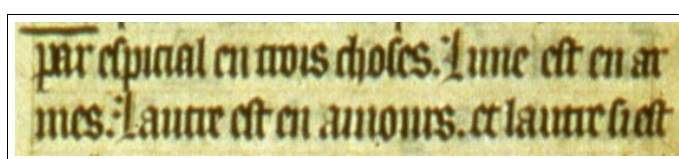


Fig. 52 : Majuscules *L* avec des points (*chasse619*, 1r).

Dans un autre cas une marque en forme de deux-points est placée en fin de ligne, en dehors d'une frontière ponctuable habituelle :

(7.25) Et sont ¶ || moult *perilleuses* bestes . (*chasse619*, 4b:7).

256 Cf. le sigle **GastPhébChasse** sous lequel le traité de Phébus est représenté dans la bibliographie du DEAF.

Manuscrits du XV^e siècle

Cette marque est certainement utilisée pour remplir le vide à la fin d'une ligne et ne joue pas le rôle de ponctuation syntaxique.

L'extrait dépouillé contient enfin une occurrence d'une majuscule seule devant un syntagme nominal apposé :

(7.26) Je gaston par la grace de dieu surnom^m||me febus **C**omte de foyz seigneur de be||arn .
(*chasse619, 1b:2*).

Il peut s'agir dans ce cas non pas d'une ponctuation syntaxique, mais d'une mise en relief d'un titre de noblesse.

L'extrait intégré dans notre corpus contient une citation d'un dicton qu'on peut assimiler à un discours direct :

(7.27) Et pour ce dit on **.** apres |[*fol. 4v*] le sangler le mire **.** et apres le cerf la || bierre **.** quar trop fort fiert einsi *comme* || vn coup de garrot . (*chasse619, 4d:10-4v:3*).

Une ponctuation faible est utilisée au début et à la fin du dicton, et la même marque se trouve à nouveau au milieu de celui-ci. Le changement de plan énonciatif n'est donc pas traité ici comme une frontière syntaxique forte, mais la ponctuation est toutefois bien présente.

Les frontières de propositions autonomes sont ponctuées assez régulièrement (c'est le cas de 42 occurrences sur 43, soit 98%). La ponctuation forte est utilisée dans cette position un peu plus souvent que la ponctuation faible (23 occurrences contre 19). Le choix entre la ponctuation forte ou faible ne semble pas dépendre de la forme du sujet, ni de la distance de la ponctuation forte précédente (la longueur des segments délimités par des ponctuations fortes varie de 5 à 86 mots). On peut toutefois remarquer que le scribe est assez sensible à la figure stylistique de l'anaphore : il met souvent une ponctuation forte, lorsque la construction d'une proposition est similaire à celle d'une proposition précédente, qui comportait déjà une ponctuation forte. Dans l'exemple suivant, il y a une répétition du pronom sujet initial suivi d'une courte forme verbale :

(7.28) Ilz sont legieres || bestes et fortes et sachantes a grant || merueille **.** Ilz vont en leur amour *que* |[*col. b*] on apelle le ruyt uers la sainte crois || de septembre . (*chasse619, 4a:8-b:2*).

On peut aussi remarquer une certaine corrélation entre la force de la ponctuation et le connecteur utilisé. Ainsi, la ponctuation est forte à ce type de frontière dans 14 cas sur 20 après la conjonction *et*, et dans 2 occurrences sur 10 après la conjonction *quar*.

Dans l'unique cas d'absence de ponctuation à une frontière entre propositions autonomes, ces propositions sont courtes, sans subordonnées, et situées à une faible distance d'une ponctuation forte précédente :

(7.29) Et de cela || ne vueill ie pas parler quar poy les chas||se l'en ¶ et pou de chiens sont qui les chas||sent ¶ (*chasse619, 1c:14*).

Les propositions ayant des éléments communs sont trop peu nombreuses dans notre extrait pour qu'on puisse déterminer avec certitude une tendance dominante. La règle générale dégagée dans la majorité des autres textes de notre corpus semble néanmoins se confirmer ici : sur 16 occurrences, 9 ne sont pas ponctuées, 6 portent une ponctuation faible et il y a une seule occurrence de ponctuation forte. Cette dernière peut s'expliquer par le fait que cette proposition (dont le sujet est commun avec deux propositions principales précédentes) entretient une relation exclusive de causalité avec la proposition suivante :

(7.30) Et lors sont il fiers . et courent || sus a l'omme eisi comme feroit vn san||glier qui fust bien eschaufe ¶ Et sont : || moult perilleuses bestes ¶ quar a grant || poine vn homme garira s'il est fort bles||sie d'un cerf . (*chasse619, 4b:7*).

Il se peut aussi que la présence d'une relative subordonnée à la deuxième proposition de ce groupe et qu'une certaine similitude de structure entre la première et la troisième proposition aient favorisé le renforcement de la ponctuation.

Les frontières entre les propositions subordonnées et principales sont ponctuées dans 6 occurrences sur 46, toujours au moyen d'une ponctuation faible. Dans un cas, l'apparition d'un point en fin de ligne peut s'expliquer par l'impératif de justification. Dans les autres cas, on peut constater que les constructions syntaxiques sont chaque fois complexes. Ainsi, dans l'exemple suivant, la subordonnée relative (*que l'en chasse...*) est séparée de son antécédent (*bestes*) par un syntagme coordonné avec un syntagme prépositionnel dont cet antécédent fait partie :

Manuscrits du XV^e siècle

- (7.31) et me||tray par chapitres de toutes natures || de bestes . et de leurs manieres et uie . que
|| l'en chasse *communement* . (*chasse619, 1c:10*).

L'emploi d'une ponctuation permet de marquer ici la rupture dans un enchaînement syntaxique ordinaire.

Dans deux cas, un point apparaît à la fin d'une relative (ou d'une chaîne de subordonnées) placée devant le noyau de la proposition principale avec postposition du sujet :

- (7.32) mes du tiers office de qui || ie ne doute que i'aye nul mestre || *combien* que ce soit
venitance . de celuy || voudray ie parler . c'est de chasse . (*chasse619, 1c:7*) ;
- (7.33) Mes des autres bestes que l'en chas||se *communement* et chiens chassent vou||lentiens .
entens ie a *parler* (*chasse619, 1c:18*).

La ponctuation facilite certainement la compréhension de ces constructions complexes. Lorsque les constructions avec la reprise de la principale après une subordonnée sont plus simples, la ponctuation est absente :

- (7.34) ceulz qui les voudront suyr a+leur droit || y aprendront mieulx (*chasse619, 1c:3*) ;
- (7.35) trestouz les autres cerfs a qui il a tolu || le ruyt . li courent sus et le tuent . (*chasse619, 4c:24*).

Bien entendu, un échantillon beaucoup plus important serait nécessaire pour vérifier si cette tendance se confirme sur l'ensemble du texte. Il semble néanmoins que la fonction distinctive de la ponctuation, qui consiste à aider le lecteur à segmenter des constructions complexes, commence à prendre une part considérable dans le système graphique de **chasse619**.

Les syntagmes coordonnés à l'intérieur d'une proposition sont ponctués dans 8 cas sur 31. Dans un cas, il s'agit d'une ponctuation forte, qui survient dans une énumération de dédicataires au début du texte :

- (7.36) **EN** non et en hon||neur de dieu . cre||ateur et seigneur || de toutes choses || et du
beneoist son || filz ihesu christ . et || du saint esperit . de toute la sainte trini||te et de la
uierge marie . Et de touz les || sains et saintes qui sont en la grace de ||[*col. b*] dieu [...] (*chasse619, 1a:1-b:1*).

Comme nous l'avons déjà constaté à plusieurs reprises, la formule (voire la page) initiale constitue un contexte particulier où la fréquence et la force de la ponctuation peuvent être supérieures au reste du texte. Le scribe a sans doute jugé nécessaire de mettre une ponctuation forte au bout de cinq compléments initiaux, devant un sixième, qui contient lui-même une coordination de niveau inférieur (*sains et saintes*) et comporte une subordonnée relative.

Les coordinations avec seulement deux membres ne sont jamais ponctuées dans l'extrait dépouillé. En revanche, la ponctuation apparaît dans certaines énumérations plus longues :

(7.37) Quar aucu||nes gens chassent . lions . liepars . che||uaux . || et buefs sauuaiges .
(*chasse619, 1c:12-13*).

Le fait que les noms d'animaux seraient porteurs d'information particulièrement importante dans un traité de chasse favorise sans doute l'emploi de la ponctuation dans cet exemple.

Nous avons relevé 8 occurrences de ponctuation (dont deux fortes) à des frontières de syntagmes particuliers à l'intérieur d'une proposition. Les deux occurrences de la ponctuation forte surviennent dans la formule initiale (devant le sujet *Je* qui suit l'énumération des dédicataires et devant le titre de noblesse *Comte* en apposition). Ces occurrences ont déjà été citées et analysées dans les exemples (7.26) et (7.36). Notons que cette formule initiale contient par ailleurs une occurrence d'une ponctuation faible devant une apposition (*dieu . creatour*).

Dans trois cas, l'emploi d'une ponctuation s'explique le plus probablement par la justification en fin de ligne (comme dans l'exemple (7.30) ci-dessus). Il convient toutefois de noter que dans les deux cas qui restent, un point placé en fin de ligne est suivi, sur une nouvelle ligne, par un groupe prépositionnel en *par* :

(7.38) tout mon temps me suy delite . || par espicial en trois choses . (*chasse619, 1b:3-4*) ;

(7.39) il ne sera ia sayson *que* touz || iours le plus grant cerf ne soit tue . || par touz les autres .
(*chasse619, 4c:27-28*).

On peut supposer que les syntagmes introduits par la préposition *par* étaient perçus par le scribe comme étant dotés d'un certain degré d'autonomie par rapport au reste de la proposition, ce qui favorisait l'emploi d'une ponctuation de justification.

Le dernier type de frontières syntaxique où nous avons relevé des occurrences de marques de ponctuation dans **chasse619** est le début de syntagmes introduits par *c'est* (au sens 'c'est-à-dire') qui contiennent des précisions ou des reformulations :

- (7.40) Ilz tuent *et* bles||sent et se combatent l'un a l'autre quant || sont au ruyt . c'est en leur amour . *et* chan||tent en leur lengaige einsi que fet vn || homme bien amoureux .
(*chasse619, 4c:6*)²⁵⁷.

Quelques lignes plus bas, la ponctuation est absente dans un contexte pourtant très similaire :

- (7.41) Et enco||re quant ilz sont au ruyt c'est en leur || amour en forest ou il ait trop petit de || biches et foison de cerfs alors se tuent || il et se blessent et se combatent . (*chasse619, 4c:15*).

À l'origine, ces syntagmes sont des propositions indépendantes, mais dans ce contexte *c'est* est en train de se grammaticaliser en locution conjonctionnelle. L'ambiguïté créée par cette grammaticalisation progressive peut être à l'origine de l'hésitation dans l'emploi de la ponctuation devant ces reformulations.

De façon générale, **chasse619** présente un taux de ponctuation des plus élevés pour un manuscrit du XV^e siècle dans notre corpus. La pratique de la ponctuation s'inscrit dans les grandes tendances de la ponctuation française médiévale avec l'usage d'une seule marque, dont la majorité des occurrences concerne les frontières entre des propositions autonomes. L'aspect décoratif (par exemple, la justification des fins de ligne) semble jouer un rôle non négligeable dans ce manuscrit très soigné. Une ponctuation distinctive destinée à faciliter la compréhension de certaines constructions complexe apparaît dans un certain nombre d'occurrences.

²⁵⁷ Cf. également la fin de l'exemple (7.32).

7.4. Lyon, BM, 765 (*Livre de chasse de Gaston Phébus*)

Ce manuscrit de la Bibliothèque municipale de Lyon semble n'avoir jamais été étudié de près. Il est daté très approximativement du XV^e siècle, son écriture cursive laisse supposer qu'il s'agit plutôt de la deuxième moitié du siècle. Le codex est composé de 199 feuillets²⁵⁸ en papier encarté de parchemin, de 282×200 mm. L'écriture, sans miniatures, est disposée sur une ou deux colonnes par page selon que le texte est en prose ou en vers.

Le *Livre de chasse* de Gaston Phébus occupe dans ce manuscrit les folios 98 à 177. Il est précédé du *Roman des deduis* (en vers) de Gace de la Buigne, un autre traité sur la chasse, mi-technique, mi-moral, mi-allégorique (Hasenohr *et al.* 1994, p. 477)²⁵⁹. La fin du volume contient un traité sur le dressage des oiseaux (folios 178 à 195) et quelques documents juridiques d'une écriture plus récente (folios 196 à 199).

Selon G. Tilander (1971), ce manuscrit (présenté sous le sigle **c** dans sa nomenclature) ne présente pas d'intérêt particulier pour la tradition textuelle du traité de Gaston Phébus, ce qui ne rend pas pour autant moins intéressante la pratique de la ponctuation de son scribe.

Comme dans le cas de **chasse619**, l'échantillon que nous avons transcrit et intégré dans notre corpus est composé de deux sous-extraits (situés au folio 100r et 102v-103r) afin d'éviter le biais qu'aurait constitué l'intégration d'une partie du prologue énumérant les thèmes abordés dans le traité et dont la structure syntaxique et la mise en forme très particulières²⁶⁰. L'extrait intégré dans notre corpus porte l'identifiant **chasse765** et comporte 584 occurrences-mots.

Avec 25 occurrences de ponctuation forte et 9 occurrences de ponctuation faible seulement, le taux de ponctuation ne dépasse pas les 5,8% dans cet extrait. Ce taux serait encore plus bas (moins de 5%) si on ne prenait pas en compte les 8 occurrences de la majuscule seule.

258 Et non de 177, contrairement à (Tilander 1971, p. 26).

259 Il convient de noter que ces deux œuvres ont plusieurs fois été imprimées ensemble au XVI^e siècle.

260 Nous avons tout de même inclus dans l'extrait le début de cette énumération (jusqu'au deuxième thème) et nous analyserons brièvement son organisation syntaxique et sa ponctuation ci-dessous.

Manuscrits du XV^e siècle

La principale marque de ponctuation utilisée est la barre oblique, dont on compte 9 occurrences suivies d'une minuscule (ponctuation faible) et 11 occurrences suivies d'une majuscule (ponctuation forte).

L'extrait présente également 9 occurrences du pied-de-mouche, ce qui signifie que la fréquence relative de cette marque « extra-forte » est assez élevée. Les pied-de-mouches sont parfois utilisés seuls (6 occurrences), parfois précédés d'une barre oblique (3 occurrences). Ils sont écrits avec la même encre (et sans doute en même temps) que le corps du texte et prennent souvent une forme simplifiée, ce qui les oppose aux pied-de-mouches colorés, ajoutés par des rubricateurs, que nous avons rencontrés dans d'autres manuscrits. Ceci rapproche, au niveau de la perception visuelle, les pied-de-mouche de la ponctuation « syntaxique » ordinaire. Une encre rouge est en revanche utilisée pour surligner la plupart des majuscules :



Fig. 53 : Pied-de-mouche et majuscule surlignée (*chasse765, 102v*).

La présence ou l'absence de surlignage sur les majuscules initiales d'unités ponctuables ne semble pas avoir de signification linguistique. On peut supposer que le rubricateur n'a simplement pas repéré toutes les occurrences. En revanche, les majuscules utilisées comme des variantes positionnelles de certaines lettres en dehors de frontières ponctuables (comme, par exemple, *R* ou *I/J*) ne sont presque jamais surlignées²⁶¹, ce qui signifie qu'il ne s'agissait pas d'une opération purement mécanique.

Le pied-de-mouche est régulièrement utilisé dans l'énumération des thèmes du traité qui occupe une grande partie du prologue. Il est intéressant de noter que le contenu de cette énumération est pratiquement identique à celui de la table des matières située juste avant le prologue. La différence entre les deux est dans la structuration syntaxique : la table des matières est une succession de propositions indépendantes qui commencent toutes par la formule *Ci devise...* ('Ici, il s'agit de...'), tandis que l'énumération des thèmes prend la forme d'une séquence textuelle narrative composée

²⁶¹ Nous avons toutefois relevé deux exceptions à la ligne 38 du folio 101r, où un *p* dans un *pas* forclusif (une minuscule élargie plutôt qu'une majuscule) est surligné en début d'une ligne. Il s'agit très probablement dans ce cas d'une erreur occasionnelle. Une autre exception sera analysée plus bas.

de plusieurs longues propositions avec un centre prédicatif *je parlerai* et une série de compléments organisés à l'aide de connecteurs énumératifs (*premierement, secondement, puis, apres, etc.*) :

- (7.42) ¶ Et parleray premierement des bestes douces qui viennent pource || qu'elles sont plus gentilz et plus nobles ¶ Et premierement du cerf || et de toute sa nature ¶ Secondement du Rangier et de toute sa nature [...] (*chasse765, 100r:23-25*).

Malgré la différence de structuration syntaxique, la mise en forme « ponctuationnelle » sert dans les deux cas à mettre en relief chaque élément énuméré grâce à un usage extensif du pied-de-mouche.

L'extrait dépouillé contient une citation d'un proverbe qu'on peut considérer comme un discours direct :

- (7.43) Et pour||ce dit on ¶ apres le sanglier le mire et apres le cerf la bierre ¶ Car trop || fort fiert [...] (*chasse765, 102v:27*).

Le proverbe est précédé d'un verbe de parole avec un pronom indéfini. La frontière ponctuable entre l'expression *dit on* et le proverbe ressemble beaucoup à celle de la fin d'une incise. L'absence de ponctuation à cet endroit n'est donc pas du tout étonnante. Le proverbe est suivi d'une proposition autonome introduite par la conjonction *car*, ce qui favorise l'emploi d'une ponctuation forte (majuscule seule). Il convient toutefois de noter que la majuscule de *Car* n'est pas surlignée.

Les frontières entre propositions autonomes sont ponctuées dans la majorité des cas (15 occurrences de la ponctuation forte et 6 occurrences de la ponctuation faible contre 7 occurrences d'absence de ponctuation).

Six occurrences d'absence de ponctuation sur sept surviennent sur une même page du manuscrit (102r), dans un espace d'une douzaine de lignes. Il s'agit donc d'un passage particulièrement peu ponctué où les règles habituelles ne semblent plus fonctionner²⁶².

262 Ce passage présente par ailleurs une occurrence de ponctuation forte à une frontière qui n'est généralement pas considérée comme « ponctuable », nous reviendrons sur ce cas dans l'exemple (7.49).

Il est difficile d'expliquer ce changement de tendance à partir de la transcription du bref extrait dont nous disposons. Une étude basée sur un échantillon plus représentatif sera nécessaire pour résoudre cette énigme. Un changement de main ou la fatigue du scribe peuvent être à l'origine de ce phénomène.

La ponctuation apparaît rarement aux frontières de propositions ayant des éléments communs : une seule occurrence sur 11 est ponctuée. Il s'agit d'une ponctuation faible séparant deux propositions relatives coordonnées :

(7.44) Mais || des aultres que l'en chace *communement* / et que chiens chacent volentiers ||
entens je a parler / pour aprendre moult de gens qui veulent chacier || *qui* ne le saient
mie faire ainsi *comme* ont par aventure la voulente (*chasse765, 100r:20-22*).


L'extrait étudié présente par ailleurs une autre occurrence du même type de frontière syntaxique (située deux lignes plus bas dans le même exemple). Dans ce deuxième cas, la frontière entre les deux propositions coïncide avec un saut de ligne, qui peut parfois, comme on le sait, se substituer à une marque de ponctuation²⁶³. Il est donc possible de supposer que la frontière entre subordonnées coordonnées est plus favorable à la ponctuation dans **chasse765** que les autres types de frontières de cette catégorie (propositions avec un sujet, un complément ou un circonstant commun). Bien entendu, cette hypothèse devra être vérifiée sur un échantillon plus large, voire sur l'ensemble du texte.

Une seule occurrence de ponctuation a été relevée à une frontière entre les propositions principale et subordonnée (sur un total de 34 occurrences). Il s'agit donc ici d'une tendance très nette à ne pas ponctuer ce type de frontière syntaxique. L'occurrence ponctuée concerne une subordonnée infinitive avec la préposition *pour*, citée dans l'exemple (7.44) ci-dessus. La première lettre de la préposition est légèrement plus grande que les autres, ce qui fait supposer qu'il s'agit peut-être d'une majuscule (qui n'est pourtant pas surlignée). Dans le doute, nous considérons cette lettre comme

²⁶³ Il convient toutefois de noter dans l'exemple (7.44) où la ligne devant la frontière ponctuable est plus courte que les lignes précédente et suivante, l'absence d'une marque de ponctuation ne peut s'expliquer par des raisons de justification des fins de ligne. Nous verrons qu'au contraire ce manuscrit contient des occurrences de barres obliques qui ne servent apparemment qu'à « rallonger » des lignes courtes (7.48).

une minuscule élargie et classons cette occurrence comme une ponctuation moyenne. La seule hypothèse que nous puissions avancer pour expliquer cette occurrence est qu'elle sert à séparer deux syntagmes prépositionnels avec des infinitifs (*a parler* et *pour apprendre*), et à indiquer que le second se rattache à l'ensemble de la proposition précédente et se situe à son tour en tête d'une construction complexe comprenant plusieurs subordonnées relatives.

Les syntagmes coordonnés à l'intérieur d'une proposition ne sont ponctués que dans 3 occurrences sur 27. Deux de ces occurrences concernent l'énumération des thèmes du traité, ponctuée avec les pied-de-mouches dont nous avons déjà parlé dans l'exemple (7.42). La troisième occurrence survient un peu plus haut dans le même passage :

(7.45) mectray par chapitres de toutes natures || de bestes  et de leurs manieres et vie que l'en chace communement (*chasse765, 100r:17*).

Il s'agit d'une petite barre oblique presque « collée » à la perluète qui la suit et il est possible que cette marque ait été ajoutée *a posteriori* :

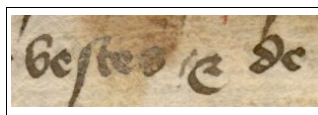


Fig. 54 : Petite barre oblique devant une perluète (*chasse765, 100r*).

Quoi qu'il en soit, la marque est utilisée dans une construction assez particulière : devant un deuxième (des trois) compléments de prédicat coordonnés, dont le premier contient l'antécédent d'une relative placée à la fin de la construction (*bestes... que l'en chace*). La présence de cette relative détachée change le statut du deuxième et du troisième termes coordonnés qui deviennent en quelque sorte un complément d'information qu'on aurait pu placer entre parenthèses dans le système moderne. Il se peut que le scribe (ou un correcteur) du manuscrit ait été sensible à cette particularité sémantico-syntaxique, ce qui l'a poussé à utiliser (non sans hésiter) une marque de ponctuation à cet endroit.

L'emploi de la ponctuation à des frontières de syntagmes particuliers est rare, mais pas exceptionnel dans **chasse765**. On y trouve 7 occurrences de ce type au total.

Manuscrits du XV^e siècle

Dans deux cas il s'agit d'une ponctuation forte devant des précisions introduites par l'expression *c'est* ('c'est-à-dire'), comme dans l'exemple suivant :

(7.46) quant il sont en Ruit **C**est en leur amour (*chasse765, 102v:30*).

Quelques lignes plus bas, la ponctuation est absente dans un contexte identique. Il est intéressant de noter que cette alternance de présence et d'absence de ponctuation est exactement la même dans **chasse765** qu'aux endroits correspondants dans **chasse619** (cf. les exemples (7.40) et (7.41)). Nous pouvons donc proposer la même explication à ce phénomène : il s'agit d'un noyau de proposition autonome (*c'est...*) qui est en train de se grammaticaliser comme simple connecteur de reformulation, d'où une hésitation dans l'utilisation de la ponctuation.

Deux autres occurrences de marques de ponctuation (majuscules seules) à des frontières de syntagmes particuliers sont enregistrées dans la formule initiale, devant le sujet pronominal qui suit une longue série de circonstants antéposés et devant un titre de noblesse en apposition :

(7.47) **E**n Nom et en souuenz de dieu createur [...] **J**e gaston || par la grace de dieu surnomme phebus **C**onte de foys seigneur de || Bran Ru (*chasse765, 100r:1-5*).

Comme nous l'avons déjà indiqué, les formules initiales présentent des constructions syntaxiques très particulières où les règles d'usage de la ponctuation peuvent être différentes de celles du reste du texte. La présence de noms propres et de titres de noblesse favorise sans doute également l'emploi des majuscules.

L'extrait analysé présente trois occurrences de ponctuation en dehors de frontières « ponctuables » habituelles. Dans deux cas, il s'agit de barres obliques placées en fin de deux lignes consécutives entre respectivement un déterminant numéral et un nom, et une préposition et un nom :

(7.48) me suis delite par especial en trois **/** || choses / L'une est en amours / L'autre est en armes / et l'autre si est en **/** || chace / (*chasse765, 100r:6-7*).

Les lignes dont il s'agit dans cet extrait sont légèrement plus courtes que celles leur entourage ; et l'emploi des barres obliques peut être motivé par la justification. Un

examen de plusieurs folios de **chasse765** non transcrits dans cette étude permet de constater qu'une telle pratique n'est pas exceptionnelle. On peut même supposer que ces barres jouent plutôt le rôle de marques de liaison, de la même façon qu'en cas de coupure de mots en fin de ligne.

La dernière occurrence « atypique » qui demande une explication est également associée à un saut de ligne. Il s'agit d'une majuscule seule (surlignée) en début de morphème prépositionnel introduisant un complément du prédicat :

(7.49) ilz vont || **E**n leur amour que l'en appelle le Ruit (*chasse765, 102v:21*).

Nous avons déjà indiqué que ce passage du texte (folio 102v) présente une pratique de ponctuation particulière (avec notamment l'absence de ponctuation entre des propositions autonomes). La raison de l'apparition de la majuscule ici n'est ni syntaxique, ni sémantique, il s'agit plutôt d'un emploi décoratif en début de ligne (peut-être, sous l'influence de textes en vers), qui se trouve juste au-dessous d'une grande lettrine en couleur.

De façon générale, la ponctuation dans **chasse765** est conforme aux tendances générales dégagées dans le reste du corpus, même si ce manuscrit présente un certain nombre de particularités intéressantes (comme l'usage « syntaxique » des pied-de-mouches et la ponctuation « de liaison » en fin de ligne).

7.5. Paris, BnF, fr. 13318 (*Sermon pour la fête de la Sainte Trinité de Jean Gerson*)

Ce manuscrit de la BnF ne porte pas d'indices particuliers permettant de le dater précisément. Les 180 feuillets de ce codex sont en parchemin et mesurent 185×140 mm, la reliure est en veau brun, l'écriture est disposée sur une colonne par page. L'écriture est très soignée et appartient, de toute apparence, à une seule main.

Le manuscrit entier est occupé par 14 sermons de Jean Gerson (1363 – 1429) composés au début du XV^e siècle. Le sermon pour la fête de la Sainte Trinité (souvent référencé par son premier mot *Vidimus*) a été sans doute composé en 1402 (Mourin

Manuscrits du XV^e siècle

1946b, p. 92). Il appartient, bien entendu, au domaine religieux, selon la grille typologique de la BFM. Dans le manuscrit BnF, fr. 13318, ce sermon occupe les folios 165r à 175r.

Le catalogue de la BnF ne propose pour ce manuscrit qu'une datation très approximative du XV^e siècle. L. Mourin arrive dans son étude de la tradition manuscrite des sermons français de Jean Gerson (1946a, p. 260) à la conclusion qu'« il faut considérer le beau manuscrit B.N. français 13318 comme une copie sur la datation de laquelle nous n'avons pas d'indication, mais qui remonte, pour les sermons qu'il possède en commun avec le manuscrit victorin, à une tradition plus correcte ». Il émet également une hypothèse concernant le commanditaire de la copie : « Probablement, ce manuscrit en parchemin, si soigné, fut-il primitivement écrit pour quelque personnage de l'aristocratie » (*idem*, p. 249).

Dans le stemma construit pour son édition du sermon pour la fête de la Sainte Trinité (Mourin 1946b), le chercheur place ce manuscrit (H, selon sa nomenclature) près de l'archétype, ce qui laisse supposer que la date du manuscrit n'est pas très éloignée de la date de composition des sermons. Nous avons toutefois préféré, faute de preuves solides, conserver pour ce manuscrit une fourchette de datation large (de 1403 à 1499) et de retenir la date moyenne de 1451 pour le tri chronologique des textes du corpus.

L'extrait intégré dans notre corpus correspond grossièrement à la première moitié du sermon pour la fête de la Sainte Trinité (folios 165v à 170r) et comporte 2102 occurrences-mots. Nous nous y référerons sous l'identifiant **gerson13318**²⁶⁴.

Le taux de ponctuation général y atteint les 5,6% avec 50 occurrences de la ponctuation forte, 59 occurrences de la ponctuation faible et 6 occurrences de la ponctuation moyenne (minuscules élargies). Les marques de ponctuation utilisées dans ce manuscrit sont assez variées. On trouve 52 occurrences du point, 29 occurrences de la barre oblique ordinaire (*virgula*), 16 occurrences d'une barre oblique avec un point au milieu (*virgula suspensiva*) et deux occurrences du crochet adlinéaire (*paragraphus*).

²⁶⁴ Dans la bibliographie du DEAF, toutes les œuvres de Jean Gerson sont référencées sous le sigle **JGerson**. La seule édition qui y est mentionnée est celle des *Œuvres complètes* (Glorieux 1960-1973).

L'extrait comporte également 15 occurrences de la majuscule seule en tête d'unité ponctuable et une occurrence de minuscule élargie seule.

Le point est utilisé dans la ponctuation faible aussi souvent que dans la ponctuation forte (26 occurrences contre 24 respectivement), tandis que la barre oblique et la *virgula suspensiva* ont tendance à être employées dans la ponctuation faible plutôt que dans la ponctuation forte (21 contre 5 et 12 contre 4 occurrences respectivement).

Outre l'usage des marques de ponctuation, **gerson13318** présente une autre technique de marquage des segments textuels. Certains passages sont en effet surlignés d'un trait de couleur rouge. Il s'agit de citations en latin et des références de la source (le plus souvent, un texte biblique) :

(7.50) car malice empesche a || congnoistre dieu et verite . sapiencie .ij. Excauit || eos malicia eorum . (*gerson13318, 167r:12-13*).

Nous n'allons pas jusqu'à affirmer que le système graphique de **gerson13318** présente réellement une triple opposition systématique de caractères majuscules, minuscules « ordinaires » et minuscules « élargies », ces dernières étant des marques de ponctuation « moyenne ». Nous avons simplement relevé un certain nombre d'occurrences (7 au total) où la taille et la forme d'un caractère ne permettent pas de dire avec certitude s'il s'agit d'une majuscule ou d'une minuscule. C'est par exemple le cas du deuxième < p > surligné dans l'illustration ci-dessous :

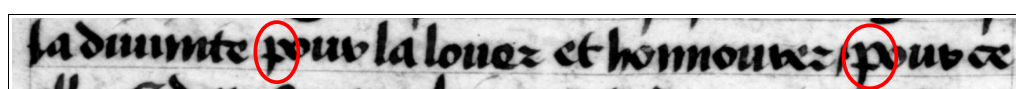


Fig. 55 : P minuscule « ordinaire » et « élargie » (*gerson13318, 166v*).

Le caractère est légèrement plus grand qu'un < p > ordinaire (premier surligné), mais les tracés des deux caractères sont très similaires. Quatre des six occurrences de minuscules élargies dans notre extrait sont en effet des < p >, les deux autres occurrences sont un < l > plus long que d'ordinaire situé sur la première ligne d'une page (167r) et un < n > (168v:22). Nous avons reproduit dans l'illustration ci-dessous quatre occurrences de du mot *nous* avec, selon notre interprétation, une minuscule « ordinaire » (première à gauche), une minuscule « élargie » (deuxième à gauche) et deux majuscules :



Fig. 56 : N minuscule « ordinaire », « élargie » et majuscule (gerson13318, 168v, 169v et 170r).

La difficulté d'interprétation de la « casse »²⁶⁵ de certains caractères existait sans doute dès l'époque de la création du manuscrit. La grande majorité des majuscules dans **gerson13318**²⁶⁶ ont en effet été surlignées en rouge par un rubricateur²⁶⁷, ce qui n'est pas le cas de 4 minuscules élargies sur 6 (notamment, du deuxième < p > dans Fig. 55 et du deuxième < n > dans Fig. 56).

La marque de ponctuation sous forme d'une barre oblique avec un point au milieu < / · >, dont **gerson13318** est le seul témoin dans notre corpus, est répertoriée dans (Parkes 1992, p. 307) en tant que variante de la *virgula suspensiva* (barre oblique). En suivant le choix terminologique de (Haugen éd. 2006a, p. 139), nous utiliserons le terme de *virgula suspensiva* pour désigner spécifiquement cette variante de la barre oblique. Le chercheur indique que ce signe était utilisé par certains écrivains humanistes du XIV^e siècle pour marquer une disjonction plus importante que celle notée par une simple barre oblique < / > et moins importante que celle notée par un comma < ; >.

Faute d'occurrences du comma, nous pouvons seulement comparer l'usage de la *virgula suspensiva* avec celui de la barre oblique ordinaire. Premièrement, il convient de noter que la barre oblique ordinaire (courte ou longue) est presque deux fois plus fréquente que la *virgula suspensiva* (29 occurrences contre 17). Les deux marques sont utilisées plutôt dans la ponctuation faible, et cette tendance est légèrement plus marquée pour la *virgula suspensiva* (82% contre 72% pour la barre oblique). La *virgula suspensiva* est utilisée habituellement entre propositions autonomes. C'est le cas de 14 occurrences sur 17. Les trois autres occurrences concernent les frontières de propositions ayant des éléments communs. La barre oblique ordinaire est utilisée à des frontières de types plus variés (par exemple, entre principale et subordonnée ou entre des syntagmes coordonnés), même si les frontières de propositions autonomes restent

265 Nous utilisons ce terme d'imprimerie dans un sens figuré, bien entendu.

266 A l'exception de < R > et de < J > variantes calligraphiques hors frontières ponctuables et d'une occurrence de < E > (*Excauit* au début d'une citation latine, 167r:12).

267 Dans nos citations du corpus, ces lettres surlignées seront présentées en gras, sur un fond rose.

majoritaires (18 occurrences sur 29). La différenciation fonctionnelle entre la barre oblique et la *virgula suspensiva* ne semble donc pas significative, et ces deux marques peuvent être considérées comme des allographes d'un même graphème.

Le crochet adlinéaire se distingue clairement par sa forme du pied-de-mouche, que nous avons déjà rencontré dans d'autres manuscrits : il s'agit d'une marque en forme de ligne noire brisée à angle droit, ressemblant à la lettre grecque *gamma* < Γ > (tandis que le pied-de-mouche ressemble à une grande lettre < C >)²⁶⁸ :



Fig. 57 : Crochets adlinéaires (gerson13318, 166r et 168v)

L'une des deux occurrences se trouve en début de ligne devant une citation latine soulignée en rouge. Il s'agit d'une reprise de la citation biblique qui se trouve au début du texte et qui détermine le thème du sermon : *Videmus nunc per speculum in enigmate* (I Cor., XIII^o). Cette reprise correspond dans la structure logique du sermon au début de l'*Introduction du thème*, tandis que le texte qui précède constitue le *Prothème* (Mourin 1946b, p. 151-152).

L'autre occurrence du crochet adlinéaire se situe en milieu de ligne et marque également une frontière dans la structure du sermon : le passage de la *Division* (où l'auteur explique ses intentions) au *Premier dialogue* (où l'*Âme dévote* parle à sa *Sœur Raison*).

En revanche, une autre frontière structurelle importante (le début de la *Division*), n'est marquée dans le manuscrit que par une simple ponctuation forte (point suivi d'une majuscule surlignée).

La fonction du crochet adlinéaire dans **gerson13318** est donc similaire à celle du pied-de-mouche dans d'autres textes et consiste à marquer les débuts de divisions importantes.

Le sermon pour la fête de la Sainte Trinité contient dans sa première partie un nombre important de citations bibliques, tandis que sa deuxième partie est construite en

²⁶⁸ Il convient de noter que dans certains manuscrits la forme de la marque utilisée ne permet pas de juger avec certitude s'il s'agit d'un pied-de-mouche ou d'un crochet adlinéaire. Ce n'est pas le cas de **gerson13318**.

Manuscrits du XV^e siècle

forme de dialogue et est donc essentiellement constituée de passages au discours direct. La frontière la plus régulièrement ponctuée dans le cadre de cette catégorie est le changement de locuteur : trois des quatre occurrences relevées comportent une ponctuation forte et une ponctuation faible :

- (7.51) [La Raison :] Je le apperceoy dedans toy et au dessus de toy |[fol. 169v] aucunement nostre dieu *comme* en vng vmbraige || et en miroir | bien est vray dit l'ame humaine
(*gerson13318, 169v:2*).

La reprise du récit à la fin d'un discours direct ou d'une citation est également ponctuée dans la majorité des cas. Dans un cas, il n'y a pas de marque de ponctuation à la fin d'une citation latine (suivie d'une traduction et de l'explication), mais le découpage des unités est assuré par le soulignement du fragment latin :

- (7.52) sy fais icy l'ar||gument de ihesu+crist Si terrena etc se Je vous dy || choses terriennes dedans vous
(*gerson13318, 168r:19*).

Dans un autre cas, il semble y avoir à la fin d'une citation un espace blanc plus long que d'ordinaire :

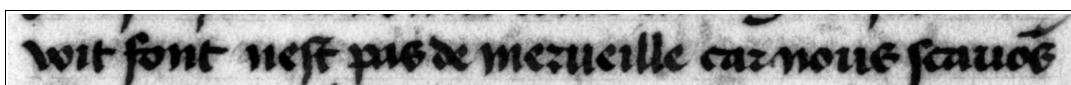


Fig. 58 : Espace blanc allongé (*gerson13318, 170r*)

- (7.53) **C**ar dit il et plus y pense et moins le puis || Je describe ou nommer *comme* en vng abisme on n'y || voit font | n'est pas de merueille car nous scauons || que incomprehensible ne puet estre compris et || autrement ne seroit il point tel /
(*gerson13318, 170r:6*).

Cette dernière occurrence est intéressante, car il s'agit de la fin d'un discours direct cité à l'intérieur d'un long discours direct de l'un des personnages du dialogue. L'éditeur du texte (Mourin 1946b) a même jugé nécessaire d'insérer un saut de paragraphe à cet endroit. Cependant, aucune marque linguistique n'indique la fin de la citation. La proposition suivante n'a pas de sujet, et on peut être tenté de le chercher dans le contexte gauche. Dans la proposition suivante, le sujet prend la forme de la première personne du pluriel, ce qui contribue à donner l'impression de la continuité du discours direct. On peut supposer que le scribe a hésité dans l'interprétation de la

structure de ce passage et a sans doute laissé la place pour qu'un correcteur puisse éventuellement insérer une marque de ponctuation, ce qui n'a jamais été fait.

Les débuts de citations en latin prennent également une ponctuation forte, qui est renforcée davantage par le soulignage (7.52). En revanche, le discours direct précédé d'une proposition introductive contenant un verbe de parole n'est jamais ponctué :

(7.54) tout c'est²⁶⁹ escrie a haulte || voix dieu nous a faiz / (*gerson13318, 168v:21*).

Ce type de frontière est proche de celle de l'incise, où l'absence de ponctuation est typique pour la grande majorité de manuscrits médiévaux. **Gerson13318** s'inscrit parfaitement dans cette tendance, à l'exception d'une seule occurrence :

(7.55) belle suer Je te diray **R**espond Raison || **J**'ay cerche par mer et par terre par l'air et par || le ciel **J** (*gerson13318, 168v:15-16*).

Ici l'incise *respond raison* est précédée d'une petite barre oblique et suivie, sur une nouvelle ligne, d'une majuscule < J > surlignée en rouge. Les majuscules < R > sont dans **gerson13318** de simples variantes positionnelles : tous les *r* en position initiale prennent la forme de majuscule dans ce manuscrit, quelle que soit leur position syntaxique. Ils ne sont par ailleurs jamais soulignés. L'usage de la barre oblique devant l'incise peut s'expliquer par le souhait de séparer deux verbes de parole appartenant à des instances énonciatives différentes qui se sont trouvés côte à côte. La majuscule surlignée²⁷⁰ après l'incise, que nous interprétons comme une marque de ponctuation forte, peut être due au fait que l'incise est placée entre deux propositions autonomes qui auraient pu être ponctuées en l'absence d'incise.

Les apostrophes, dont on trouve 8 occurrences dans l'extrait dépouillé, ne sont jamais ponctuées ni à gauche, ni à droite :

(7.56) **N**ennil voir nostre ame *nostre* || dieu n'est Riens de celles choses (*gerson13318, 169v:10*).

269 On écrirait *s'est* en orthographe moderne.

270 Cette précision est importante, car < J > initial peut également être une variante positionnelle sans valeur ponctuationnelle. A la différence de < R >, cette lettre peut être surlignée quand elle se trouve au début d'une unité ponctuable et la minuscule < j > est utilisée en position initiale de certains mots. Il convient toutefois de noter que toutes les occurrences du pronom personnel JE (24 au total) portent une majuscule.

Manuscrits du XV^e siècle

Dans l'exemple cité ci-dessus, la ponctuation est absente malgré le risque de confusion créé par la répétition du déterminant *nostre* utilisé d'abord dans une apostrophe et puis devant le sujet de la proposition.

L'extrait dépouillé contient également trois propositions incidentes, qui ne sont pas non plus marquées par une ponctuation quelconque :

(7.57) **E**n || l'autre annee **b**ien m'en souuient **J**e parlay de || trois mirouers de nature de l'escripture et de hu||mai~~n~~ne || creature (*gerson13318, 168r:8*).

La majuscule de *Je* (non surlignée) dans ce cas ne peut être considérée comme une marque de ponctuation forte, car, comme nous l'avons déjà remarqué, toutes les occurrences de ce pronom personnel dans **gerson13318** portent une majuscule.

A l'intérieur des passages homogènes du point de vue énonciatif, ce sont les frontières de propositions autonomes qui portent le plus souvent une marque de ponctuation, même s'il ne s'agit pas d'une majorité écrasante des occurrences. Si on enlève les 5 occurrences de frontières liées à des divisions importantes dans la structure générale du texte, qui portent toujours une ponctuation forte (et notamment les deux crochets adlinéaires), il y a à ce type de frontière 25 occurrences de la ponctuation forte, 3 occurrences de la ponctuation moyenne, 28 occurrences de la ponctuation faible et 33 occurrences d'absence de la ponctuation. Autrement dit, près des deux tiers des frontières de propositions autonomes sont ponctuées avec une répartition pratiquement égale entre la ponctuation forte et faible.

En cas de présence dans la deuxième proposition d'un sujet nominal, la proportion des occurrences ponctuées s'élève à 75%, et la ponctuation forte est deux fois plus fréquente que la faible dans ces conditions. Il est difficile de dégager un facteur unique qui détermine l'emploi ou non d'une ponctuation dans chaque cas précis.

La distance entre les ponctuations fortes est loin d'être constante. Elle peut varier d'une vingtaine à plus de 200 mots dans certains cas. Les passages les plus longs sans ponctuation forte se trouvent aux folios 166v (explication de l'unicité de la nature humaine) et à la fin du folio 169r – au début du folio 169v (la Raison explique à sa sœur Âme pourquoi elle n'est pas Dieu). L'unité thématique de ces passages fait que l'absence de la ponctuation forte y paraît assez naturelle.

En ce qui concerne les ponctuations faibles entre des propositions autonomes, nous pouvons supposer que la longueur de la première proposition joue un certain rôle : plus elle est courte, moins il y a de chances de rencontrer une ponctuation. Dans l'exemple suivant, une ponctuation faible apparaît seulement devant la troisième des 4 propositions interrogatives successives :

(7.58) est ce chose coulouree est || ce chose enluminee / est ce chose que on puisse tou||cher taster ou embrasser vient elle dedans par la || porte des yeulz ou des oreilles ou de aucuns des || v. sens de mon corps . (*gerson13318, 169v:6-10*).

Dans cet exemple, la répétition du syntagme *est ce chose* au début des propositions remplit déjà la fonction d'un « ponctuant lexical » et permet de repérer facilement les frontières syntaxiques et d'identifier la modalité interrogative.

Certains mots-connecteurs semblent favoriser l'absence de la ponctuation. Par exemple aucune²⁷¹ des quatre occurrences de *neantmoins* n'est précédée d'une ponctuation, même si ce connecteur met en relation des constructions assez complexes comprenant plusieurs subordonnées, comme dans l'exemple suivant :²⁷²

(7.59) **N**ous ne *sommes* choses que on puisse || ouir ou oudourer ou assauouer ou taster / neant||moins quant Je congnois nostre dieu Je *congnois* || vne lumiere toute autre que n'est celle que les || yeulz voyent (*gerson13318, 169v:16*).

Les limites entre les propositions avec des éléments communs sont ponctuées assez rarement, c'est le cas de 6 occurrences sur 30 seulement (soit 20%), et il s'agit alors toujours d'une ponctuation faible.

La ponctuation apparaît, par exemple, après la cinquième des six subordonnées juxtaposées (avec l'omission du verbe *estre*). En l'absence de ponctuation, cette séquence demande un certain effort pour être comprise²⁷³, ce qui nous amène à fournir

271 Il existe un doute concernant une occurrence à la ligne 10 du folio 169r où on semble apercevoir une toute légère barre oblique. Un examen attentif de l'image dont nous disposons nous amène à conclure qu'il s'agit d'un élément du tracé du caractère < n >. Du point de vue syntaxique, l'apparition d'une ponctuation serait d'autant plus étonnante dans ce cas concret qu'il s'agit, selon notre classification, d'une reprise de principale après une subordonnée, une position où la ponctuation est très rare dans **gerson13318**.

272 Cf. aussi l'exemple (7.60) ci-après.

273 La difficulté d'interprétation de la séquence est liée, entre autres, au fait que l'adjectif attribut passe de la position pré-sujet à la position post-sujet à partir de la quatrième subordonnée.

Manuscrits du XVe siècle

dans l'exemple donné ci-dessous la ponctuation proposée par l'éditeur, à côté de celle du manuscrit :

- (7.60) J'ay Regarde et || aduise du tout en tout quelle tu es [> ,] comme grande || est ta memoire [> ,] clere ton intelligence [> ,] franche ta || volente [> ,] ton essence immortelle [/ > ,] ta vie espirituelle [> ;] || neantmoins tu n'es pas dieu (*gerson13318, 169r:14-18*).

Dans un cas, la présence de la ponctuation peut être expliquée par le fait que la frontière entre propositions ayant un sujet commun coïncide avec la fin d'une série de syntagmes coordonnés (7.61). Un autre exemple (7.62) témoigne toutefois de l'absence de toute ponctuation dans des conditions similaires :

- (7.61) J'ay cerche par mer et par terre par l'air et par || le ciel ! et ay demande a la terre se elle estoit mon || dieu (*gerson13318, 168v:17*) ;
- (7.62) saluons le saint temple la sale et habitacle || singulier de+la+diuinite la vierge glorieuse et luy || disons deuotement . Ave maria . (*gerson13318, 166r:8-10*).

Quoi qu'il en soit, la présence d'éléments communs à plusieurs propositions semble avoir une incidence assez importante sur le choix de la ponctuation dans **gerson13318**.

Les limites entre les principales et subordonnées sont ponctuées encore plus rarement que limites entre propositions partageant des éléments communs. C'est le cas de 5 occurrences sur 113 seulement (moins de 5%). Il s'agit dans tous les cas de subordonnées circonstancielles.

Une occurrence de la ponctuation forte devant une subordonnée comparative survient dans le contexte particulier de l'insertion d'une citation latine, la citation étant mise en évidence par la ponctuation forte et par le soulignement. La rupture créée par cette citation nous amène à recatégoriser cette frontière comme une limite entre propositions autonomes, d'autant plus que la comparative concernée a elle-même une structure assez longue et complexe :

- (7.63) Sur toutes choses les pechiez de || charnalite Rendent obscure l'ame et+la+font impuissante || a Riens congnoistre de dieu et de ses choses spiritu||elles / Patet ibidem . Comme par exemple nous veons || en vng miroir ou en oeil se ilz sont chassieux de || boue et d'ordure / l'oyeil ne puet Riens veoir / et ou || miroir on ne puet Riens apparceuoir (*gerson13318, 167r:13-19*).

Il convient de noter qu'aucune des huit autres occurrences de comparatives en *comme* n'est ponctuée dans notre extrait.

Une occurrence de ponctuation se situe à une frontière dont le statut peut changer selon l'interprétation qu'on donne des relations caractérisant les constituants syntaxiques qui composent un long passage vers la fin de la partie initiale du texte (le *Prothème*). Nous présentons cet exemple avec la ponctuation du manuscrit, mise en parallèle avec celle de l'éditeur, car il illustre bien à quel point les choix de l'éditeur (Mourin 1946b) en matière de ponctuation peuvent être arbitraires et imposer une interprétation qui n'est qu'une possibilité parmi d'autres :

- (7.64) **M**aintenez || vous doncques [> ,] o deuot pueple crestien [> ,] a grant ||
 Reuerence [> ,] a+grant paix [> ,] a grant silence [> ,] a sainte [> ,] ||[fol. 166r] et
 humble [> ,] pure deuocion [> ,] autant ou plus *comme* || vous feriez en la presence
 du benoit corps ihesu+crist [> ,] || *non*+mie pour ma personne [> , -] hors vne telle
 presumpcion [> ! -] || mais pour la maieste de dieu [/ > .] et pour empetrer || grace de
 dire chose sans erreur de la benoite diuinite || et trinite [> ,] pour nous enluminer [/ >]
 chose aussy fruc|tueuse [> ,] pour vous esmouuoir a bien ouurer et dieu || amer [. > ,]
 saluons le saint temple [...] (*gerson13318, 165v:23-166r:8*).

Si on suit l'interprétation de l'éditeur, ce passage contient deux phrases complexes, dont la deuxième commence par *et pour empetrer*. La première « phrase » comprend une apostrophe et une énumération de compléments du prédicat (dont un est doté de trois adjectifs épithètes coordonnés). Elle contient également un syntagme *hors une telle presumption*, que l'éditeur interprète comme une sorte d'incidente comprenant une remarque émotionnelle et qu'il ponctue avec des tirets autour et un point d'exclamation à la fin. Cependant, l'absence de toute ponctuation dans le manuscrit permet également une seconde interprétation : il s'agit d'un syntagme prépositionnel servant de modificateur du groupe nominal *ma personne*. La seule marque de ponctuation à l'intérieur de cette « phrase » dans le manuscrit est un point entre *devotion* et *autant*, qui signale la fin d'une énumération de compléments et le début d'une série de circonstanciés. La séquence qui précède cette ponctuation peut en effet être interprétée comme une phrase complète, tandis que ce qui suit se rattacherait à la deuxième « phrase » de l'éditeur.

Manuscrits du XV^e siècle

La limite supposée de phrase n'est marquée dans le manuscrit que par une très légère barre oblique... très probablement rajoutée par une autre personne que le scribe à l'origine du manuscrit. Vu la répétition de la préposition *pour*, on peut être tenté de considérer tous les constituants qui la contiennent comme des circonstants de but coordonnés, exprimés en partie par des syntagmes prépositionnels nominaux et en partie par des propositions infinitives. L'ensemble de ces circonstants pourrait se rattacher aussi bien à la principale qui précède (*Maintenez vous...*) qu'à celle qui suit (*saluons le saint temple*).

Selon qu'on privilégie telle ou telle interprétation, la frontière « ponctuable » devant *saluons* peut soit « monter » au niveau de propositions autonomes, soit n'être qu'un début de principale après une série de propositions infinitives coordonnées. La présence dans le manuscrit de deux points de ponctuation faible aux frontières de ce groupe de constituants « à rattachement incertain » peut témoigner de la difficulté dans l'interprétation de ce passage qu'éprouvait déjà le scribe.

Notons enfin qu'une légère barre oblique sépare dans le manuscrit *enlumine* et *chose*. L'éditeur ne place aucune ponctuation à cet endroit et interprète *chose* comme objet du verbe *enluminer*. Il est pourtant possible de voir dans ce syntagme *chose aussy fructueuse* une remarque de même type que celle illustrée par le groupe *hors une telle presumption* cité plus haut dans le même passage. Dans ce cas, l'apparition d'une ponctuation dans ce contexte devient moins étonnante.

Les syntagmes coordonnés ou juxtaposés à l'intérieur d'une proposition sont très nombreux dans **gerson13318** : on compte plus de 120 occurrences de frontières syntaxiques de cette catégorie, dont 8 seulement portent une ponctuation faible (soit près de 7%). Trois de ces occurrences concernent une énumération de noms propres de philosophes célèbres de l'Antiquité :

(7.65) ainsy l'ont prouue et tesmoingnie tous les excel||lens et esleuez philosophes |platon | aristote |her||mes tuelles et autres | (*gerson13318, 166r:14-15*).

Comme dans de nombreux autres manuscrits, l'énumération de noms propres crée ici des conditions particulières qui demandent l'emploi de la ponctuation. Notons

toutefois l'absence de ponctuation devant *tulles* (Marc Tulle Cicéron), le dernier des noms cités.

Une autre occurrence survient devant le troisième des quatre éléments d'une énumération exprimant les différents aspects de l'inconstance de l'âme humaine :

(7.66) car tu es mu||able de ton²⁷⁴ estre a estre d'une congnoissance a || l'autre d'une affection en l'autre affection et de || volente a volente (*gerson13318, 169r: 20-22*).

Deux autres occurrences surviennent dans des conditions similaires, dans l'énumération de plusieurs causes possibles introduites avec le connecteur *soit* :

(7.67) les aucuns [*ne sont disposés pour voir Dieu*] pour l'empesche||ment du corps qui oste l'usaige de Raison soit a || cause de aage comme es enfans soit a cause de ma||ladie comme es+demoniaques et hors du sens et || soit par mauuaise complexion (*gerson13318, 166v:22-25*).

Il s'agit ici de barres obliques très légères, à peine visibles sur la page du manuscrit.

Il existe donc des conditions qui semblent plus favorables à l'emploi de la ponctuation entre des syntagmes coordonnés, mais la tendance à l'absence de ponctuation à ce type de frontière est fortement dominante.

Les occurrences de ponctuation aux frontières de groupes particuliers à l'intérieur de propositions sont extrêmement rares dans **gerson13318**. Outre l'occurrence d'une légère barre oblique entre le verbe et l'objet (ou devant une remarque) que nous avons analysée dans l'exemple (7.64), nous avons relevé une seule occurrence d'une majuscule surlignée, sans doute due à une erreur du rubricateur (sachant que *je* est toujours noté avec une majuscule dans ce manuscrit) :

(7.68) n'en demande plus || car autrement Je ne puis declairer ou monstre || bien disoit platon que [...] (*gerson13318, 169v:25*).

On peut supposer que ce *Je* était interprété comme le début d'une citation de Platon, tandis qu'en réalité le discours rapporté suit le verbe de parole.

274 Une erreur scribale, à corriger en *non*.

Bien que le taux général de ponctuation dans **gerson13318** soit relativement peu élevé, ce manuscrit présente un système de marques assez diversifié, utilisé en conformité avec les niveaux de frontières dans la hiérarchie des unités syntaxiques : l'usage de la ponctuation est peu fréquent en dehors des frontières entre propositions autonomes et de certains éléments du discours direct. Un signe de ponctuation en forme de barre oblique avec un point au milieu apparaît avec ce manuscrit pour une première fois dans notre corpus. Ce texte, dont certains passages sont riches en citations latines accompagnées de l'indication de leurs sources présente également la pratique du soulignement des passages cités.

7.6. Lyon, BM, 769 (Fragment d'une description de l'Orient)

Ce manuscrit de la Bibliothèque municipale lyonnaise n'est pas suffisamment étudié à ce jour. Sa datation d'après le catalogue des manuscrits des bibliothèques publiques de France est très approximative et s'étend sur l'ensemble du XV^e siècle. Les 145 folios en papier de ce codex mesurent 280×190 mm. Le manuscrit a certainement été recomposé à partir d'éléments hétérogènes. Selon les parties, l'écriture est répartie sur une ou deux colonnes. Outre le fragment d'une description de l'Orient (dont l'auteur et la date de composition demeurent inconnus) qui occupe les 77 premiers folios, le manuscrit contient une note en vers burlesques datée du 6 octobre 1456 et signée «De Basticia» (fol. 78) et une *Histoire de Ponthus de Galice* en prose (fol. 80-145). Le folio 79 est blanc.

La date figurant sur la note en vers burlesques permet de supposer que le manuscrit a pu être écrit aux alentours du milieu du XV^e siècle, mais le caractère composite de cet ouvrage ne permet pas d'appuyer suffisamment cette hypothèse. Un examen paléographique minutieux (et notamment une étude des filigranes) devrait permettre dans l'avenir d'avancer une date plus précise pour la réalisation de chacune des parties de ce manuscrit.

La description de l'Orient du manuscrit 769 de la BM de Lyon n'a jamais été, à notre connaissance, ni étudiée ni éditée. La classification typologique de ce texte ne

pose pourtant pas de problème, il s'agit du domaine didactico-scientifique et du genre de la description géographique.

L'écriture de la description de l'Orient peut être qualifiée de cursive, elle n'est pas particulièrement soignée, mais est assez propre. L'aspect décoratif de ce texte se limite à l'usage de lettrines de couleur rouge, au remplissage d'un ornement très simple (de couleur rouge également) des fins de lignes sans texte précédant les lettrines et de majuscules parfois décorées en début de page (*cf.* Fig. 59 et l'explication ci-dessous).

Le fragment que nous avons transcrit se situe entre les folios 4v et 5v et comporte 613 occurrences-mots. Il porte l'identifiant **egypte769** dans notre corpus. Avec 17 occurrences de la ponctuation forte et 15 occurrences de la ponctuation faible, le taux de ponctuation général s'élève dans ce texte à 5,2%. Le système de marques de ponctuation utilisé dans ce texte est extrêmement simple : une majuscule seule (surlignée habituellement d'une tache à l'encre ocre) utilisée comme ponctuation forte²⁷⁵ et une légère barre oblique remplissant la fonction de ponctuation faible. Il n'est pas exclu que ces barres obliques (ou au moins certaines d'entre elles) aient été rajoutées après la phase de copie initiale.

L'extrait dépouillé comporte six occurrences de majuscules non surlignées, dont deux < R > hors frontières ponctuables qu'on interprète comme de simples variantes calligraphiques (qui ne sont toutefois pas obligatoires, car on trouve également des < r > minuscules en position initiale). Une occurrence de < J > dans le nom géographique *Jerusalem* peut sans doute également être classée dans cette catégorie. Ceci prouve que le travail du rubricateur qui a surligné les majuscules n'était pas purement mécanique, mais s'appuyait sur une certaine analyse syntaxique. En revanche, les deux occurrences de < E > et l'unique occurrence de < M > en tête de proposition ont sans doute simplement été oubliées par le rubricateur.

Une majuscule, toujours surlignée et souvent décorée, se trouve au début de chaque page de **egypte769** :

275 Avec une exception, où il semble y avoir une barre oblique devant une majuscule non surlignée, *cf.* l'exemple (7.70) ci-dessous.



Fig. 59 : Majuscule décorée en début d'une page (egipte769, 5r).

Ces majuscules ne coïncident pas forcément avec des limites de frontières ponctuables, il s'agit donc d'un usage purement décoratif dans certains cas et d'un usage combiné – décoratif et syntaxique – dans d'autres.

L'extrait étudié présente par ailleurs une occurrence d'une lettrine de couleur rouge (nettement différente de celle utilisée pour le surlignage des majuscules) occupant un espace laissé par deux lignes en retrait.

Egipte769 ne contient aucun passage au discours direct, ce qui n'est pas étonnant pour une description géographique.

Les limites entre des propositions autonomes sont ponctuées dans 25 cas sur 32, avec une légère dominance de la ponctuation forte (14 occurrences contre 11 pour la ponctuation faible).

La ponctuation forte est largement dominante en cas de changement de thème ou d'introduction d'un nouveau sujet nominal (13 occurrences contre 6 pour la ponctuation faible et 2 absences de ponctuation). Si, en revanche, le sujet de la deuxième proposition est pronominal ou nul, c'est l'absence de ponctuation qui l'emporte (10 occurrences contre 1 ponctuation forte et 4 ponctuations faibles).

Les deux occurrences où la ponctuation est absente devant une proposition autonome avec un sujet nominal surviennent dans des contextes assez similaires :

- (7.69) **S**y est ainsi celle terre appellee || feu / *et* vers mydy est etyope / *et* vers septemtrion est le+desert || quj dure jusques a+tirie **et ainsi** est fort le+pays de egipte || a+tous coustes (egipte769, 4v:13) ;
- (7.70) moult fort cite souloit estre amyete / Mays elle || fust troys foys guagnie des chrestiens **et pour tant** abatirent || les sarazins les murs *et* les fourteresses / *et* firent vne || aultre cite plus loing de+la+mer qu'ilz appelloyent la || nouvelle damiete (egipte769, 4v:22-26).

Dans (7.69), la ponctuation est absente devant une proposition qui sert de conclusion à une série de propositions précédentes (séparées, elles, par des ponctuations faibles). Cette conclusion est introduite par un connecteur composé *et ainsi* qui déclenche l'inversion du sujet.

Dans (7.70), la ponctuation est absente devant une autre proposition introduite par un connecteur composé *et pour tant* ('et à cause de cela'), avec également la postposition du sujet.

C'est dans ce même exemple que nous trouvons la seule occurrence de ponctuation forte devant une proposition, dont le sujet pronominal renvoie à un argument de la proposition précédente (*Mais elle fust...*). Ici la majuscule < M > n'est pas surlignée, ce qui peut être une erreur du rubricateur, mais peut être aussi un choix intentionnel. C'est par ailleurs la seule occurrence où la barre oblique est utilisée devant une majuscule²⁷⁶, cette marque semblant de façon générale être réservée à la ponctuation faible dans **egipte769**. Il n'est pas exclu que la barre oblique ait été ajoutée par une autre personne que le scribe, ce qui permet de formuler l'hypothèse que nous avons affaire dans ce cas à un désaccord entre scribe et rubricateur concernant la force de la ponctuation à cette frontière.

Les frontières entre propositions ayant des éléments communs sont peu nombreuses dans l'extrait dépouillé (moins d'une dizaine d'occurrences au total), mais une tendance à l'absence de ponctuation semble se dessiner nettement : 2 occurrences ponctuées contre 7 sans ponctuation.

La première des occurrences ponctuées apparaît dans l'exemple (7.70) présenté ci-dessus (devant *et firent une autre cite*). Comme nous l'avons déjà remarqué, le passage en question accumule plusieurs « anomalies » dans l'usage de la ponctuation.

La deuxième occurrence ponctuée (une ponctuation forte) survient dans le cadre d'un récit de la rencontre entre un ermite et un monstre cornu :

(7.71) Es desers de || egipte jadiz encontra²⁷⁷ vng saint preudomme hermicte quj la habitoyt ||
en vng moustier sicomme vng homme a+deulx grans cornes tranchans || en son front /

276 À une première vue, cette barre oblique peut être perçue comme un élément du tracé de la lettre < M >, mais un examen attentif montre qu'il s'agit d'un trait indépendant.

277 Correction en marge de la forme *avoyt* (barrée).

Manuscrits du XV^e siècle

et avoit corps de+homme jusques a+la boutoniere || **E**t desoubz avoyt tiel cops *comme* vne chieure / (*egipte769, 5r:5-9*).

Il convient de noter dans le même passage l'usage d'une ponctuation faible devant une proposition dont le sujet implicite est l'objet de la proposition précédente (*ung homme a deulx cornes*). La ponctuation dans ce contexte peut participer à la création d'un effet dramatique : une pause est marquée devant chaque nouvel élément de la description du monstre.

Sur les 32 occurrences de limites entre propositions principales et subordonnées la ponctuation faible apparaît dans un seul cas :

(7.72) *et pria* a || l'ermite qu'ilz volisse dieu *prier pour* ly *comme* a+celluy *qui pour* || humain ligniage sauuer / descendit au+corps de+la benoyte vierge || marie de+quielle il nasquit *et* souffrit mort *et* passion (*egipte769, 5r:14*).

Il s'agit de la suite de l'histoire de l'ermite et du monstre dont le début est cité dans l'exemple (7.71). Une barre oblique apparaît à la fin d'une proposition infinitive située au milieu d'une construction très complexe avec de multiples subordonnées successives. C'est la seule frontière dans ce passage qui ne soit pas marquée par un mot-connecteur, et l'usage d'une ponctuation aide à la repérer.

L'extrait analysé ne présente pas beaucoup d'occurrences de frontières entre des syntagmes coordonnés (12 au total, dont 2 comportent une ponctuation faible). Dans les deux cas où la ponctuation apparaît, il s'agit de la coordination de syntagmes nominaux avec des modificateurs, comme dans l'exemple suivant :

(7.73) **E**Gipte a xv journées de+lonc *et* troys journées || de large / *et* plus de xv journées de desers (*egipte769, 4v:16*).

Chacun des syntagmes coordonnés véhicule une information nouvelle et importante, et la ponctuation peut servir ici à la mettre en valeur.

En ce qui concerne la ponctuation intrapropositionnelle non liée à la coordination, elle se limite à deux occurrences de majuscules décoratives de début de page (*cf.* Fig. 59 et son analyse).

De façon générale, le système de la ponctuation dans **egipte769** est simple et économique : l'usage de deux marques (barre oblique et majuscule seule) qui se limite, à quelques exceptions près, à l'indication de frontières entre des propositions autonomes. C'est un bel exemple de « ponctuation minimale », système qui s'oppose dans la tradition médiévale à un type de ponctuation plus sophistiqué, employant un nombre plus important de marques différentes et utilisées dans des conditions variées. Rappelons qu'il n'y a pas à notre connaissance de manuscrits médiévaux français qui soient privées de toute forme de ponctuation.

7.7. Paris, BnF, fr. 11610 (*Roman du Comte d'Artois*)

Le manuscrit français 11610 de la BnF est entièrement occupé par le *Livre du treschevalereux conte d'Artois*. C'est un codex composé de 119 feuillets (dont 3 blancs à la fin) en papier de 295×205 mm. Le texte est disposé sur une colonne avec 27 lignes par page pleine et est décoré de rubriques et de miniatures à l'aquarelle. L'écriture est assez soignée. Les miniatures de ce manuscrit, vraisemblablement de la main du maître de Jean de Wavrin, sont d'ailleurs assez célèbres et citées notamment dans (BnF 1955).

D'après les recherches de J.-Ch. Seigneuret (1966, p. XX-XXI), ce manuscrit aurait pu être réalisé entre 1415 et 1467. Étant donné que la date de composition du *Roman du comte d'Artois* se situe au milieu du XV^e siècle, et vraisemblablement, selon le même chercheur, entre 1453 et 1467, on peut faire l'hypothèse que le manuscrit n'a pas pu être exécuté avant 1453, ce qui ramène à 1460 la date « moyenne » utilisée dans notre classement chronologique des manuscrits.

Le *Roman du comte d'Artois* est une œuvre anonyme appartenant, bien entendu, au domaine littéraire. Le titre traditionnel de ce texte contient l'indication du genre, mais sa structure, basée sur celle du conte populaire, le rapproche plutôt du genre de la nouvelle (Bossuat 1951).

L'extrait que nous avons transcrit et intégré dans notre corpus s'étend du folio 1r à 4r et contient 1119 occurrences-mots. Il porte l'identifiant **artois11610** que nous

Manuscrits du XV^e siècle

utiliserons désormais pour les références²⁷⁸. Le taux de ponctuation général y atteint les 4,5% avec 36 occurrences de la ponctuation forte (3,2%) et 14 occurrences de la ponctuation faible (1,3%).

Les marques de ponctuation utilisées dans l'extrait dépouillé sont assez variées : nous avons relevé 27 occurrences de la barre oblique²⁷⁹ (13 dans la ponctuation forte et 14 dans la ponctuation faible), 10 occurrences de la majuscule seule (à une frontière ponctuable), 9 occurrences du point (toujours dans la ponctuation forte) et 4 occurrences d'une marque en forme de lettre *v* ou d'une petite barre oblique avec un crochet, suivies de deux points disposés plus ou moins horizontalement vers le haut d'une ligne. Cette marque apparaît toujours en fin de paragraphe :

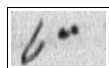


Fig. 60 : Marque de « fin de section » (artois 11610, 4r).

Aucune marque ayant cette forme ne figure dans le répertoire de Parkes (1992), ni dans les autres travaux consacrés à l'histoire de la ponctuation que nous connaissons. Compte tenu de son emploi réservé aux fins de paragraphes (ponctuation « extra-forte »), cette marque peut provenir de la *positura* « en forme de 7 » (Parkes 1992, p. 43). Selon Parkes, cette marque pouvait prendre une forme ressemblant au chiffre « 9 » dans les manuscrits du XV^e siècle. Une autre hypothèse est que cette marque est une forme réduite de la décoration de fin de ligne souvent utilisée au Moyen Âge pour remplir le vide. Quelle que soit son origine, le terme de « marque de fin de section » nous semble convenir pour désigner ce signe de ponctuation.

L'extrait étudié contient plusieurs unités structurelles de haut niveau : il commence par un prologue précédé d'une miniature et d'un titre écrit en rouge occupant un paragraphe entier de trois lignes. Une grande initiale de quatre lignes réalisée en couleurs bleue, rouge et dorée et décorée par l'image d'un blason bleu qui inclut lui-même, en abîme, l'image d'un petit blason de couleur plus foncée, ouvre le texte du prologue.

²⁷⁸ Le *Roman du comte d'Artois* est référencé sous le sigle **ComteArt** dans la bibliographie du DEAF.

²⁷⁹ Dans la majorité de ces occurrences, le bas de la barre oblique forme un petit crochet grâce à un léger retour de plume. Nous considérons que cette particularité graphique est due uniquement au style d'écriture et n'en tenons pas compte dans la transcription.

La fin du prologue est suivie, à quelques lignes d'intervalle, du titre du premier chapitre. Le texte du chapitre, qui commence sur une nouvelle page, est précédé d'une nouvelle miniature. Le chapitre est divisé en plusieurs paragraphes, dont chacun commence par une grande initiale de 3 lignes et finit par une marque de fin de section. La division graphique en paragraphes correspond parfaitement à la structure logique du texte, ce qui fait que les limites des paragraphes ont été respectées dans l'édition moderne (Seigneuret 1966).

Outre les divisions structurelles marquées dans le manuscrit par une ponctuation « extra-forte », nous avons repéré deux frontières qui correspondent à des changements d'épisodes, selon notre classement. Cette interprétation s'appuie sur la présence d'expressions permettant de marquer une rupture dans la continuité de la narration : *Et dist l'istoire que...* (*artois11610, 3r:4*) et *Or est ainsy que...* (*artois11610, 3r:10*). Ces frontières sont marquées par de simples ponctuations fortes, ce qui s'explique sans doute par le fait que les épisodes en question sont courts et se succèdent immédiatement.

L'extrait étudié contient la représentation d'un dialogue où sont combinés le discours direct et le discours indirect :

(7.74) Le conte de boulongne || demanda *au* conte d'artois se jl auoit jntencion || de tournoyer
 A+quoy jl Respondy / sans dire || vvy / ou nanil / par foy sire jl souffira assez || se je
 voy ce qui se fera / et ce disoit le conte [...] (*artois11610, 4r:8-12*).

Une seule occurrence est bien-sûr insuffisante pour arriver à quelque conclusion que ce soit, mais l'usage qu'elle présente mérite néanmoins d'être observé.

Le passage commence par une proposition principale servant à introduire les participants du dialogue, suivie d'une complétive exprimant une question indirecte. Aucune ponctuation n'est utilisée à l'intérieur de cet ensemble. La réponse est introduite par une nouvelle proposition principale (contenant toutefois le pronom relatif *quoi* utilisé comme anaphorique) qui commence par une ponctuation forte. Le circonstant de cette principale est exprimé par une proposition infinitive *sans dire* où la fonction d'objet est remplie par deux « mots-phrases » *wi* ('oui') et *nannil* ('non') coordonnés. La proposition infinitive et le deuxième objet sont précédés d'une ponctuation faible.

On peut supposer que la rencontre de deux verbes de paroles et de deux discours – la réplique au discours direct et le discours hypothétique représenté par les « mots-phrases » – a entraîné cette densité de ponctuation exceptionnelle pour **artois11610**. Le discours direct commence et finit par une ponctuation faible. Il est composé d'une apostrophe et de trois propositions courtes, dont deux subordonnées « en chaîne ». Aucune ponctuation n'apparaît à l'intérieur de ce discours direct. Nous voyons donc une différence dans la ponctuation du discours direct et du discours indirect et nous constatons que les mots 'oui' et 'non' sont peut-être traités d'une façon particulière.

Les limites entre propositions autonomes sont ponctuées dans la grande majorité des occurrences (21 sur 22, dont 20 ponctuations fortes sans compter les frontières de divisions logiques importantes). La ponctuation forte apparaît même lorsqu'une proposition autonome vient s'insérer entre une subordonnée et une principale comme une sorte d'incidente :

(7.75) le noble conte d'artois || aprez ce qu'il eubt par longue espanse || de tempz poursieuy
lez armez et que || plus ne trouuoit a soy employer ¶ Car aussy || ne puet paz tousiours
estre guerre ¶ Il se mist || au Retour vers sa conte d'artois (*artois11610*, 2v:22-27).

Dans cet exemple, le sujet de la proposition principale²⁸⁰ est suivi de deux circonstancielle subordonnées et d'un commentaire sur les informations contenues par ces subordonnées exprimé par une proposition formellement autonome par rapport à son entourage. La principale recommence après ce commentaire avec la reprise du sujet par un pronom anaphorique. Les deux ponctuations fortes utilisées au début et à la fin du commentaire sont les seules marques de ponctuation qu'on trouve dans cette construction complexe. On peut ainsi supposer que les relations syntaxiques immédiates entre deux propositions voisines jouent un rôle plus important pour le choix de la ponctuation que la structure générale d'une construction complexe.

Dans un autre cas, le choix de la ponctuation semble toutefois être motivé par le souhait de représenter une hiérarchie de proximité logique entre des propositions autonomes successives. La seule occurrence de la ponctuation faible devant une

280 La principale de cette construction est en effet à son tour une subordonnée complétive par rapport à la formule initiale d'épisode *L'ystoire dit que...*

proposition autonome survient en effet à l'endroit où plusieurs épisodes (selon nos critères, *cf.* plus haut) se succèdent à une faible distance :

(7.76) Et dist l'histoire qu'il se tenoit volentiers en || la ville d'arraz ou jl se donnoit du bon tempz || avecquez sez *hommez* et amis qui le venoient || visiter *honnourablement* pour la valeur de sa || *personne* et jl ne failloit a eulx de Riens || Receullir *chascun* selon son degre et festoyer sy || a point que tous en estoient contempz || Or est || ainsy que [...] (*arfois11610, 3r:8*).

Il semblerait que, puisque les frontières d'épisodes ne sont marquées ici que d'une simple ponctuation forte, l'usage de la ponctuation faible devant une proposition autonome à l'intérieur d'un épisode permet de signaler qu'il s'agit d'une frontière logique moins importante.

La seule occurrence d'absence de ponctuation à une frontière entre propositions autonomes survient également dans un contexte assez particulier (qui suit immédiatement l'exemple précédent). Nous présentons ce passage avec la ponctuation du manuscrit et de l'édition (Seigneuret 1966) pour mettre en évidence la différence de structuration du texte dans ces deux sources :

(7.77) Or est || ainsy que [> ,] dura ce tempz [> ,] auoit vng conte || a boulongne [> ,] *homme* de grant lieu et bien || *Renomme* [/ > ,] Lequel [> ,] tant pour veoir le vasselage || tant dez noblez de son pays *comme* pour baillier || congnoissance d'une sienne fille [> ,] qui son heri||tiere estoit [> ,] tant belle et tant parfaite en || tous biens que dieu et nature n'y auoient || Riens laissie a faire [/ > ,] Auoit fait publyer vng || tournoy par tout lez Royalmez de france [> ,] d'en||gleterre [> ,] d'escoce et d'allemaigne et jour assigne || pour venir a boulongne quiconque voudroit || faire parler de luy et acquerir *honneur* [> ;] || desquellez nouuellez oyr eult parfaite || plaisance [/ >] Le noble conte d'arfois [> ,] qui trop || enuis se fust tenu d'y+aller [. > ;] (*arfois11610, 3r:10-25*).

Nous observons ici une longue construction contenant une dizaine de propositions liées entre elles par des relations de subordination, puis un autre groupe formé d'une principale et d'une subordonnée, où la principale commence par un déterminant relatif *desquelles* (qui indique une liaison forte entre les deux groupes). L'éditeur a utilisé 9 virgules à l'intérieur du premier groupe et a séparé les deux groupes par un point-virgule.

Dans le manuscrit, deux ponctuations fortes apparaissent à l'intérieur du premier groupe : devant une relative introduite par le pronom *lequel* et devant la reprise d'une

principale après une série de quatre subordonnées. Il convient de noter que le pronom *lequel* est le sujet de cette proposition séparé par une série de subordonnées de son prédicat qui prend du coup une majuscule (*vng conte a boulongne [...] Lequel [...] Avoit fait publier [...]*). Du point de vue informationnel c'est cette proposition qui joue le rôle le plus important dans l'ensemble de la construction, même si syntaxiquement elle est subordonnée. La ponctuation forte permet d'attirer l'attention sur le centre prédicatif de cette proposition.

La frontière entre les deux groupes (propositions autonomes) ne comporte pas de marque de ponctuation dans le manuscrit, et c'est la seule occurrence d'absence de ponctuation à ce type de frontière dans **artois11610**. Toutefois, cette frontière coïncide avec un saut de ligne, qui peut, comme nous le savons, se substituer à une marque de ponctuation dans certains cas. On peut également supposer que le scribe n'avait pas repéré l'importance de cette frontière, qui peut ressembler à celle d'une subordonnée du fait de la présence du pronom relatif. En revanche, une ponctuation forte apparaît devant le syntagme nominal sujet (*Le noble conte d'artois*) qui est placé dans cette proposition après le prédicat. Là aussi deux hypothèses sont possibles : une erreur du scribe qui a mal interprété les frontières syntaxiques et a pris le syntagme nominal pour le début d'une nouvelle proposition, ou bien un choix intentionnel visant à mettre en relief le sujet et à exprimer le respect pour le personnage. Si la deuxième hypothèse se confirme, on peut parler pour ce passage d'une ponctuation « sémantique » servant à marquer les éléments-clefs de la structure informationnelle, qui se substitue à la ponctuation « syntaxique ». Bien entendu, il faudrait trouver plusieurs autres attestations de ce phénomène pour arriver à une conclusion définitive.

Les frontières entre propositions ayant des éléments communs sont ponctuées dans 5 cas sur 14, et il s'agit toujours d'une ponctuation faible. On voit donc une nette différence par rapport aux propositions autonomes.

Nous n'avons pas réussi à dégager un facteur unique qui puisse expliquer les cinq occurrences ponctuées dans cette catégorie. Il est toutefois possible, à chaque fois, de trouver une explication *ad hoc*. Dans le contexte suivant, par exemple, une petite

barre oblique apparaît devant la deuxième et la quatrième proposition dans une série de quatre propositions avec un sujet commun :

- (7.78) Il adouchoit lez || furieux / humilioit lez orgueilleux appaisoit || lez discordez / et brief faisoit tant que par sez || meritez jl estoit ame et loe dez hommez et || ame de dieu (*artois11610, 2v:17-18*).

En l'absence d'un connecteur, la ponctuation est particulièrement utile à la compréhension du texte à la première frontière. A la deuxième frontière, devant une troisième proposition construite sur un même modèle (un verbe à l'imparfait suivi d'un déterminant défini au pluriel et d'un adjectif substantivé en guise d'objet), on arrive très bien à repérer la frontière sans l'aide de la ponctuation. La quatrième proposition sert de conclusion à la série des trois premières, et la ponctuation permet de souligner cette opposition logique.

En ce qui concerne les limites entre les principales et les subordonnées, la tendance à l'absence de ponctuation est très nette : 86 occurrences contre 6 occurrences de la ponctuation forte et 5 occurrences de la ponctuation faible. Deux des occurrences de la ponctuation forte ont déjà été présentées dans l'exemple (7.77) : la ponctuation y sert sans doute à mettre en relief le sujet et le prédicat de la proposition importante du point de vue informationnel. Dans un autre cas, il s'agit sans doute d'une « ponctuation de respect » (devant *Le conte de boulogne, artois11610, 4r:8*). Un autre cas (une ponctuation faible devant une infinitive dans un dialogue rapporté) a été analysé dans (7.74). Dans deux cas, enfin, une ponctuation apparaît devant des connecteurs exprimant la concession (*toutesvoies* et *Non obstant que...*). Il se peut que ces connecteurs attirent la ponctuation dans **artois11610**, hypothèse qu'il faudrait vérifier sur l'ensemble du manuscrit. Quoi qu'il en soit, la ponctuation reste un phénomène marginal à ce type de frontière syntaxique.

La fréquence relative de la ponctuation est encore plus basse aux frontières de syntagmes coordonnées à l'intérieur d'une proposition. Seules deux occurrences sur 54 sont dotées d'une ponctuation faible. L'une de ces occurrences concerne la coordination de « mots phrases » *wi* et *nannil* (7.74). L'autre survient devant un modificateur qui s'applique à un groupe de deux syntagmes nominaux :

Manuscrits du XV^e siècle

(7.79) vng liuret qui fait menciomn || dez haultez entrepriez amours et beaulx || fais d'armez /
d'un Conte d'artois (*artois11610, 1v:10*).

Très probablement, il s'agit dans ce cas encore d'un exemple de la « ponctuation de respect ».

La ponctuation aux frontières de syntagmes particuliers est également très rare dans **artois11610** : nous n'avons relevé que quatre occurrences ponctuées dans cette catégorie. Dans trois de ces cas, la ponctuation apparaît devant le syntagme *le (noble) conte*. Il s'agit donc de la « ponctuation de respect », incontestablement représentée dans ce manuscrit.

La dernière occurrence est une majuscule seule à la frontière entre un verbe et son objet direct :

(7.80) Il se mist || au Retour vers sa conte d'artois ou dieux scet |[fol. 3r] La grant joye que
lez siens firent de sa+venue (*artois11610, 3r:1*).

Il s'agit d'une ponctuation décorative liée non pas à la structure syntaxique, mais à la position en début de page : nous avons repéré des occurrences similaires sur d'autres pages de ce manuscrit (en dehors de l'extrait dépouillé), même si cette mise en majuscule de la première lettre d'une page n'est pas systématique.

De façon générale, malgré son taux relativement peu élevé, la ponctuation dans **artois11610** est assez logique et conséquente. Une concurrence entre deux principes – syntaxique et sémantico-pragmatique – apparaît clairement dans ce texte.

7.8. Paris, BnF, fr. 2701 (*Proposition... par devant... le conte d'Eu de Jean Juvenal des Ursins*)

Le manuscrit français 2701 de la Bibliothèque nationale de France est entièrement occupé par différents écrits de Jean Juvenal des Ursins (épîtres, discours, sermons, etc.). Le codex est composé de 140 feuillets de papier de 38,5×28,5 cm. organisés en 12 cahiers reliés sous une couverture de vélin. Le manuscrit n'est pas daté,

mais un certain nombre d'indices permettent de situer la date de sa réalisation dans le troisième quart du XV^e siècle (Lewis & Hayez 1978, p. 17).

P. S. Lewis et A.-M. Hayez ont pu identifier quatre mains différentes dans l'écriture de cet ouvrage, dont la deuxième couvre les folios 118r à 125r. Le texte est disposé sur deux colonnes dans l'ensemble du codex, le nombre de lignes est variable, mais toujours supérieur à 50 lignes (68 à 69 lignes dans l'extrait transcrit). Au début du manuscrit (jusqu'au folio 118r seulement), les majuscules sont marquées de légers traits de plume jaunes et les citations sont soulignées en rouge. L'écriture de l'ensemble des mains peut être qualifiée de cursive et irrégulière. L'aspect décoratif du manuscrit se limite à l'usage de grandes initiales rouges au début de certains textes, dont celui qui nous intéresse.

La Proposicion faicte par Jehan Juvenal des Ursins, evesque et conte de Beauvais par devant hault et puissant prince le conte d'Eu occupe les folios 121r à 121v²⁸¹. Cette petite œuvre de 1578 mots a pu être transcrite et intégrée dans notre corpus dans sa totalité. Elle appartient, bien entendu, au domaine politique et son genre ne peut être mieux défini que par son nom : proposition. Il s'agit en effet d'une demande de protection contre les tribulations que les soi-disants « hommes du roi » font subir à la population du comté de Beauvais. La date de composition de la *Proposition* remonte à 1438. Ce texte porte l'identifiant **ursins2701** dans notre corpus.

Le taux de ponctuation général dans **ursins2701** atteint à peine les 3,9%, la ponctuation faible étant dominante avec 38 occurrences contre 13 occurrences de la ponctuation forte et 8 occurrences de la ponctuation moyenne (avec des minuscules élargies).

Le seul symbole de ponctuation utilisé dans l'extrait étudié est la barre oblique, dont nous avons relevé 52 occurrences. Ce fragment présente également 6 occurrences de la majuscule seule, 3 occurrences de la minuscule élargie seule et 3 occurrences d'un espace blanc légèrement plus long que les autres, qui peut être assimilé à une marque de ponctuation.

Dans l'exemple suivant, un grand blanc est placé après une apostrophe utilisée par l'auteur à plusieurs reprises pour évoquer le destinataire du texte :

281 Et non 120r-120v, contrairement à (Lewis & Hayez 1978, p. 19).

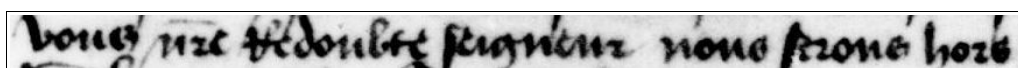


Fig. 61 : Blanc allongé (ursins2701, 121b).

- (7.81) je disoie que [par]²⁸² || vous *nostre* Redoubte seigneur | nous serons hors
(ursins2701, 121b:64).

Il convient de noter que les trois occurrences du grand blanc sont liées aux frontières « énonciatives » (apostrophe, début ou fin du discours direct) et à chaque fois le mot *seigneur* se trouve dans un contexte très proche. Il est également intéressant de noter que deux autres occurrences du syntagme *nostre redoubté seigneur* en apostrophe sont suivies d'une barre oblique, ce qui est plutôt étonnant, car la tendance générale constatée dans la majorité des manuscrits de notre corpus est à l'absence de ponctuation aux frontières des apostrophes. On peut sans doute parler dans ce texte politique de « ponctuation de respect » envers la personnalité à qui s'adresse le message.

La « grande » longueur des espaces blancs « ponctuants » est toute relative, et nous ne pouvons pas affirmer avec certitude qu'il s'agisse effectivement de gestes intentionnels du scribe qui les a laissés à certains endroits, mais le phénomène mériterait d'être étudié attentivement sur un corpus plus important dans l'avenir. Quoi qu'il en soit, la faible fréquence de ce phénomène dans notre corpus fait penser qu'il n'occupe qu'une place très périphérique dans le système de la ponctuation.

Notre extrait présente également deux occurrences d'un très grand espace blanc (d'une longueur de 5 à 7 caractères environ) aux endroits où un mot ou un syntagme manque dans une proposition :

- (7.82) et pour ce je [] la+premiere partie || de mon theume / (ursins2701, 121b:30) ;
(7.83) et est ad plus [] quant jl en a+prins || la charge / (ursins2701, 121c:12).

Très vraisemblablement, le scribe a laissé des blancs à ces endroits pour insérer par la suite un ou plusieurs mots qu'il n'avait pas réussi à comprendre lors de son travail initial²⁸³.

282 Correction proposée dans (Lewis & Hayez 1978).

283 P. S. Lewis et A.-M. hayez (1978) proposent d'ajouter *prends* dans (7.82) et *grand prix* dans (7.83). Il convient de préciser que ces corrections ne s'appuient pas sur les leçons des autres manuscrits.

Tout comme dans **gerson13318**, nous avons été amené à recourir à la notion de « minuscule élargie », non pas que nous affirmions l'existence d'une opposition de trois « forces » de ponctuation dans le système graphique de **ursins2701**, mais parce que dans un certain nombre de cas, il est difficile de déterminer si une lettre donnée est une majuscule ou une minuscule : elle semble légèrement plus grande que son entourage, mais son tracé est le même que celui d'une minuscule ordinaire. Prenons, par exemple, trois occurrences de < a > dans le déterminant composé *au* situées sur une même page du manuscrit :

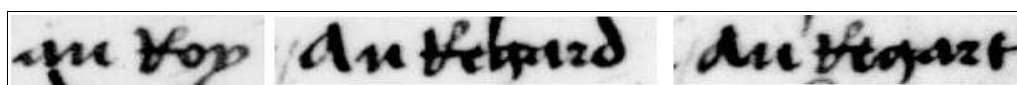


Fig. 62 : A minuscule et majuscule (*ursins2701*, 121r).

Dans le premier cas, il s'agit très certainement d'une minuscule. Dans le deuxième cas, nous avons probablement affaire à une majuscule, tandis que dans le troisième cas un doute subsiste.

Dans 5 cas sur 9, les grandes minuscules sont précédées d'une barre oblique, il est donc certain que la frontière syntaxique à laquelle elles se trouvent est ponctuée. Ce sont les quatre occurrences de la minuscule élargie seule qui laissent planer le doute sur la présence d'une ponctuation²⁸⁴. Le nombre de ces cas d'incertitude est trop réduit pour avoir une incidence sérieuse sur le taux global de la ponctuation. Toutefois, pour éviter toute confusion, nous excluons ces occurrences de nos calculs ultérieurs.

La particularité de **ursins2701** dans notre corpus est qu'il s'agit d'un texte entier. Nous pouvons donc observer dans ce texte le fonctionnement des marques de ponctuation qui servent à organiser la structuration globale du texte.

Dans le manuscrit BnF, français 2701 les différentes œuvres de Jean Juvénal des Ursins se succèdent en continu : une nouvelle œuvre commence sur la même colonne où finissait l'œuvre précédente s'il y reste de la place. C'est notamment le cas de la *Proposition... par devant le conte d'Eu* qui commence par un long titre après la 12^e ligne du folio 192r (colonne a). Ce titre est précédé d'un interligne équivalent à une ligne

²⁸⁴ Deux de ces occurrences se situent à des frontières de propositions autonomes, et la troisième ouvre une « référence bibliographique » (*Ad Romanos .xii^o*).

blanche, et son premier mot est écrit en caractères nettement plus grands que le reste du texte. Le titre se termine par une barre oblique et un saut de ligne :

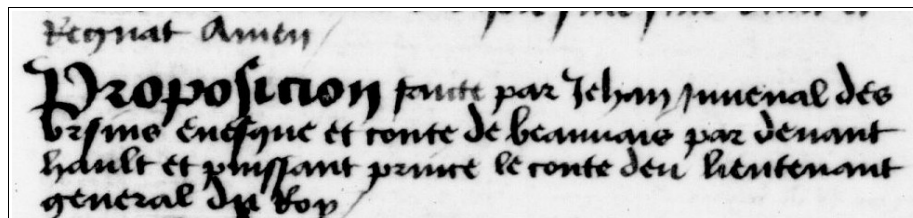


Fig. 63 : Mise en forme du titre (ursins2701, 121a).

Le corps du texte commence par une initiale rouge de deux lignes de hauteur. Le reste du premier mot est écrit en grands caractères, comme le premier mot du titre.

Le texte est très bien structuré du point de vue logique : on peut clairement distinguer une introduction qui contient, entre autres, le résumé du message qui suit, divisé en trois parties (ou *points*) – structuration parfaitement respectée par la suite. Trois de ces quatre divisions logiques sont marquées par des ponctuations fortes. La dernière partie commence par une grande minuscule précédée d'une barre oblique (cf. Fig. 62 à droite) qui peut sans doute être assimilée dans ces conditions à une ponctuation forte.

Le texte finit par le mot *Amen* qui sert de « ponctuant lexical », et est un indicateur de la fin d'une unité textuelle (de même que *Explicit* dans d'autres textes). Ce mot commence par une majuscule et est précédé d'un espace blanc prolongé.

Nous voyons donc que des marques graphiques spécifiques sont utilisées pour la mise en forme du texte dans sa globalité. En revanche, la ponctuation forte « ordinaire » est utilisée pour les grandes divisions logiques internes.

Ursins2701 contient un nombre important de citations en latin accompagnées de références et de la traduction en français de l'époque. Les frontières de ces citations sont assez souvent ponctuées (18 occurrences, dont 5 fortes), mais l'absence de la ponctuation n'est pas exceptionnelle (13 occurrences). L'exemple suivant est assez typique en ce qui concerne le traitement de ces citations :

- (7.84) et si auons *in c. administratores eadem q. iterum* / *administratores* || *secularium dignitatum* qui ad *ecclesiarum tuicionem pupillorum* || et *viduorum Rapacium* que Refrenacionem constituti procul||dubio esse debent quociens ab *episcopis* et *viris*

clesiasticis || conuenti fuerunt querimonas audiant et *secundum quod* || necessitas absque negligencia diligenti studio corrigant || **C**est a+dire en francois que les administrateurs de || seculieres dignites sont ordonnes a+la tuicion et+deffences || des eglises et personnes ecclesiastique des pupilles et veufuez || et a+Refrener les mauuaises volentes et les exceps || de ceulx qui les leur vouldroient faire / et quant jlz sont || Requis sont tenus a toute diligence de y+Remedier et de || corriger et punir ceulx qui le feroient / (*ursins2701, 121c:28-40*).

La citation est introduite par la formule *si avons* suivie de la référence de la source : un livre d'Isidore de Séville avec l'indication du chapitre (*c.* = *capitulo*) et du paragraphe (*q.* = *quaestione*). L'usage d'abréviations avec des points crée déjà un repère graphique qui permet de mettre en évidence la citation (même si le point n'a pas ici la fonction d'une marque de ponctuation). La citation proprement dite est précédée d'une petite barre oblique et comprend plusieurs propositions étalées sur 6 lignes. Aucune marque de ponctuation n'est utilisée à l'intérieur de la citation. La citation latine est suivie d'une traduction en français introduite par la formule *C'est a dire en françois* (avec une majuscule). Une petite barre oblique apparait à l'intérieur de la traduction à la frontière de propositions autonomes et une autre barre oblique (suivie d'une minuscule) sépare la traduction de la suite du texte.

Nous voyons donc que les limites des citations et des traductions sont marquées à la fois par des moyens lexicaux (formules introductives) et graphiques (abréviations spécifiques, ponctuation).

Les limites entre les propositions autonomes sont ponctuées dans 14 cas sur 52 (sans compter les limites des grandes unités logiques). C'est l'un des taux les plus bas de notre corpus. Seules quatre occurrences comportent une ponctuation forte.

Une de ces occurrences est placée devant une proposition servant de formule d'introduction d'une citation, une position particulièrement favorable à la ponctuation dans **ursins2701**. Les trois autres occurrences surviennent toutes dans des contextes où l'auteur évoque d'une façon ou d'une autre le destinataire de sa proposition, comme dans l'exemple suivant :

(7.85) car en vous || auons esperence que ainsy le feres / Et par ce aures || louange et honneur en ce monde et a+la fin la+gloire || de paradis / (*ursins2701, 121c:45*).

Il n'est donc pas impossible que le « facteur du respect » ait joué un rôle dans le choix d'utiliser une ponctuation forte.

Parmi les occurrences de la ponctuation faible, on trouve l'évocation des trois parties logiques de la *Proposition* dans le résumé fourni dans l'introduction. Les autres occurrences ne présentent pas, à notre avis, de particularités qui les distinguent de la majorité non ponctuée.

Les frontières entre les propositions ayant des éléments communs sont ponctuées (toujours faiblement) dans 3 cas sur 29 seulement. Il s'agit dans les trois cas de propositions avec un sujet commun, dont la première contient soit juste une relative (7.86), soit une construction plus complexe avec une participiale suivie d'une relative, le tout faisant partie d'une subordonnée circonstancielle (7.87) :

- (7.86) car vous estes prochain parent du Roy et de plus*ieurs* || de son sanc et linage feu monsieur vostre pere estoit tenu et || Repute l'ung des plus vaillans seigneur et prince qui || fut en ce Royaulme **¶** et fut en sarasinisme contre les || mescreans et y mourout ou ala de vie a+trespassement / (*ursins2701, 121a:67*) ;
- (7.87) et combien || que les ennemis voyans les duretes et tirannies que || nous faisoient les+dessus*dis* par diuers moyens ont fait || plusjeurs entreprinses sur ceste cite **¶** et par parolles || deceptiues les ont cuide Reduire a leur obeissance || toutesuoie comme bon vrais et loyaulx subgetz se || sont fermement tenus en la vraye obeissance du || Roy / (*ursins2701, 121b:49*).

Les frontières entre principales et subordonnées sont assez nombreuses dans **ursins2701** et très rarement ponctuées : 3 occurrences seulement sur 84 sont ponctuées²⁸⁵. Il s'agit dans un cas d'une relative introduite par le pronom relatif *lequel*, qui est séparée de son antécédent par une autre relative, elle-même contenant une relative de deuxième niveau. Ce n'est donc pas non plus, à proprement parler, une frontière entre une principale et subordonnée, mais entre deux subordonnées non coordonnées. Ce type de frontière est peut-être proche de celui des propositions autonomes. Notons également que l'occurrence se situe dans le « résumé » du deuxième point abordé dans la *Proposition*, et il se peut que la ponctuation serve à séparer le thème de l'information plus détaillée :

285 Sans compter une occurrence « douteuse » d'une grande minuscule seule (*ursins2701, 121b:3*).

(7.88) le deux*ieme* point || sera pour expeller les tribulacions et afflictions que || nous auons et mesment par ceulx qui nous deussent || garder / lesquelles semblent estre perpetuelles toutesuoie || nous auons esperance qu'elles ne le seront mie par *vostre* || moyen (*ursins2701, 121a:57*).

Une autre occurrence précède une relative qui n'a pas d'antécédent direct, mais qui sert de sorte de conclusion à une longue série de propositions décrivant les exploits du comte d'Eu :

(7.89) je vous congnois || de jeunesse *vostre* hault courage que aues tousiours || et pour+seruir le Roy [...] pour vous emploier ou *seruice* du Roy / qui sont toutes || choses et signes de vaillant et noble corrage de prince (*ursins2701, 121b:1-11*).

Ici non plus on ne peut pas parler d'une simple subordination entre les deux propositions qui « touchent » à la frontière ponctuable.

La troisième occurrence apparait à nouveau dans le résumé des parties du texte, à une frontière qui peut être interprétée comme celle d'une proposition infinitive (même si le prédicat de la principale n'est qu'un verbe auxiliaire *estre*) :

(7.90) le tiers point || si est / vous Requerir de en faire justice et Raison (*ursins2701, 121a:60*).

Ici, selon toute vraisemblance, la motivation de l'emploi d'une marque de ponctuation n'est pas syntaxique, mais pragmatique : le syntagme qui la suit a une importance toute particulière dans ce texte. Toute la *Proposition* de Jean Juvénal des Ursins est en effet une demande de protection auprès du comte d'Eu, et c'est le verbe performatif *requerir* qui sert à exprimer cette action. L'expression *vous requerir* est répétée trois fois au fil du texte – au début, dans le « résumé » et à la fin – et à chaque fois elle est associée à une ponctuation. Dans deux cas, une ponctuation est placée devant le pronom personnel (7.90) et (7.92) et dans un cas elle sépare le verbe du pronom :

(7.91) et disons tous ensemble que vous soiez le || tres bien venu et en sonmes tous Resiouys en aiant || esperance en dieu et en vous / Requerens que jl vous || plaise les tenir pour excuses (*ursins2701, 121a:25*) ;

(7.92) et pour ce moy || qui suis personne ecclesiastique et les+gens+d'eglise nobles || et peuple enffens pupilles et veufuez / vous Requerons || que jl vous plaise nous aider et conforter et Remedier || aux tribulacions (*ursins2701, 121c:42*).

Les frontières syntaxiques auxquelles la ponctuation apparaît dans les derniers exemples sont *a priori* très peu favorables à son emploi : il s'agit d'une frontière entre le groupe du sujet et le groupe du prédicat (7.92) et entre l'objet pronominal et le verbe (7.91). L'explication pragmatique nous paraît donc la plus probable (voire la seule possible) dans ces cas. Il convient de rappeler également l'importance dans **ursins2701** du « facteur respect » envers le destinataire du texte, que nous avons déjà évoqué à propos d'autres exemples.

Une seule occurrence de ponctuation a été enregistrée parmi les 80 frontières de syntagmes coordonnés à l'intérieur de propositions. Il s'agit d'une coordination de sujets dans la traduction d'une citation latine :

- (7.93) C'est a+dire || que vng chascun doit penser a+augmenter son pais / et || mesmement ceulx aux quieulx la chose publicque || est baillee en gouuernement (*ursins2701, 121c:9*).

Il est intéressant de noter qu'une marque ponctuation identique est employée à l'endroit correspondant de la citation originale devant la conjonction *sed* ('mais') :

- (7.94) decet cunctos patrie sue augmenta cogitare / sed eos || maxime quo res publica summis sibi honoribus || obligauit (*ursins2701, 121c:5*).

On peut supposer que l'emploi du connecteur exprimant une opposition a provoqué l'emploi de la barre oblique dans le texte latin et que cette barre a ensuite été reprise dans la traduction.

La coordination mise à part, **ursins2701** présente 5 occurrences de ponctuation à des frontières intrapositionnelles, dont deux occurrences « pragmatiques » liées à l'emploi du verbe *requerir* que nous venons d'analyser dans les exemples (7.91) et (7.92). Dans un cas, il s'agit très probablement d'une nouvelle occurrence de la « ponctuation du respect » devant une construction comparative non propositionnelle qui évoque le destinataire du texte :

- (7.95) jl en a+prins || la charge / comme vous nostre Redoubte seigneur (*ursins2701, 121c:13*).

Un autre cas est similaire à l'exemple (7.90) : la ponctuation marque la fin du syntagme qui introduit une grande division logique :

(7.96) Au Regard || du premier point / de la noblesse elle est toute notaire (*ursins2701*, 121a:63).

La dernière occurrence (une barre oblique suivie, sur une nouvelle ligne, d'une grande minuscule) figure à l'intérieur d'un groupe nominal, entre un nom et son modificateur introduit par la locution prépositionnelle *au regard* :

(7.97) *nostre* loyaulte / || au Regart des anglois (*ursins2701*, 121b:34).

Or les autres occurrences de cette locution prépositionnelle dans **ursins2701** se trouvent toutes au début de grandes divisions logiques (7.96). En rencontrant cette expression, le scribe a sans doute machinalement reproduit la ponctuation qu'il venait d'utiliser dans des conditions syntaxiques bien différentes.

Dans leur édition du texte qui nous intéresse, P. S. Lewis et A.-M. Hayez (1978) ont le plus souvent procédé à un ajout de ponctuation (133 occurrences). Cette intervention concerne en premier lieu les limites de propositions autonomes et le début de certaines relatives. La ponctuation du manuscrit a été supprimée dans 9 cas, notamment dans les cas que nous avons interprétés comme étant des « ponctuations de respect ». La ponctuation du manuscrit a été maintenue dans 55 cas et 145 frontières ponctuables sont restées sans ponctuation dans l'édition aussi bien que dans le manuscrit.

Ursins2701 fait donc partie des manuscrits les moins ponctués de notre corpus. Si la plupart des marques de ponctuation qui y sont utilisées apparaissent aux frontières ponctuables habituelles, on perçoit clairement dans ce texte l'importance de facteurs n'ayant pas de rapport direct avec la syntaxe : la mise en forme des références des citations et l'expression du respect envers le destinataire du texte.

7.9. Rouen, BM, 1052 (*Quinze joies de mariage*)

Le manuscrit 1052 de la Bibliothèque municipale de Rouen est daté de novembre 1464, et il n'y a pas de raison pour mettre cette indication en doute. Ce codex de 130 feuillets en papier de 272×202 mm contient, outre les *Quinze joies de mariage* (fol. 84 – 130), l'*Histoire de la mort du roi Richard* (fol. 1 – 40) et les *Enseignements d'un père à son fils* (fol. 41 – 129). L'écriture du manuscrit est cursive et relativement peu soignée. L'aspect décoratif se limite à l'emploi de lettrines en couleur au début des grandes unités de la structure textuelle (chacune des quinze « joies » dans le texte qui nous intéresse).

Les chercheurs s'accordent pour situer l'origine géographique du manuscrit dans la France occidentale et proposent une localisation plus précise soit dans la région d'Angers (Coville 1935), soit dans le Poitou (Rychner 1963).

Le manuscrit présente des traces d'annotations clairement postérieures à l'époque de sa rédaction : il s'agit de notes placées en marge et écrites en cursive post-médiévale. Une marque de renvoi en forme de croix est souvent insérée dans le corps du texte pour signaler l'endroit auquel la note s'applique, dans d'autres cas les mots ou syntagmes concernés sont soulignés.

L'état de la reproduction du manuscrit dont nous disposons ne permet malheureusement pas, dans certains cas, de déterminer avec certitude si une marque de ponctuation appartient à l'écriture originale du manuscrit ou si elle a été ajoutée au cours de l'annotation postérieure. Il s'agit des barres obliques utilisées dans la ponctuation faible. Le nombre peu élevé des occurrences « douteuses » (4 au total) permet toutefois de ne pas remettre en cause la validité des tendances constatées.

Les *Quinze joies* est une œuvre anonyme composée sans doute à la fin du XIV^e ou dans les premières années du XV^e siècle (cf. Rychner 1963, p. xxxviii-xlvi pour la discussion). Elle appartient bien évidemment au domaine littéraire et son genre peut être défini comme celui de la nouvelle.

L'extrait que nous avons transcrit et intégré dans notre corpus se situe au début de la « Première joie » (fol. 87r – 88r) et compte 659 occurrences-mots. Cette transcription porte l'identifiant **qjm1052** dans notre corpus²⁸⁶.

L'extrait dépouillé présente 23 occurrences de la ponctuation forte et 9 occurrences de la ponctuation faible, ce qui porte le taux de ponctuation global à 4,9%. Les marques de ponctuation utilisées dans ce fragment sont le point (23 occurrences, dont 18 dans la ponctuation forte), la barre oblique (4 occurrences « douteuses » dans la ponctuation faible, *cf.* plus haut) et la majuscule seule (5 occurrences).

L'extrait que nous avons analysé ne contient pas de frontières d'unités appartenant à la macrostructure du texte, à l'exception du début de la première « joie ». Celui-ci est marqué par une ligne blanche, une lettrine de deux lignes de hauteur, et ses deux premiers mots sont écrits en caractères nettement plus grands que le reste du texte et sont suivis d'un point à une frontière syntaxique normalement « non ponctuable » (*cf.* (7.102) ci-dessous). Ce point participe donc, selon nous, au marquage graphique du début d'une grande division du texte. Des techniques similaires pour structurer graphiquement les grandes unités de sens sont utilisées tout au long du texte.

Une partie importante du fragment dépouillé est occupée par un dialogue entre un mari et sa femme, il contient donc un certain nombre de frontières ponctuelles liées à un changement du plan énonciatif. Le début et la fin d'un discours direct, ainsi que le changement de locuteur semblent être ponctués assez régulièrement (c'est le cas de 7 occurrences sur 8, et il s'agit toujours d'une ponctuation forte). La seule occurrence non ponctuée survient devant une réponse courte du mari à une longue réplique de sa femme :

(7.98) Par Dieu fait elle *sire* il n'est ia mestier que je || le+vous dye car c'est vne chose puis que je || la vous auroye dite *vous* n'en feriez compte *et* || il vous sembleroit que ie le feisse pour autre || chose *v*rayement fait il vous le me direz (*qjm1052, 88r:13*).

Soit le scribe n'avait pas repéré cette frontière qui n'est pas renforcée par un connecteur quelconque, soit il a jugé que la réponse du mari était trop courte pour être marquée d'une ponctuation forte. Aucune ponctuation autour des incisives ou des noms en

²⁸⁶ *Cf.* le sigle **Qjoyes** sous lequel cette œuvre est référencée dans la bibliographie du DEAF.

Manuscrits du XV^e siècle

apostrophe n'a été relevée dans notre extrait (sur 17 occurrences de ce type de frontières).

Dans deux cas, le début du discours direct est marqué, en plus d'une ponctuation forte ordinaire (point et majuscule), d'une marque supplémentaire sous la forme d'une parenthèse ouvrante ou d'une barre oblique inversée :

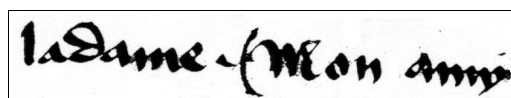


Fig. 64 : Une « parenthèse » devant le discours direct (*qjm1052, 88r*).

(7.99) Lors *commence et* dit || ainxin la+dame .(Mon amy *lessez* moy car je suis || a grant malaise . (*qjm1052, 88r:3*).

Il s'agit ici sans doute d'une tentative de marquage spécifique de ce type de frontière entre plans énonciatifs. L'unicité d'une telle pratique dans notre corpus nous fait penser qu'il s'agit très probablement d'un ajout postérieur.

Les frontières entre les propositions autonomes sont ponctuées dans 14 cas sur 26 (dont 12 ponctuations fortes). Le nombre limité d'occurrences dont nous disposons ne permet pas de tester les différentes hypothèses qui peuvent expliquer pourquoi la ponctuation est utilisée dans un cas et non dans un autre. La position de la frontière à l'intérieur du récit ou du discours direct ne semble pas avoir incidence sur la ponctuation : il y a autant de frontières ponctuées que non ponctuées dans les deux cas. Les mêmes connecteurs sont utilisés tantôt avec, tantôt sans ponctuation, mais la piste de l'incidence sur la ponctuation du premier mot de l'unité ponctuable mériterait d'être explorée sur un échantillon plus important. Nous avons par exemple constaté qu'aucune des 5 occurrences de *car* n'est précédée d'une ponctuation forte : il y a deux ponctuation faibles et trois absences de ponctuations. Il s'agit d'un phénomène assez inattendu, car dans la plupart des manuscrits ce connecteur tend plutôt à « attirer » la ponctuation.

La longueur des segments délimités par des ponctuations fortes est très variable (de 3 à 73 mots), mais il semble que dans certains cas l'usage d'une ponctuation forte à

une courte distance de la précédente puisse être motivé par la répétition d'un même mot (connecteur ou non) :

(7.100) **Et** a+l'auenture || el aura le+cuer bon **Et**+goy et auisa l'autre jour || vne feste (*qjm1052, 87v:16*) ;

(7.101) **Je** || m'en rapporte a+ceulx qui sceuent les lignees || **Je** ne le+dy pas pour mon estat (*qjm, 88r:24*).

Une seule des 14 frontières de propositions ayant des éléments communs est ponctuée dans **qjm1052**, et il s'agit d'une ponctuation forte :

(7.102) **LA premiere** . joye de mariage si est || quant le+jeune homme est en+sa+belle jeunesse || qu'il est frois net *et* plaisant **¶** Et ne s'esmoye || fors de+tirer esgulletes faire ballades et jcelles || chanter / (*qjm1052, 87r:12*).

L'occurrence est située à une vingtaine de mots du début de l'unité structurale du texte consacrée à la première « joie », à une frontière entre propositions coordonnées avec un sujet commun qui peuvent par ailleurs être analysée comme des subordinées coordonnées entre elles (*quant le jeune homme... qu'il est frois... et ne s'esmoye...*). L'absence d'une marque formelle de subordination dans la dernière proposition permet toutefois de la considérer comme autonome.

L'apparition de la ponctuation dans ces conditions peut donc être expliquée par plusieurs facteurs : la proximité du début d'une grande division textuelle, l'autonomie relative de la nouvelle proposition et l'envie de distinguer la coordination d'attributs (*net et plaisant*) de celle de propositions.

L'usage de la ponctuation est encore plus rare aux frontières entre les principales et les subordinées, car deux occurrences seulement de ponctuation faible ont pu être relevées sur 49 frontières de ce type :

(7.103) aultres femmes de+son estat **¶** qui+estoint || abillees a+la nouvelle fasson . (*qjm1052, 57v:18*).

Il s'agit d'une grosse barre oblique devant une proposition relative. Sa forme est différente des autres occurrences de la barre oblique qui sont plus fines et plus longues, ce qui fait penser qu'il s'agit plutôt d'une correction (lettre barrée) que d'une marque de

Manuscrits du XVe siècle

punctuation. L'état de conservation du manuscrit ne permet malheureusement pas de vérifier cette hypothèse.

Pour l'autre occurrence, il s'agit également d'une barre oblique placée à la frontière entre une relative et une infinitive qui dépendent d'un même terme dans la principale :

(7.104) la+haste qu'il a / de taster du || past (*qjm1052, 87v:5*).

Comme nous l'avons déjà indiqué, il n'est pas impossible que toutes les barres obliques soient des ajouts tardifs dans **qjm1052**. Si néanmoins on retient l'hypothèse que la punctuation dans cet exemple date de l'époque de la rédaction du manuscrit, on peut supposer que celle-ci sert à marquer la division entre deux subordonnées très courtes et peut-être à mieux distinguer la forme verbale *a* de la préposition qui suit.

Quoi qu'il en soit, la tendance à l'absence de punctuation entre les principales et les subordonnées est très nette dans ce manuscrit.

Les frontières entre syntagmes coordonnés à l'intérieur d'une proposition sont également ponctuées dans quelques cas isolés seulement. Dans deux cas sur trois, il s'agit d'une punctuation faible marquée par une barre oblique, et il n'est pas exclu que ces barres aient été ajoutées *a posteriori*. La dernière occurrence (présentée dans l'exemple) est également incertaine : il s'agit d'une majuscule entre deux adjectifs épithètes coordonnés : *le cuer bon Et goy*. Nous avons expliqué cette occurrence par la tendance à la répétition de la punctuation devant un même connecteur. En effet, l'écriture du manuscrit est peu lisible à cet endroit, et il se peut qu'il ne s'agisse pas en réalité d'une majuscule.

Les occurrences de la punctuation aux frontières de syntagmes particuliers sont elles aussi très peu nombreuses. Dans deux cas, il s'agit de précisions introduites par *c'est* :

(7.105) ilz ont l'apast empres eux dedens la nasse || C'est assauoir la+femme qui est belle bien pa||ree || et bien abillee de tieulx abille _mens que |[fol. 87v] a+l'aventure son mari n'a pas paieez . (*qjm1052, 87r:24*) ;

(7.106) Et volentiers elles || deuroient parler de+leurs choses *especialles* la || ou leurs mariz sont plus subgitz et doiuent || estre plus enclins pour octrier **.C**est ou lit (*qjm1052*, 87v:25).

Même si nous interprétons le *c'est* introductif comme un connecteur grammaticalisé ne formant pas une proposition, il en a l'apparence. Dans (7.106) la précision vient après une relative circonstancielle, il s'agit donc de toute façon d'une frontière interpropositionnelle. Dans (7.105), la précision est assez longue et comporte plusieurs relatives enchâssées.

Une autre occurrence classée dans cette catégorie correspond à un point placé après les premiers mots d'une division textuelle (7.102). Nous interprétons cette marque comme un élément du dispositif de ponctuation macro-textuelle.

La dernière occurrence – un point en fin de ligne entre deux éléments d'une locution en voie de grammaticalisation (*pour ce*) – peut sans doute être réinterprétée comme une marque supplémentaire d'abréviation (les deux dernières lettres de *pour* sont représentée par une marque suscrite) :

(7.107) ce lui semble *pour* **!** || ce qu'ilz ont l'apast (*qjm1052*, 87r:22).

Compte-tenu du taux de ponctuation général peu élevé dans **qjm1052**, il n'est pas étonnant que l'éditeur (Rychner 1963) ait souvent procédé à l'ajout d'une marque de ponctuation là où il n'y en avait pas dans le manuscrit (54 occurrences). La ponctuation du manuscrit a été maintenue dans l'édition dans 26 cas et supprimée dans 8 cas seulement (dont 3 barres obliques).

En général, la ponctuation dans **qjm1052** semble suivre d'une façon assez conséquente le principe syntaxique : assez fréquente à la frontière de propositions autonomes, elle est pratiquement inexistante dans les cas où les liens entre des propositions voisines sont plus forts (présence d'éléments communs ou subordination), sans parler du niveau intrapropositionnel. Il semble par ailleurs y avoir une tendance au marquage régulier des frontières du discours direct (la fréquence relative de la ponctuation forte y est supérieure à celle qu'on trouve entre propositions autonomes à l'intérieur du récit).

7.10. Paris, BnF, fr. 2682 (*Chroniques d'Enguerrand de Monstrelet*)

Le manuscrit français 2682²⁸⁷ de la BnF constitue avec le manuscrit 2681 du même fonds un ouvrage unique en deux volumes. Les deux codex partagent les mêmes propriétés physiques : dimensions in-folio, feuillets en papier avec une écriture cursive en deux colonnes et à rubriques, reliure en maroquin rouge, aux armes de France sur les plats. Le manuscrit 2681 est entièrement occupé par le premier volume de la *Chronique* d'Enguerrand de Monstrelet. L'autre manuscrit, composé de 268 feuillets numérotés contient, outre le second volume de la *Chronique*, une petite pièce connue sous le nom de *Recouvrement de la duché de Normandie et de Guienne* (à partir du fol. 244).

Le catalogue de la BnF n'indique pour les deux volumes qu'une date très approximative, le XV^e siècle. On peut toutefois restreindre cet empan temporel, puisque les événements décrits dans la *Chronique* de Monstrelet vont jusqu'en 1445. La date du manuscrit est donc forcément postérieure. Si on accepte les dates de 1445 et de 1499 comme limites de la période possible d'exécution du manuscrit, la date moyenne retenue pour le classement chronologique des textes de notre corpus sera de 1472.

Selon la grille typologique de la BFM, l'œuvre de Monstrelet est classée dans le domaine historique et son genre est celui de la chronique. Les formes graphiques des manuscrits français 2681 et 2682 présentent de nombreux traits de la scripta picarde (que L. Douët d'Arcq a pris soin d'éliminer de son édition pourtant basée sur ces manuscrits).

L'extrait que nous avons transcrit s'étend du folio 233r au folio 233v, ce qui correspond aux événements des années 1443 et 1444, soit aux chapitres 271 à 273 dans le tome 6 de l'édition de L. Douët d'Arcq (1862), et contient 753 occurrences-mots. Cet extrait porte l'identifiant **monstre2682** dans notre corpus²⁸⁸.

Le nombre d'occurrences de signes de ponctuation relevés dans notre extrait est particulièrement bas : on ne compte que 21 occurrences du point, 4 occurrences de la

287 Anciennement 8346. C'est sous cette cote que ce manuscrit est référencé dans l'édition de L. Douët d'Arcq (1857).

288 La *Chronique* de Monstrelet n'est pas référencée dans la base bibliographique du DEAF.

barre oblique et deux occurrences du pied-de-mouche. Si on ne prend en compte que ces occurrences, le taux général de ponctuation ne dépasse pas dans ce texte les 3,3%. Cependant, puisque nous considérons les emplois de la majuscule seule à une frontière ponctuable comme des marques de ponctuation, les 28 occurrences de ces majuscules font monter le taux de ponctuation à 7,3%.

Les majuscules seules étant prises en compte, la ponctuation forte devient dominante avec ses 38 occurrences contre 17 pour la ponctuation faible.

Le point est utilisé à tous les types de frontières syntaxiques, le plus souvent dans la ponctuation faible (15 occurrences contre 6 pour la ponctuation forte).

Toutes les occurrences de la barre oblique sont concentrées sur une colonne de l'extrait transcrit (233c), sans qu'il y ait un changement de main apparent par rapport à la colonne précédente. S'il ne s'agit pas d'un ajout tardif, l'apparition d'une nouvelle marque peut être liée à un changement de type de séquence textuelle : une copie d'un document (une déclaration de Pierre de Buisemont, seigneur de Chargni, annonçant la tenue d'un tournoi) commence en effet à la huitième ligne de cette colonne. Il n'est pas exclu que cette ponctuation particulière provienne du document original (ou de sa copie) utilisé par l'auteur de la chronique.

Deux de ces barres obliques sont placées devant des titres de noblesse en apposition :

(7.108) An||thoine de vaudray / segneur de laigle || Guillaume de vauldray / segneur de ||
collaon Jaques de challant seigneur || de ameuille Mesire ame segneur || d'espirey Et
jehan de Chaingny (*monstre2682, 233c:40-44*).

Il convient de noter qu'aucune ponctuation n'est employée dans les mêmes conditions dans la suite de la liste de personnes. On peut supposer que la ponctuation sert ici à mettre en relief les noms propres et, en l'occurrence, à mieux distinguer deux personnalités appartenant sans doute à une même famille : Antoine et Guillaume de Vaudray.

Une autre occurrence de la barre oblique survient un peu plus haut dans la même liste de personnes, devant un nouveau nom :

(7.109) Guillaume || de viane segneur de monbis et || gilly / Jehan segneur de walengon
(*monstre2682, 233c:35*).

Il s'agit formellement d'une ponctuation forte, mais l'emploi de la majuscule s'explique ici plutôt par la mise en relief d'un nom propre plutôt que par le marquage d'une frontière syntaxique.

Enfin, la première occurrence de la barre oblique de notre extrait est relevée au début de la déclaration citée, juste après le verbe conjugué (qui suit la longue dédicace initiale) et devant un infinitif :

(7.110) **E**n l'onneur de *nostre+segneur* et de sa tres||glorieuse mere . de ma dame || sainte *anne* et de mon*segneur* saint geor||ge Je pierre de baufremont *segneur* || de chargni . de monlyet et de mon||fort . *chevalier* consillier et chambellan || de *tres+hault* . tres+puissant et tres+exe||lent || prinche mon *tres+redoubte* et || souuerain *segneur* mon*segneur* le duc || de bourg*ongne* Fay / scaouir a+tous prin||ches barons *chevaliers* et escuyers sans || reprouche [...] (*monstre2682, 233c:9-20*).

L'emploi d'une ponctuation entre le verbe *faire* et un infinitif dans une construction où le verbe conjugué peut être considéré comme un auxiliaire servant à composer une forme factitive peut sembler étonnant. On peut toutefois avancer plusieurs explications à ce phénomène. Premièrement, l'infinitif avec ses compléments pouvait être perçu comme une proposition : l'extrait transcrit présente en effet un autre énoncé dans lequel une proposition infinitive est séparée par une ponctuation (majuscule seule) de la proposition principale postposée :

(7.111) *pour* || augmenter et acroistre le *tres+noble* || mestier et exercite des armes || **M**a voulente et mon intencion est [...] (*monstre2682, 233c:26*).

Néanmoins, étant donné que les propositions subordonnées ne sont que rarement ponctuées dans **monstre2682**, comme nous le verrons plus bas, une explication plus sémantique et moins syntaxique nous paraît plus probable : dans cette longue formule protocolaire, le scribe a voulu (en s'inspirant peut-être des pratiques de certaines chancelleries) mettre en relief les éléments principaux de la proposition, car le sujet *je* et le verbe *faire* sont écrits avec des majuscules. La barre oblique après la forme verbale pouvait servir de moyen supplémentaire de mise en relief du prédicat (de la même façon que des points placés avant et après des chiffres).

Bien entendu, une étude d'un nombre important de documents du même type serait nécessaire pour pouvoir vérifier la solidité de ces hypothèses.

En ce qui concerne les deux occurrences du pied-de-mouche, elles apparaissent dans des conditions syntaxiques qui peuvent également paraître surprenantes à première vue. La première occurrence figure dans une proposition introduite par une forme du déterminant relatif *lequel* :

(7.112) Et y heubt entre les || deux parties de tres grosses escarmu||ches aux_uelles en y heubt de mors || et de naures tant d'un coste *comme* || de l'autre ¶ Ou_ quel temps le || conte de sombreset assambla bien || jusques au nombre de .vi^M. comba||tans ou enuiron (*monstre2682, 233a:28*).

Normalement, ce type de déterminant suppose une relation étroite avec le contexte précédent, ce qui entre en contradiction avec la fonction du pied-de-mouche qui sert habituellement à marquer des divisions importantes dans la structure textuelle. Dans ce cas cependant, le scribe (voire l'auteur) de la chronique a peut-être utilisé cette contradiction consciemment, à des fins stylistiques : pour souligner en même temps la simultanéité et le caractère indépendant des deux événements.

La seconde occurrence survient dans le « document inséré » que nous avons déjà évoqué en parlant des barres obliques :

(7.113) Ma voulente et mon intencion est || auoecq .xij. *chevaliers* et escuyers gen||tilz hommes de .iiij. costes ¶ et des||quelz les noms chi apres s'*ensieuent* || ¶ Ch'est assauoir ¶ thiebault segneur || de rougmont et de mussy *Mesire* || guillaume de brefremont segneur || de sees et de souuegnon (*monstre2682, 233c:30*).

Le pied-de-mouche est placé ici en début de ligne devant la locution *ch'est assauoir* ('c'est-à-dire'), dont la fonction est normalement d'introduire une précision ou une reformulation. Dans le cas en question, cette expression sert à introduire une liste des personnes à l'origine de l'organisation du fait d'armes annoncé. Cette liste est d'une importance capitale au sein du document et elle est mise en relief par l'usage de l'expression introductive précédée d'un pied-de-mouche et suivie d'un point. Encore une fois, des considérations sémantiques et pragmatiques prévalent ici sur les simples relations syntaxiques dans les choix de ponctuation.

Les propositions autonomes sont ponctuées dans toutes les occurrences recensées dans l'extrait dépouillé, et il s'agit d'une ponctuation forte dans 22 cas sur 24.

Manuscrits du XV^e siècle

Les deux occurrences de la ponctuation faible se succèdent dans un passage qui décrit les conséquences d'une bataille mal préparée :

(7.114) Et pour che furent yceulz || franchoix mis en desroy **|** et en y || heubt de .xx. a .xxx. que mors que || pris **|** et les aultres se sauuerent au || mieulx qu'ilz peurent *par* forche de || fuyr (*monstre2682, 233b:12-17*).

On peut considérer dans ce cas que le connecteur causal *pour che* ('pour cela') s'applique aux trois propositions successives, ce qui ferait passer les frontières ponctuables dans la catégorie des propositions avec des éléments communs.

L'exemple (7.114) mis à part, cette dernière catégorie n'est représentée que par 6 occurrences dans l'extrait transcrit, dont 3 portent une ponctuation faible, les 3 autres n'étant pas ponctuées. Dans les 3 cas d'absence de ponctuation, on a affaire à une coordination de propositions courtes avec un sujet commun, comme dans l'exemple suivant :

(7.115) Et puis s'en ala logier a poussay || et y fu bien deux mois (*monstre2682, 233a:38*).

La ponctuation est pourtant présente dans un autre cas assez similaire :

(7.116) Si ala au deuant d'eulz **|** et les vint || rencontrer qu'ilz ne leur donnoient || de garde (*monstre2682, 233b*).

Il s'agit peut-être dans ce cas d'une variation libre ou bien d'un facteur que le petit nombre d'occurrences dont nous disposons ne permet pas de révéler. Il est toutefois intéressant de noter que dans les deux exemples donnés ci-dessus, l'éditeur a fait un choix de ponctuation contraire à son manuscrit : il a utilisé une virgule dans (7.115) et a supprimé la ponctuation dans (7.116).

Dans un cas, une ponctuation faible apparaît devant une proposition relative coordonnée au modificateur d'un nom : *gentilz hommes de .iiij. costes **|** et desquelz les noms chi apres s'ensieuent* (7.113). C'est sans doute la coordination inhabituelle entre des unités syntaxiques de niveaux différents qui a provoqué l'emploi d'une ponctuation dans cet énoncé.

Une ponctuation faible est utilisée également dans la coordination de deux propositions infinitives :

(7.117) **A**u *commen*chement de || chest an Le roy de fran||ce fist grande *assam*blee || de gens d'armes sur || intencion d'*entr*er en normendie || l'este ensieuant **¶** et aussy pour || baillier secours a ceulz de dyepe (*monstre2682, 233a*).

Ici l'emploi d'une ponctuation peut être favorisé par la présence de l'adverbe *aussy* qui fait considérer la deuxième proposition coordonnée comme une information supplémentaire, ajoutée après-coup. Le fait que les infinitifs dans les deux propositions sont introduits par des prépositions différentes (*sur l'intention de* et *pour*) peut également avoir joué un rôle dans l'apparition d'une ponctuation.

Comme nous l'avons déjà mentionné, les limites entre les propositions principales et subordonnées ne sont que très rarement ponctuées dans **monstre2682** : c'est le cas de 2 occurrences sur 33 seulement. Il convient de noter toutefois que dans 2 cas sur 3 il s'agit d'une ponctuation forte (majuscule seule).

Une occurrence de ce type a déjà été présentée dans l'exemple (7.111). L'autre occurrence survient juste avant celle-là, dans le même passage du « document inséré » :

(7.118) **J**e pierre de baufremont [...] **F**ay / scaouvoir a+tous prin||ches [...] **Q**ue *pour* || augmenter et accroistre le *tres*+noble || mestier et exercite des armes || **M**a voulente et mon intencion est [...] (*monstre2682, 233c:12-26*)²⁸⁹.

Comme nous l'avons déjà vu, dans cette longue formule « protocolaire », le scribe a voulu utiliser des ponctuations fortes pour mettre en évidence les principaux constituants de la proposition principale (le sujet *Je* et le verbe *Fay*). Du coup, l'usage d'une ponctuation forte s'impose de manière quasi-automatique pour les frontières entre propositions, d'autant plus que celles-ci correspondent à des éléments importants dans la structure sémantique du document.

Les syntagmes coordonnés ou énumérés à l'intérieur d'une proposition sont assez nombreux dans **monstre2682** : 57 occurrences relevées dans l'extrait dépouillé. Dans la majorité des cas les frontières entre ces syntagmes ne sont pas ponctuées (38 occurrences), mais le nombre de ponctuations n'est pas négligeable : 8 occurrences de la ponctuation faible et 11 occurrences de la ponctuation forte.

²⁸⁹ Pour rendre l'exemple lisible, nous avons remplacé par des points de suspension entre crochets plusieurs lignes de texte contenant des subordonnées en chaîne ou des éléments coordonnés. Le texte complet du début de cette séquence est cité dans l'exemple (7.110).

Manuscrits du XV^e siècle

En réalité, toutes les occurrences de la ponctuation forte surviennent dans la liste des organisateurs du fait d'armes, devant le nom ou le titre d'un nouveau personnage (cf. l'extrait de cette liste dans l'exemple (7.108)). Ceci n'est qu'une confirmation de la tendance générale à l'usage de la ponctuation forte dans des listes de noms propres comme nous l'avons constatée déjà dans plusieurs manuscrits de notre corpus (par exemple, **im574**).

La majorité des occurrences de la ponctuation faible se trouve également dans des énumérations de noms propres, dans des séquences moins longues et moins complexes que celle que nous venons d'évoquer. L'emploi et l'absence de ponctuation alternent souvent dans une même énumération. Parfois, la règle moderne semble se profiler (une ponctuation entre éléments juxtaposés et l'absence de ponctuation devant la conjonction *et*), mais dans d'autres passages la pratique est strictement inverse :

(7.119) *segneur* || de chargni . de monlyet et de mon||for (*monstre2682, 233c:13*) ;

(7.120) pa||lys d'angou de traonnoix . et de || chastrangonnoix (*monstre2682, 233a:42*).

Le premier cas semble toutefois dominant, car la majorité des occurrences d'absence de ponctuation (32 occurrences sur 38) concerne les coordinations avec la conjonction *et*.

L'usage de la ponctuation aux frontières de syntagmes particuliers est assez rare mais pas exceptionnel : on compte 9 occurrences de ce type dans l'extrait dépouillé. La quasi-totalité de ces occurrences ont déjà été analysées plus haut : les barres obliques devant les titres de noblesse en apposition (7.108), la ponctuation forte devant le sujet et le verbe dans la première proposition du « document inséré » (7.110), les précisions introduites par *ch'est assauoir* (7.113).

Une occurrence d'une majuscule seule suit un complément circonstanciel initial, devant le sujet et son déterminant : *Le roy de france* (7.117). S'il ne s'agit pas d'une simple variante calligraphique, cette majuscule peut servir à mettre en valeur la mention d'une personnalité importante. C'est donc plutôt une ponctuation « de respect » qu'une ponctuation syntaxique.

La seule occurrence qui se prête difficilement à une explication est le point placé entre un nom et son modificateur :

(7.121) le || duc de bourgongne estant en || son pays | de bourgongne y heubt plui||seurs gentilz
hommes de son hostel (*monstre2682, 233b:35*).

Il peut s'agir dans ce cas isolé d'un simple incident d'écriture : l'absence d'espace blanc quelconque devant et après cette marque témoigne en faveur de cette hypothèse. Si on suppose néanmoins que le point a été ajouté intentionnellement, on peut penser à la mise relief d'un syntagme prépositionnel contenant un nom géographique.

De façon générale, **monstre2682** s'inscrit entièrement dans les tendances dominantes de la ponctuation française médiévale avec l'usage plus ou moins systématique de la ponctuation aux frontières de propositions autonomes et l'usage nettement moins fréquent dans les autres cas. Le fragment dépouillé contenant un « document inséré », nous avons pu observer l'influence du type de séquence textuelle sur la pratique de la ponctuation. Dans plusieurs cas, la ponctuation dans ce texte est motivée par des considérations sémantiques et pragmatiques, plutôt que par la structuration syntaxique formelle.

7.11. Paris, BnF, fr. 15217 (*Le Saint voyage de Jherusalem du seigneur Ogier d'Anglure*)

Le manuscrit 15217 du fonds français de la BnF a été soigneusement décrit par Léopold Pannier à la fin du XIX^e siècle (Paris & Pannier 1872, p. 332-336). Cette description reste à ce jour la plus précise et la plus détaillée.

Il s'agit d'un codex de 95 feuillets en papier, de 252 × 192 mm, assemblés sous une reliure en parchemin. L'écriture est cursive et le texte est disposé sur une colonne. La datation du manuscrit n'a pu établir de date plus précise que la deuxième moitié du XV^e siècle, ce qui porte la date « moyenne » retenue pour le classement chronologique à 1475.

Les œuvres contenues dans ce manuscrit sont assez variées mais liées toutes d'une manière ou d'une autre au domaine religieux : le *Dialogue entre l'esprit de Guy de Tarno et le prieur...* (fol. 2 – 13), *La vie de st Alexis* (fol. 14 – 24), *l'Exemple sur le*

Manuscrits du XV^e siècle

secret de la confession (fol. 25 – 26), *La prinse de Constantinoble* (fol. 67 – 72) ;
L'exposicion des vigiles de mors... (fol. 73 – 96).

Le *Saint voyage* d'Ogier d'Anglure y occupe les folios de 27 à 66. Il s'agit du récit d'un pèlerinage en Terre Sainte entrepris par l'auteur en 1395 et mis par écrit quelques années plus tard. Cette œuvre se situe à la fois dans les domaines littéraire et religieux. L'extrait que nous avons transcrit couvre les folios 27r à 30v et comporte 1962 occurrences-mots. Nous nous y référerons sous l'identifiant **anglure15217**²⁹⁰.

Le taux de ponctuation général atteint les 7,8% dans l'extrait dépouillé. Dans la quasi-totalité des occurrences (152 sur 155) il s'agit d'une ponctuation forte (dont 112 occurrences de la majuscule seule).

En général, la forme des lettres dans **anglure15217** permet de faire une distinction claire entre les majuscules et les minuscules. Une seule lettre, le < d >, pose parfois un problème d'interprétation : les deux variantes ont à peu près la même forme, mais la majuscule est en général plus large que la minuscule (cf. les graphies *deing* 'd'Ain' et *dillec* 'de là-bas' présentés ci-dessous).

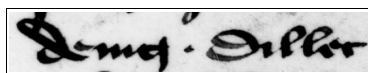


Fig. 65 : Lettre < d > majuscule et minuscule (*anglure15217*, 27v).

Dans certains cas, la différence est moins évidente, mais nous avons pris le parti de considérer comme des majuscules toutes les occurrences de < d > visiblement plus grandes que le reste de la graphie d'un mot au début d'une unité ponctuable.

Si on ne prenait en compte que les occurrences des signes de ponctuation proprement dits, le taux de ponctuation ne dépasserait donc pas les 2,1% dans ce texte.

Le point est utilisé 41 fois (sans compter les occurrences liées à la mise en relief des chiffres). Dans un cas, on peut apercevoir une très légère barre oblique suivie d'une majuscule. Il est difficile de dire s'il ne s'agit pas d'un ajout tardif, mais de toute façon cette marque ne peut jouer qu'un rôle très périphérique dans **anglure15217**.

Certaines (mais pas toutes) les majuscules sont surlignées d'un trait à l'encre brune. Le manuscrit présente un nombre important d'annotations : des passages

²⁹⁰ Dans la bibliographie du DEAF ce texte figure sous le sigle **Anglure**.

soulignés et des notes en marges (notamment, des dates et des noms de villes). Il s'agit sans doute d'ajouts tardifs.

Le fragment que nous avons transcrit se situe au début de l'œuvre et consiste essentiellement en une description du trajet parcouru par les pèlerins entre la France et les îles grecques de la Mer Adriatique (en passant par l'Italie).

La première page du texte (27r) contient un prologue (une sorte de long titre descriptif), un paragraphe traitant du jour du départ et de l'objectif du voyage et le début de la description de l'itinéraire parcouru. Chacune de ces divisions textuelles commence par une grande initiale en couleur, dont la hauteur diminue d'une ligne à chaque nouvelle division : 4 lignes pour le titre, 3 lignes pour l'objectif du voyage et deux lignes pour le début de l'itinéraire.

Le prologue se caractérise par une densité exceptionnelle de majuscules : un peu moins d'une majuscule pour tous les quatre mots, tandis que le taux moyen est deux fois moins élevé. On discerne ici, selon nous, la même tendance que celle qui amène à la « surponctuation » des premières pages de certains manuscrits.

Les premiers mots de la deuxième division (*Et premiers*) sont présentés comme un titre : ils sont écrits à l'encre rouge et placés au milieu d'une ligne. Le reste du passage étudié ne présente pas d'occurrences de grandes initiales, mais est structuré en paragraphes « modernes » avec des sauts de lignes et des espaces interlinéaires augmentés.

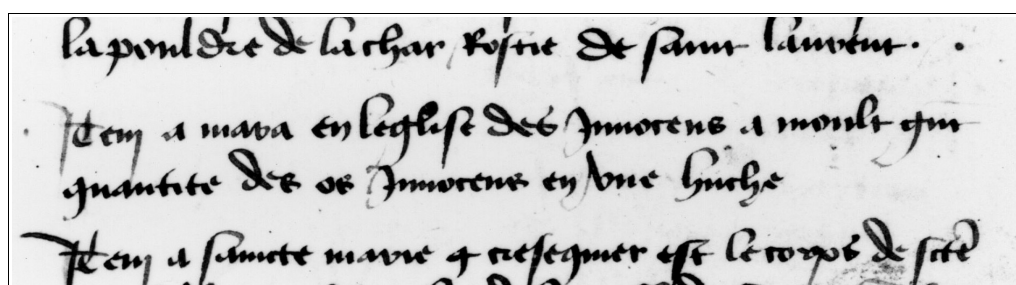


Fig. 66 : Mise en forme des paragraphes (anglure15217, 28v).

La plupart de ces paragraphes commencent par l'une des trois expressions jouant le rôle d'organisateur textuel : le mot *item*, l'énumératif *premiers* ou *premierement*, ou encore les expressions *nous partismes* ou *nous arrivames*.

Manuscrits du XV^e siècle

On peut donc conclure que **anglure15217** présente un système bien organisé de structuration macrotextuelle avec une mise en forme particulière de la première page du texte.

L'extrait étudié ne contient pas de discours direct, nous ne pouvons donc rien dire sur le traitement des frontières de type énonciatif dans **anglure15217**. Notre extrait présente en revanche une citation latine précédée d'une ponctuation forte et soulignée (le soulignement pouvant être tardif) :

(7.122) le fust de || sa lence estoit aussi gros **Q**uasi liciatorium texencium . (*anglure15217*, 29v:4)²⁹¹.

Du point de vue syntaxique, cette citation joue le rôle d'une comparative elliptique. On peut se demander pourquoi elle n'est pas traduite. Il se peut que l'auteur ait éprouvé une difficulté dans la traduction des termes techniques dans ce passage et qu'il ait noté cette citation pour y revenir par la suite. Si le sens du passage n'était entièrement clair ni pour l'auteur ni pour le scribe, l'usage de la ponctuation forte peut être lié à l'hésitation sur le degré de liaison entre cette citation et son contexte gauche.

Les frontières entre les propositions autonomes sont ponctuées assez régulièrement dans ce manuscrit (83 occurrences sur 95 sans compter les limites de divisions textuelles). Une analyse plus détaillée montre que certains connecteurs servant à introduire des constructions fréquemment répétées dans ce texte ont une nette affinité avec la ponctuation, tandis que d'autres « permettent » son absence. Ainsi, les 32 occurrences de *Item* sont précédées d'une ponctuation forte. L'adverbe assertif *Sy* est également toujours ponctué s'il se trouve au début d'une proposition autonome (6 occurrences) :

(7.123) **Sy** lui vint adont voulente || de combatre celui jayant dont ses freres et autres || le blasmoient tres fort **Sy** fut dauid+arme des armes || mesmes du Roy saul (*anglure15217*, 29v:8-10).

Tel est aussi le cas de la préposition *de* introduisant le point de départ : 7 occurrences dans des propositions autonomes (dont 3 suivies de l'adverbe *illec* 'en ce

291 Il s'agit de la description de l'armure de Goliath : la hampe de sa lance est comparée à l'ensouple d'un métier de tisserand (*I Samuel XVII*, 7).

lieu-là') et 25 occurrences dans des énumérations de points de départ et d'arrivée que nous interprétons comme des propositions avec ellipse du verbe²⁹² :

(7.124) **De** Chasteillon a la maison de froit || mantel **. D'illec** a flore sur ource **De** flore a ||
Beaune **De** Beaune a challon sur la sone **De** chalon || ou peage de tornus **D'illec** a Saint
thiuer (*anglure15217, 27v:1-4*).

La conjonction *et* en début de proposition autonome est ponctuée dans 18 cas sur 22. En revanche, 2 occurrences sur 3 de *car* et 3 sur 5 de *mais* ne sont pas ponctuées.

Une seule occurrence de frontière entre propositions autonomes présente l'usage d'une ponctuation faible :

(7.125) le+fer de+sa lence pesoit .vj^C. sicles[!] le fust de || sa lence estoit aussi gros
(*anglure15217, 29v:3*).

Il s'agit d'une partie de la description détaillée de l'armure de Goliath comportant notamment l'indication de son poids, et le point apparaît juste après le mot *sicle*²⁹³ précédé d'un chiffre. Il se peut que cette occurrence soit motivée non pas par la frontière syntaxique, mais par le marquage de l'unité de mesure. Il convient toutefois de noter que les autres occurrences de ce mot et des autres unités de mesure ne sont pas ponctuées.

Les propositions ayant des éléments communs sont ponctuées dans 35 occurrences sur 51 (68%). Les marques de ponctuation utilisées sont la majuscule seule (23 occurrences), le point suivi d'une majuscule (9 occurrences), le point suivi d'une minuscule (2 occurrences) et, peut-être, une barre oblique suivie d'une majuscule (1 occurrence, où la barre oblique peut être un ajout tardif).

La plupart des occurrences ponctuées sont des propositions elliptiques indiquant les lieux de départ et d'arrivée, telles que nous les avons présentées dans l'exemple (7.124). La ponctuation forte est largement dominante dans ces constructions (25 occurrences contre seulement deux non ponctuées)²⁹⁴. Si ces constructions ne sont pas

292 Il convient de rappeler toutefois que dans certains cas il est difficile de faire la distinction entre le < d > majuscule et minuscule.

293 Mesure de poids et monnaie hébraïque (équivalent de 12 g).

294 Il convient de noter toutefois que quelques unes des 16 occurrences de la majuscule seule peuvent être remises en question suite à l'absence de distinction nette entre < d > majuscule et minuscule.

Manuscrits du XV^e siècle

autonomes syntaxiquement, elles véhiculent des informations importantes dans la structure du texte. Le fait qu'elles contiennent des noms géographiques peut être un facteur supplémentaire favorisant l'emploi de la ponctuation forte.

Le scribe paraît plus hésitant dans la ponctuation des propositions elliptiques introduites par *après*, qui s'ajoutent parfois à celles qu'on vient d'examiner et indiquent un point d'arrivée ultérieur :

(7.126) D'illec || au monlt du chat . apres a chambry en Savoie || D'illec a monlt melian .
Après a aigüe belle (*anglure15217*, 27v:9-10).

Sur 5 occurrences, deux portent une ponctuation faible (très rare dans **anglure15217** !), deux autres sont dotées une ponctuation forte (une majuscule seule et une majuscule précédée d'un point), la dernière occurrence n'étant pas ponctuée. L'importance informationnelle de la frontière entre ici en contradiction avec la brièveté de l'unité « ponctuable ».

Un mode d'expression d'une nouvelle destination encore plus elliptique consiste en la juxtaposition de plusieurs syntagmes prépositionnels en *a* :

(7.127) D'illec a saint || anthony . A villaines a monlt callier A lyer en || pymont
(*anglure15217*, 27v:16).

Selon notre classement des frontières ponctuables, nous ne sommes plus au niveau de la proposition, mais de syntagmes énumérés. Le rapprochement basé sur un critère sémantique s'impose toutefois. Parmi les 6 frontières entre éléments énumérés de ce type, 4 comportent une ponctuation forte et deux ne sont pas ponctuées. Le traitement de ces séquences est très similaire à celui des propositions elliptiques.

Si l'on exclut les constructions exprimant les points de départ et d'arrivée, la proportion des frontières ponctuées entre propositions ayant des éléments communs baisse à moins d'un tiers des occurrences (6 sur 19).

Les limites entre propositions principales et subordonnées sont ponctuées dans 11 cas sur 96. 8 de ces 11 occurrences sont des subordonnées relatives, mais ce type de subordonnée est également le plus fréquent pour les frontières non ponctuées (52

occurrences sur 85). Une analyse plus détaillée permet de conclure que la probabilité d'apparition d'une ponctuation est nettement plus élevée en cas d'usage du pronom (ou déterminant) relatif LEQUEL²⁹⁵, surtout s'il est précédé d'une préposition : dans ce cas, 5 occurrences sur 8 sont ponctuées (6 sur 14, si prend en compte les occurrences sans préposition). Cette particularité de la ponctuation peut signifier que les relatives introduites par LEQUEL étaient perçues à l'époque comme plus autonomes que celles en *que* ou *qui*, par exemple.

Dans un cas, deux occurrences très similaires se succèdent à quelques lignes d'intervalle, on peut supposer donc que le facteur de répétition joue un certain rôle dans le choix de la ponctuation de la deuxième occurrence :

(7.128) Entre paula et corfo a+vne isle deshabee que l'en appelle || cazopoly **En la quelle** isle il y+a vne chappelle+de *nostre* dame que || l'en appelle *nostre*+dame de cazopoly **En la quelle** chappelle deuant || l'ymage a+vne lampe plaine d'uille (anglure15217, 30v).

Il convient de noter tout de même que LEQUEL joue dans les deux cas le rôle d'un déterminant et peut être interprété comme un simple anaphorique n'établissant pas de lien de subordination (*en la quelle isle* = 'sur cette ile'). Dans d'autres occurrences ponctuées, il s'agit bien au contraire de LEQUEL pronom.

Deux des 31 occurrences de subordonnées en *qui* sont ponctuées. Dans un cas, il s'agit très probablement d'une ponctuation plutôt décorative que syntaxique, car la marque se trouve à la première ligne d'un paragraphe qui est en même temps la dernière ligne d'une page. Sur cette même ligne, on observe une majuscule marquant la frontière entre un sujet pronominal et le verbe (*Nous Partismes*) ; la ligne suivante (la première d'une nouvelle page) commence également par une majuscule coïncidant avec la frontière entre plusieurs épithètes énumérées. Une majuscule apparaît également dans la forme *Jour* qui suit un adjectif ordinal exprimé sous la forme d'un chiffre romain avec un *e* suscrit. Ces trois occurrences de ponctuation à des frontières « peu ponctuables » confirment à notre avis l'hypothèse « décorative » :

295 Nous utilisons les majuscules, car il s'agit d'un lemme représentant l'ensemble des formes de ce pronom.

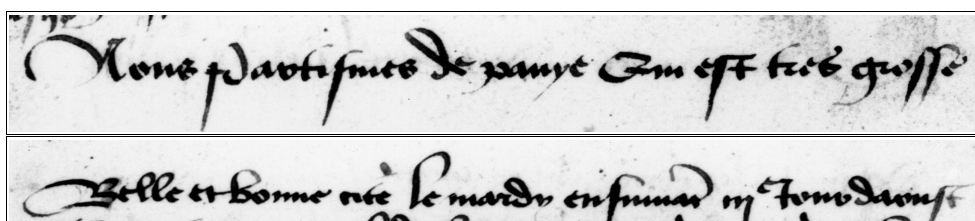


Fig. 67 : Majuscules « décoratives » à la dernière et à la première ligne d'une page (anglure15217, 27v-28r).

- (7.129) Nous Partismes de pauye Qui est tres grosse ||[fol. 28r] Belle et bonne cite le mardy ensuiuant iije Jour d'aoust (anglure15217, 27v:30-28r:1).

La même pratique d'usage de majuscules décoratives peut être retrouvée à la dernière ligne du folio 29v et à la première ligne de la page suivante. Il convient de noter tout de même que ces majuscules apparaissent le plus souvent à des frontières potentiellement ponctuables et non à l'intérieur de syntagmes soudés.

La deuxième occurrence figure dans une construction où la relative substantive placée devant la principale permet de « topicaliser » un argument (sujet « logique » occupant la position d'objet indirect) :

- (7.130) Et qui n'a || lettres de past Sy fault il payer la+gabelle partout . (anglure15217, 28r:22).

La particularité de cette construction qui sert à accentuer le contraste entre le thème et le rhème (exprimé sans doute à l'oral par un schéma intonatif spécifique) a motivé le marquage graphique de la frontière syntaxique et surtout informationnelle. Ne disposant pratiquement pas d'autre marque que la majuscule, le scribe a tout de même hésité à poser une marque forte et a écrit une lettre en forme de majuscule mais de la taille d'une minuscule :

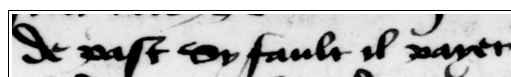


Fig. 68 : « Petite majuscule » (anglure15217, 28r).

Il s'agit de l'occurrence unique de ce type de « ponctuation moyenne » dans l'extrait dépouillé. L'examen minutieux de l'ensemble du manuscrit devrait permettre de savoir dans l'avenir s'il s'agit d'un cas isolé ou d'une pratique, certes, peu fréquente, mais consciente.

La dernière occurrence de frontière subordonnée/principale ponctuée se trouve au milieu d'une construction de 8 propositions reliées par des relations de subordination avec 6 niveaux d'imbrication. Nous présentons cet énoncé ci-dessous à l'aide de crochets avec l'indication du niveau de subordination syntaxique :

(7.131) [⁰Car+la sainte escripture fait mention [¹que || pour le temps du Roy saul premier Roy des hebrieux II || auoit assemble grant host contre les philistiens [²entre || lesquelz estoit Icellui goliast [³qui aloit disant *chascun jour* || *et* criant aux ebrieux [⁴que [⁵s'il y+auoit nul ebrieu [⁶qui+vousist || combatre a+lui seul a+seul⁶]] Cellui d'eulx [⁵qui auroit victoire⁵] || l'*autre partie seroit serue et* subgecte au vainqueur⁴]³]²]¹]⁰]
(*anglure15217, 29r:24*).

La ponctuation apparait devant la reprise de la quatrième proposition (*que... Cellui... l'autre partie seroit serve...*) après deux subordonnées imbriquées. Cette frontière syntaxique correspond également à une frontière informationnelle importante entre le préambule et la partie rhématique principale de l'énoncé.

Il convient de noter, dans le même passage, l'usage d'une majuscule dans le déterminant démonstratif *Icellui* à la frontière verbe/sujet. Si l'on exclut l'hypothèse d'une simple variante positionnelle (qui n'est pourtant pas improbable), on peut supposer qu'elle sert de marqueur « détaché » du nom propre. Nous avons déjà rencontré un phénomène similaire dans des listes de personnalités et de noms géographiques.

Les frontières entre des syntagmes coordonnés ou énumérés à l'intérieur d'une proposition sont ponctuées dans 8 cas sur 44 (18%). La moitié des occurrences ponctuées concernent les énumérations de destinations évoquées dans (7.127). Trois autres occurrences apparaissent dans le prologue, qui se caractérise, comme nous l'avons remarqué déjà, par une densité exceptionnelle de la ponctuation. La dernière occurrence, enfin, coïncide avec le début d'une page et s'explique selon nous par des raisons décoratives (7.129).

L'extrait dépouillé présente 6 occurrences de ponctuation à des frontières de syntagmes particuliers à l'intérieur d'une proposition. Deux de ces occurrences surviennent dans le prologue et trois autres soit à la première, soit à la dernière ligne d'une page, d'où leur explication probable par le facteur décoratif.

La dernière occurrence apparaît devant un sujet pronominal après un long circonstant antéposé dans le cadre de la construction syntaxique complexe présentée dans l'exemple (7.131). Il se peut que cette majuscule, comme celle qui la suit quelques lignes plus bas, s'explique par l'analyse plutôt pragmatique que grammaticale de cette construction par le scribe. De ce point de vue, une ponctuation à la fin d'un circonstant posant le cadre temporel de l'énoncé est tout-à-fait à sa place.

La comparaison entre la ponctuation du manuscrit **anglure15217** et celle de son édition (Paris 1879) permet de constater que dans 146 cas une marque de ponctuation est présente à une frontière syntaxique à la fois dans l'édition et dans le manuscrit. Dans 79 occurrences, une frontière ponctuable n'est ponctuée dans aucune des sources. 140 marques de ponctuation ont été ajoutées par les éditeurs par rapport au manuscrit et seulement 9 marques du manuscrit ont été supprimées.

De façon générale, **anglure15217** présente un système de ponctuation très simple (avec la dominance de la majuscule seule), mais appliqué d'une manière conséquente. Le premier facteur est syntaxique (avec la ponctuation quasi systématique des propositions autonomes), mais d'autres facteurs entrent en jeu à des niveaux inférieurs : la pragmatique (énumération de destinations, séparation du thème et du rhème) et l'esthétique (majuscules dans le prologue et sur les premières et dernières lignes des pages).

7.12. Paris, BnF, fr. 1357 (*Recueil des plus célèbres astrologues de Simon de Phares*)

Le manuscrit français 1357 de la BnF est précisément décrit et daté. Il a été écrit en 1494 sous la dictée ou d'après les fiches de Simon de Phares et a ensuite été corrigé par l'auteur à plusieurs reprises entre 1496 et 1498. J.-P. Boudet (1999) y a identifié 5 strates de corrections autographes. Il est hautement probable que le manuscrit a été composé à Paris, où Simon de Phares habitait en 1496 et en 1499.

Dans son œuvre, Simon de Phares retrace l'histoire de l'astrologie en présentant au lecteur une centaine de notices biographiques d'astrologues (dont la sienne).

L'ouvrage était destiné à Charles VIII et l'auteur cherchait avant tout à se justifier devant le roi après sa condamnation par des théologiens. Si le contenu et la forme du *Recueil* suggèrent son classement dans le domaine littéraire et le genre de la biographie, la visée polémique le rapproche dans une certaine mesure du domaine politique. L'auteur lui-même définit le genre de son œuvre comme *Elucidaire*. Le type de séquence textuelle dominant dans le passage transcrit est sans aucun doute argumentatif.

Le manuscrit est composé de 167 feuillets de papier, dont la largeur varie de 190 à 197 cm, et la hauteur de 287 à 292 cm. Le texte est disposé sur une seule colonne par page. Deux écritures – celle d'un scribe professionnel et celle de l'auteur (dans ses corrections) – s'opposent très nettement dans la copie du texte. Il s'agit toutefois dans les deux cas d'une écriture cursive et rapide. Le texte initial peut être barré et les corrections sont ajoutées entre les lignes ou en marges. Des marques d'insertion ayant la forme de ν renversé $\langle \wedge \rangle$ apparaissent le plus souvent à l'endroit où le texte doit être inséré et devant le texte à insérer. Nous ne les considérons pas comme des marques de ponctuation, même si elles servent *de facto* dans une certaine mesure à la structuration graphique du texte. Nous n'analyserons donc que les marques de ponctuation présentes soit dans le texte initial, soit à l'intérieur des corrections. Nous conserverons en revanche les marques d'insertion et mettrons en évidence les « strates » de corrections dans les exemples cités, en suivant l'analyse de J.-P. Boudet (1997)²⁹⁶.

Les modifications de la structure syntaxique apportées par les corrections ne sont prises en compte que dans l'étiquetage des frontières ponctuables à l'intérieur des corrections. Il est bien-sûr possible (et même probable en ce qui concerne les doubles barres obliques) que des marques de ponctuation ont été ajoutées dans le corps du texte pendant les corrections, mais elles relèvent de toute façon de la pratique contemporaine à la rédaction du manuscrit et sont pertinentes pour notre analyse. Par ailleurs, nous verrons que la ponctuation forte est dominante dans ce texte et les corrections remplaçant une minuscule par une majuscule ou *vice versa* auraient été visibles, or nous n'en avons pas repéré dans notre extrait.

296 En suivant les conventions adoptées par J.-P. Boudet, nous entourerons de chevrons les fragments ajoutés avec l'indication, en exposant, de la « strate ». Le texte barré lors des corrections sera barré et présenté en couleur rouge dans nos citations.

Il convient de préciser que la syntaxe de Simon de Phares est très complexe : dès la « première strate », son texte abonde en subordonnées juxtaposées et enchâssées de tous types, de longues incidentes, et les corrections ne font que complexifier davantage ces constructions. Nous tenterons néanmoins d'appliquer à ce texte notre méthode d'analyse habituelle et nous verrons que l'étiquetage « aplati » des frontières ponctuables permet de lever un nombre important de difficultés.

L'extrait que nous avons transcrit s'étend du folio 1v à 4v et compte 2168 occurrences-mots. Il porte l'identifiant **phares1357**²⁹⁷ dans notre corpus.

Avec 149 occurrences de la ponctuation forte et 29 occurrences de la ponctuation faible, le taux général de ponctuation s'élève dans **phares1357** à 8,2%. La marque la plus répandue est la majuscule seule (76 occurrences), le point est également assez fréquent (50 occurrences, dont 34 dans la ponctuation forte). Les autres marques sont nettement plus rares : la barre oblique²⁹⁸ simple (18 occurrences, dont 7 dans la ponctuation forte) ou double (15 occurrences, toujours dans la ponctuation forte), le deux-points (11 occurrences, dont 9 dans la ponctuation forte). Notre extrait présente par ailleurs une occurrence d'un point suivi d'une barre oblique (ou plutôt verticale) et d'une majuscule (Fig. 69) et une occurrence d'une marque en forme de losange composé de quatre points et suivi d'une minuscule (Fig. 70).

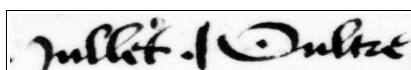


Fig. 69 : Point suivi d'une barre verticale (*phares1357*, 3v).

(7.132) Et de ce ordonna || xij moys desquelz en *per*petuelle memoire de son nom || jl *nom*ma l'un et le plus cler jullet . | Oultre plus jl *com*posa || le kallendrier et nombre d'or (*phares1357*, 3v:28).

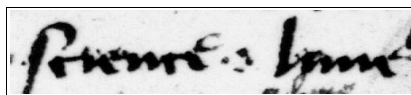


Fig. 70 : Losange de quatre points (*phares1357*, 1v).

297 Cf. le sigle **SimPhares** utilisé dans la bibliographie du DEAF.

298 Certaines barres obliques sont très peu inclinées, au point qu'il serait plus juste de les appeler des barres verticales. Il s'agit néanmoins toujours de la même marque que la barre oblique sur le plan graphématique.

(7.133) *Comment Astrologie || est vraye science ¶ l'une des sept ars liberaulx Λ <2et que elle a este 2> Aprise || sceue et pratiquee (phares1357, 1v:15).*

Les frontières ponctuables correspondant à ces occurrences n'ont rien d'exceptionnel : propositions autonomes avec un sujet pronominal dans la deuxième d'entre elles (7.132) et syntagmes nominaux juxtaposés (7.133). L'usage des marques exceptionnelles s'explique donc sans doute par des raisons extralinguistiques.

La variété des signes de ponctuation et une légère tendance à la spécialisation des différentes marques (le point et le deux-points dans la ponctuation forte et la simple barre oblique dans la ponctuation faible) annoncent les changements à venir dans le système de la ponctuation liés au développement de l'imprimerie.

L'extrait dépouillé ne contient que deux frontières de divisions textuelles majeures marquées par le scribe (qui travaillait, rappelons-le, sous la dictée ou d'après les notes de l'auteur) : le début de la dédicace initiale et le passage d'une argumentation plutôt personnelle liée aux mésaventures de l'auteur à la défense de l'astrologie en tant que « vraie science ». Dans les deux cas, le scribe a respecté un retrait de cinq lignes afin qu'un rubricateur puisse y dessiner une grande initiale, ce qui n'a jamais été fait dans ce manuscrit « de travail ».

L'extrait dépouillé contient quelques citations de textes faisant autorité et un « discours direct » où l'auteur reproduit les propos de ses détracteurs :

(7.134) *Mais || vsans d'auctorite empruntee ¶ Et de la Raison contenue || en ce mectre de Juuenal ¶ Sic volo sic jubeo ¶ sit pro || Racione voluntas ¶ ont calumpniee de supersticion (phares1357, 2v:5) ;*

(7.135) *Car com dit le+sage || nemo dat quod non habet pour_quoy s'il y+a quelque erreur (phares1357, 1v:34)²⁹⁹ ;*

(7.136) *Et la ou <2ce 2> aucuns detracteurs || dient nous confessions que astrologie est vraye science || [fol. 4v] Mais elle ne deuroit traicter que du mouuement et cours || du ciel et des corps celestes ¶ je leur-<2lui 2> demanderoie³⁰⁰ || volentiers quel fruit (phares1357, 4r:32-4v:2).*

299 Ce passage fait partie de la « deuxième strate » d'ajouts, placées en l'occurrence sur la marge basse de la page.

300 J.-P. Boudet transcrit *demanderoit*, ce qui nous semble peu probable vu la forme du dernier caractère manuscrit et la désinence habituelle de la première personne du conditionnel.

La première observation qu'on peut tirer de ces exemples est que le début du discours direct ou de la citation n'est ponctué qu'en absence du verbe de parole dans la proposition précédente (7.134). La frontière entre une proposition avec un verbe de parole et le discours direct se rapproche de celle d'une incise, et il n'y a pas de ponctuation en (7.135) et (7.136). En revanche, le début de la proposition contenant le verbe de parole (ou sa principale) comporte une ponctuation forte. Il convient de remarquer toutefois que dans (7.135) le début de la citation coïncide avec un saut de ligne, qui peut, comme on le sait, remplacer une marque de ponctuation, surtout à l'intérieur d'une correction écrite en marge.

Notons au passage la structure syntaxique complexe de (7.134), où le discours direct « s'insère » entre une proposition participiale et sa principale (*usans d'auctorité... [DD] L'ont calumpniee...*). L'apparition de plusieurs ponctuations fortes dans ce passage montre que les relations « immédiates » entre les unités « ponctuables » sont parfois plus importantes pour le choix de la marque que la structure syntaxique globale.

Dans (7.136), il convient de noter l'usage de la ponctuation faible (une barre oblique simple) à la reprise du discours de l'auteur (qui continue pourtant à relater son dialogue imaginaire avec un détracteur).

Les limites entre propositions autonomes sont ponctuées dans la grande majorité des cas. Il convient de noter que la distance entre les frontières de propositions autonomes est très longue dans **phares1357**, à cause de la multitude de constructions complexes comportant un grand nombre de subordonnées.

Une seule occurrence sur 33 n'est pas ponctuée : il s'agit de deux propositions très courtes dans une citation latine : *Sic volo sic jubeo* (7.134). Le lien fort entre ces deux propositions ne fait aucun doute, même si elles sont entièrement indépendantes sur le plan formel. Notons que la seule occurrence de la ponctuation faible (point suivi d'une minuscule) à ce type de frontière se trouve dans la même citation (*sic jubeo . sit pro Racione voluntas*).

Les propositions ayant des éléments communs sont séparées par une ponctuation dans un peu plus de la moitié des cas (27 occurrences sur 47). Il s'agit dans tous les cas d'une ponctuation forte. La ponctuation apparaît plus souvent quand la frontière

« horizontale » entre propositions avec des éléments communs coïncide avec la frontière finale d'un groupe syntaxique de niveau inférieur (une subordonnée, une incidente, une énumération). Dans ce cas, nous avons relevé 11 occurrences ponctuées contre une sans ponctuation. En revanche, en cas de succession de propositions relativement simples avec un sujet commun, l'absence de ponctuation est majoritaire (11 occurrences contre 5).

Les remarques de l'auteur prenant la forme d'incidentes ne semblent pas présenter de marques de ponctuation spécifiques, mais il s'agit dans les trois cas relevés dans notre extrait d'ajouts de la « deuxième strate » placés dans un espace interlinéaire :

- (7.137) Qu'il || y fut plus profond que nul qui deuant lui ne apres ^{<2} ce croy je ^{2>} aye || este
(*phares1357, 3v:17*) ;
- (7.138) jl anathematiza tous || ceulx qui presumeroyent oster ou diminuer aucuns || d'iceulx liures ^{<2} ce qu'il n'eust este fait s'ilz eussent este autres que bons ^{2>} ; (*phares1357, 3r:24*) ;
- (7.139) vostre || grand pere qui tousiours ^{<2}a ^{2>} eut a+l'entour de lui les || plus experts astrologiens qu'il pouoit finer **Λ** ^{<2} de+l'un desquieulx parle gaguin en sa+Cronique ^{2>}
Et || vouloit bien oyr leurs oppinions (*phares1357, 3r:26*).

Du fait de leur position détachée, les limites de ces incidentes sont très bien marquées sans qu'on ait recours à des signes de ponctuation. Il convient de noter qu'aucune de ces incidentes ne commence par une majuscule, ce qui témoigne du lien fort que l'auteur voyait entre elles et leur contexte.

Les propositions subordonnées sont très nombreuses dans **phares1357**, leurs frontières avec les principales sont assez souvent ponctuées, même si l'absence de la ponctuation est majoritaire : sur 136 occurrences, 43 (32%) portent une ponctuation forte et 8 (6%) sont dotées d'une ponctuation faible. Il est difficile de dégager un ou plusieurs facteurs précis qui puissent expliquer cette variation. La fréquence relative des occurrences ponctuées est plus élevée pour les subordonnées circonstancielle (13 frontières ponctuées contre 11 sans ponctuation), et cette tendance semble encore plus prononcée en cas de préposition de la subordonnée (les quatre occurrences relevées dans notre extrait sont ponctuées). Bien entendu, il convient de relativiser ce résultat, étant donné le faible nombre total des occurrences.

Manuscrits du XV^e siècle

Les syntagmes coordonnés ou juxtaposés à l'intérieur de propositions sont extrêmement fréquents dans notre fragment où nous avons relevé 204 frontières ponctuables de ce type. La ponctuation est absente dans 163 occurrences (80%) ; dans 31 cas (15%) on trouve une ponctuation forte et dans 10 cas (5%), une ponctuation faible. Nous retrouvons ici la tendance à ponctuer les longues listes de personnalités importantes (même s'il s'agit en l'occurrence non pas de noms propres, mais de titres de noblesse ou de hiérarchie ecclésiastique :

(7.140) [Astrologie a este] Aprise || sceue et pratiquee \wedge ^{<2 et leue >} Par *plusieurs* sains patriarches || prophetes Papes CCardinaulx : Arceuesques E_uesques || Empereurs ! Roys . Ducz Contes ! Et ^{<2 de >} plusieurs autres || nobles hommes sages \wedge ^{<2 graues par_sonnes >} et grans docteurs philosophes || et clerics jusques a *present* (*phares*1357, 1v:16-20).

Nous voyons dans cet exemple que la ponctuation forte exprimée par une majuscule seule alterne avec des occurrences renforcées par des marques comme le point ou le deux-points. Au début de l'exemple cité, nous avons une série de quatre participes passés coordonnés (dont un a été ajouté lors de la « deuxième strate » de correction) sans aucune ponctuation. La succession de compléments du prédicat (« sujets logiques » dans cette construction passive) commence par une majuscule, le deuxième complément (*patriarches*) n'étant pas ponctué (mais il se trouve en début de ligne). Les 8 compléments suivants comportent tous une majuscule. Une deuxième liste, qui fait mention de « plusieurs autres » personnalités, moins haut placées dans la hiérarchie sociale, s'ajoute à cette première après une ponctuation forte. Aucune ponctuation n'apparaît à l'intérieur de la deuxième liste.

De simples listes de noms propres ne contiennent en général que très peu de ponctuations :

(7.141) Comme adam || noe abraham moise dauid job ! salomon saint denis || et *plusieurs* autres (*phares*1357, 4r:23-25).

En dehors des longues énumérations, la ponctuation apparaît le plus souvent en cas de coordination ou de juxtaposition de groupes syntaxiques complexes, avec parfois une imbrication de coordinations à plusieurs niveaux :

(7.142) la science de || astrologie fait a louer et a Recommander en tant de \wedge \langle si tres vaillans et saiges \rangle || empereurs et Roys dessusdicts Et mesmement ou glorieux || et saint empereur charlemaigne . (phares1357, 4r:16-19).

En ce qui concerne les frontières entre syntagmes particuliers à l'intérieur d'une proposition, la ponctuation n'y apparait pas très souvent, même si le nombre d'occurrences n'est pas négligeable : 22 au total, dont 13 ponctuations faibles. Les constructions comparatives (non propositionnelles) tendent nettement à être ponctuées : c'est le cas de 5 occurrences sur 6. Le connecteur *comme* semble d'ailleurs attirer la ponctuation forte dans ce texte, quel que soit le niveau d'unité syntaxique qu'il introduit :

(7.143) Comme jl sera || deduit cy apres plus amplement de chacun en son || temps ordre et aage et par especial trois portans |[fol. 3r] vostre nom Comme ce vertueulx victorieulx et glorieux || empereur Charlemaigne premier jnstituteur+et fondateur || de l'universite de paris . (phares1357, 2v:39-3r:3).

Les marques de ponctuation apparaissent sporadiquement (une ou deux occurrences dans chaque catégorie) avant ou après des appositions, compléments circonstanciels, petites reformulations ou groupes adjectivaux étendus.

Comme nous l'avons déjà indiqué, la syntaxe de Simon de Phares est très complexe, il faut parfois faire un effort considérable pour comprendre les relations entre les différents syntagmes. La ponctuation permet alors de marquer les mots ou les syntagmes qui avaient sans doute une importance particulière pour l'auteur. Dans le passage suivant, la ponctuation apparait devant le syntagme prépositionnel *de france* et, sur la ligne suivante, après *en france*. Le complément circonstanciel de lieu *par especial a paris* est entouré de marques de ponctuation :

(7.144) Et mesmement que les dessusdicts || Roys de france \wedge \langle et+empereurs \rangle ou temps desquelz ont este les grans clerks \langle et flory \rangle || en france profonds en toutes sciences plus+sans comparoison || que a present par especial a paris Eussent souffert en eulx || autour de eulx ne en leur Royaume des+lors appelle || tres+crestien \wedge \langle liures ne autre \rangle chose qui eust este supersticieuse mauuaise || ne contre la foy (phares1357, 4r:7-13).

Il convient de noter qu'à l'exception de quelques passages particuliers la distance entre deux ponctuation fortes dans **phares1357** varie de 10 à 25 mots : c'est le

Manuscrits du XV^e siècle

cas de près de 50% des séquences (38% sont plus courtes et 12% seulement plus longues) :

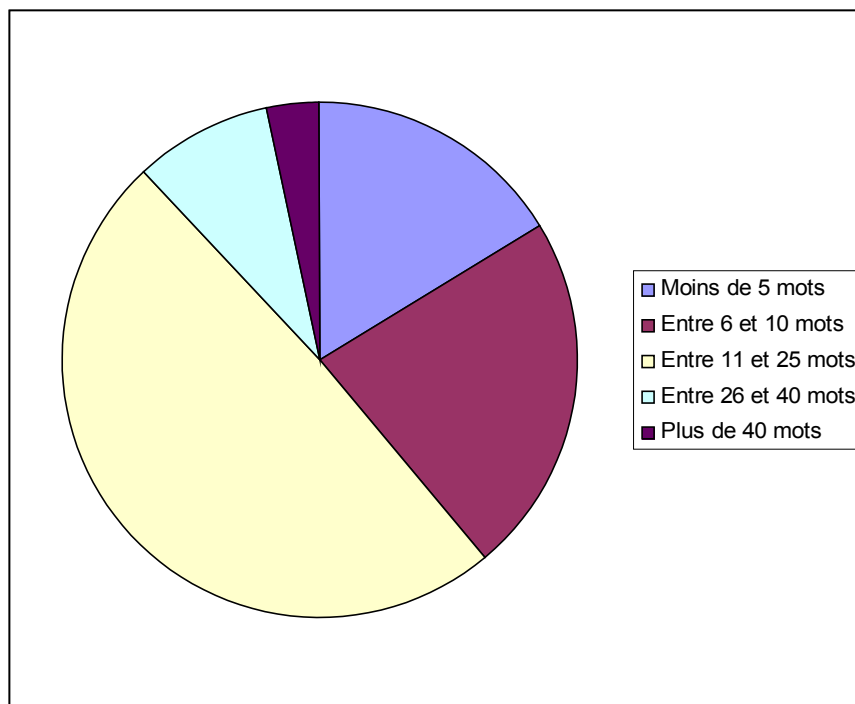


Fig. 71 : Longueur de séquences délimitées par des ponctuations fortes (phares 1357).

Il n'est donc pas exclu qu'à partir d'une certaine distance de la ponctuation forte précédente, la probabilité d'emploi d'une ponctuation à une frontière ponctuable devient plus importante quel que soit le type de frontière.

La comparaison de la ponctuation du manuscrit **phares1357** et de son édition (Boudet 1997) permet de constater que dans 133 cas l'éditeur a jugé utile de préserver la ponctuation du manuscrit. Il s'agit le plus souvent de frontières entre propositions autonomes ou avec des éléments communs (49 occurrences). L'absence de ponctuation à une frontière ponctuable a été maintenue dans 169 cas. Dans 153 cas, l'éditeur a procédé à un ajout de ponctuation (toujours faible), et il s'agit, pour le tiers des occurrences, de frontières entre syntagmes coordonnés ou juxtaposés. La ponctuation scribale a été supprimée dans 45 cas, et il est intéressant de noter que comme pour les ajouts, il s'agit le plus souvent de la coordination de syntagmes.

De façon générale, la ponctuation dans **phares1357** est assez riche et variée. Plusieurs marques différentes sont utilisées, même si leur spécialisation ne se dessine que vaguement. Les frontières de propositions autonomes (très espacées à cause de la multitude de constructions syntaxiques complexes chez Simon de Phares) sont ponctuées très régulièrement, la ponctuation est utilisée entre la plupart des propositions avec des éléments communs et apparaît assez souvent entre une subordonnée et sa principale. Les syntagmes coordonnés ou juxtaposés, très nombreux dans ce texte, sont rarement ponctués, et d'autres facteurs, comme le respect des personnalités importantes, la mise en relief d'informations importantes et la distance entre les ponctuations fortes commencent à jouer un rôle.

Chapitre 8

Livres imprimés

Les livres imprimés constituent une partie périphérique, mais importante de notre corpus. La façon dont les pratiques de ponctuation se sont progressivement stabilisées et normalisées depuis les premiers incunables en langue française jusqu'aux livres richement ponctués du milieu du XVI^e siècle est un sujet de recherche en soi. En intégrant dans notre corpus deux incunables parisiens³⁰¹ et deux livres du XVI^e, nous avons cherché à ouvrir des pistes de recherche et à tester certaines hypothèses, plutôt qu'à constituer un sous-corpus comparable à celui des manuscrits. Ceci explique la brièveté relative du présent chapitre comparé aux deux précédents.

8.1. *Lancelot du Lac*, tome 2 (Paris, Jehan du Pré, 1488)

Ce bel incunable in-folio contient la deuxième partie du cycle de *Lancelot – Graal*³⁰². La *Mort au roy artus*, dont nous avons tiré un extrait à étudier, y occupe les pages 357 à 462. La date d'achèvement de l'impression est précisée à la dernière page : le 16 septembre 1488.

301 Ces incunables sont en réalité à l'origine de notre corpus, car nous les avons transcrits pour une première fois en 1994 dans le cadre de notre D.E.A.

302 La première partie avait été imprimée la même année à Rouen, par Jehan et Gaillard Le Bourgeois.

Un facsimilé de l'exemplaire de la British Library de cet ouvrage a été publié en Grande Bretagne (Pickford 1973), et une reproduction en noir et blanc de l'exemplaire de la BnF est disponible dans la bibliothèque numérique Gallica³⁰³. C'est l'exemplaire de la BnF (Rés. Y² 47) que nous avons utilisé dans notre travail.

L'extrait que nous avons transcrit et annoté pour notre corpus se situe à la dixième page du roman³⁰⁴ et porte l'identifiant **artu1488**³⁰⁵. Il contient 595 occurrences-mots. Avec ses 28 occurrences de la ponctuation faible et 17 occurrences de la ponctuation forte, le taux de ponctuation y atteint les 7,6%. L'extrait dépouillé présente des occurrences de la barre oblique (19 occurrences, toujours suivies d'une minuscule), du deux-points (11 occurrences, toujours toujours suivies d'une minuscule), du point (12 occurrences suivies d'une majuscule et 1 suivie d'une minuscule) et de la majuscule seule (5 occurrences). On voit donc une nette tendance à la spécialisation des marques selon la force de la ponctuation.

Nous disposons par ailleurs pour cet incunable de certaines données issues d'un extrait beaucoup plus important (25 000 occurrences-mots environ), non annoté et transcrit d'une façon simplifiée dans le cadre de notre mémoire de DEA. Dans cet extrait, le taux général de la ponctuation s'élève à 6,9% (sans tenir compte des majuscules seules, qui pourraient augmenter le taux de quelques dixièmes de pour cent). Dans cet échantillon plus représentatif, le point apparaît comme la marque la plus fréquente (avec 750 occurrences), suivi de la barre oblique (537 occurrences) et du deux-points (435 occurrences). La tendance à la spécialisation des marques se confirme, même si elle semble plus nuancée ici que dans l'échantillon de 595 mots. Le point est utilisé dans la ponctuation faible dans 3,6% des occurrences (soit 27 sur 750), tandis que le deux-points apparaît dans la ponctuation forte dans 8,7% des occurrences (38 sur 435). En revanche, l'usage de la barre oblique semble réservé exclusivement à la ponctuation faible.

303 Identifiant Gallica : <<http://gallica.bnf.fr/ark:/12148/bpt6k111060r>>.

304 Les pages de cet incunable ne sont pas numérotées. Nous utiliserons pour nos références les numéros d'images correspondantes dans la version accessible sur Gallica. La 10^e page de la *Mort au roy Artus* correspond à la page 366 de la version numérisée de l'ensemble du volume.

305 Cf. le sigle DEAF **MortArtu**.

Livres imprimés

Notre grand extrait permet par ailleurs de constater des phénomènes de basses fréquences invisibles sur le petit échantillon. Il s'agit par exemple de l'emploi du comma « droit », dont nous avons relevé 3 occurrences :

- (8.1) ie vueil fist || il et commande que vous y alliez et que par||ties le matin et ie demouray ¶ mais deuant *que* vous reueniez ie seray tout guery se dieu plaist (*artu1488, 359a:7*)³⁰⁶ ;
- (8.2) Je vueil dist le roy que pour la grant desloy||aulte qu'elle a faicte enuers moy ¶ qu'on en fa/||ce si grant iustice *que* toutes dames qui en or/||ront parler en soient chastiees (*artu1488, 404a:2*) ;
- (8.3) mais Sagremor estoit si na/||ure parmi le corpus ¶ *que* a peine se pouoit il soub||stenir en selle (*artu1488, 454b:7*).

Toutes ces occurrences (dont deux se trouvent à l'intérieur du discours direct) surviennent dans des contextes très similaires à ceux où on pouvait rencontrer le *comma* dans les manuscrits du XIII^e et du XIV^e siècle (*cf.* par exemple **qgraal77**). Ce qui frappe ici, c'est l'extrême rareté de cette marque. Nous n'avons trouvé aucune condition linguistique ou textuelle qui explique l'emploi du comma dans ces occurrences et son absence ailleurs. La seule hypothèse que nous puissions proposer est que ce caractère s'était glissé dans la casse de l'imprimeur et mélangé avec un autre caractère de ponctuation, sans doute la barre oblique. Le compositeur ne faisait pas de différence entre la barre oblique et le *comma*, il utilisait donc ce dernier caractère quand il lui « tombait sous la main » à la place de la barre oblique.

Les conditions matérielles de l'imprimerie de l'époque, voire des erreurs de composition peuvent expliquer également les 2 occurrences du « point haut » que nous avons relevées dans l'extrait élargi. L'image présentée ci-dessous permet de comparer visuellement un point « haut » (à gauche) à un point « normal » (à droite) :

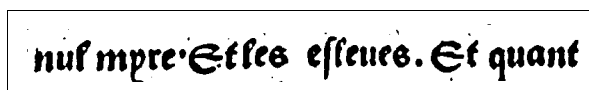


Fig. 72 : Point « haut » (*artu1488, l.4516, 4476*).

On peut supposer que le caractère a simplement été retourné ou mal aligné, comme cela arrivait parfois à des lettres (*cf. m* dans *m'avez* ci-dessous) :

³⁰⁶ Les pages de cet incunable ne sont pas numérotées. Nous utiliserons pour nos références les numéros d'images correspondantes dans la version accessible sur Gallica.

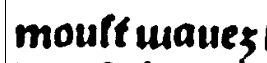


Fig. 73 : Caractère < m > retourné (*artu1488, l.317*).

Il convient de souligner que les phénomènes introuvables dans l'extrait restreint sont rares et n'occupent qu'une place très périphérique dans l'ensemble du système graphique. Les tendances générales qu'on peut essayer de dégager à partir d'un échantillon restreint se confirment sur l'extrait élargi, même si certaines fréquences relatives peuvent varier.

L'extrait dépouillé contient un nombre important de passages au discours direct qui sont apparemment ponctués d'une façon très cohérente. Le début et la fin d'un discours direct, ainsi que le changement de locuteur sont toujours marqués par une ponctuation forte (12 occurrences au total). Les incises, en revanche, ne sont jamais ponctuées (19 occurrences). On voit donc ici l'aboutissement logique des tendances déjà perceptibles dans les manuscrits (comme, par exemple, **qgraal77**). Un échantillon plus important sera bien entendu nécessaire pour vérifier si cette régularité se confirme sur l'ensemble du texte.

On constate par ailleurs que dans 5 cas la ponctuation utilisée pour le marquage du discours direct est la majuscule seule. Il s'agit d'ailleurs de la totalité des occurrences de ce type de marque :

- (8.4) *et nompourtant || se vous me aymiez bien par amours il m'en || peseroit plus pour vous que pour autre chose || Pour moy dist il damoiselle pourquoi vous || en peseroit il .*
(*artu1488, 366a: 39*) ;
- (8.5) *Et comment a il a nom dist messire gau||uain Sire dist elle son nom ne vous diray || ie pas /* (*artu1488, 366b: 24*).

Dans 4 occurrences sur 5 l'absence d'un signe de ponctuation devant une majuscule coïncide avec un saut de ligne (8.4). On peut donc l'expliquer par la justification des fins de lignes, impérative dans les imprimés. Mais il existe aussi des cas où la majuscule seule est employée au milieu d'une ligne (8.5). Un sondage rapide effectué sur près de la moitié de l'échantillon élargi (12 800 occurrences-mots) montre que le facteur de la justification en fin de ligne joue certainement un rôle important : 27

Livres imprimés

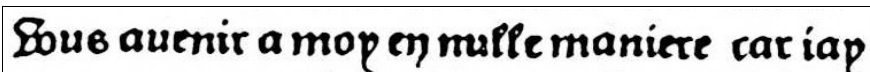
des 35 occurrences relevées de majuscule seule se trouvent dans cette position. Le facteur du discours direct paraît moins déterminant, car il intervient dans 19 occurrences (54%) seulement.

On peut constater que la fréquence des occurrences de ce type augmente considérablement dans certaines colonnes (dont la colonne 466b intégrée dans notre échantillon restreint), alors que de façon générale l'usage de la majuscule seule en tant que marque de ponctuation est extrêmement rare.

Les limites de propositions autonomes sont ponctuées presque régulièrement : une seule occurrence (sur 18) échappe à cette règle :

- (8.6) si ne pourriez || vous auenir a moy en nulle maniere **car** i'ay||me par amours vng cheualier (*artu1488, 366a:42*).

En regardant attentivement l'original, on s'aperçoit que l'espace blanc devant *car* est plus long qu'entre les autres mots de la même ligne :



Vous auenir a moy en nulle maniere car iay

Fig. 74 : Espace blanc allongé (*artu1488, l.828*).

On peut supposer qu'une marque de ponctuation avait été prévue à cet endroit et qu'elle a été effacée, ou bien qu'elle n'a pu être imprimée suite à un problème matériel.

La ponctuation faible est plus fréquente dans cette position que la ponctuation forte. Même si nous disposons de trop peu d'occurrences pour pouvoir rechercher des corrélations entre la force de la ponctuation et les différents facteurs qui jouent à ce niveau de frontière syntaxique, nous pouvons néanmoins constater que toutes les occurrences de la ponctuation forte sont suivies de la conjonction *et*. Dans 4 cas sur 5, cette conjonction est suivie d'un connecteur temporel ou causal (*quant, lors, pour ce*) qui marque une frontière importante dans la narration ou dans l'argumentation :

- (8.7) Gaheriet tira le seigneur || son hoste a part : affin que messire gauvain par||last plus priueement a la damoiselle s'il vou||loit **¶** Et quant messire gauvain se veit en si bon || point de parler a la damoiselle : il la requist d'a||mours . (*artu1488, 366a:23*) ;
- (8.8) car se dieu m'aist il n'est pas mains bon che||ualier que vous ne mains beau ne mains pri||se d'armes **¶** Et pour ce vous dis+ie que ce se||roit peine perdue de moy prier d'amours / (*artu1488, 366b:5*).

Dans les deux exemples cités la ponctuation forte marque une limite entre deux constructions syntaxiques (on dirait ‘phrases’ s’il s’agissait du français moderne) complexes impliquant des propositions subordonnées et principales.

Dans l’exemple suivant (qui se croise partiellement avec (8.7)), les deux propositions concernées contiennent du discours rapporté avec un changement de locuteur entre les deux :

(8.9) il la requist d'a||mours ¶ Et elle lui demanda qui il estoit (*artu1488, 366a:25*).

On peut supposer que la ponctuation a été influencée ici par le cas similaire du changement de locuteur dans un discours direct.

Parmi les occurrences de la ponctuation faible, nous avons 5 cas d’usage de la conjonction *et* (dans des constructions apparemment moins complexes que celles où on trouve des majuscules), 5 emplois de *car*, une occurrence de la conjonction *mais* et une occurrence sans connecteur lexical :

(8.10) Comment sire dist elle cuidez vous que ie le || feisse / ie pourroie bien faire mourir les ii. plus || preudhommes du monde ¶ ou l'un d'eulx / (*artu1488, 366b:15*).

Nous avons ici affaire à un discours direct contenant une question « rhétorique » suivie d’une hypothèse sur les conséquences possibles de l’action imaginée. Les deux propositions sont prononcées par le même locuteur et entretiennent des relations logiques très fortes, ce qui a sans doute empêché l’apparition d’une ponctuation forte. Le besoin de démarquer la séquence interrogative a en revanche motivé l’apparition d’une barre oblique.

Notons que l’extrait dépouillé contient 3 autres occurrences d’énoncés interrogatifs, dont deux précèdent immédiatement un changement de locuteur (qui implique la ponctuation forte) et une est suivie d’une incise, puis d’une autre interrogation :

(8.11) il m'en || peseroit plus pour vous que pour autre chose || Pour moy ¶ dist il damoiselle pourquoi vous || en peseroit il ¶ (*artu1488, 366a:39*).

Livres imprimés

L'absence de ponctuation après la première séquence interrogative peut s'expliquer par le fait que la modalité ne change pas ensuite et que les incises ne sont jamais séparées du discours direct auquel elles se rattachent.

De toute évidence, l'imprimeur de **artu1488** ne connaissait pas l'usage d'un signe de ponctuation spécial réservé à l'interrogation (aucune occurrence même dans l'extrait élargi), tandis qu'une telle marque existait dans les manuscrits au XIII^e siècle déjà, même si son usage n'a jamais été systématique.

Revenons à présent à la ponctuation entre propositions autonomes. Les 5 occurrences de *car* sont toutes précédées d'une barre oblique. Nous avons pourtant vu que dans beaucoup de manuscrits ce connecteur était associé plutôt à la ponctuation forte. L'emploi de la ponctuation faible dans **artu1488** peut être une trace du changement de la relation syntaxique exprimée par ce connecteur. Si dans les manuscrits où *car* est associé à la ponctuation forte on peut parler du marquage d'une frontière logique importante et de l'autonomie des deux propositions (relation de parataxe), la ponctuation faible peut servir à souligner le lien de causalité, et la relation syntaxique peut se rapprocher de l'hypotaxe.

Les données de l'extrait élargi permettent de faire une seconde observation intéressante concernant l'usage de la ponctuation en lien avec la conjonction *car*. 31 des 39 occurrences de cette conjonction sont précédées du deux-points sur les trois premières pages du roman. Sur les trois pages suivantes, c'est la barre oblique qui domine avec 32 occurrences sur 54. Il s'agit donc d'un changement net dans l'usage de la ponctuation dans des conditions syntaxiques qui restent stables.

Les raisons de ce changement sont sans doute extralinguistiques : changement de compositeur ou manque de caractères dans une caisse.

Les propositions ayant un élément commun sont fort peu nombreuses dans l'extrait dépouillé : 8 occurrences au total, dont 5 avec une ponctuation faible et 3 sans ponctuation. Les cas d'absence de ponctuation concernent la coordination de propositions courtes à sujet commun (8.12), tandis que la ponctuation sépare des propositions dotées de circonstants ou d'autres éléments qui les rendent plus complexes (8.13) :

- (8.12) que vous le laissez *et* me prenez (*artu1488, 366b:13*) ;
 (8.13) messire gauvain le fist asseoir || empres lui au coste dextre *et* a senestre fist as||seoir la damoiselle (*artu1488, 366a:16*).

Dans l'exemple (8.13) (qui suit immédiatement (8.12) dans le texte), une ponctuation apparaît entre deux groupes, structures syntaxiques que nous avons interprétées comme des propositions subordonnées coordonnées avec un prédicat (*estoit assis*) commun. Le fait qu'elles aient des actants en partie différents justifie, à notre avis, cette analyse. Chacune de ces propositions contient une coordination de compléments du prédicat avec la conjonction *et*. L'usage d'une barre oblique devant la même conjonction coordonnant les propositions permet de mettre en évidence la hiérarchie des relations syntaxiques et la structure logique de la phrase :

- (8.14) si qu'elle estoit assise entre || lui *et* mordrec *et* l'oste entre gaheriet *et* messire || gauvain. (*artu1488, 366a:18*).

La ponctuation entre les propositions subordonnées et principales est variable. La ponctuation est absente dans près de trois quarts des cas (35 occurrences sur 47), et une ponctuation faible est utilisée ailleurs. Dans 9 cas sur 12, la marque utilisée est le deux-points. Il est difficile de dire si l'explication de ce fait est purement extralinguistique (comme le passage du deux-points à la barre oblique devant *car*) ou si on a affaire ici à une spécialisation de ce signe dans le marquage des relations d'hypotaxe. Dans toutes les sous-catégories de la subordination, l'absence de ponctuation domine nettement. La seule exception est la frontière entre une subordonnée circonstancielle et sa principale postposée, où le nombre d'occurrences avec ou sans ponctuation est égal (*cf.* deux constructions très similaires qui se succèdent dans l'exemple suivant) :

- (8.15) Et quant il *vint* la *il* trouua messire gauvain || *et* ses compagnons qui la se deduisoient . Et || quant ilz le veirent venir *ilz* se leuerent au de||uant de lui / (*artu1488, 366a:14*).

Les syntagmes coordonnés à l'intérieur de propositions sont rares dans l'extrait dépouillé. Une seule occurrence sur 7 est ponctuée. Il s'agit de celle qui apparaît dans l'exemple (8.10) ci-dessus (devant *ou l'un d'eulx*). On voit qu'il ne s'agit pas d'une

Livres imprimés

simple coordination avec la conjonction *et*, mais d'une précision dans le cadre d'un discours argumentatif. On peut penser que la ponctuation correspond ici à une pause rhétorique dans le discours oral du personnage.

L'extrait analysé ne contient aucune occurrence de ponctuation servant à démarquer des syntagmes particuliers à l'intérieur d'une proposition.

L'incunable **artu1488** présente un système de ponctuation novateur à plusieurs égards par rapport aux pratiques que nous avons constatées dans les manuscrits étudiés jusqu'à présent. L'usage régulier de trois marques distinctes avec une nette spécialisation dans la force de la ponctuation, la régularité de la ponctuation dans le discours direct et l'absence de ponctuation à l'intérieur des propositions sont des traits caractéristiques qui se stabiliseront dans l'avenir pour former le système des règles modernes. Par d'autres aspects, la ponctuation de **artu1488a** reste cependant proche de celle des manuscrits. Il s'agit notamment de l'absence de norme soigneusement respectée et de la possibilité de variation à presque tous les niveaux. Les conditions matérielles de l'imprimerie constituent un nouveau facteur, dont il convient de tenir compte dans l'analyse de la ponctuation.

8.2. *Image du monde* (Paris, Jean Trepperel, entre 1492 et 1498)

Cette édition de *l'Image du monde* fait partie des premières réalisations de l'imprimeur-libraire parisien Jean Trepperel³⁰⁷ qui a exercé son activité de 1492 jusqu'à sa mort en 1511 (Renouard 1965). La date précise de l'impression n'est pas indiquée, mais elle est vraisemblablement antérieure à 1498. Nous avons par commodité choisi la date moyenne de 1495 pour forger l'identifiant de ce texte dans notre corpus (**im1495a**). Cet imprimeur « était spécialisé » dans la production d'ouvrages « populaires » (calendriers, récits d'événements politiques, mais aussi des romans de chevalerie, mystères, facéties) (Öhlund-Rambaud 2006, p. 124).

³⁰⁷ Dans l'édition le nom de l'imprimeur apparaît ainsi : « Jehan treperl librayre demourant sur le pont nostre dame ». La notation du nom est peut-être abrégée.

Nous avons travaillé sur l'exemplaire conservé à la Bibliothèque nationale de France (Rés. D. 4722). L'extrait que nous avons transcrit en mode multi-facettes et dont nous avons étiqueté les « frontières ponctuables » se situe aux folios 2r et 2v et comporte 576 occurrences-mots.

Nous disposons par ailleurs pour ce texte de données supplémentaires issues d'une transcription simplifiée sans étiquetage linguistique effectuée dans le cadre de notre DEA. Cette transcription est composée de deux extraits pris au début et au milieu du texte et comporte au total environ 10750 occurrences-mots.

Le taux général de ponctuation n'atteint que 4,3% dans l'extrait annoté, tandis qu'il monte à 5,0% en moyenne dans la transcription élargie.

Le point est la marque de ponctuation largement dominante dans ce texte avec 459 occurrences dans l'échantillon élargi (contre une centaine d'occurrences de la majuscule seule, 33 occurrences de la barre oblique et 10 occurrences du deux-points). Il convient de noter que 9 sur 10 occurrences du deux-points se trouvent dans le deuxième extrait de l'échantillon élargi. Ce changement de fréquence d'une marque (qui reste de toute façon basse) ne peut sans doute être expliqué que par des conditions matérielles.

La ponctuation faible est nettement plus fréquente que la ponctuation forte (366 occurrences contre 237, soit 60%). La barre oblique est utilisée uniquement dans la ponctuation faible, la majuscule étant pour nous par définition une ponctuation forte. En ce qui concerne le point, il est suivi d'une majuscule dans près de 29% des cas (soit 132 occurrences sur 459), tandis que la barre oblique l'est dans 4 occurrences sur 10.

L'extrait dépouillé ne contient aucun passage au discours direct. La frontière syntaxique la plus ponctuée est sans surprise celle de propositions autonomes. On y trouve une marque de ponctuation dans un peu plus de 50% des cas (soit 20 occurrences sur 38). Dans cette position, la ponctuation forte est utilisée dans 9 occurrences sur 20.

On peut constater que la majorité des occurrences ponctuées (14 sur 21 au total et 9 sur 9 pour la ponctuation forte) concernent les frontières accompagnées d'un changement de sujet nominal. En revanche, les propositions avec un sujet pronominal ne sont ponctuées que dans 6 cas sur 22.

Livres imprimés

La ponctuation faible peut être employée même si la division logique entre les deux propositions semble importante, comme c'est le cas devant *dieu* dans l'exemple suivant :

- (8.16) Car de tant que le mal est plus || noir *et*+obscur de tant a il plus de douleur
et+amurtume || en *en*fer avec les dyables . dieu peult tout faire *et*+tout de||faire *sans* en
riens muer en soy en quelque maniere que || ce soit : (*im1495a*, 2v:18).

Le manuscrit **im574** utilise dans ce même contexte une ponctuation extra-forte (point suivi d'une initiale en couleur).

Les propositions possédant des éléments communs sont relativement nombreuses dans l'extrait dépouillé, et la tendance est nettement à l'absence de ponctuation (14 occurrences sur 16). Les deux cas où la ponctuation est présente entre des propositions à un sujet commun surviennent dans un long passage qui décrit les différents aspects de la puissance de Dieu :

- (8.17) car dieu fut est *et* sera tousiours en || vng point ne ia pour chose qu'il fist oncques le
monde ilz || n'en valut ne pis ne mieulx . Car oncques n'eust necessari||te de chose qui
soit il voit tout il tient tout en sa main || il n'eust oncques fain ne soif temps ne ans iour
ne heure . || tousiours demeure en vng estat il scet toutes les choses || qui furent
oncques qui sont et qui seront . (*im1495a*, 2r:16-23) ;

La première occurrence (forte, qui plus est) peut s'expliquer par la présence de la conjonction *car* qui est connue pour sa capacité à « attirer » la ponctuation. En effet, dans l'échantillon élargi près de 70% des occurrences de cette conjonction (65 sur 91) sont précédées d'une ponctuation. Il est toutefois intéressant de noter que toutes les occurrences de la ponctuation forte sont situées dans le premier extrait. On peut supposer que cette variation dans la pratique de la ponctuation est due au changement de compositeur.

La proposition qui précède la deuxième occurrence contient trois paires de compléments coordonnées (*fain ne soif*, *temps ne ans* et *iour ne heure*), et l'emploi d'une marque de ponctuation permet de marquer la limite de cette énumération qui coïncide avec la fin d'une proposition.

Notons par ailleurs qu'aucune ponctuation n'est utilisée dans ce même passage devant les propositions formellement autonomes qui commencent par un sujet pronominal *il*.

Les limites entre les propositions principales et subordonnées ne sont ponctuées que dans deux cas sur 36, ces deux occurrences se succédant à trois lignes d'intervalle dans le passage suivant (il s'agit du point devant *aussi* à la ligne 23) :

(8.18) il scet toutes les choses || qui furent oncques qui sont et qui seront . aussi bien loinh || comme pres . Et aussi bien veoit il le monde auant qu'il || fust fait comme il fait maintenant si n'eust oncques fait || le monde aussi bien fust il comme il est . et aussi bon ou || autrement ne seroit il pas dieu s'il ne scauoit et veoit . || tout ce qui a este est et sera . (*im1495a, 2r:22-28*).

Dans le premier cas, une ponctuation faible (un point entre *qui seront* et *aussi bien*) est employée après une série de trois relatives « insérées » entre l'objet direct et un complément circonstanciel de la principale. La proposition principale peut dans ce cas être perçue comme complète sans le complément circonstanciel. Il se peut que le compositeur ait hésité en analysant la structure de cette construction (à savoir à quelle proposition se rattache le groupe *aussi bien loinh comme de pres*), d'autant plus que la proposition suivante commence par une séquence très similaire : *Et aussi bien veoit il...*

Dans le deuxième cas, le point entre *il est* et *aussi bon* est utilisé à la fin d'une très courte comparative insérée dans une proposition principale devant un deuxième attribut du sujet rajouté « après coup » à une proposition dans laquelle l'attribut est placé devant le verbe et le sujet (modèle XVS). Il s'agit là d'une certaine rupture dans une structure syntaxique, que la ponctuation sert à mettre en évidence.

Une autre hypothèse qui peut expliquer les deux occurrences de ponctuation non situées entre propositions autonomes est l'usage répétitif du connecteur *aussi* : *aussi bien... Et aussi bien... et aussi bon*, qui a pu être perçu comme un repère dans la structuration du texte indépendamment des relations syntaxiques entre les unités dont il fait partie.

Il convient de noter que l'exemple (8.18) contient également la seule occurrence dans l'extrait dépouillé d'une ponctuation placée en dehors des frontières « ponctuables » habituelles relevée. Il s'agit du point entre le verbe *veoit* et son complément d'objet direct *tout*. Située en fin de ligne, cette occurrence peut tout simplement être liée à la justification typographique. D'autres occurrences du même type présentes à l'extérieur de l'extrait dépouillé semblent confirmer cette hypothèse (7r:4, 7v:5, 7v:25, etc.). Il se peut également que le compositeur ait considéré

Livres imprimés

l'ensemble du groupe *tout ce qui* comme faisant partie d'une proposition subordonnée, ce qui a favorisé l'emploi du « point de justification ».

La ponctuation entre syntagmes coordonnés est plutôt rare dans **im1495a** : 3 occurrences ponctuées seulement sur 38 frontières de ce type. Deux de ces occurrences se succèdent dans une énumération des joies du paradis :

- (8.19) Cent mille ne luy montent mie a cent mille partie d'u||ne heure ne a tous ceulx qui avec luy *sont* dont le pire *qui* || demeure avec luy a+plus de bien dedens vne heure ioye || bonneur / soulas / *et*+deduit que nul homme ne pourroit || dire en cent mille ans (*im1495a*, 2v:25).

Il ne s'agit pas d'un exemple isolé, car des occurrences similaires (avec des barres obliques dans des énumérations) ont pu être trouvées ailleurs dans le texte, en dehors de l'extrait dépouillé (*cf.* notamment 18v:19 et 23v:16), mais cette pratique n'est pas systématique, car tous les éléments énumérés ne sont pas ponctués et on trouve également des énumérations sans aucune ponctuation :

- (8.20) le mal *descend* en bas en lieu moult horrible *et* tenebreux || plain de toutes paines de douleurs *et* de tourmens (*im1495a*, 2v:12-13).

La seule occurrence de la ponctuation entre syntagmes particuliers a été analysée ci-dessus dans le cadre de l'exemple (8.18).

L'analyse de la ponctuation dans **im1495a** nous mène à la conclusion que cet incunable est plus proche dans ses pratiques de la ponctuation des manuscrits « peu soignés » du XV^e siècle que de celles des premiers livres (en latin) de l'atelier de la Sorbonne et de **artu1488**. Une seule marque de ponctuation (le point) est nettement majoritaire, et elle est utilisée pour la ponctuation faible aussi bien que pour la ponctuation forte. La ponctuation est rare en dehors des frontières de propositions indépendantes, et même dans cette position la moitié seulement des occurrences sont ponctuées.

8.3. *Image du monde* (Paris, Jean Janot, vers 1522)

Cette nouvelle édition de l'*Image du monde* se situe en principe en dehors de la période choisie pour notre étude de la ponctuation médiévale. Nous avons néanmoins jugé utile de l'ajouter à notre corpus, parce qu'elle permet d'observer l'évolution des pratiques en comparant un même texte imprimé par des libraires-imprimeurs appartenant à une même famille (Jean Janot était en effet le gendre de Jean Trepperel).

La date précise de la parution de cette nouvelle édition est inconnue, mais des éléments extérieurs permettent de la situer dans les années 1521-1522 : soit avant, soit peu après la mort de Jean Janot, dont l'activité a été reprise par sa veuve.

Notre transcription a été effectuée à partir de la reproduction disponible sur Gallica³⁰⁸. Il s'agit d'une numérisation d'un microfilm de la collection *French books before 1601* (British Library 1990) réalisé à partir de l'exemplaire conservé à la Public Library de New York (*KB 152- (*Image du monde*. (L)*Image du monde*))³⁰⁹. Le volume est composé de 58 pages et le texte (à l'exception du prologue) est disposé en 2 colonnes. Les pages ne sont pas numérotées, et nous utilisons pour nos références la numérotation automatique proposée par Gallica.

L'extrait dépouillé correspond exactement au passage annoté dans **im1495a**, ce qui nous a permis de mettre en évidence les changements dans l'usage de la ponctuation. Dans notre corpus, cet extrait porte l'identifiant **im1522a**.

Le premier constat est que le taux de ponctuation monte nettement dans cette nouvelle édition : 6,6% (contre 4,3% sur le même passage dans **im1495a**), et cette augmentation concerne surtout la ponctuation faible (27 occurrences dans **im1522a** contre 16 dans **im1495**), tandis que la ponctuation forte progresse plus légèrement (11 occurrences contre 9).

Les fréquences relatives des marques sont plus équilibrées et une tendance à l'usage spécialisé se dessine. La marque de ponctuation dominante dans ce texte est la barre oblique, mais cette dominance est moins marquée que celle du point dans

308 L'identifiant Gallica 2 : <<http://gallica.bnf.fr/ark:/12148/bpt6k53560z>>.

309 Il convient de noter que l'information concernant l'exemplaire à l'origine du microfilm n'est pas facilement accessible ni sur Gallica ni dans le catalogue de la collection des microfilms.

- une ponctuation forte ou faible est maintenue dans 10 cas ;
- l'absence de la ponctuation est maintenue dans 6 cas ;
- une ponctuation faible est « renforcée » dans 2 cas ;
- une ponctuation forte est « affaiblie » dans 3 cas ;
- une ponctuation forte est ajoutée dans 2 cas ;
- une ponctuation faible est supprimée dans 1 cas.

On peut constater tout d'abord une certaine stabilité : la présence ou l'absence d'une marque est maintenue dans 16 cas (ou dans 21, si on compte aussi les cas où la marque est maintenue mais change de force).

Pour représenter des exemples des changements de ponctuation, nous utiliserons la même notation que celle employée dans la comparaison des manuscrits aux éditions modernes : les deux marques³¹⁰ sont placées entre crochets (d'abord celle de **im1495a**, puis celle de **im1522a**) et séparées par le symbole >. Les formes graphiques sont celles de **im1522a**.

L'ajout d'une ponctuation faible est le changement le plus fréquent :

(8.21) car oncques n'eut || necessite de chose qui y soit [**>/**] il voyt || tout il tient tout en sa main [**> ;**] il n'eut || oncques fain ne soif (*im1522a, 8a:13-16*).

Ici deux des trois propositions courtes avec un sujet pronominal *il* [Dieu] qui n'étaient pas ponctuées dans **im1495a** prennent une marque de ponctuation faible dans **im1522a**. La proposition qui se trouve au milieu de cette séquence (*il tient tout...*) reste toutefois non ponctuée.

Il est intéressant de noter que les propositions introduites par la conjonction *car* sont concernées à la fois par le renforcement et par l'affaiblissement de la ponctuation :

(8.22) QUant dieu fist || le monde premie||rem^{ent} il ne luy || estoit point de || mestier [**. car >** / **Car**] || autant vault || il apres comme || deuant [**.> ;**] car dieu fut est *et* sera tous||iours en vng point : ne ia pour chose que il fist oncques le monde il n'en vault || ne pis ne mieulx [**. Car > / car**] oncques n'eut || necessite de chose qui y soit (*im1522a, 8a:3-14*).

Dans cet exemple (situé dans le texte juste avant (8.21)) la répétition du *car* initial crée une figure rhétorique d'anaphore, et il semble que son effet serait renforcé

310 On ajoute éventuellement le premier mot de l'unité suivante en cas de changement de force (remplacement d'une majuscule par une minuscule ou inversement).

Livres imprimés

encore si les trois propositions étaient ponctuées de la même façon. Or ce n'est le cas ni dans **im1495a**, ni dans **im1522a**. Dans **im1495a**, l'usage de la ponctuation paraît plus « mécanique » que « logique » : après deux ponctuations faibles vient une ponctuation forte. La ponctuation forte marque le début de la figure, deux ponctuations faibles viennent après. Parallèlement, les signes de ponctuation alternent : une barre oblique, puis un deux-points, puis de nouveau une barre oblique.

Les limites de propositions avec des éléments communs sont ponctuées dans 4 cas sur 16 dans **im1522a** (contre 2 occurrences ponctuées seulement dans cette position dans **im1495a**). Le nombre d'occurrences dont nous disposons est trop faible pour juger si l'augmentation de la fréquence de la ponctuation est significative de ce type de frontière.

Les deux occurrences où la ponctuation a été ajoutée présentent toutes les deux une particularité intéressante : dans **im1495** une marque de ponctuation dans une position « atypique » est employée dans le même contexte, à quelques mots de distance (cf. l'analyse de l'exemple (8.18) dans la section précédente) :

- (8.23) il scet toutes les cho||ses *qui* furent oncques qui sont [> :] et *qui* || seront [. >]aussi bien loing *comm_*e pres . (*im1522a*, 8a:19) ;
- (8.24) ou aultrement ne || seroit il pas dieu s'il ne scauoyt [> /]et || voit [. >] tout ce *qui* a este est *et* sera (*im1522a*, 8a:26).

En lisant ces exemples, on a l'impression que les marques de ponctuation « aberrantes » de **im1495a** ont été déplacées vers une frontière plus « convenable » dans **im1522a**. On peut alors supposer que le compositeur de 1522 travaillait à partir de la version de 1495, en corrigeant ses erreurs typographiques apparentes. Bien entendu, une analyse approfondie d'échantillons plus importants (voire de la totalité des textes) serait nécessaire pour pouvoir confirmer cette hypothèse.

La frontière entre une proposition principale et sa subordonnée est ponctuée dans un seul cas (sur 33) :

- (8.25) Cent mille ne luy || montent mie a cent mille+partie de || vne heure ne a *tous* ceulx qui auec||ques || luy sont [> :] dont le pere³¹¹ *qui* demoure ||[p. 9] auecques luy a plus de

311 *pire* dans **im1495a**. Il s'agit vraisemblablement d'une erreur dans **im1522a**, car *pire* est plus logique dans ce contexte. Les leçons que donnent **im574** et **im587** sont *li mendres* et *le plus petit*

bien dedans || vne heure ioye honneur [/ >] soulas [/ >] et || desduyt que nul homme ne pourroit || dire en cent mille ans (*im1522a, 8b:36-9a:4*).

Il s'agit d'une relative en *dont* se trouvant au milieu d'une « cascade » de subordonnées qui peut être représentée par le schéma suivant :

[P [R1 : [R2 [R3] ... [Cpr]]]]

où P est la principale (*Cent mille... tous ceulx*), R1 la première relative (*qui avecques luy sont*), R2 la deuxième relative précédée d'une ponctuation (*dont le p[ri]re... a plus de... desduyt*), R3 la troisième relative insérée dans R2 (*qui demeure*) et Cpr une comparative dépendant de l'adverbe *plus* (*que nul homme...*). Notons que la relative R2 est séparée de son antécédent par le prédicat, ce qui produit un enchaînement syntaxique un peu inhabituel et peut justifier l'emploi d'une ponctuation. Une autre hypothèse (qui n'est pas forcément incompatible avec la première) est que les marques de ponctuation qui étaient utilisées dans le même passage entre des termes coordonnés (*ioye honneur / soulas / et desduyt*) dans **im1495a** ont été « déplacées » à une frontière de niveau supérieur, comme c'était sans doute le cas dans les exemples analysés précédemment.

Les syntagmes coordonnés à l'intérieur d'une proposition sont rarement ponctués. Sur 40 frontières de ce type seules trois portent une marque de ponctuation (toujours faible). Il s'agit chaque fois d'une barre oblique séparant deux paires de termes coordonnés avec *et* ou *ne* :

(8.26) il n'eut || oncques fain ne soif / temps ne ans / || iour ne heure (*im1522a, 8a:16*) ;

(8.27) Le mal est lait *et* obscur / noir *et* hor||rible (*im1522a, 8b:11*).

La ponctuation permet dans ce cas de représenter la hiérarchie des relations entre les éléments coordonnés. Dans **im1495**, aucune ponctuation n'est utilisée dans ces passages.

Aucune ponctuation aux frontières de syntagmes particuliers n'a été relevée dans l'extrait étudié.

respectivement.

De façon générale, la ponctuation dans **im1522a** est assez proche de celle de la plupart des manuscrits du XV^e siècle. Dans certains de ses aspects (comme la dominance d'une seule marque utilisée à la fois pour la ponctuation faible et forte) elle est plus conservatrice que celle de **artu1488**. Cependant, par rapport à l'édition du même texte réalisée un quart de siècle auparavant par un imprimeur de la même famille (im1495) elle présente une nette évolution : les marques sont plus variées et plus « spécialisées », les limites de propositions autonomes sont plus régulièrement ponctuées, les marques de ponctuation à des frontières « atypiques » disparaissent.

8.4. La manière de bien traduire d'une langue en aultre (Lyon, Estienne Dolet, 1540)

Notre analyse des pratiques de la ponctuation dans les sources primaires se termine par l'étude le célèbre ouvrage d'Étienne Dolet contenant notamment son traité sur la ponctuation présenté dans le chapitre 1 (section 1.1.3.). Même s'il reprend certains éléments déjà évoqués dans le *Dialogue* de Jehan Heynlin (Lapide) de 1471, le traité de Dolet marque une rupture radicale avec les pratiques de la ponctuation « courante » dans les manuscrits médiévaux français.

Il nous a paru intéressant d'appliquer au texte de Dolet la même méthode d'analyse de l'emploi de la ponctuation qu'à notre corpus de transcriptions de manuscrits et d'incunables, afin de vérifier si la rupture dans la pratique est aussi importante qu'elle le paraît en théorie. Nous avons cependant choisi un extrait du traité sur l'art de la traduction (et non du traité sur la ponctuation), afin d'éviter le biais lié aux emplois métalinguistiques des marques de ponctuation.

L'extrait se situe aux pages 12 à 14 du livre et comprend 613 occurrences-mots. Dans notre corpus il porte l'identifiant **dolet1540**. L'extrait a été transcrit à partir du facsimilé numérisé accessible sur Gallica³¹².

Le taux de ponctuation dans l'extrait dépouillé atteint les 17,8% et dépasse ainsi non seulement le taux le plus élevé enregistré dans les manuscrits et incunables de notre

312 L'identifiant Gallica : <<http://gallica.bnf.fr/ark:/12148/bpt6k106073c>>.

corpus (**im574** avec ses 13,3%), mais aussi le taux moyen de la ponctuation modernisée utilisée dans les éditions critiques.

Le passage dépouillé présente l'usage des marques de ponctuation suivantes (par ordre de fréquence) :

- virgule (« incisum » ou « point à queue ») < , > – 66 occurrences, toujours ponctuation faible ;
- point (« colon » ou « point rond ») < . > – 18 occurrences, toujours ponctuation forte ;
- deux-points (« comma ») < : > – 12 occurrences, toujours ponctuation faible ;
- parenthèses < (> et <) > – 5 paires d'occurrences ;
- point d'interrogation (« interrogant ») < ? > – 3 occurrences, dont 2 dans la ponctuation forte et une faible³¹³.

Le passage analysé ne contient aucune occurrence du « point admiratif » < ! > décrit dans le traité de Dolet.

Le passage analysé ne contient pas de discours direct attribué à des instances énonciatives autres que l'auteur, mais le traité entier étant construit comme un discours à la première personne adressé au lecteur, il utilise un certain nombre d'incises comme dans l'exemple suivant :

(8.28) Quant à la difference (dy ie) de ces || dictions animus , *et* anima , il ne s'i fault point arre||ster : (*dolet1540, 12:4*).

Conformément à la règle énoncée dans le traité, ces incises sont ponctuées à l'aide d'une paire de parenthèses. Bien entendu, ces incises « sans discours direct » ne sont pas utilisées pour indiquer le locuteur, mais elles jouent le rôle d'une remarque ou d'une précision. Ces remarques, toujours ponctuées par des parenthèses, peuvent avoir une structure plus complexe et contenir à l'intérieur des marques de ponctuation syntaxique et énonciative :

(8.29) Et par ainsi || c'est superstition trop grande (diray ie besterie , ou || ignorance ?) de commencer sa traduction au commen||cement || de la clausule : (*dolet1540, 13:16-17*).

313 Il s'agit dans ce dernier cas de la fin d'une incidente placée entre parenthèses, cf. (8.31).

Livres imprimés

Notons dans les deux exemples l'absence de mise en évidence typographique des mots utilisés avec une fonction métalinguistique (citations de mots latins dans (8.28) et proposition de synonymes contextuels dans (8.29)).

Du point de vue strictement formel, ces remarques peuvent entretenir des relations syntaxiques différentes avec les structures « de base » : une subordonnée relative comme dans (8.30), une apostrophe (sous forme d'un pronom personnel avec une relative, comme dans (8.31)), etc. :

- (8.30) Et si aucun le fait, cela luy procede de paure, et deffault d'esprit. Car s'il a les qualites dessusdictes, lesquelles il est besoing estre en ung bon tra ducteur, sans auoir esgard à l'ordre des mots il s'arrestera aux sentences, et fera en sorte, que l'intention de l'auteur sera exprimee, gardant curieusement la propriété de l'une, et l'autre langue. (dolet1540, 13:9-15) ;
- (8.31) Dy moy, toy qui entends Latin, estoit il possible de bien traduire ce passage, sans une grande intelligence du sens de Ciceron ? (dolet1540, 12:13).

On voit donc que dans ce cas le choix de la ponctuation est déterminé par la sémantique (voire la pragmatique) plutôt que par la pure syntaxe.

Les limites de propositions autonomes sont toujours ponctuées, et il s'agit de la ponctuation forte dans 19 cas sur 24. En cas de ponctuation faible, la marque utilisée est toujours le deux-points, une marque plus forte que la virgule dans la théorie de Dolet. Sur les 5 occurrences relevées dans le passage dépouillé, deux concernent des propositions introduites par *car*, et deux par *mais*, des connecteurs qui marquent un lien logique fort :

- (8.32) Lesquelles si le traducteur ignore, il fait tort à l'auteur, qu'il traduit, et aussi à la langue, en laquelle il le tourne, car il ne represente, et n'exprime la dignité, et richesse de ces deux langues, desquelles il prend le maniment. (dolet1540, 13:2-6).

Le connecteur *car* peut toutefois être associé à une ponctuation forte si la proposition qu'il introduit entre dans une structure syntaxique complexe avec des propositions subordonnées, incidentes, etc., comme c'était le cas dans l'exemple (8.30) décrit ci-dessus.

Dans son traité, Dolet s'arrête assez longuement sur la notion de période (sans en donner pour autant une définition formelle). Une période (ou *clausule*) est délimitée

par la ponctuation forte (point ou point d'interrogation suivi d'une majuscule). Elle doit constituer un sens complet et peut être composée de deux ou trois membres (*sentences*). La définition de la période est donc avant tout logique, mais sa construction doit prendre en compte des considérations rhétoriques : sa longueur ne doit pas « excéder l'haleine de l'homme ».

L'analyse des périodes graphiques (segments textuels délimités par des ponctuations fortes) dans notre extrait montre que leur longueur est malgré tout assez variable : elle va de 10 à 52 mots avec une moyenne de 28 mots environ. Les longues périodes (de plus de 40 mots) contiennent habituellement des incidentes entre parenthèses qui permettent sans doute d'introduire une « pause respiratoire » (*cf.* la période la plus longue de notre extrait citée dans l'exemple (8.30)).

Le principe de ne pas introduire dans une période plus de trois propositions autonomes (séparées par un deux-points) est en général respecté dans le passage analysé, même si nous avons relevé une exception en quatre parties :

(8.33) Et est cer||tain , que animus est dict de anima ; *et* que anima est || l'organe de animus ;
comme si tu uoulois dire la uer||tu , *et* instruments uitaux estre origine de l'esprit ; || *et*
iceluy esprit estre ung effect de ladicte uertu uita||le . (*dolet1540, 12:8-12*).

Les propositions avec des éléments communs sont toujours séparées par une ponctuation faible : dans 2 cas, par un deux-points et dans 4 cas par une virgule. Le nombre d'occurrences dont nous disposons est trop bas pour pouvoir dégager les facteurs qui jouent dans le choix entre ces deux marques, mais on peut supposer que la présence d'une occurrence du deux-points dans le contexte de gauche favorise le choix de la virgule, ce qui permet à l'auteur de respecter formellement la « règle des trois membres ». Dans l'exemple (8.32) donné ci-dessus, une virgule est utilisée dans une coordination de prédicats (*ne represente, et n'exprime*), tandis que dans (8.33), le deux-points apparaît entre deux subordonnées coordonnées (*que animus est dict... et que anima est...*), ce qui provoque d'ailleurs la violation de la règle des trois membres.

Les limites entre les propositions principales et subordonnées sont ponctuées dans la majorité des cas : sur un total de 48 occurrences 1 porte une ponctuation forte, 40 une ponctuation faible et 7 occurrences seulement ne sont pas ponctuées.

Livres imprimés

La seule occurrence de la ponctuation forte concerne une relative introduite par le pronom *lequel* suivi immédiatement de la conjonction *si* introduisant une subordonnée circonstancielle (8.32). On peut dire à propos de cet exemple que la relation de subordination des relatives en *lequel* est moins forte que celle des relatives en *que* et *qui*, certains chercheurs parlant dans ces cas de juxtaposition plutôt que de subordination. Le fait que la proposition introduite par *lequel* soit elle-même une principale pour plusieurs subordonnées (la circonstancielle *si le traducteur ignore* et la relative *qu'il traduit*) favorise certainement le renforcement de la ponctuation. Notons qu'une autre relative en *lequel*, dont la structure interne est plus simple, a été considérée comme une remarque et placée entre parenthèses (8.30).

L'absence de ponctuation entre les principales et les subordonnées semble être limitée à deux cas de figure : les frontières de propositions infinitives et participiales d'une part et les subordonnées placées immédiatement après le premier mot d'une principale, comme c'est le cas de la relative *qui entend le Latin* dans (8.31) et de la conditionnelle *si ung homme...* dans l'exemple suivant :

(8.34) Cuydes tu , que si || ung homme n'est parfaict en la langue Latine , et || Francoyse , il puisse bien traduire en Francoys quel||que oraison de Ciceron ? (*dolet1540, 12:25*).

En ce qui concerne les propositions infinitives et participiales, celles-ci peuvent parfois être ponctuées, mais la pratique est moins systématique que dans le cas de propositions comportant des verbes conjugués :

(8.35) mais contente toy du || commun , sans innouer aulcunes dictions follement , et || par curiosite reprehensible . (*dolet1540, 14:11*) ;

(8.36) et fera en sorte , que l'intention || de l'auteur sera exprimee , gardant curieusement la || propriete de l'une , et l'aultre langue (*dolet1540, 13:14*).

Dans l'exemple (8.35) une virgule est utilisée devant une infinitive en *sans*, tandis qu'une proposition similaire ne porte pas de ponctuation spécifique (8.30) : la parenthèse y sert à délimiter l'incidente qui précède, et la fin de l'infinitive n'est pas ponctuée du tout. Il convient de noter que l'usage de la préposition *sans* favorise peut-être l'apparition d'une ponctuation, même lorsqu'elle sert à introduire des syntagmes non propositionnels, comme c'était le cas dans (8.31).

En ce qui concerne les limites de propositions participiales, l'absence de ponctuation paraît être la règle. Dans (8.36), cela est sans doute motivé par le besoin de distinguer le participe passé faisant partie de la forme passive du prédicat, du participe présent doté de ses propres compléments.

Les syntagmes coordonnés et juxtaposés sont ponctués d'une façon très régulière (qu'ils soient ou non introduits par une conjonction de coordination) : toutes les 25 occurrences de ce type de frontière sont marquées par une virgule. On trouve une marque plus forte (le deux-points) dans un cas (8.32) où le premier des compléments coordonnés contient une relative et où la coordination porte le sens de complémentation (présence du connecteur *aussi*). Pour les frontières de ce type, Dolet adopte donc une approche strictement formelle, sans chercher à distinguer les différents types de coordination et de juxtaposition.

Les syntagmes particuliers à l'intérieur d'une proposition ne sont que très rarement ponctués : 4 occurrences seulement dans l'extrait dépouillé. Deux de ces occurrences concernent une remarque placée entre parenthèses qui est formellement une apposition du complément d'objet :

(8.37) S'il ad||uient *doncques* , que tu traduises quelcque Liure La||tin en ycelles
 (mesmement en la Francoyse) il te || fault garder d'usurper mots trop approchans du ||
 Latin (*dolet1540, 14:8*).

Comme nous l'avons déjà remarqué, l'usage des parenthèses chez Dolet est avant tout dicté par la sémantique, et le rôle formel des éléments qu'elles entourent dans la structure syntaxique a peu d'importance.

Nous avons déjà mentionné une autre occurrence de la ponctuation à la frontière d'un syntagme particulier : une virgule devant un complément circonstanciel introduit par la préposition *sans* (8.31). C'est la présence de la préposition *sans* qui a, à notre sens, favorisé cet usage.

La dernière occurrence d'une ponctuation devant un syntagme particulier est se justifie aussi aisément : il s'agit d'une virgule devant un complément de comparatif introduit par *que* :

Livres imprimés

(8.38) Tu te garderas diligem[^]ment de || ce uice : qui ne demonstre aultre chose , que l'igno||
rance du traducteur . (*dolet1540, 13:28*).

Comme *sans* dans l'exemple précédent, la conjonction *que* (habituellement utilisée dans les propositions subordonnées) peut à elle seule « provoquer » l'usage d'une ponctuation. Il est également possible que Dolet ait considéré ce type de complément comme une petite proposition.

Notre brève analyse de la pratique de la ponctuation chez Estienne Dolet montre qu'il attache effectivement beaucoup d'attention à l'usage de ces marques et applique scrupuleusement les principes qu'il a lui-même formulés, même si on peut parfois constater de légers écarts (notamment par rapport à la règle « des trois membres »). Selon les types de frontières, la motivation de l'emploi d'une marque peut être avant tout formelle (comme dans le cas de la coordination ou de l'usage de certains connecteurs) ou sémantique (comme dans le cas des parenthèses). Quoi qu'il en soit, la ponctuation de Dolet se distingue très fortement par son rationalisme des pratiques « impressionnistes » caractéristiques des manuscrits médiévaux et des premiers imprimés. A certains égards, la ponctuation de Dolet est plus rationnelle (voire plus « mécanique ») que celle que la littérature normative recommande pour le français contemporain.

Chapitre 9

Synthèse des tendances de la ponctuation

Après avoir analysé l'usage de la ponctuation dans chacun des extraits de manuscrits et de livres imprimés qui composent notre corpus, nous tenterons dans ce chapitre de dégager à partir de ces données, ainsi que des résultats obtenus par les autres chercheurs ayant travaillé sur la ponctuation du français médiéval, les pratiques dominantes, les cas « aberrants » et les tendances de l'évolution dans le temps.

Comme nous l'avons déjà mentionné dans la section 3.4.3 où nous avons présenté les caractéristiques générales de notre corpus, ce corpus est loin d'être représentatif tant sur le plan des variables externes de la typologie textuelle que sur les dimensions des échantillons transcrits et annotés pour chaque texte. Il permet néanmoins de formuler des hypothèses concernant les tendances dominantes pour les phénomènes les plus fréquents et d'ouvrir des pistes pour l'approfondissement des recherches sur des corpus plus importants.

9.1. Observations générales

Dans l'état actuel de notre corpus une opposition semble se profiler entre les manuscrits antérieurs au XV^e siècle, ceux du XV^e et les imprimés, où le système se stabilise et se normalise progressivement.

Siècle	Domaine	Texte	Taux de ponctuation				
			faible		forte	général	
<i>Manuscrits</i>							
11	juridique	strasb9768	8,70	67%	4,35	33%	13,04
13	didactique	im587	0,47	14%	2,96	86%	3,43
		tyr828	6,33	60%	4,22	40%	10,54
	historique	villehardouin12203	6,99	58%	5,17	43%	12,16
		beauma11652	2,46	48%	2,70	52%	5,15
	littéraire	lancelot77	7,27	80%	1,77	20%	9,04
		qgraal77	5,52	67%	2,74	33%	8,26
qgraal1119		3,08	76%	0,98	24%	4,06	
14	didactique	im574	6,01	45%	7,47	55%	13,48
		dictier840	1,39	19%	6,10	81%	7,49
	historique	joinville13568	5,56	58%	4,10	42%	9,65
		villehardouin587	5,97	60%	3,98	40%	9,94
		froissart869	5,33	95%	0,29	5%	5,61
religieux	joachim768	4,33	54%	3,63	46%	7,96	
15	didactique	chasse619	6,13	61%	3,87	39%	10,00
		chasse765	1,71	27%	4,62	73%	6,34
		egipte769	2,47	48%	2,64	52%	5,11
	historique	monstre2682	2,39	32%	5,18	68%	7,57
	littéraire	quadrilogue126	5,39	71%	2,18	29%	7,57
		artois11610	1,43	31%	3,13	69%	4,56
		qjm1052	1,38	28%	3,52	72%	4,89
		anglure15217	0,10	1%	7,76	99%	7,86
	phares1357	1,52	19%	6,64	81%	8,16	
politique	ursins2701	2,71	77%	0,83	23%	3,54	
religieux	gerson13318	3,04	56%	2,43	44%	5,47	
<i>Imprimés</i>							
15	littéraire	artu1488a	5,21	65%	2,86	35%	8,07
	didactique	im1495a	3,13	64%	1,74	36%	4,87
16	didactique	im1522a	5,07	71%	2,10	29%	7,17
		dolet1540	14,03	80%	3,43	20%	17,46

Légende :

Le taux de ponctuation correspond au nombre moyen des occurrences de la ponctuation pour 100 mots du texte.

Pour les ponctuations forte et faible, la première colonne indique le taux absolu et la deuxième, le pourcentage par rapport à l'ensemble des ponctuations.

Les chiffres « exceptionnels » sont surlignés : dans la gamme du rouge pour les valeurs élevées et dans la gamme du vert pour les valeurs basses.

Tableau 7 : Taux général et force de la ponctuation dans l'ensemble du corpus

Synthèse des tendances de la ponctuation

Dans le Tableau 7 ci-dessus, nous avons rassemblé les données concernant le taux général et la répartition entre les ponctuations fortes et faibles dans l'ensemble du corpus. Les textes sont classés par siècle, puis par domaine. Même si le nombre de textes dans chaque domaine est trop bas pour établir une corrélation, cette information nous paraît utile.

9.1.1. Variations du taux général

Dans les manuscrits en prose antérieurs au XV^e siècle, trois groupes se forment selon le taux de ponctuation général. Les plus nombreux (7 au total) sont les manuscrits où ce taux se situe entre 7 et 10% (**tyr828**, **lancelot77**, **qgraal77**, **dictier840**, **joinville13568**, **villehardouin587** et **joachim768**). Dans 4 manuscrits (**im587**, **beuma11652**, **qgraal1119** et **froissart869**), le taux de ponctuation est nettement plus bas (entre 3 et 5%). Deux manuscrits (**villehardouin12203** et **im574**), enfin, présentent un taux de ponctuation exceptionnellement élevé, entre 12 et 13%. Tel est également le cas de **strasb9768**, manuscrit beaucoup plus ancien que presque toutes les variables externes opposent au reste du corpus.

Au XV^e siècle, la majorité des manuscrits présente un taux de ponctuation qui varie entre 4,5 et 7,5% et est donc nettement moins élevé que la moyenne de la période antérieure. Un texte (**ursins2701**) est particulièrement peu ponctué (3,5%), mais l'écart par rapport à la moyenne n'est pas très important. Le manuscrit le plus ponctué du XV^e siècle (**chasse619**) présente un taux de ponctuation de près de 10%. On peut être tenté de conclure que la variation du taux de ponctuation baisse au cours du XV^e siècle et que la normalisation qui viendra avec la typographie commence à « se préparer » déjà. Bien entendu, une telle hypothèse ne pourra être confirmée (ou infirmée) que sur un corpus beaucoup plus important et représentatif.

Le taux de ponctuation dans les deux incunables de notre corpus (**artu1488** et **im1495**) atteint les 8,1% et 4,9% respectivement, il ne diffère donc pas sensiblement de celui des manuscrits. Dans **im1522**, le taux de ponctuation augmente par rapport à **im1495** et atteint les 7,2%. La différence est significative, puisqu'il s'agit de deux « rédactions » d'un même texte, quasi identiques au niveau lexical.

Le traité d'E. Dolet (**dolet1540**) présente un taux de ponctuation exceptionnel de 17,5% (plus élevé que dans les imprimés modernes), ce texte marque donc une vraie rupture par rapport à la tradition manuscrite et aux premiers imprimés en langue française.

9.1.2. Ponctuation forte vs. ponctuation faible

En ce qui concerne la répartition entre les ponctuations forte et faible, deux manuscrits s'opposent clairement au reste du corpus : **froissart869** avec son usage presque exclusif (95%) de la ponctuation faible et **anglure15217**, où, au contraire, c'est la ponctuation forte (majuscule seule le plus souvent) qui est utilisée dans 99% des occurrences.

Le reste du corpus des manuscrits se divise en trois groupes de dimensions comparables. Le plus important est composé de 10 manuscrits (**tyr828**, **villehardouin12203**, **beauma11652**, **im574**, **joinville13568**, **villehardouin587**, **joachim768**, **chasse619**, **egypte769** et **gerson13318**) avec un usage équilibré de la ponctuation forte et faible. La proportion de chacune des forces de ponctuation y varie de 40 à 60% environ. Dans 7 manuscrits sur 10, la ponctuation faible est légèrement plus fréquente que la forte.

Sept autres manuscrits (**im587**, **dictier840**, **chasse765**, **monstre2682**, **artois11610**, **qjm1052** et **phares1357**) forment un deuxième groupe, caractérisé par la dominance de la ponctuation forte (entre 69 et 81% des marques de ponctuation). Cinq des sept manuscrits de ce groupe datent du XV^e siècle. La dominance de la ponctuation forte s'explique en partie par l'usage fréquent de la majuscule seule dans la plupart de ces manuscrits.

Le troisième groupe est composé de 5 manuscrits (**strasb9768**, **lancelot77**, **qgraal77**³¹⁴, **qgraal1119**, **quadrilogue126** et **ursins2701**). La ponctuation faible y est dominante avec entre 66 et 80% des occurrences des marques de ponctuation.

Dans les incunables, la ponctuation faible domine avec près des deux tiers des occurrences. Ce taux augmente encore au XVI^e siècle avec 71% des ponctuations faibles

314 Rappelons que les extraits **lancelot77** et **qgraal77** représentent des œuvres différentes tirées d'un même manuscrit.

Synthèse des tendances de la ponctuation

dans **im1522a** et 80% chez **dolet1540**, c'est donc dans le domaine de la ponctuation faible que l'évolution est la plus significative dans les livres du XVI^e par rapport aux incunables.

9.1.3. Forme de la marque

Les données quantitatives concernant les fréquences des différentes formes des marques de ponctuation sont présentées dans le Tableau 8 ci-dessous. Nous avons distingué 3 marques principales (le point, la barre oblique et la majuscule seule) et additionné toutes les autres occurrences dans la colonne « autre », car leur fréquence relative n'est jamais très élevée dans notre corpus.

Les recherches menées par nos prédécesseurs sur la ponctuation des manuscrits français en prose ont montré que souvent le point était la seule marque employée (Careri *et al.* 2001, p. xxxv)³¹⁵. Nos données confirment de façon générale cette conclusion, tout en permettant d'y apporter quelques précisions.

Dans la majorité des textes du corpus, on constate une nette dominance d'une seule forme de marque : pour 14 manuscrits et un incunable (**im1495a**) il s'agit du point, pour 3 manuscrits et un livre du XVI^e (**im1522a**), de la barre oblique et pour un manuscrit (**anglure15217**), de la majuscule seule. Le point et la barre oblique s'utilisent dans ces cas aussi bien dans la ponctuation forte que dans la ponctuation faible. En même temps, une autre forme de marque de ponctuation apparaît, au moins occasionnellement, dans chaque extrait³¹⁶.

Dans deux manuscrits (**villehardouin587** et **gerson13318**) et dans deux imprimés (**artu1488** et **im1522**), plusieurs signes de ponctuation sont utilisés avec une fréquence comparable. Tel est également le cas de **dolet1540** où un lien solide entre la forme et la fonction de la marque est établi.

315 En vers, en revanche, la ponctuation serait moins dense, mais « meilleure », grâce à la distinction de plusieurs signes (Careri 2008, p. 215).

316 A l'exception, peut-être, de **froissart869**, où la seule occurrence de la majuscule seule peut être accidentelle.

Siècle	Dom.	Texte	Point	Barre	Autre	M.s.	Total	
<i>Manuscrits</i>								
11	J	strasb9768	93% (14)	(1)			15	
13	D	im587	86% (18)		(1)	(2)	21	
	H	tyr828	81% (50)	(6)	(1)	(4)	61	
		villehardouin12203	95% (76)			(3)	(1)	80
	J	beauma11652	92% (59)			(3)	(2)	64
	L	lancelot77	96% (49)			(2)		51
		qgraal77	97% (164)			(3)	(2)	169
qgraal1119		93% (27)		(1)		(1)	29	
14	D	im574	83% (62)			(12)	74	
		dictier840	44% (68)	(8)	(4)	48% (75)	155	
	H	joinville13568	95% (100)			(5)		105
		villehardouin587	65% (22)			32% (11)	(1)	34
		froissart869			98% (57)		(1)	58
	R	joachim768	91% (42)			(1)	(3)	46
15	D	chasse619	97% (73)			(1)	(1)	75
		chasse765	(1)	57% (21)	(7)		(8)	37
		egipte769		52% (16)	(1)	45% (14)		31
	H	monstre2682	36% (21)	(4)	(4)	49% (28)		57
	L	quadrilogue126	94% (117)			(4)	(4)	125
		artois11610	18% (9)	53% (27)		(5)	20% (10)	51
		qjm1052	72% (23)	(4)			(5)	32
		anglure15217	25% (39)	(2)			73% (111)	152
		phares1357	28% (51)	19% (33)		(15)	44% (78)	177
	P	ursins2701	(2)	80% (51)		(4)	(7)	64
R	gerson13318	44% (51)	39% (45)		(3)	(16)	115	
<i>Imprimés</i>								
15	L	artu1488	42% (750)	31% (537)	25% (435)	(~30)	1752	
	D	im1495	76% (459)	(33)	(10)	17% (~100)	602	
16	D	im1522	(7)	55% (22)		(9)	(2)	40
		dolet1540	17% (18)	60% (65)	22% (24)			107

Légende :

Le pourcentage est calculé par rapport au nombre total des occurrences des marques de ponctuation dans l'extrait dépouillé. Le nombre absolu est indiqué entre parenthèses. Seuls les pourcentages supérieurs à 10% sont indiqués.

Tableau 8 : Fréquences des formes des marques de ponctuation dans l'ensemble du corpus

Synthèse des tendances de la ponctuation

Enfin, dans quatre manuscrits (**dictier840**, **egipte769**, **monstre2682** et **phares1357**), l'usage d'un signe de ponctuation est concurrencé par celui de la majuscule seule.

Les données dont nous disposons ne permettent pas d'affirmer qu'une forme particulière d'une marque « se spécialise » dans une fonction donnée dans les manuscrits médiévaux français : les différentes marques sont le plus souvent polyvalentes et interchangeable.

Bien entendu, cette observation ne concerne pas les marques destinées à la ponctuation macrotextuelle, comme le pied-de-mouche, le crochet adlinéaire ou la « fin de section » (attestée dans **artois11610**).

Une étude attentive des phénomènes peu fréquents sur des textes intégraux peut amener à nuancer davantage l'hypothèse sur la polyvalence des marques. Ainsi, le *comma*, dont le nombre d'occurrences dans notre corpus est trop peu élevé pour en tirer quelque conclusion que ce soit, peut en effet servir à marquer certains types de constructions. Selon l'analyse de l'emploi de cette marque dans le texte intégral de la *Queste del Saint Graal* (dans la version du même manuscrit Lyon, BM, P.A. 77, d'où proviennent nos extraits **qgraal77** et **lancelot77**) effectuée par Ch. Marchello-Nizia (2007), le *comma* apparaît dans des constructions corrélatives ou entre subordonnée et principale ou encore derrière l'interjection *ha(a)*, *hé*, etc. Toutes ces constructions pourraient avoir en commun un certain modèle intonatif. Les occurrences du *comma* provenant des autres manuscrits de notre corpus (en particulier, **im574**) ne contredisent pas cette hypothèse. Il semble néanmoins que le *comma* puisse toujours être remplacé par la marque dominante (à l'exception de la position très caractéristique après les interjections au début du discours direct).

Le *punctus interrogativus*, dont les seules occurrences dans notre corpus proviennent du traité d'E. Dolet, doit aussi être considéré comme une marque spécialisée (dans le marquage de la modalité interrogative ou de l'intensité émotionnelle). Son usage n'est pourtant pas systématique non plus : l'usage de la marque dominante est toujours possible (*cf.* nos remarques sur le texte intégral de **qgraal77**).

Une spécialisation des marques selon leur forme commence à se profiler dans **artu1488** et **im1522** : le point dans la ponctuation forte, la barre oblique et le deux-points dans la ponctuation faible, même si des exceptions sont toujours possibles. Dans **im1495**, le système est plus proche de la polyvalence des manuscrits pour le point et le deux-points, alors que la barre oblique se spécialise dans la ponctuation faible. Dans **dolet1540**, la spécialisation des marques atteint (voire dépasse) la régularité du système moderne.

Après cet aperçu général, nous pouvons passer à l'examen des tendances de la ponctuation aux différents types de frontières ponctuables que nous passerons en revue dans le même ordre qu'au cours de l'analyse des textes individuels.

9.2. Ponctuation selon le type de frontière ponctuable

La définition des frontières d'unités de la macrostructure textuelle n'est pas entièrement indépendante de leur marquage effectif par la ponctuation (*cf.* la section 2.6.1.). Il est donc inutile de chercher des corrélations entre des variables liées par définition. Une étude de la ponctuation à ce niveau devrait se faire sur un corpus différent du nôtre : il devrait inclure des textes intégraux (pas forcément transcrits au niveau allographique) de plusieurs manuscrits de la tradition d'une même œuvre médiévale.

Les frontières liées au discours direct et aux citations sont présentes dans certains textes de notre corpus seulement, et leur fréquence totale est trop faible pour tenter de procéder à des généralisations. Il semblerait néanmoins que les débuts et fins des passages au discours direct, ainsi que les changements de locuteurs attirent la ponctuation forte ou faible (dans les cas où cette dernière est dominante dans un texte donné). Le taux effectif des frontières de ce type ponctuées dépend du taux général de ponctuation dans le texte et est similaire (sinon supérieur) à celui des frontières entre propositions autonomes. L'absence d'une marque de ponctuation spécifique pour

Synthèse des tendances de la ponctuation

signaler le discours direct fait que le rôle des marques segmentales (verbes de parole, interjections, etc.) devient particulièrement important.

Les incises et les apostrophes ne sont, quant à elles, que très rarement ponctuées, et à chaque occurrence d'autres facteurs peuvent expliquer l'apparition d'une marque.

Les frontières entre propositions autonomes sont les plus favorables à la ponctuation. Dans certains textes, elles sont ponctuées dans plus de 90% des cas³¹⁷. Dans les manuscrits les moins ponctués, ce sont tout de même les frontières de ce type qui attirent la majorité des occurrences des marques de ponctuation. Le choix entre la ponctuation forte ou faible dépend de la répartition globale de ces forces de ponctuation dans le texte, mais quoi qu'il en soit, les chances de rencontrer une ponctuation forte à ce type de frontières sont plus élevées qu'ailleurs.

Les facteurs supplémentaires permettant d'expliquer l'emploi ou non d'une marque à un endroit donné sont divers. Ils incluent la complexité de la construction (nombre de subordonnées, de divers éléments coordonnés, etc.), la distance de la ponctuation forte précédente, l'importance de l'information véhiculée, etc.

Les frontières entre propositions avec des éléments communs sont moins fréquentes que celles de propositions autonomes, et le nombre d'occurrences dans certains textes est trop peu élevé pour qu'il soit analysable. Dans les textes où le nombre d'occurrences est suffisant, on voit une nette baisse du taux de ponctuation par rapport aux propositions indépendantes. Par exemple, le taux de frontières ponctuées passe de 100% à 25% dans **artu1488a** et de 75% à moins de 10% dans **im1495a** pour les deux types de frontières respectifs. La ponctuation forte est très rare aux frontières de ce type (à l'exception des manuscrits comme **dictier840** ou **anglure15217**, où celle-ci est largement dominante de façon générale).

Les frontières entre propositions subordonnées et principales sont, elles aussi, nettement moins souvent ponctuées que les frontières entre propositions indépendantes. Le taux d'absence de ponctuation à ce type de frontière est légèrement supérieur par rapport aux frontières entre propositions avec des éléments communs (84% contre 62%

³¹⁷ Voir à 100% dans les extraits dépouillés **villehardouin12203**, **villehardouin587**, **im574**, **monstre2682**, **artu1488** et **dolet1540**.

respectivement sur l'ensemble du corpus³¹⁸). Dans certains manuscrits (**lancelot77/qgraal77**, **joinville13568**, **anglure15217**), la différence de taux est plus importante : les frontières entre propositions avec des éléments communs y sont deux ou trois fois plus souvent ponctuées que celles entre principale et subordonnée. Un seul manuscrit (**froissart869**) échappe apparemment à cette tendance : 29% des frontières principale/subordonnée y sont ponctuées contre 17% pour les frontières de propositions avec des éléments communs. Si les frontières de ce type sont ponctuées, il s'agit d'une ponctuation faible dans 62% des cas en moyenne. Sur ce paramètre, la différence par rapport aux propositions avec des éléments communs (où la ponctuation faible est utilisée dans 57% des cas) est très légère. Il semblerait néanmoins que dans certains manuscrits, où la ponctuation forte est dominante de façon générale, les frontières principale/subordonnée « attire » la ponctuation faible dans les rares cas où celles-ci sont ponctuées. C'est le cas de **dictier840** et de **phares1357**.

Il convient dans tous les cas d'observer une grande prudence dans l'interprétation de ces données quantitatives, car le nombre d'occurrences de frontières entre propositions avec des éléments communs n'est pas élevé (moins de 10 pour certains textes), ce qui augmente considérablement la marge d'erreur.

Le manuscrit qui ponctue le plus (14 occurrences sur 30, soit 47%) les frontières des subordonnées est **im574**. On y trouve par ailleurs un nombre relativement élevé de *commas*. Le manuscrit en question présente un des taux de ponctuation général les plus élevés de notre corpus, ce qui se répercute à tous les types de frontières.

Les incunables **artu1488a** et **im1495a**, ainsi que **im1522a** ne présentent pas de différence significative par rapport aux manuscrits dans le traitement de ce type de frontière.

Il est difficile de dégager une tendance générale qui puisse expliquer les occurrences où la ponctuation est effectivement présente. On y trouve assez souvent les frontières de subordonnées circonstancielles et celles de principales, mais il ne s'agit en aucun cas d'une régularité absolue.

Le traité **dolet1540** s'oppose radicalement sur ce point au reste du corpus, car les frontières entre principales et subordonnées y sont ponctuées dans 85% des cas (41

318 A l'exception de **dolet1540** qui s'oppose très nettement au reste du corpus sur ce paramètre.

Synthèse des tendances de la ponctuation

occurrences sur 48). En effet seules certaines propositions infinitives et participiales, ainsi que des subordinées « imbriquées » à une courte distance de la ponctuation précédente ne sont pas ponctuées (*cf.* la section 8.4 pour une analyse détaillée).

Les frontières de syntagmes coordonnés ou juxtaposés à l'intérieur d'une proposition sont de façon générale peu ponctués. Le taux moyen d'absence de ponctuation y atteint les 80% sur l'ensemble du corpus (à l'exception du traité de Dolet). En réalité, la situation varie considérablement selon les textes. Dans **im587**, **qgraal1119**, **froissart869**, **artois11610**, **ursins2701** et **gerson13318**, l'absence de ponctuation est quasi-systématique (plus de 93% des cas). Un seul texte (**villehardouin12203**) présente une tripartition parfaite : 9 cas d'absence de ponctuation contre 8 occurrences de ponctuation faible et 9 occurrences de ponctuation forte. La fréquence extraordinaire de la ponctuation forte s'explique par la présence dans l'extrait dépouillé d'une longue liste de noms propres et de titres de noblesse. La ponctuation dans ce cas joue à la fois un rôle dans la structuration d'un type particulier de séquence textuelle et permet de mettre en relief les noms propres et de témoigner du respect envers des personnages importants.

Le taux relativement élevé (entre 30% et 40%) des occurrences ponctuées dans les textes comme **im574**, **dictier840** et **monstre2682** s'explique également par la présence d'énumérations plus ou moins longues. La ponctuation forte apparaît dans les cas où il s'agit d'énumération de noms propres de personnes, de noms géographiques ou de termes importants (comme les lettres utilisées dans un sens métalinguistique dans **dictier840**).

L'usage de la ponctuation entre des termes de même nature coordonnés ou juxtaposés fait partie de la tradition médiévale, y compris dans les textes en vers qui peuvent être très peu ponctués par ailleurs (Roques 1952, p.194 ; Llamas Pombo 2007, p. 22 ; Careri 2008, p. 220-222)³¹⁹. Notre recherche a permis de démontrer que la nature des termes énumérés est importante, ainsi que la longueur des énumérations : la grande majorité des séquences ne dépassant pas les trois ou quatre termes restent sans ponctuation.

319 E. Llamas Pombo insiste par ailleurs sur le fait que dans ces cas le point sert plutôt à indiquer la liaison des termes qu'à les séparer.

En ce qui concerne le traité d'E. Dolet, la ponctuation de syntagmes coordonnés ou juxtaposés y est systématique : les 26 frontières de ce type relevées portent toutes une ponctuation faible.

Les frontières de syntagmes particuliers à l'intérieur de la proposition ne sont que très rarement ponctuées. Six textes seulement de notre corpus (**tyr828**, **villehardouin12203**, **joinville13568**, **quadrilogue126**, **monstre2682** et **phares1357**) présentent plus de 5 occurrences ponctuées, toutes les sous-catégories de ce type confondues.

Nous avons identifié un type de séquence textuelle particulier où la probabilité de rencontrer de telles occurrences est plus élevée qu'ailleurs dans le texte, il s'agit de longues formules introductives (ou dédicaces) qui ouvrent souvent les œuvres médiévales.

Ailleurs dans les textes, des raisons décoratives (justification des fins de ligne, « capitalisation » des premières lettres d'une page, etc.) ou des erreurs dans l'interprétation de passages difficiles peuvent expliquer la plupart des occurrences.

Chez E. Dolet, la ponctuation autour de syntagmes particuliers est tout aussi rare que dans le reste du corpus : 4 occurrences seulement, dont deux parenthèses autour d'une précision.

Même si notre corpus est trop restreint pour qu'on puisse procéder à des généralisations plus ou moins significatives, l'analyse des quatre catégories de frontières ponctuables suffisamment représentées (propositions autonomes, propositions avec des éléments communs, propositions principale/subordonnée et syntagmes coordonnés ou énumérés) permet de voir une nette différence entre les propositions autonomes et les trois autres catégories où l'absence de ponctuation se présente comme une tendance dominante dans la majorité des textes du corpus. Les facteurs qui peuvent déclencher l'emploi d'une ponctuation sont différents pour chaque catégorie, mais le taux global ne varie pas beaucoup entre les trois catégories en fin de compte.

Avant de conclure notre synthèse par une tentative de typologie des textes médiévaux basée sur l'analyse factorielle de l'emploi de la ponctuation aux quatre types de frontières les plus répandus, nous présenterons quelques observations concernant la

ponctuation « normalisée » des éditions scientifiques des textes médiévaux et ses relations avec la ponctuation des sources primaires.

9.3. Éditeurs scientifiques face à la ponctuation médiévale

Nous considérons la ponctuation modernisée des éditions scientifiques des textes médiévaux comme un outil supplémentaire qui sert à révéler la spécificité de la ponctuation médiévale par rapport à la norme moderne et à évaluer le degré d'interventionnisme des différents éditeurs face aux données primaires de leurs manuscrits de base. Par ailleurs, la comparaison automatisée entre les données linguistiques (graphies, morphologie, ordre des constituants, etc.) permet d'obtenir un « indice de fidélité » de l'édition qui peut aider le chercheur à décider si telle ou telle édition est utilisable pour un certain type d'étude (*cf.* Lavrentiev 2007a).

Nous présentons dans le Tableau 9 les données quantitatives concernant le traitement de la ponctuation dans les éditions de référence que nous avons utilisées pour constituer le niveau normalisé de nos transcriptions³²⁰. Les références des éditions sont indiquées dans une section spéciale de notre bibliographie.

Nous avons distingué quatre relations possibles entre la ponctuation de la source primaire et celle de l'édition : maintien d'absence de la marque (M–), ajout d'une marque dans l'édition à une frontière non ponctuée dans la source (A)³²¹, suppression dans l'édition d'une marque de ponctuation présente dans le manuscrit ou dans l'imprimé ancien (S) et, enfin, le maintien de la présence d'une marque (M+), quelles que soient sa forme et sa force. Pour pouvoir comparer les fréquences dans les échantillons de tailles différentes, nous présentons dans le tableau le taux des interventions (on non interventions) par rapport au nombre total des frontières ponctuables dans l'extrait dépouillé.

320 A l'exception des textes inédits **egipte769** et **joachim768** où nous avons procédé nous-même à la normalisation de la ponctuation, en respectant les règles précises que nous nous étions données (*cf.* 4.4.1.7). Dans **dolet1540** la ponctuation originale a été préservée au niveau normalisé, ce texte est donc exclu du corpus pour l'étude des interventions éditoriales.

321 Nous excluons de nos décomptes les occurrences d'ajout des guillemets au début et à la fin du discours direct, car une telle marque n'existait simplement pas dans le système graphique de nos sources.

Texte	F.P.	M-	A	M+	M+/Pé	S	S/M+
strasb9768	27	22%	19%	48%	62%	11%	0,23
im587	136	51%	34%	15%	30%	1%	0,05
tyr828	122	41%	10%	30%	50%	20%	0,67
villehardouin12203	134	27%	13%	42%	57%	19%	0,45
beauma11652	243	51%	21%	27%	55%	2%	0,06
lancelot77	119	32%	26%	36%	53%	6%	0,16
qgraal77	418	36%	22%	41%	65%	0%	0
qgraal1119	155	32%	48%	14%	20%	6%	0,43
im574	118	31%	8%	48%	70%	14%	0,28
dictier840	495	37%	31%	28%	45%	4%	0,14
joinville13568	185	28%	13%	41%	56%	18%	0,45
villehardouin587	65	32%	12%	46%	68%	9%	0,2
froissart869	191	43%	27%	22%	39%	8%	0,36
joachim768	107	43%	13%	34%	59%	10%	0,31
chasse619	153	35%	16%	41%	62%	9%	0,23
chasse765	114	52%	18%	25%	53%	4%	0,17
egipte769	108	55%	17%	28%	61%	1%	0,03
monstre2682	154	36%	27%	34%	53%	3%	0,1
quadrilogue126	311	36%	24%	29%	45%	12%	0,42
artois11610	212	55%	21%	19%	43%	5%	0,27
qjm1052	144	40%	37%	18%	30%	6%	0,31
anglure15217	374	21%	37%	39%	49%	2%	0,06
phares1357	500	34%	31%	27%	40%	9%	0,34
ursins2701	342	42%	39%	16%	28%	3%	0,16
gerson13318	437	38%	36%	25%	41%	1%	0,05
im1495a	134	46%	36%	17%	32%	1%	0,09
artu1488a	119	33%	27%	37%	55%	3%	0,09
im1522a	137	44%	28%	27%	48%	1%	0,03
Total/Moyenne	5754	38%	25%	30%	48%	7%	0,2

Légende :

F.P.	nombre de frontières ponctuables	M+	taux de maintien des marques de ponctuation de la source dans l'édition
M-	taux de maintien d'absence de ponctuation (par rapport au nombre de frontières ponctuables)	M+/Pé	taux de maintien des marques de ponctuation de la source (par rapport au nombres de frontières ponctuées)
A	taux d'ajout de marques de ponctuation par l'éditeur	S/M+	taux de suppression des marques de la source (par rapport à leur maintien)
S	taux de suppression des marques de ponctuation de la source par l'éditeur		

Tableau 9: Traitement de la ponctuation scribale dans les éditions scientifiques

Synthèse des tendances de la ponctuation

Le taux de maintien d'absence de ponctuation n'est pas une donnée très significative pour ce genre d'analyse : il évolue principalement en fonction du taux de ponctuation général de la source (le taux de maintien d'absence le plus élevé caractérise les manuscrits les moins ponctués) et de la complexité des structures syntaxiques du texte (le taux est plus bas dans les textes avec un grand nombre de subordinations, comme certaines œuvres historiques, telles que **villehardouin12203** ou **joinville13568**).

Le taux d'ajout des marques de ponctuation est également fortement lié au taux de ponctuation général d'un texte : il est le plus élevé pour les sources les moins ponctuées. Le taux d'ajout le plus bas (8%) est enregistré pour **im574** (édition de référence Prior 1913), tandis que le taux le plus élevé (48%) caractérise **qgraal1119** (édition de référence Pauphilet 1923). Le taux d'ajout moyen sur l'ensemble du corpus s'élève à 25%.

Les frontières ponctuables les plus concernées par l'ajout de la ponctuation sont celles des subordinées circonstancielles et relatives, ainsi que celles des syntagmes juxtaposés.

Le taux de maintien de la ponctuation caractérise le respect par l'éditeur de la ponctuation de sa source primaire, mais aussi le degré d'acceptabilité des marques de ponctuation sribales dans le système normalisé moderne. Il est plus intéressant d'évaluer ce taux par rapport au nombre de frontières effectivement ponctuées dans au moins une des sources (source primaire ou édition) que par rapport au nombre total des frontières ponctuables (qui inclut les cas de maintien d'absence de ponctuation). Ce taux est indiqué dans la colonne « M+/Pé ». Le taux le plus élevé (près de 70%) est enregistré dans **im574**. Ce beau manuscrit du XIV^e siècle présente en effet un usage de ponctuation remarquablement riche et régulier. Par ailleurs, l'édition de ce texte réalisée par O. H. Prior (1913) dans le cadre de sa thèse peut être considérée comme exemplaire en matière de respect des données linguistiques de son manuscrit de base. L'édition de **villehardouin587** (Faral 1961) présente également un taux de maintien de ponctuation important (68%), mais l'extrait transcrit et annoté dont nous disposons pour ce texte est trop bref pour être représentatif³²².

322 Il n'est pas exclu que le taux soit dû en partie à la « surponctuation » de la première page du manuscrit.

Dans l'édition de **qgraal77** (Marchello-Nizia 2008) où le respect de la ponctuation scribale est un principe éditorial, le taux de maintien de la ponctuation du manuscrit est suffisamment élevé (65%), mais n'est pas extraordinaire. Ceci s'explique par un nombre important d'ajouts de ponctuations (22%) aux incises et aux autres frontières liées au discours direct.

Le taux de suppression de la ponctuation est un paramètre très intéressant qui caractérise la densité des marques sribales jugées comme « aberrantes » du point de vue des règles modernes. De façon générale, ce taux est nettement moins élevé (7% sur l'ensemble du corpus) que le taux d'ajout (25%) et de maintien (30%) des marques de ponctuation. Une édition de **qgraal77** (Marchello-Nizia 2008) se démarque des autres sur ce point : aucune marque de ponctuation présente dans le manuscrit n'a été enlevée dans l'édition conformément au principe déclaré.

La frontière la plus concernée par la suppression de la ponctuation est celle des syntagmes coordonnés avec la conjonction *et*. Ce fait est assez révélateur de la spécificité de la pratique médiévale : la ponctuation pouvait apparaître aussi bien entre les syntagmes juxtaposés que coordonnés, alors que dans le système moderne la règle est de ponctuer les syntagmes juxtaposés et la ponctuation n'est utilisée que rarement, sous conditions précises, en cas de coordination.

Le taux de suppression le plus élevé (entre 18% et 20%) est enregistré dans **tyr828**, **villehardouin12203** et **joinville13568**. Ces textes se caractérisent par un nombre relativement important de ponctuations à des frontières de syntagmes particuliers. Ces trois textes, ainsi que **im574**, sont les seuls de notre corpus où le taux de suppression est supérieur au taux d'ajout de la ponctuation³²³.

Le dernier indice qui peut être intéressant pour caractériser à la fois la source et son édition est le rapport entre le nombre d'occurrences de suppression et de maintien des marques de ponctuation sribales (colonne « S/M+ »). La valeur moyenne de ce rapport est de 0,2, c'est-à-dire qu'il y a en moyenne 1 occurrence de suppression de la ponctuation pour 5 occurrences de maintien. Cela montre bien que la ponctuation

323 En ce qui concerne **im574**, ce fait s'explique par le taux d'ajout particulièrement bas.

Synthèse des tendances de la ponctuation

scribale est loin d'être arbitraire et qu'il existe un « socle commun » important avec le système moderne.

Un texte se démarque clairement du reste du corpus. Il s'agit de **tyr828** où le rapport entre la suppression et le maintien de la ponctuation atteint le chiffre de 0,67. Comme nous l'avons déjà indiqué, ce manuscrit présente un taux important d'emploi de la ponctuation à des frontières de syntagmes particuliers. L'une des hypothèses que nous avons avancées pour expliquer la particularité de ce manuscrit est liée au fait qu'il avait été réalisé dans un scriptorium à Acre, loin de la France métropolitaine, et pouvait ainsi subir les influences des traditions d'écriture étrangères. Quoi qu'il en soit, ce manuscrit conservé à la Bibliothèque municipale de Lyon mérite une étude approfondie et une comparaison avec d'autres manuscrits provenant de la même région.

Le fait que le taux sommaire du maintien d'absence et de présence de ponctuation à une frontière ponctuable dépasse en général les 60% est rassurant sur le plan de l'utilisabilité des éditions scientifiques en tant que sources de données pour les recherches linguistiques³²⁴.

9.4. Vers une typologie des ponctuations médiévales

L'un des objectifs initiaux de notre recherche était de construire une typologie des textes de français médiéval basée sur les pratiques de la ponctuation en relation avec les différentes variables de la typologie « externe ». Malheureusement, la saisie et l'annotation des transcriptions multi-facettes des sources primaires se sont avérées beaucoup plus longues et coûteuses que prévu, ce qui ne nous a pas permis de constituer un corpus suffisamment représentatif pour obtenir des résultats statistiquement exploitables.

Nous avons néanmoins voulu tester certains outils d'analyses statistiques pour vérifier si les résultats et les représentations qu'ils apportent concordent avec les conclusions de l'analyse qualitative.

³²⁴ Cf. la remarque de G. Nunberg (1990, p. 131) selon laquelle la « modernisation » de la ponctuation entraîne l'altération des structures linguistiques.

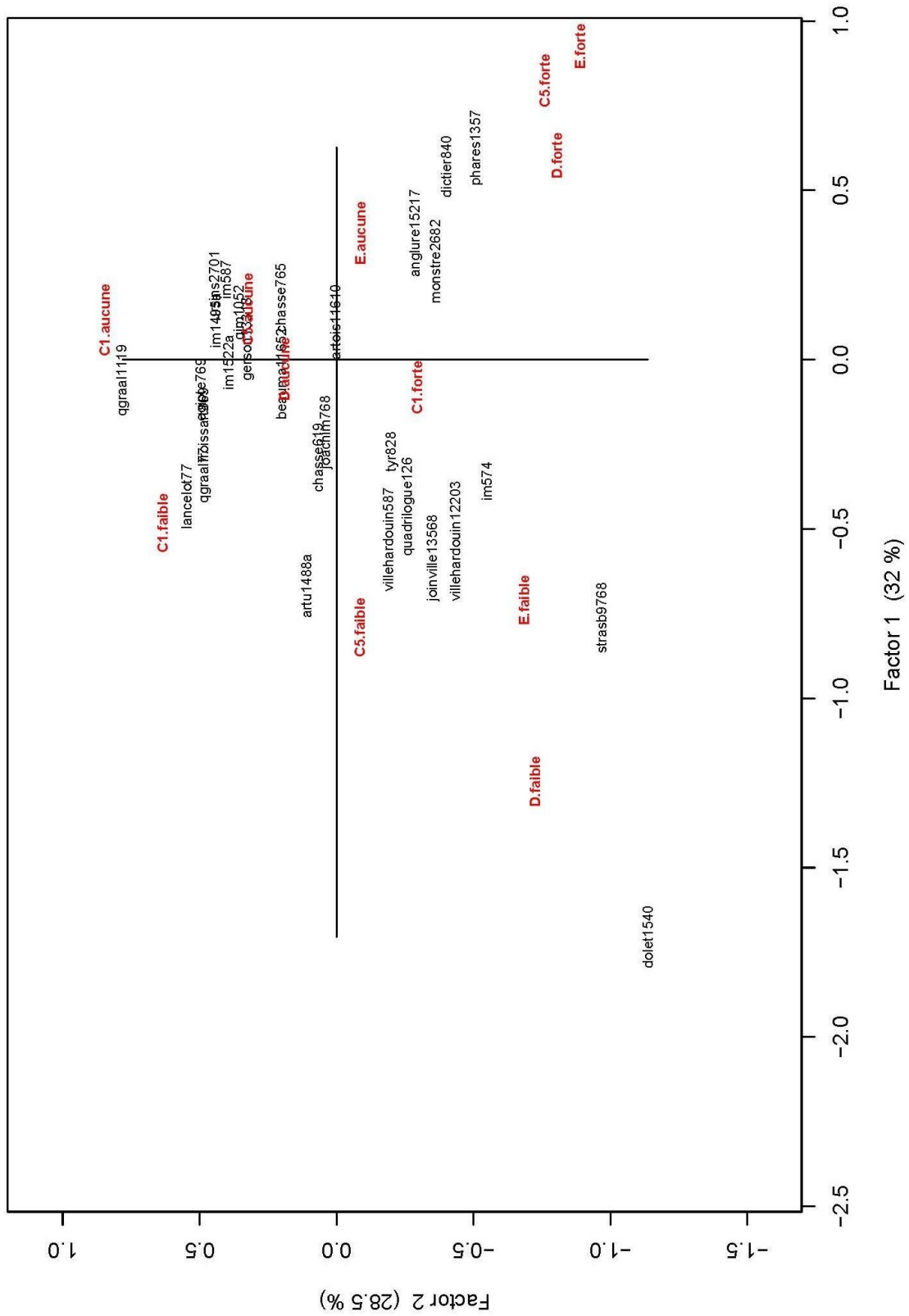


Fig. 76 : Analyse factorielle : distribution des textes et des types de ponctuation

Synthèse des tendances de la ponctuation

Concrètement, nous avons voulu utiliser un plan factoriel pour évaluer dans un premier temps les liens entre les types de frontières ponctuables et les différentes forces de la ponctuation (forte, faible ou aucune). Dans un deuxième temps, le plan factoriel devrait permettre de vérifier si les textes du corpus se regroupent par rapport à ces variables qui nous intéressent.

Il convient de souligner que l'analyse quantitative dont nous présentons les résultats ci-dessous ne visait qu'à tester une piste pour des recherches ultérieures. Il s'agissait d'appliquer à nos données une procédure « clé en main » intégrée dans un logiciel d'analyse statistique (« R ») et de tenter d'interpréter les résultats à partir d'une connaissance sommaire de la méthode³²⁵ et surtout des hypothèses formulées lors de l'analyse qualitative. Nous tenons à remercier Sylvain Loiseau pour son aide dans l'utilisation du logiciel et dans l'interprétation des résultats.

Pour effectuer notre test, nous avons utilisé les données concernant le nombre d'occurrences des trois types de ponctuation (forte, faible ou aucune) sur les quatre catégories de frontières ponctuables les plus répandues (propositions autonomes, propositions avec des éléments communs, principale/subordonnées et syntagmes coordonnés ou juxtaposés) dans chacun des textes de notre corpus.

L'association d'une force de la ponctuation et d'un type de frontière constitue une *variable* (ce qui donne un nombre total de 12 variables) et les textes du corpus sont les *individus* qui constituent la *population*.

Très grossièrement, le principe de l'analyse factorielle peut être résumé de la façon suivante :

L'analyse factorielle traite des tableaux de nombres et elle remplace un tableau difficile à lire par un tableau plus simple à lire qui soit une bonne approximation de celui-ci (Cibois 1983, p. 5).

³²⁵ Pour nous familiariser avec la méthode, nous avons consulté les ouvrages (Cibois 1983 ; Lemerrier & Zalc 2008 ; Baayen 2008). Nous nous sommes également inspiré de l'application des méthodes statistiques dans les recherches en linguistique diachronique (Prévost 2001b ; Mazziotta 2007b).

En d'autres termes, cette analyse permet de réduire (typiquement, jusqu'à 2) le nombre de dimensions de la représentation des données qui correspond initialement au nombre de variables prises en compte.

La démarche au fond de l'analyse consiste à calculer l'écart pondéré (« chi-2 ») entre les valeurs observées et les valeurs « attendues »³²⁶ (effectif théorique en cas d'indépendance entre les variables croisées) et à rechercher les facteurs (au sens mathématique) permettant d'apprécier la « distance » entre les différentes variables.

Cette analyse peut être considérée comme « un simple outil permettant d'éviter de produire, de lire ou de commenter des dizaines de tableaux croisés assortis de leur chi-2 » (Lemercier & Zalc 2008, p. 60). Les résultats d'une analyse factorielle sont en général fournis sous la forme d'un ou de plusieurs graphiques qui peuvent être interprétés de la façon suivante :

Pour simplifier, on peut dire que les modalités des variables que l'on retrouve souvent chez les mêmes individus (modalités liées statistiquement) sont représentées par des points proches sur le graphique, tandis que les modalités de variables que l'on ne retrouve que rarement associées apparaissent éloignées (*ibidem*, p. 61).

Le graphique fourni par le logiciel à partir de nos données est présenté dans la Fig. 76.

La représentation conventionnelle sur deux axes ne peut bien entendu pas faire apparaître la totalité des liens entre les modalités de toutes les variables analysées. L'analyse factorielle vise en effet à « mettre en avant ceux de ces liens qui sont les plus structurants, qui résument le mieux l'information contenue dans l'ensemble des données : d'autres éléments sont nécessairement perdus » (*ibidem*, p. 63). La mesure de cette perte est indiquée par le logiciel sous la forme du pourcentage de l'*inertie* (de l'information totale) pris en compte. Dans le cas de nos données ce pourcentage atteint 32% pour le facteur 1 et 28,5% pour le facteur 2.

La question de savoir si les « facteurs » mis en évidence par l'analyse factorielle correspondent à des variables existant réellement est difficile. Dans notre cas, on

326 La valeur « attendue » correspond à une distribution strictement homogène de l'une des variables du tableau en fonction d'une autre. La valeur attendue dans une cellule du tableau est la somme des valeurs de la ligne correspondante, multipliée par la somme des valeurs de la colonne correspondante, divisé par l'effectif total du tableau.

pourrait dire que le facteur 1 caractérise dans une certaine mesure l'opposition entre la ponctuation faible et forte, tandis que le facteur 2 met en avant l'opposition entre la présence et l'absence de la ponctuation. Plus la position d'une variable ou d'un individu (texte) est éloignée du croisement des axes, plus sa valeur est marquée.

Bien entendu, il faut interpréter les graphiques avec beaucoup de prudence : les distances affichées sont approximatives, les modalités pour lesquelles les résultats ne sont pas interprétables doivent être exclues. Néanmoins, même un premier regard sur les groupements visuels des textes et des variables permet d'apprécier l'intérêt de cette méthode d'analyse et de présentation des résultats.

En analysant la position des variables, on voit très clairement une opposition entre les frontières de propositions autonomes et les trois autres catégories. Sur la Fig. 77, nous avons dessiné des traits en couleur entre les points correspondant à chaque type de frontière ponctuable et avons ainsi obtenu quatre triangles (rouge pour les propositions autonomes, jaune pour les propositions avec des éléments communs, vert pour les frontières subordonnée/principale et bleu pour les syntagmes coordonnés ou juxtaposés). Les trois derniers triangles se superposent en grande partie tandis que le triangle rouge est orienté différemment. Cet effet graphique traduit le fait que pour les propositions autonomes l'absence de ponctuation est un phénomène marqué, tandis que la ponctuation forte y est ordinaire (le point se situe relativement près du centre du graphique). Pour les autres types de frontières la tendance est inverse : l'absence de ponctuation est normale, tandis que les ponctuations forte et faible sont inhabituelles.

Les regroupements des textes sont tout aussi révélateurs (Fig. 78). La plupart des textes se situent dans l'un des trois groupes distincts. Un groupe de quatre textes (entouré de bleu) se caractérise par la prédominance de la ponctuation forte qui peut être utilisée à tous les types de frontières. Le deuxième groupe composé de six textes est caractérisé par la préférence pour la ponctuation faible et par un taux de ponctuation général relativement élevé. Le troisième groupe est le plus nombreux (14 textes). Il se caractérise par un taux de ponctuation général peu élevé et par un relatif équilibre entre les ponctuations forte et faible. Les extraits **lancelot77** et **qgraal77** tirés d'un même manuscrit sont très proches, ce qui confirme en quelque sorte la pertinence de l'analyse.

Synthèse des tendances de la ponctuation

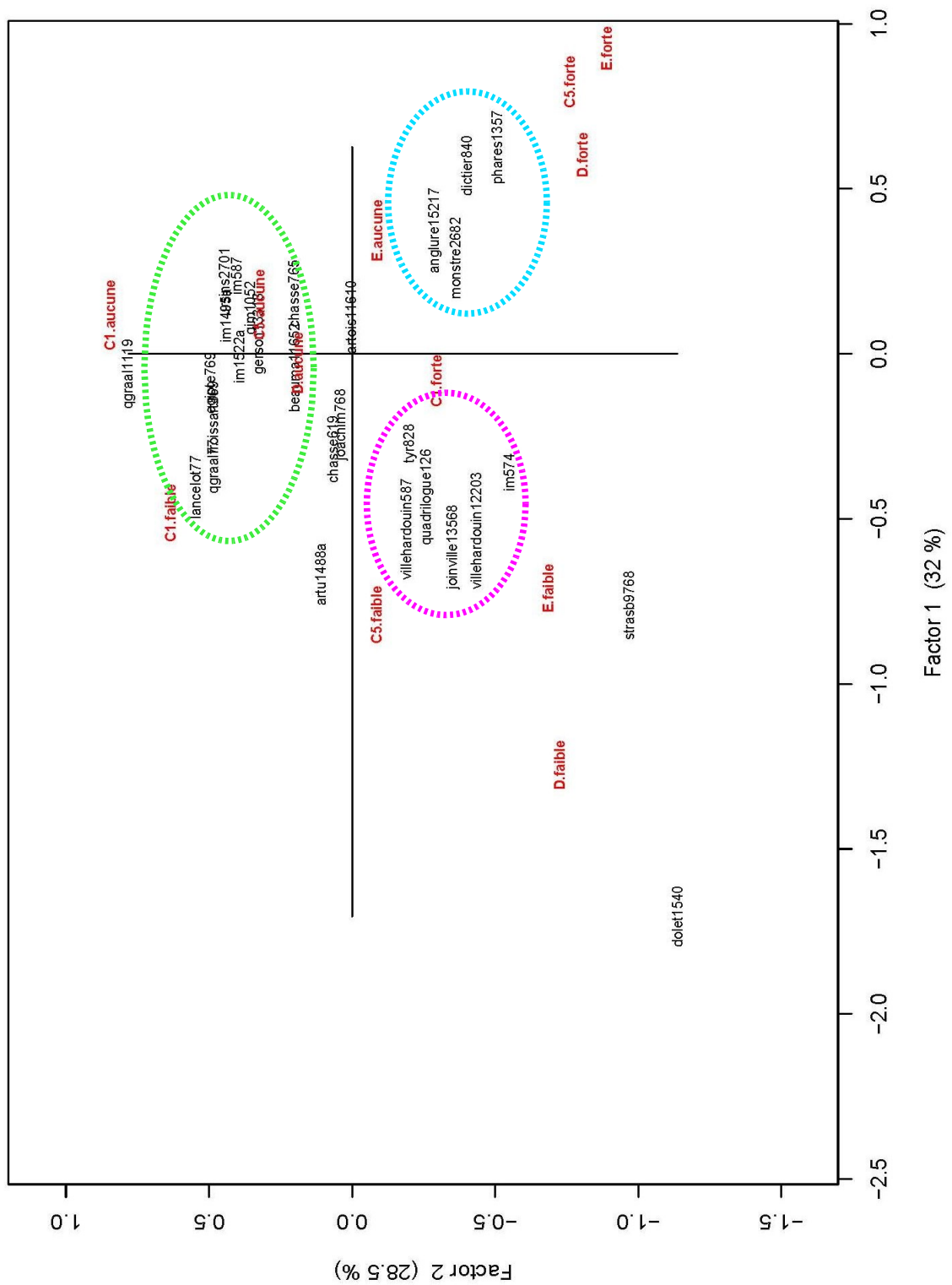


Fig. 78 : Analyse factorielle : groupements de textes

Quelques textes seulement échappent aux regroupements représentés par des cercles dans la Fig. 78. Trois d'entre eux (**artois11610**, **joachim768** et **chasse619**) se situent très près du centre du graphique, ce qui signifie qu'ils représentent des modalités qui ne sont pas concernées par les oppositions qui définissent les axes. Trois autres textes sont situés à l'écart des groupements, il s'agit de **artu1488a**, **strasb9768** et **dolet1540**. En ce qui concerne **artu1488a**, cet incunable présente un système de ponctuation où on commence à entrevoir les tendances de la normalisation qui s'opère dans les imprimés du XVI^e siècle. La ponctuation y est plus dense et plus diversifiée que dans **im1495a** et **im1522a** (qui se situent dans le groupe principal des manuscrits peu ponctués). En même temps, cet incunable reste pour l'essentiel dans les traditions de la ponctuation médiévale, ce qui se traduit dans le graphique par sa position qui n'est pas très éloignée de celle des groupes principaux. Les *Serments de Strasbourg* (**strasb9768**) occupent une position particulière dans notre corpus, du fait que le manuscrit est très ancien et rédigé essentiellement en latin. Le texte en question est trop bref (18 unités ponctuables seulement) pour que les données quantitatives puissent être réellement exploitables, mais le fait que ce texte très particulier du point de vue des variables externes se trouve à l'écart des autres textes sur le graphique est rassurant.

La position la plus excentrée sur le graphique est celle de **dolet1540**. Elle se caractérise par une affinité forte avec la variable « D-faible », ce qui correspond entièrement à la réalité observée (Dolet ponctue régulièrement les propositions subordonnées). Comme nous l'avons indiqué, le traité de Dolet constitue une véritable rupture vis-à-vis des pratiques de la ponctuation médiévale et sa position excentrée sur le graphique en est une excellente illustration.

Bien entendu, ces observations très impressionnistes devraient être vérifiées et précisées selon une méthode plus rigoureuse et sur un corpus beaucoup plus important et représentatif. Il sera notamment intéressant d'étudier les affinités entre les différentes variables de la typologie textuelle « externe » et les variables liées à l'usage de la ponctuation aux différents types de frontières syntaxiques. Nous pouvons néanmoins conclure que la piste de l'analyse factorielle se révèle très intéressante pour

Synthèse des tendances de la ponctuation

l'établissement de la typologie des ponctuations médiévale et que les recherches dans ce sens mériteraient d'être poursuivies.

Conclusion

Au terme de cette étude qui s'achève une trentaine années après les premières recherches linguistiques sur la ponctuation française médiévale et plus d'un demi-siècle après le célèbre appel de Mario Roques³²⁷, nous sommes toujours obligé de conclure qu'une histoire générale de la ponctuation médiévale en France mettant en évidence les tendances générales et les différents « courants » qui se sont formés au cours des siècles sous l'influence de multiples facteurs (ou *variables*) linguistiques et extralinguistiques... cette histoire reste à écrire.

Les données accumulées et analysées par les chercheurs (auxquelles nous espérons avoir apporté une modeste contribution) permettent certes d'entrevoir certaines « constantes » et certaines « variables », mais on n'arrive pas toujours à expliquer pourquoi, à une même époque, un scribe met des marques de ponctuation deux ou trois fois plus souvent qu'un autre...

L'étude de la ponctuation demande un travail minutieux sur les sources primaires (manuscrits ou imprimés anciens). Une personne seule ou une petite équipe de chercheurs ne pourra jamais rassembler et traiter suffisamment de données pour pouvoir procéder à des généralisations incontestables. Les technologies numériques qui permettent, si elles sont bien utilisées, de constituer des éditions et des corpus pérennes, réutilisables par d'autres, extensibles et interopérables, sont à notre avis le seul moyen pour dépasser le stade des hypothèses et pour dire qu'il ne reste plus de « taches

327 (Roques 1952, p. 196), *cf.* la section 1.3 pour la citation.

blanches » sur la « carte » de la ponctuation française médiévale, qui était une *terra incognita* il y a quelques dizaines d'années encore.

En particulier, si les nouvelles éditions de textes médiévaux acceptaient d'intégrer au moins deux niveaux de représentation des données primaires – que nous appelons *diplomatique* (ou encore *graphématique*) et *normalisé* – elles pourraient augmenter considérablement la base empirique des recherches et servir non seulement aux passionnés de la ponctuation, mais à bon nombre de linguistes qui veulent savoir si une graphie donnée provient du manuscrit ou si elle contient des lettres substituées par l'éditeur à une abréviation médiévale.

L'élaboration de la méthode d'une part, et la technologie de la représentation des données primaires des manuscrits dans des transcriptions électroniques d'autre part, ont représenté une part importante dans notre travail. Basés sur des standards internationaux et indépendants des logiciels propriétaires, nos protocoles de transcription et les transcriptions elles-mêmes, ainsi que les outils de leur exploitation, pourront, nous l'espérons, être utiles à la communauté des chercheurs et contribuer à l'évolution des pratiques éditoriales.

En ce qui concerne notre contribution à l'histoire de la ponctuation française, nous espérons que notre méthode d'analyse en *unités et frontières ponctuables*, qui allie une relative simplicité avec la prise en compte des divers facteurs pouvant expliquer l'apparition (ou non) d'une marque à un endroit donné dans un texte, pourra être utilisée pour l'annotation de corpus plus importants que le nôtre, et qu'elle permettra un jour d'établir une véritable typologie des pratiques de la ponctuation médiévales, et, au-delà, d'étudier ses rapports avec les différentes variables « externes » de la typologie textuelle. Une première tentative d'utilisation de l'analyse factorielle pour mettre en évidence les groupements de textes et les variables liées à l'usage de la ponctuation s'est révélée très intéressante.

Comme notre dernier chapitre (chapitre 9) présente une synthèse de nos observations sur les textes que nous avons étudiés et peut faire office sur ce point de conclusion, nous tenterons ici de passer à un niveau de généralisation plus élevé en prenant en compte les résultats des autres recherches publiées à ce jour, - sans oublier

Conclusion

néanmoins qu'il s'agit encore et toujours le plus souvent d'hypothèses qui restent à confirmer.

Le signe de ponctuation le plus souvent utilisé dans les manuscrits en français avant le XV^e siècle est un point placé à la mi-hauteur de la ligne. Il peut être suivi d'une majuscule (ponctuation forte) ou d'une minuscule (ponctuation faible). A partir de la fin du XIV^e siècle, dans certains manuscrits, ce rôle de « ponctuant universel » est rempli par la barre oblique. Le point et la barre oblique sont parfois utilisés dans un même manuscrit (et forment même parfois une marque « composite »), mais le plus souvent l'une des marques est largement dominante.

Nous considérons l'emploi d'une majuscule seule à une frontière ponctuable comme une marque de ponctuation forte. On rencontre des occurrences de la majuscule seule dans la majorité des manuscrits dès le XIII^e siècle. Dans certains manuscrits la fréquence de son emploi est comparable à celle de la marque dominante, et même, dans un manuscrit du XV^e siècle, c'est la majuscule seule qui devient la marque dominante (avec 73% de toutes les occurrences de la ponctuation).

Certains manuscrits présentent un usage dominant soit de la ponctuation forte, soit de la faible, mais dans la majorité des cas les deux « forces » de la ponctuation sont utilisées d'une façon relativement équilibrée.

A côté de la marque dominante, tous les manuscrits ou presque présentent des occurrences de marques « occasionnelles », dont la fréquence reste souvent très basse (ne dépasse pas quelques centièmes d.). Les échantillons restreints que nous avons utilisés sont peu adaptés à l'étude de ces marques peu fréquentes, qui jouent pourtant un certain rôle dans le système de la ponctuation³²⁸. Le *comma* (ou *punctus elevatus*) qui prend la forme d'un point surmonté d'une virgule renversée ou d'une petite barre oblique (< † > ou < / >), est le plus souvent suivi d'une minuscule et s'utilise parfois (jamais systématiquement)³²⁹ à des frontières syntaxiques importantes (entre des propositions assez longues, par exemple), indiquant un lien fort entre les deux unités

328 Nous nous demandons si ce n'est pas la taille des échantillons qui a conduit (Careri *et al.* 2001, p. xxxv) à conclure que la moitié des manuscrits étudiés ne présentaient qu'une seule marque de ponctuation.

329 Les auteurs de l'*Album des manuscrits du XIIIe siècle* (Careri *et al.* 2001, p. xxxv) signalent tout de même un manuscrit où le *comma* est la marque dominante.

séparées. Lors de la lecture, ce lien pouvait être signalé par une intonation montante (Marchello-Nizia 2007). Un autre usage typique du *comma*, après les interjections *ha* et *hé* au début du discours direct, ne contredit pas cette hypothèse, même si l'analyse rythmique des vers montre que le comma après les interjections ne pouvait pas correspondre à une pause (Llamas Pombo 2007a, p. 29). La forme même de la marque semble suggérer une interprétation « mélodique » (la voix qui monte), ce qui est confirmé par certains traités latins (Hubert 1972, p. 121), même si les recommandations de ces traités étaient peu suivies dans les manuscrits français.

Il existe également des manuscrits qui utilisent occasionnellement le point d'interrogation (ꝛ) en fin d'énoncés interrogatifs ou marqués d'une forte charge émotionnelle. La forme de cette marque est parfois difficile à différencier de celle du *comma*. Son usage n'est jamais systématique.

D'autres marques, comme le *periodus* (ou *punctus versus*), marque ressemblant au point-virgule moderne < ; > et destinée selon les traités de l'époque à marquer la « pause finale », sont extrêmement rares dans les manuscrits français à partir du XIII^e siècle³³⁰. Des marques « exceptionnelles », formées par exemple de plusieurs points, apparaissent parfois dans des manuscrits isolés.

Par ailleurs, un symbole appelé en France le « pied-de-mouche »³³¹ (¶), le plus souvent tracé en bleu ou en rouge, peut être utilisé comme une marque de division textuelle ou de « mise en scène » des dialogues de personnages³³². Cette marque occupe une position transitoire entre la ponctuation « syntaxique » et la structuration globale du texte.

L'usage des marques de ponctuation est de façon générale moins dense dans les manuscrits médiévaux et dans les incunables français que dans les textes modernes, mais la variation du « taux de ponctuation » est assez importante dans les textes anciens (de 2 à 14 marques de ponctuation pour 100 occurrences-mots). La faible fréquence de la ponctuation est en partie compensée par une « abondance » de mots-connecteurs (surtout de la conjonction *et*).

330 On en trouve en revanche un certain nombre dans les manuscrits les plus anciens (Baddeley 2001).

331 *Paraph* pour M.B. Parkes (1992) et *paraffo* pour les auteurs de (Mortara Garavelli éd. 2008).

332 « Rubrique énonciative » (Llamas Pombo à paraître).

Conclusion

Il est pratiquement toujours possible d'expliquer l'usage d'une marque de ponctuation par la présence d'une frontière syntaxique ou pragmatique d'un certain type. C'est entre des propositions autonomes (juxtaposées ou coordonnées) qu'on a le plus de chances de trouver une marque de ponctuation. Les frontières de propositions ayant des éléments communs, de propositions subordonnées et de syntagmes coordonnés ou juxtaposés à l'intérieur d'une proposition sont nettement moins souvent ponctuées. Des facteurs sémantiques (mise en évidence des noms propres et d'informations importantes), pragmatiques (ponctuation « de respect » mettant en relief les références à des personnalités importantes) ou esthétiques (ponctuation servant à justifier les fins de ligne ou à décorer certains éléments sur la page) interviennent souvent dans ces cas.

Comme nous pensons l'avoir montré, une étude contrastive de la ponctuation du manuscrit et de celle introduite dans le même texte par un éditeur scientifique moderne peut aussi conduire à des résultats intéressants : la spécificité du système ancien apparaît plus clairement ainsi. Par ailleurs on peut évaluer dans quelle mesure la structuration du texte est le fruit des choix d'interprétation du texte par son éditeur.

Notre étude a montré entre autres choses que, paradoxalement, certaines éditions du XIX^e et du début du XX^e siècle, parfois très interventionnistes au niveau des graphies et des lexies, reproduisent la ponctuation des manuscrits plus fidèlement que les éditions plus récentes. Ainsi, dans les éditions des *Coutumes de Beauvaisis* de Philippe de Beaumanoir (éd. Amédée de Salmon, 1899) et de la *Chronique* de Monstrelet (éd. Louis Douët d'Arcq, 1857), la ponctuation de l'éditeur correspond à celle du manuscrit dans 48% des occurrences, tandis que dans des éditions plus récentes ce taux se situe entre 27 et 38% (Lavrentiev 2007a).

Notre analyse des textes du corpus et des études précédentes de la ponctuation médiévale permet de conclure que la ponctuation était présente sous une certaine forme dans tous les manuscrits français médiévaux. De façon générale, la ponctuation apparaît en premier lieu aux frontières des unités de structure textuelle de haut niveau (unités thématiques, propositions indépendantes). La variation dans les pratiques de ponctuation est liée à une multitude de facteurs, mais des traits communs semblent se

dégager dans les manuscrits de certaines époques, de certains types de textes et de styles d'écriture.

Une étude plus étendue, sur un corpus plus large, de la ponctuation et de l'ensemble des systèmes graphiques des manuscrits et incunables français médiévaux est nécessaire pour confirmer ces résultats, et ainsi pour mieux comprendre le fonctionnement de la langue écrite de l'époque. Une autre voie pour poursuivre ces recherches, voie fort possible et qui nous paraît nécessaire, est la prise en compte de la perspective pan-romane des pratiques de la ponctuation médiévale. Mais il s'agit là d'une ambition à bien plus long terme...

Bibliographie

Travaux cités

- ABNEY, Steven P. (1991). « Parsing by chunks », in : Robert Berwick *et al.* (éds.). *Principle based parsing*, Dordrecht : Kluwer Academic Publishers, p. 257-278.
[Également disponible en ligne] <<http://www.vinartus.net/spa/90e.pdf>>.
- ACHER, Jean (1913). « Sur l'x finale des manuscrits : Lettre ouverte à M. le Directeur de la "Romania" », *Revue des langues romanes*, vol. 56, 1913, p. 148-158.
- ADAM, Jean-Michel (1992). *Les Textes : types et prototypes. Récit, description, argumentation, explication et dialogue*, Paris : Nathan, 223 p.
- (1999). *Linguistique textuelle : Des genres de discours aux textes*, Paris : Nathan, 208 p.
- ANDRIEUX-REIX, Nelly et MONSONÉGO, Simone (1997). « Écrire les phrases au Moyen Âge. Matériaux et premières réflexions pour une étude des segments graphiques observés dans des manuscrits français médiévaux », *Romania*, vol. 115, n° 459-460, 1997, p. 289-336.
- (1998). « Les unités graphiques du français médiéval : mots et syntagmes, des représentations mouvantes et problématiques », *Langue française*, n° 119, 1998, p. 30-21.

- ANIS, Jacques (1983). « Pour une graphématique autonome », *Langue française*, n° 59, 1983, p. 31-44. [Également disponible en ligne] <<http://www.persee.fr>>.
- (1988). « Pour une graphématique autonome », in : Nina Catach (éd.). *Pour une théorie de la langue écrite*, Paris : Éditions du CNRS, p. 213-223.
- (2004). « Les linguistes français et la ponctuation », *L'Information grammaticale*, n° 102, juin 2004, p. 5-10.
- ANIS, Jacques, CHISS, Jean-Louis et PUECH, Christian (1988). *L'écriture : théories et descriptions*, Bruxelles : De Boeck, 252 p.
- ANTOINE, Gérald (1958). *La coordination en français*, Paris : Éditions d'Artrey, 701 p.
- BAAYEN, R. H. (2008). *Analyzing linguistic data : a practical introduction to statistics using R*, Cambridge, UK ; New York : Cambridge University Press, 353 p. [Également disponible en ligne] <<http://www.msu.edu/course/lin/875/BaayenCUPstats.pdf>>.
- BADDELEY, Susan (1998). « Théorie et pratique de la segmentation graphique dans les textes français du premier tiers du XVI^e siècle », *Langue française*, n° 119, Septembre 1998, p. 52-68.
- (2001). « La ponctuation de manuscrits français du IX^e au XII^e siècle », *Liaisons HÉSO/AIROÉ*, n° 32-33, Septembre 2001, p. 139-149.
- BADDELEY, Susan et BIEDERMANN-PASQUES, Liselotte (2004). « Histoire des systèmes graphiques du français à travers des manuscrits et des incunables (IX^e - XV^e siècle). Segmentation graphique et faits de langue », *Revue de linguistique romane*, n° 269-270, 2004, p. 181-201.
- BAKER, Alfred T. (1937). « Le futur des verbes *avoir* et *savoir* », *Romania*, vol. 63, n° 249, 1937, p. 1-30.

Bibliographie

- BARBANCE, Céline (1995). « La ponctuation médiévale : quelques remarques sur cinq manuscrits du début du XV^e siècle », *Romania*, vol. 113, n° 455-456, 1995, p. 505-525.
- BAUDE, Olivier, BLANCHE-BENVENISTE, Claire, CALAS, Marie-France, CAPPEAU, Paul, CORDEREIX, Pascal, GOURY, Laurence, JACOBSON, Michel, DE LAMBERTERIE, Isabelle, MARCHELLO-NIZIA, Christiane et MONDADA, Lorenza (2006). *Corpus oraux : Guide des bonnes pratiques 2006*, Olivier Baude (éd.), Orléans : Presses Universitaires d'Orléans ; CNRS Éditions, 203 p.
- BEAULIEUX, Charles (1927). *Histoire de l'orthographe française. Tome second. Les accents et autres signes auxiliaires*, Paris : Champion, 134 p.
- BECKER, Siegfried (1972). *Untersuchungen zur Redaktion der Straßburger Eide*, Bern; Frankfurt/M. : Herbert Lang ; Peter Lang, 154 p.
- BÉDIER, Joseph (1928). « La tradition manuscrite du *Lai de l'Ombre*, réflexions sur l'art d'éditer les anciens textes », *Romania*, vol. 54, 1928, p. 161-196 ; 236-356.
- BELTRAN, Évencio (1985). « Un traité inconnu de Guillaume Fichet sur la ponctuation », *Scriptorium*, vol. 39, n° 2, 1985, p. 284-291.
- BIBER, Douglas (1989). « A typology of English texts », *Linguistics*, vol. 27, 1989, p. 3-43.
- (1990). « Methodological issues regarding corpus-based analyses of linguistic variation », *Literary and Linguistic Computing*, vol. 19, n° 2, 1990, p. 257-270.
- (1993). « Using register-diversified corpora for general language studies », *Computational linguistics*, vol. 19, n° 2, 1993, p. 243-258.
- BIBER, Douglas, CONRAD, Susan P. et REPPEN, Randi (1998). *Corpus linguistics: investigating language structure and use*, Cambridge : Cambridge University Press, 300 p.

- BIEDERMANN-PASQUES, Liselotte (1998). « Des segmentations particulières d'un incunable (1488) à l'écriture du français en unités lexicales et grammaticales », *Langue française*, n° 119, 1998, p. 69-87.
- (2001). « Approche du système graphique de la *Séquence de sainte Eulalie* (deuxième moitié du IX^e siècle) », in : Isabel Uzcanga Vivar *et al.* (éds.). *Presencia y renovación de la lingüística francesa*, Salamanca : Universidad de Salamanca, p. 25-39.
- BISCHOFF, Bernhard (1993). *Paléographie de l'Antiquité romane et du Moyen Âge occidental*, Hartmut Atsma et Jean Vezin (trad.), Paris : Picard, 325 p.
- BnF [Bibliothèque nationale de France] (1955). *Les manuscrits à peintures en France du XIII^e au XVI^e siècle*, Paris : Bibliothèque nationale, 190 p.
- BONDARKO, Aleksandr (1991). *Functional grammar: a field approach*, Amsterdam : John Benjamins, 207 p.
- BOSSUAT, Robert (1951). *Manuel bibliographique de la littérature française du Moyen âge*, Melun : Librairie d'Argences, 638 p.
- BOUDET, Jean-Patrice (éd.) (1999). *Le Recueil des plus célèbres astrologues de Simon de Phares. Tome II. Présentation et commentaire*, Paris : H. Champion, 491 p.
- BOURGAIN, Pascale (1989). « Qu'est-ce qu'un vers au Moyen Âge ? », *Bibliothèque de l'École des chartes*, n° 147, 1989, p. 231-282. [Également disponible en ligne] <<http://www.persee.fr/>>.
- BOURGAIN, Pascale et VIELLIARD, Françoise (coord.) (2002). *Conseils pour l'édition des textes médiévaux. Fascicule III, Textes littéraires*, Paris : CTHS, École nationale des chartes, 253 p.
- BRANCA-ROSOFF, Sonia (1996). « Retour aux genres », in : Sylvain Auroux *et al.* (éds.). *Histoire et grammaire du sens. Hommage à Claude Chevalier*, 4, Paris : Armand Colin, Coll. Collection U. Série Linguistique, p. 189-203.
- British Library (1990). *French books before 1601*, Kentucky; Watertown, Mass. : University of Kentucky Libraries, Microfilm Center; General Microfilm Co.

Bibliographie

- BRONCKART, Jean-Paul (1996). « Genres de textes, types de discours et opérations discursives », *Enjeux*, n° 37-38, 1996, p. 31-47.
- BURIDANT, Claude (1980). « Le strument *et* et ses rapports avec la ponctuation dans quelques textes médiévaux », in : Anne-Marie Dessaux-Berthonneau (éd.). *Théories linguistiques et traditions grammaticales*, Villeneuve-d'Asq : Presses Universitaires de Lille, p. 13-53.
- (2007). *Grammaire nouvelle de l'ancien français*, Paris : SEDES, 800 p.
- BUSBY, Keith (1993). « *Variance and the Politics of Textual Criticism* », in : *Towards a Synthesis?: Essays on the New Philology*, Amsterdam : Rodopi, p. 29-45.
- CAPPELLI, Adriano (¹1949). *Lexicon abbreviaturarum : dizionario di abbreviature latine ed italiane usate nelle carte e codici, specialmente del Medio-evo*, Milan : U. Hoepli, 4^e éd. 531 p.
- CARERI, Maria (2008). « Manoscritti provenzali e francesi. Dalle origine alla fine del XIII secolo », in : Bice Mortara Garavelli (éd.). *Storia della punteggiatura in Europa*, Roma : Laterza, p. 213-232.
- CARERI, Maria, FERY-HUE, Françoise, GASPARRI, Françoise, HASENOHR, Geneviève, LABORY, Gillette, LEFÈVRE, Sylvie, LEURQUIN, Anne-Françoise et RUBY, Christine (2001). *Album de manuscrits français du XIII^e siècle. Mise en page et mise en texte*, Rome : Viella, 238 p.
- CARTON, Fernand (²1974). *Introduction à la phonétique du français*, Paris : Bordas, 2^e éd. revue et corrigée, 250 p.
- CATACH, Nina (éd.) (1980a). *La ponctuation, Langue française*, n° 45, Paris : Larousse, 150 p.
- (1980b). « La ponctuation », *Langue française*, n° 45, 1980, p. 16-27.
- (éd.) (1988). *Pour une théorie de la langue écrite*, Paris : Éditions du CNRS, 259 p.
- (1994). *La ponctuation*, Paris : Presses universitaires de France, 127 p.

- CATACH, Nina et TOURNIER, Claude (éds.) (1979). *La ponctuation. Recherches historiques et actuelles. Fascicule 2. Actes de la Table ronde internationale C.N.R.S. de mai 1978*, Paris ; Besançon : C.N.R.S. - Groupement de recherches sur les textes modernes, 293 p.
- CERQUIGLINI, Bernard (1989). *Éloge de la variante. Histoire critique de la philologie*, Paris : Seuil, 122 p.
- CIBOIS, Philippe (1983). *L'analyse factorielle*, Paris : Presses universitaires de France, 128 p.
- CID [Commission internationale de diplomatique] (1997). *Vocabulaire international de la diplomatique*, València : Universitat de València, 308 p.
- CID et CIS [Commission internationale de diplomatique et Commission internationale de sigillographie] (1984). *Diplomatica et Sigillographica : travaux préliminaires de la Commission internationale de diplomatique et de la Commission internationale de sigillographie*, Zaragoza : Cátedra Zurita, Institución Fernando el Católico, 221 p.
- COLIGNON, Jean-Pierre (1975). *La ponctuation : art et finesse*, Paris : J.-P. Colignon, Impr. Lecoœur, 95 p.
- COLOMBO, Maria (2008). « Il cinquecento », in : Bice Mortara Garavelli (éd.). *Storia della punteggiatura in Europa*, Roma : Laterza, p. 242-254.
- COMBETTES, Bernard (2000). « La ponctuation et l'énoncé complexe au XVI^e siècle (l'usage des deux points chez Jean de Léry) », *La Licorne*, n° 52, 2000, p. 63-79.
- CONNOCHIE-BOURGNE, Chantal (1999). *L'Image du monde, une encyclopedie du XIIIe siecle. Édition critique et commentaire de la première version*. Thèse de Doctorat d'État non publiée, Paris, Université de Paris IV, 1138 p.
- COUSIN, Jean (1975). *Quintilien. Institution oratoire. Tome I. Livre I*, Paris : Les Belles Lettres, 187 p.
- (1979). *Quintilien. Institution oratoire. Tome VI. Livres X et XI*, Paris : Les Belles Lettres, 375 p.

Bibliographie

- COVILLE, Alfred (1935). *Recherches sur quelques écrivains du XIV^e et du XV^e siècle*, Paris : Droz, 340 p.
- CROFT, William (2002). *Typology and universals*, New York : Cambridge University Press, 341 p.
- CURTIUS, Ernst R. (éd.) (1911). *Li quatre livres des reis*, Dresden : Gesellschaft, XCV-244 p.
- DAMOURETTE, Jacques (1939). *Traité moderne de ponctuation*, Paris : Larousse, 144 p.
- DAVISS, Bennett (1999). « Paper goes electric », *New scientist*, n° 2186, publié en ligne le 15/05/1999, 15 octobre 2007. (Consulté le 15/01/2008).
- DE CRÉCY, Marie-Claude (1997). *Le roman de Ponthus et Sidoine*, Genève : Droz, 459 p.
- DE WAILLY, Natalis (1838). *Éléments de paléographie*, Paris : Imprimerie Royale, 716 p.
- DEES, Anthonij (1987). *Atlas des formes linguistiques des textes littéraires de l'ancien français*, Tübingen : Max Niemeyer Verlag, 684 p.
- DELISLE, Léopold (1852). « Notice sur l'origine des *i* pointés », *Bibliothèque de l'École des chartes*, vol. 3, 1852, p. 563-564.
- DEMONET, Marie-Luce (2000). « Ponctuation et narration chez Rabelais et ses contemporains », *La Licorne*, n° 52, 2000, p. 37-62.
- DESBORDES, Françoise (1988). « La prétendue confusion de l'écrit et de l'oral dans les théories linguistiques de l'Antiquité », in : Nina Catach (éd.). *Pour une théorie de la langue écrite*, Paris : Éditions du CNRS, p. 27-33.
- (1990). *Idées romaines sur l'écriture*, Villeneuve-d'Ascq : Presses universitaires de Lille, 295 p.
- DOPPAGNE, Albert (1978). *La Bonne ponctuation*, Paris : Duculot, 112 p.
- DRILLON, Jacques (1991). *Traité de la ponctuation française*, Paris : Gallimard, 472 p.

- DUBOIS, Jean, GIACOMO, Mathée, GUESPIN, Louis, MARCELLESI, Christiane, MARCELLESI, Jean-Baptiste et MÉVEL, Jean-Pierre (1999). *Dictionnaire de linguistique et des sciences du langage*, Paris : Larousse, 514 p.
- DÜRRENMATT, Jacques (2004). « La virgule entre sujet et verbe : petite histoire d'un emploi oublié », *L'Information grammaticale*, n° 102, juin 2004, p. 31-34.
- FAULHABER, Charles B. (1991). « Textual criticism in the 21st century », *Romance philology*, vol. 45, n° 1, 1991, p. 123-148.
- FAYOL, Michel (1989). « Une approche psycholinguistique de la ponctuation. Étude en production et compréhension », *Langue française*, vol. 81, 1989, p. 21-39.
- FERARI, Barbara (2008). « I secoli XIV e XV », in : Bice Mortara Garavelli (éd.). *Storia della punteggiatura in Europa*, Roma : Laterza, p. 233-242.
- FILLMORE, Charles (1992). « 'Corpus linguistics' or 'Computer-aided armchair linguistics' », in : Jan Svartvik (éd.) *Directions in Corpus Linguistics. Proceedings from a 1992 Nobel Symposium on Corpus Linguistics*, (Actes de colloque, Stockholm, 1992), Berlin : Mouton de Gruyter, p. 35-60.
- FOERSTER, Wendelin, KOSCHWITZ, Eduard et HILKA, Alfons (1932). *Altfranzösisches Übungsbuch zum Gebrauch bei Vorlesungen und Seminarübungen*, Leipzig : Reiland, 7^e éd. 326 p.
- FOULET, Alfred et SPEER, Mary B. (1979). *On Editing Old French Texts*, Lawrence : The Regents Press of Kansas, 136 p.
- FOULET, Alfred et UETTI, Karl D. (éds.) (1989). *Chrétien de Troyes. Le chevalier de la charrette (Lancelot)*, Paris : Bordas, Coll. Classiques Garnier, 409 p.
- GALLY, Michèle et MARCHELLO-NIZIA, Christiane (1985). *Littératures de l'Europe médiévale*, Paris : Magnard, 598 p.
- GEE, James P. et GROSJEAN, François (1983). « Performance structures: a psycholinguistic end linguistic appraisal », *Cognitive psychology*, vol. 15, n° 3, 1983, p. 411-458.

Bibliographie

- GEYMONAT, Mario (2008). « Grafia e interpunzione nell'antichità greca e latina, nella cultura bizantina e nella latinità medievale », in : Bice Mortara Garavelli (éd.). *Storia della punteggiatura in Europa*, Roma : Laterza, p. 25-62.
- GREVISSE, Maurice et GOOSSE, André (¹⁴2007). *Le bon usage : grammaire française*, Bruxelles ; Louvain-la-Neuve : Duculot ; de Boeck, 1600 p.
- GUILLOT, Céline, HEIDEN, Serge et LAVRENTIEV, Alexei (2007). « Typologie des textes et des phénomènes linguistiques pour l'analyse du changement linguistique avec la Base de Français Médiéval », *LINX*, n° spécial, 2007, p. 125-139.
- GUILLOT, Céline et LAVRENTIEV, Alexei (2007). *Manuel de description de textes pour la Base de Français Médiéval*, v. 1.2, Lyon : Projet BFM, 11 p. [En ligne] <http://ccfm.ens-lsh.fr/IMG/pdf/Manuel_Descripteurs_BFM_v1.2.pdf> (Consulté le 2 mai 2009).
- GUYOTJEANNIN, Olivier (coord.) (2001). *Conseils pour l'édition des textes médiévaux, Fascicule II, Actes et documents d'archives*, Paris : CTHS, École nationale des chartes, 175 p.
- GUYOTJEANNIN, Olivier, PYCKE, Jacques et TOCK, Benoît-Michel (1993). *Diplomatique médiévale*, Turnhout : Brepols, 442 p.
- HABERT, Benoît (2000). « Des corpus représentatifs : de quoi, pour qui, comment ? », *Cahiers de l'Université de Perpignan*, n° 31, 2000, p. 11-58. [Également disponible en ligne] <<http://www.limsi.fr/Individu/habert/Publications/Fichiers/habert00b.ps.gz>>.
- HASENOHR, Geneviève (1998). « Abréviations et frontières de mots », *Langue française*, n° 119, Septembre 1998, p. 24-29.
- HASENOHR, Geneviève, ZINK, Michel, BOSSUAT, Robert, PICHARD, Louis et RAYNAUD DE LAGE, Guy (éds.) (²1994). *Dictionnaire des lettres françaises. Le Moyen Âge*, Paris : Fayard, 1506 p.
- HASPELMATH, Martin (2002). *Understanding morphology*, London : Arnold, 290 p.

HAUGEN, Odd E. (2004). « Parallel Views: Multi-level Encoding of Medieval Nordic Primary Sources », *Literary and Linguistic Computing*, vol. 19, n° 1, 2004, p. 73-91.

--- (éd.) (2006a). *MUFI character recommendation. Characters in the official Unicode Standard and in the Private Use Area for Medieval texts written in the Latin alphabet. Part 1: Alphabetical order. Version 2.0*, Bergen : Medieval Unicode Font Initiative, 156 p.

--- (éd.) (2006b). *MUFI character recommendation. Characters in the official Unicode Standard and in the Private Use Area for Medieval texts written in the Latin alphabet. Part 2: Code-chart order. Version 2.0*, Bergen : Medieval Unicode Font Initiative, 200 p.

--- (éd.) (2008). *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources. Version 2.0*, Bergen : Medieval Nordic Text Archive, [En ligne] <<http://www.menota.org/guidelines>>

HEIDEN, Serge et LAVRENTIEV, Alexei (2004). « Ressources électroniques pour l'étude des textes médiévaux : approches et outils », *Revue française de linguistique appliquée*, vol. 9, n° 1, juin 2004, Dossier : Linguistique et informatique : nouveaux défis. p. 99-118.

HEIDEN, Serge et GUILLOT, Céline (2003). « Capitalisation des savoirs par le web : une application de la TEI pour l'encodage et l'exploitation des textes de la Base de Français Médiéval », in : Pierre Kunstmann *et al.* (éds.). *Ancien et moyen français sur le Web : enjeux méthodologiques et analyse du discours [Actes du colloque d'Ottawa, 4-5 oct. 2002]*, Ottawa : Éditions David, p. 77-92.

HENRY, Albert (1953). *Chrestomathie de la littérature en ancien français*, Berne : A. Francke, 175 p.

HOLTUS, Günter (1998). « Rilievi su un'edizione comparatistica dei *Giuramenti du Strasburgo* », in : József Herman (éd.). *La transizione dal latino alle lingue romanze. Atti della Tavola Rotonda di linguistica storica*, Università Ca'

Bibliographie

- Foscari di Venezia, 14-15 giugno 1996*, Tübingen : Max Niemeyer Verlag, p. 195-212.
- HOLTUS, Günter, OVERBECK, Anja et VÖLKER, Harald (2003). *Luxemburgische Skriptastudien. Edition und Untersuchung der altfranzösischen Urkunden Gräfin Ermesindes (1226-1247) und Graf Heinrichs V. (1247-1281) von Luxemburg*, Tübingen : Max Niemeyer Verlag, 699 p.
- HUBERT, Marcel (1970). « Corpus stigmatologicum minus », *Bulletin du Cange, Archivum latinitatis Medii Ævi*, n° 37, 1970, p. 5-171.
- (1972). « Le vocabulaire de la “ponctuation” aux temps médiévaux : Un cas d’incertitude lexicale », *Bulletin du Cange, Archivum latinitatis Medii Ævi*, n° 38, 1972, p. 57-167.
- IPA [International Phonetic Association] (1999). *Handbook of the International Phonetic Association : A Guide to the Use of the International Phonetic Alphabet*, Cambridge : Cambridge University Press, 204 p.
- JAKOBSON, Roman (1963). *Essais de linguistique générale. Les fondations du langage*, Nicolas Ruwet (trad.), Paris : Les éditions de Minuit, 258 p.
- JAUSS, Hans R., KÖHLER, Erich, DELBOUILLE, Maurice et GUMBRECHT, Hans U. (1972). *Grundriss der romanischen Literaturen des Mittelalters*, Heidelberg : Carl Winter Universitätsverlag, 696 p.
- JAYEZ, Jacques (1988). *L’inférence en langue naturelle. Le problème des connecteurs. Représentation et calcul*, Paris ; Londres : Hermès, 287 p.
- KEIL, Heinrich (1981). *Grammatici latini. IV. Probi Donati Servii qui feruntur de arte grammatica libri*, Hildesheim : G. Olms, reproduction de l’édition de 1864, 613 p.
- KENNEDY, Elspeth (1970). « The scribe as editor », in : *Mélanges de langue et de littérature du Moyen Age et de la Renaissance offerts à Jean Frappier*, Genève : Droz, p. 523-531.

- KOCH, Peter (1993). « Pour une typologie conceptionnelle et médiale des plus anciens documents/monuments des langues romanes », in : Maria Selig *et al.* (éds.). *Le passage à l'écrit des langues romanes*, Tübingen : Gunter Narr, p. 39-82.
- KOCH, Peter et OSTERREICHER, Wulf (1985). « Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte », *Romanistisches Jahrbuch*, vol. 36, 1985, p. 15-43.
- (2001). « Langage parlé et langage écrit », in : Günter Holtus *et al.* (éds.). *Lexikon der Romanistischen Linguistik*, Tübingen : Max Niemeyer, p. 584-627.
- KOSCHWITZ, Eduard (¹1920). *Les plus anciens monuments de la langue française publiés pour les cours universitaires*, Leipzig : O.R. Reisland, 53 p.
- LALLOT, Jean (²1998). *La grammaire de Denys le Thrace*, Paris : CNRS éditions, 2^e éd. rev. et augm. 308 p.
- LAUER, Philippe (éd.) (1926). *Nithard. Histoire des fils de Louis le Pieux*, Coll. Classiques de l'histoire de France au Moyen Âge, 172 p.
- LAVRENTIEV, Alexei (2000). « A propos de la ponctuation dans l'*Image du monde* », *La Licorne*, n° 52, 2000, p. 23-37.
- (2005). « Représentation de transcriptions diplomatiques de manuscrits français médiévaux en XML-TEI », in : Claus D. Pusch *et al.* (éds.). *Romanistische Korpuslinguistik II. Romance Corpus Linguistics. Korpora und diachrone Sprachwissenschaft. Corpora and Diachronic Linguistics*, Tübingen : Gunter Narr Verlag, p. 109-121.
- (2007a). « Base de français médiéval et transcriptions de manuscrits : recherche de complémentarité », in : David Trotter (éd.) *Actes du XXIV^e Congrès International de Linguistique et de Philologie Romanes*, (Actes de colloque, Aberystwyth, 1-6 août 2004, Tübingen : Max Niemeyer, p. 405-410.
- (2007b). « Connecteurs et ponctuation comme outils de structuration du texte à travers les manuscrits et incunables français en prose du XIII^e au début du XVI^e siècle », in : Anne Vanderheyden *et al.* (éds.). *Texte et discours en Moyen*

Bibliographie

- français, Actes du XI^e Colloque international sur le moyen français*, Turnhout : Brepols, Coll. Texte, codex & contexte, p. 149-162.
- (2007c). « Pour une méthodologie d'étude de la ponctuation médiévale basée sur une approche typologique », in : Bernard Combettes et Christiane Marchello-Nizia (éds.) *Études sur le changement linguistique en français, communications du Colloque « Diachro 2 »*, (Actes de colloque, Paris, 15, 16, 17 janvier 2004), Nancy : Presses Universitaires de Nancy, p. 191-204.
- (dir.) (2007d). *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies*, Chambéry : Université de Savoie, 151-66 p.
- (2007e). « Typologie textuelle pour l'étude linguistique de manuscrits français médiévaux », in : Alexei Lavrentiev (éd.). *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies*, Chambéry : Université de Savoie, Coll. Langages, p. 49-66.
- (2008). « A Tag for Punctuation? », in : 8 novembre 2008), Londres : TEI, [En ligne]
<http://www.cch.kcl.ac.uk/cocoon/tei2008/programme/workingpapers/wp_002.html> (Consulté le 2 mai 2009).
- LAZARD, Gilbert (1994). *L'actance*, Paris : Presses universitaires de France, 285 p.
- LEE, David (2001). « Genres, registers, text types, domains and styles: clarifying the concepts and navigating a path through the BNC jungle », *Language Learning & Technology*, vol. 5, n° 3, 2001, p. 37-72.
- LEEUEW VAN WEENEN, Andrea de (2000). *A grammar of Möðruvallabók*, Leiden : Universiteit Leiden, 366 p.
- LEMERCIER, Claire et ZALC, Claire (2008). *Méthodes quantitatives pour l'historien*, Paris : La Découverte, 120 p.
- LEWIS, Peter S. et HAYEZ, Anne-Marie (1978). *Écrits politiques de Jean Juvénal des Ursins. Tome I*, Paris : Klincksieck, 551 p.

- LI, Huei-Chen (2007). *Découpage et structuration du texte : Lettrines, Majuscules, Blancs et autres signes de ponctuation dans les versions manuscrites et imprimée du Roman de Perceforest : Étude comparative*, Thèse de Doctorat non publiée, Strasbourg, Université Marc Bloch - Strasbourg 2, 400 p.
- LLAMAS POMBO, Elena (1996a). « Écriture et oralité : ponctuation, interprétation et lecture des manuscrits français de textes en vers (XIII^e-XV^e s.) », in : Emilia Aloinsi *et al.* (éds.). *La linguistique française : grammaire, histoire et épistémologie*, Seville : Grupo Analuz de pragmática, p. 133-144.
- (1996b). *Escritura y oralidad en los Ovidiana franceses des siglo XII*, Thèse de Doctorat non publiée, Université de Salamanque, 407 p.
- (2001). « La ponctuation du vers dans un manuscrit du XIV^e siècle », *Liaisons HSO/AIRO*, n° 32-33, Septembre 2001, p. 151-171.
- (2003). « Séquences graphiques dans le *Roman de Fauvel* (Ms. BnF, f. fr. 146) », *Thélème. Revista compultense des estudios franceses*, n° numéro spécial, 2003, p. 225-236.
- (2007a). « Réflexions méthodologiques pour l'étude de la ponctuation médiévale », in : Alexei Lavrentiev (éd.). *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies*, Chambéry : Université de Savoie, Coll. Langages, p. 11-48.
- (2007b). « Réflexions sur la variation graphique dans l'écrit médiéval (XIII^e s.) », in : *Littérature et linguistique Ressource électronique. Autour des travaux de Michèle Perret*, Chambéry : Université de Savoie, Coll. Langages, p. 637-651.
- (2008). « Ponctuer, éditer, lire. État des études sur la ponctuation dans le livre manuscrit », *Syntagma. Revista del Instituto de Historia del libro y de la lectura*, n° 2, 2008, p. 129-171.

Bibliographie

- (à paraître). « Marques graphiques du discours rapporté (Manuscrits du *Roman de la Rose*, XV^e siècle) », in : *Actes du colloque « Diachro 4 », Madrid, 22-24 octobre 2008*,
- LODGE, Anthony R. (1997). *La français : Histoire d'un dialecte devenu langue*, Cyril Veken (trad.), Paris : Fayard, 382 p.
- LOWRY, Martin (1979). *The world of Aldus Manutius: business and scholarship in Renaissance Venice*, Sheila Mooney et François Dupuigrenet Desroussilles (trad.), Ithaca, N.Y. : Cornell University Press, 350 p.
- LUSCHER, Jean-Marc (1994). « Les marques de connexion : des guides pour l'interprétation », in : Jacques Moeschler *et al.* (éds.). *Langage et pertinence*, Nancy : Presses Universitaires de Nancy, p. 175-228.
- MABILLON, Jean (1709). *De re diplomatica libri VI*, Paris : C. Robustel, 680 p.
- MALRIEUX, Denise (2004). « Linguistique de corpus, genres textuels, temps et personnes », *Langages*, n° 153, 2004, p. 73-85.
- MARCHELLO-NIZIA, Christiane (1978). « Ponctuation et "unités de lecture" dans les manuscrits médiévaux ou : je ponctue, tu lis, il théorise », *Langue française*, n° 40, 1978, p. 32-44.
- (1985). *Dire le vrai : l'adverbe "si" en français médiéval : essai de linguistique historique*, Genève : Droz, 256 p.
- (éd.) (1995). *Tristan et Yseut : les premières versions européennes*, Paris : Gallimard, 1728 p.
- (2004). « Linguistique historique, linguistique outillée : les fruits d'une tradition », *Le français moderne*, n° 1, 2004, p. 58-70.
- (2007). « Le *comma* dans un manuscrit en prose du 13^e siècle : grammaticalisation d'un marqueur de corrélation, ou marquage d'intonation ? », in : Olivier Bertrand *et al.* (éds.). *Discours, diachronie, stylistique du français. Études en hommage à Bernard Combettes*, Berne : Peter Lang, p. 293-305.

- (2008). *Queste del saint Graal : Édition électronique*, Lyon : ENS Lettres et sciences humaines, [Prototype en ligne] <<http://weblex.ens-lsh.fr/pub/kq/>>
- MARNETTE, Sophie (2006). « La signalisation du discours rapporté en français médiéval », *Langue française*, vol. 149, 2006, p. 31-47.
- MARTIN, Henri-Jean (1977). « Pour une histoire de la lecture », *Revue française d'histoire du livre*, vol. 16, 1977, p. 583-609.
- MASAI, François (1950). « Principes et conventions de l'édition diplomatique », *Scriptorium*, vol. 4, 1950, p. 177-193.
- MAZZIOTTA, Nicolas (2004). « Le texte dans tous ses états, philosophie d'encodage du projet *Khartés* », in : Gérald Purnelle, Cédric Fairon et Anne Dister (éds.) *Le poids des mots*, 10 - 12 mars 2004), Louvain-la-Neuve : Presses universitaires de Louvain, p. 793-803.
- (2006). *Propositions pour l'édition des sources primaires CCFM*, (Publié en ligne le 13/02/2008), <http://ccfm.ens-lsh.fr/IMG/pdf/CCFM-codage-manuscripts_ULG.pdf> (Consulté le 3 mars 2008).
- (2007a). « Ponctuation et parties du discours diplomatique dans les actes émanés de la cour allodiale de Liège entre 1260 et 1290 », in : Alexei Lavrentiev (éd.). *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies*, Chambéry : Université de Savoie, p. 67-84.
- (2007b). *Ponctuation et syntaxe dans la langue française médiévale : Étude d'un corpus de chartes originales écrites à Liège entre 1236 et 1291*, Thèse de Doctorat non publiée, Liège, Université de Liège, 396 p.
- (2008a). « Ponctuation et syntaxe en ancien français : Étude d'un corpus de chartes écrites à Liège avant 1292. Méthode de travail », in : Jacques Durand, Benoît Habert et Bernard Lacks (éds.) *Congrès mondial de linguistique française - CMLF'08*, (Actes de colloque, Paris, 9 - 12 juillet 2008), Paris : Institut de linguistique française, p. 267-277. [En ligne] <<http://www.linguistiquefrancaise.org>>.

Bibliographie

- (2008b). « Traiter les abréviations du français médiéval. Théorie de l'écriture et pratiques d'encodage », *Corpus*, n° 7, 2008b, p. 191-218.
- (à paraître_a). « Étude linguistique de la ponctuation d'une charte originale en langue française (Liège 1272) », *Orbis. Bulletin international de documentation linguistique*, à paraître.
- (à paraître_b). « Ponctuation de la proposition dans les chartes originales écrites en français à Liège », in : Heidi Siller-Runggaldier (éd.) *Actes du XXV^e CILPR*, (Actes de colloque, Innsbruck, 3 - 9 septembre 2007), Tübingen : Max Niemeyer, à paraître.
- MICHA, Alexandre (1960). « Les manuscrits de *Lancelot en prose* », *Romania*, vol. 81, n° 322, 1960, p. 145-187.
- MÖHREN, Frankwalt (2005). Compte-rendu de Holtus *et al.* 2003, *Zeitschrift für französische Sprache und Literatur*, n° 115, 2005, p. 208-211.
- MONFRIN, Jacques (1958). « Fragments de la chanson d'*Aspremont* conservés en Italie », *Romania*, vol. 79, 1958, p. 237-260; 376 - 405.
- MOREAU-MARÉCHAL, Juliette (1968). « Recherches sur la ponctuation », *Scriptorium*, vol. 22, n° 1, 1968, p. 56-65.
- MORGAN, Margaret R. (éd.) (1982). *La continuation de Guillaume de Tyr : 1184-1197*, Paris : P. Geuthner, 220 p.
- MORPURGO DAVIES, Anna (1998). *History of linguistics. Volume IV : Nineteenth century linguistics*, Giulio Lepschy (éd.), London, New York : Longman, 434 p.
- MORTARA GARAVELLI, Bice (éd.) (2008). *Storia della punteggiatura in Europa*, Roma : Laterza, Coll. Enciclopedia del sapere, 650 p.
- MOŠIN, Vladimir A. et TRALJIĆ, Seid M. (1957). *Filigranes des XIII^e et XIV^e siècles*, Zagreb : Académie yougoslave des sciences et des beaux-arts, Institut d'histoire, 173 p.

- MOURIN, Louis (1946a). « L'œuvre oratoire de Jean Gerson et les manuscrits qui la contiennent », *Archives d'histoire doctrinale et littéraire du Moyen Age*, vol. 15, 1946a, p. 225-261.
- MURRAY, Sarah-Jane (2007). « Medieval scribes, Modern scholars : reading *Le Chevalier de la Charrette* in the twenty-first century », in : Michael Stolz *et al.* (éds.). *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien*, Bern : Germanistik.Ch, p. 140-159. [Également disponible en ligne] <<http://www.germanistik.ch>>.
- NAÏS, Hélène (1979). « La ponctuation dans le manuscrit B de Villehardouin », in : Nina Catach et Claude Tournier (éds.). *La ponctuation. Recherches historiques et actuelles. Fascicule 2. Actes de la Table ronde internationale C.N.R.S. de mai 1978*, Paris ; Besançon : C.N.R.S. - Groupement de recherches sur les textes modernes, p. 45-55.
- NICHOLS, Stephen G. (1990). « Introduction: Philology in a Manuscript Culture », *Speculum*, vol. 65, n° 1, janvier 1990, p. 1-10.
- NUNBERG, Geoffrey (1990). *The linguistics of punctuation*, Stanford, CA : Center for the Study of Language and Information, 141 p. [Également disponible en ligne] <<http://books.google.fr/books?id=Sh-sruuKjJwC>>.
- ÖHLUND-RAMBAUD, Stéphanie (2006). « L'atelier de Jean Trepperel, imprimeur-libraire parisien (1492-1511) », in : Godfried Croenen et Peter Ainsworth (éds.). *Patrons, authors and workshop*, Louvain ; Paris ; Dudley (Mass.) : Peeters, Coll. Synthesa, p. 123-141.
- OPSOMER, Carmélia (2003). « Le scribe, l'enlumineur et le commanditaire : à propos des *Tacunia sanitatis* illustrés », in : *La collaboration dans la production de l'écrit médiéval. Actes du XIII^e colloque du Comité international de paléographie latine*, (Actes de colloque, Weingarten, 22 - 25 septembre 2000), Paris : cole nationale des chartes, p. 183-192.
- OUY, Gilbert (1987). « Orthographe et ponctuation dans les premiers manuscrits autographes des humanistes français des XIV^e et XV^e siècles », in : Alfonso

Bibliographie

- Marierù (éd.). *Grafia e interpunzione del latino nel medioevo (Seminare internazionale, Roma, 27-29 settembre 1984)*, Roma : Edizioni dell'Ateneo, p. 169-205.
- PARIS, Gaston et PANNIER, Léopold (1872). *La Vie de Saint Alexis. Poème du XI^e siècle et renouvellements des XII^e, XIII^e et XIV^e siècles publié avec préface, variantes, notes et glossaires*, Paris : F. Franck, 416 p.
- PARKES, Malcolm B. (1992). *Pause and effect: an introduction to the history of punctuation in the West*, Aldershot : Scolar Press, 327 p.
- PAUPHILET, Albert (1921). *La tradition manuscrite et l'établissement du texte de la Queste del Saint Grraal, attribuée à Gautier Map*, Paris : Champion, 207 p.
- PERRET, Michèle (1981). « A la fin de sa vie ne fuz-je mie », *Revue des Sciences Humaines*, n° 183, 1981, p. 17-37.
- PERROT, Jean (1980). « Ponctuation et fonctions linguistiques », *Langue française*, vol. 45, 1980, p. 67-76.
- PÉTILLON-BOUCHERON, Sabine (2002). *Les détours de la langue. Étude sur la parenthèse et le tiret double*, Louvain ; Paris ; Dudley (Mass.) : Peeters, 363 p.
- PETRUCCI, Armando (1977). « Edizione diplomatica o/e riproduzione ? Un problema critico (con appendice Boccacciana) », *Belfagor*, vol. 32, n° 1, 1977, p. 63-71.
- PICKFORD, Cedric Edward (éd.) (1973). *Lancelot du Lac*, Menston : Scolar Press, 800 p.
- PICOCHÉ, Jacqueline et MARCHELLO-NIZIA, Christiane (³1994). *Histoire de la langue française*, Paris : Nathan, 397 p.
- PIGNATELLI, Cinzia (2007). « Présence et fréquence de la ponctuation dans les manuscrits en vers du XIII^e siècle : les huit manuscrits du *Chevalier de la Charrette* au banc d'essai », in : Alexei Lavrentiev (éd.). *Systèmes graphiques de manuscrits médiévaux et incunables français : ponctuation, segmentation, graphies*, Chambéry : Université de Savoie, Coll. Langages, p. 85-105.

- PLOUZEAU, May (1994). « A propos de *La Mort Artu* de Jean Frappier », *Travaux de linguistique et de philologie*, vol. 32, 1994, p. 207-221.
- POIRION, Daniel (1978). « Les paragraphes et le pré-texte de Villehardouin », *Langue française*, n° 49, 1978, p. 59-44. [Également disponible en ligne] <<http://www.persee.fr>>.
- PRÉVOST, Sophie (2001a). *Étiquetage morpho-syntaxique de la BFM : CATTEX (avec explicitation des choix linguistiques)*, Lyon : Projet BFM, 18 p. [En ligne] <<http://ccfm.ens-lsh.fr/IMG/pdf/cattex2001.pdf>>
- (2001b). *La postposition du sujet en français aux XV^e et XVI^e siècles*, Paris : CNRS Éditions, 139 p.
- (2005). « Constitution et exploitation d'un corpus de français médiéval : enjeux, spécificités et apports », in : Anne Condamines (éd.). *Semantique et corpus*, 4, Paris : Lavoisier, Coll. Cognition et traitement de l'information, p. 147-176.
- (2008). « Corpus informatisés de français médiéval : contraintes sur leur constitution et spécificités de leurs apports », *Corpus*, n° 7, 2008, p. 35-64. [Également disponible en ligne] <<http://corpus.revues.org/>>.
- PROU, Maurice (⁴1924). *Manuel de paléographie latine et française*, Paris : Auguste Picard, 4^e éd. 511 p.
- RAFTI, Patrizia (1988). « L'interpunzione nel libro manoscritto : mezzo secolo di studi », *Scrittura et civiltà*, vol. 12, 1988, p. 239-298.
- RAJNA, Pio (1929). « -X = -US », *Romania*, vol. 55, 1929, p. 528-528.
- RENOUARD, Philippe (1965). *Répertoire des imprimeurs parisiens libraires, fondateurs de caractères et correcteurs d'imprimerie, depuis l'introduction de l'imprimerie à Paris (1470) jusqu'à la fin du seizième siècle... Avertissement, table des enseignes et adresses, liste chronologique par Jeanne Veyrin-Forrer et Brigitte Moreau. [Nouvelle édition.]*, Paris : Lettres modernes (Abbeville, impr. F. Paillart),

Bibliographie

- RIEGEL, Martin, PELLAT, Jean-Christophe et RIOUL, René (1997). *Grammaire méthodique du français*, Paris : Press Universitaires de France, 646 p.
- ROQUES, Gilles (1982). « Étude lexicologique du *Mystère de Saint Martin* d'Andrieu de la Vigne (vers 1 à 2500) », in : Peter Wunderli (éd.) *Du mot au texte. Actes du III^e Colloque international sur le Moyen français*, (Actes de colloque, Düsseldorf, 17-19 septembre 1980), Tübingen : Gunter Narr, Coll. Tübinger Beitrage zur Linguistik, p. 103-114. [Également disponible en ligne] <<http://books.google.fr/books?id=Q1GI4y6grd8C>>.
- ROQUES, Mario (1926). « Établissement des règles pratiques pour l'édition des anciens textes français et provençaux », *Romania*, vol. 52, 1926, Rapport de la 2^e commission. p. 243-249. [Également disponible en ligne] <<http://gallica.bnf.fr>>.
- (1952). « Le manuscrit fr. 794 de la Bibliothèque Nationale et le scribe Guiot », *Romania*, vol. 73, 1952, p. 177-199.
- SAENGER, Paul (1997). *Space between words. The origins of silent reading*, Stanford, CA : Stanford university Press, 480 p.
- SAMARAN, Charles (1929). « Fragment d'une traduction en prose française du psautier composée en Angleterre au XII^e siècle », *Romania*, vol. 55, 1929, p. 161-173.
- SCHALDE, Igor, ANTOINE, Jean-Yves, LE PÉVÉDIC, Brigitte et POIRIER, Franck (2004). « SibyMot : Modélisation stochastique du langage intégrant la notion de *chunks* », in : Bernard Bel et Isabelle Marlien (éds.) *Actes du colloque TALN 2004*, (Actes de colloque, Fès, 19-24 avril), p. 1-6. (Consulté le 10/04/2009).
- SCHNEUWLY, Bernard, ROSAT, Marie-Claude et DOLZ, Joachim (1989). « Les organisateurs textuels dans quatre types de textes écrits. Étude chez des élèves de dix, douze et quatorze ans », *Langue française*, n° 81, 1989, p. 40-58.
- SERÇA, Isabelle (2004). « La ponctuation : petit tour d'horizon », *L'Information grammaticale*, n° 102, juin 2004, p. 11-17.

- SEVILLA ARROYO, Florencio (³1999). *Miguel de Cervantes Saavedra. Obras completas*, Madrid : Castalia, 1220 p.
- SHORT, Ian (2005). « Part 1. The Oxford version », in : Joseph J. Duggan (éd.). *La chanson de Roland. The song of Roland. The French corpus*, Turnhout : Brepols, p. I/1-I/338.
- SINCLAIR, John (1996). *Preliminary recommendations on corpus typology*, EAGLES (Expert Advisory Group on Language Engineering Standards), 13 p. [En ligne] <www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/corpus typ.ps.gz>.
- SLETSJÖE, Leif (éd.) (1968). *Le mystère d'Adam. Édition diplomatique accompagnée d'une reproduction photographique du manuscrit de Tours et des leçons des éditions critiques*, Paris : C. Klincksieck, Coll. Bibliothèque française et romane, 89 p.
- STANOVAĀ, Lydia A. (2003). « La standardisation en ancien français », in : Michèle Goyens et Werner Verbeke (éds.). *The Dawn of the Written Vernacular in Western Europe*, Leuven : Leuven University Press, p. 241-272.
- STASSE, Micheline (1998). « Retour aux sources : point final ! », in : Jean-Marc Defays et al. (éds.). *À qui appartient la ponctuation ?* Bruxelles : Duculot, Coll. Champs linguistiques, p. 179-185.
- STIENNON, Jacques (²1991). *Paléographie du Moyen Âge*, Paris : Armand Colin, 2^e éd. 368 p.
- STIENNON, Jacques et HASENOHR, Geneviève (1973). *Paléographie du Moyen Âge*, Paris : Armand Colin, 352 p.
- SWEETSER, Franklin P. (éd.) (1966). *Les cent nouvelles nouvelles*, Genève : Droz, 651 p.
- TESNIÈRE, Lucien (²1965). *Éléments de syntaxe structurale*, Paris : Klincksieck, 674 p.

Bibliographie

- TOURATIER, Christian (2006). « Que faut-il entendre par “connecteur” ? », in : *La connexion et les connecteurs*, Aix-en-Provence : Publications de l'université de Provence, Coll. Travaux, p. 19-40.
- TOURNIER, Claude (1980). « Histoire des idées sur la ponctuation des débuts de l'imprimerie à nos jours », *Langue française*, n° 45, 1980, p. 28-40.
- TOUSTIN, Charles F. et TASSIN, René P. (1750). *Nouveau traité de diplomatique*, Paris : Desprez & Cavelier, 719 p.
- VAN DIJK TEUN, Adrianus (1977). *Text and context explorations in the semantics and pragmatics of discourse*, London : Longman,
- VÉDÉNINA, Ludmilla (1973). « La transmission par la ponctuation des rapports du code oral avec le code écrit », *Langue française*, vol. 19, n° 1, 1973, Consulté le 27 avril 2009. p. 33-40. [Également disponible en ligne] <<http://www.persee.fr>>.
- (1980). « La triple fonction de la ponctuation dans la phrase : syntaxique, communicative et sémantique », *Langue française*, n° 45, 1980, [Également disponible en ligne] <<http://www.persee.fr>>.
- (1989). *Pertinence linguistique de la présentation typographique*, Paris : Peeters/SELAF, 153 p.
- VEYRIN-FORRER, Jeanne (1987). *La lettre et le texte. Trente années de recherches sur l'histoire du livre*, Paris : École normale supérieure de jeunes filles, 486 p.
- VIELLIARD, Françoise et GUYOTJEANNIN, Olivier (coord.) (2001). *Conseils pour l'édition des textes médiévaux, Fascicule I, Conseils généraux*, Paris : CTHS, cole nationale des chartes, 175 p.
- WAGNER, Robert-Léon (éd.) (1949). *Textes d'étude (Ancien et moyen français)*, Genève ; Paris : Droz ; Minard, 206 p.
- WAGNER, Robert-Léon et COLLET, Olivier (éds.) (1995). *Textes d'étude (Ancien et moyen français)*, Genève : Droz, 382 p.
- WALTHER, Iohann L. (1752). *Lexicon diplomaticum, abbreviationes syllabarum et vocum in diplomatibus et codicibus a seculo VIII ad XVI usque occurrentes*

exponens, junctis alphabetis et scripturae speciminibus integris, Göttingen :
Iohann Wilh. Schmidt, 282 p.

ZINK, Gaston (1986). *Phonétique historique du français*, Paris : PUF, 254 p.

ZUMTHOR, Paul (1984). *La poésie et la voix dans la civilisation médiévale*, PUF,
117 p.

Éditions de référence des textes du corpus

BONNARDOT, François et LONGNON, Auguste (1878). *Le saint voyage de
Jherusalem du seigneur d'Anglure*, Paris : Firmin-Didot, 178 p.

BOUDET, Jean-Patrice (éd.) (1997). *Le Recueil des plus célèbres astrologues de Simon
de Phares. Tome I^{er}. Édition critique*, Paris : H. Champion, 604 p.

DILLER, George T. (éd.) (1972). *Froissart. Chroniques. Début du premier livre.
Édition du manuscrit de Rome Reg. lat. 869*, Genève ; Paris : Droz ; Minard,
Coll. TLF, 1010 p.

DOUËT-D'ARCQ, Louis (1857-1862). *La chronique d'Enguerran de Monstrelet*,
Paris : Veuve J. Renouard, 2 vol., 416 et 479 p.

DROZ, Eugénie (²1950). *Alain Chartier. Le quadrilogue invectif*, Paris : Champion,
82 p.

FARAL, Edmond (éd.) (²1961). *Villehardouin. La conquête de Constantinople*, Paris :
Les belles lettres, Coll. Classiques de de l'histoire de France au Moyen Âge,
229 p.

FRAPPIER, Jean (³1964). *La mort le roi Artu*, Genève ; Paris : Droz ; Minard, 308 p.

GLORIEUX, Palémon (éd.) (1960-1973). *Jean de Gerson. Œuvres complètes*. Paris ;
Tournai, Londres : Desclée, 11 vol.

LEWIS, Peter S. et HAYEZ, Anne-Marie (1978). *Écrits politiques de Jean Juvénal des
Ursins. Tome I*, Paris : Klincksieck, 551 p.

Bibliographie

- MARCHELLO-NIZIA, Christiane (2008). *Queste del saint Graal : Édition électronique*, Lyon : ENS Lettres et sciences humaines, [Prototype en ligne] <<http://weblex.ens-lsh.fr/pub/kq/>>.
- MICHA, Alexandre (éd.) (1979). *Lancelot (D'une aventure d'Agravain jusqu'à la fin de la quête de Lancelot par Gauvain et ses compagnons)*, Genève ; Paris : Droz, 406 p.
- MONFRIN, Jacques (éd.) (1998). *Joinville. Vie de saint Louis*, Paris : Classiques Garnier, 487 p.
- MOURIN, Louis (éd.) (1946b). *Six sermons français inédits de Jean Gerson*, Paris : J. Vrin, 311 p.
- PARIS, Paulin (éd.) (1879). *Guillaume de Tyr et ses continuateurs : texte français du XIII^e siècle*, Paris : Firmin-Didot, Coll. Histoire générale des croisades par les auteurs contemporains, 559 p.
- PAUPHILET, Albert (éd.) (1923). *La Queste del Saint Graal. Roman du XIII^e siècle*, Paris : Champion, 302 p.
- PRIOR, Oliver H. (1913). *L'Image du Monde de Maître Gossouin : Rédaction en prose. Texte du Manuscrit de la Bibliothèque Nationale, Fonds Français No. 574. Avec corrections d'après d'autres manuscrits, Notes et Introduction*, Lausanne; Paris : Payot, 216 p.
- QUEUX DE SAINT-HILAIRE, Auguste-Henry-Edouard (éd.) (1878). *Eustache Deschamps. Œuvres complètes publiées d'après le manuscrit de la Bibliothèque nationale*, Paris : Firmin-Didot, Coll. Société des anciens textes français, 415 p.
- RAYNAUD, Gaston (éd.) (1891). *Eustache Deschamps. Œuvres complètes publiées d'après le manuscrit de la Bibliothèque nationale*, Paris : Firmin-Didot, 384 p.
- RYCHNER, Jean (éd.) (1963). *Les XV joies de mariage*, Genève ; Paris : Droz ; Minard, 205 p.
- SALMON, Amédée (éd.) (1899). *Philippe de Beaumanoir. Coutumes de Beauvaisis. Texte critique publié avec une introduction, un glossaire et une table analytique*,

Paris : A. Picard et fils, Coll. Collection de textes pour servir à l'étude et à l'enseignement de l'histoire, 512 p.

SEIGNEURET, Jean-Charles (éd.) (1966). *Le roman du comte d'Artois*, Genève : Droz, 211 p.

TILANDER, Gunnar (éd.) (1971). *Gaston Phébus. Livre de chasse*, Karlshamn : E. G. Johansson, 453 p.

Sources des transcriptions

anglure15217 – Ms. Paris, BnF, fr. 15217 (2e m. XV^e s.), f^o 27r-30v : Ogier d'Anglure, *Le Saint voyage de Jherusalem*.

- Édition de référence (Bonnardot & Longnon 1878).

artois11610 – Ms. Paris, BNF, fr. 11610 (entre 1453 et 1467), f^o 1r-4r, *Roman du comte d'Artois*.

- Édition de référence (Seigneurret 1966).

artu1488a – Imprimé à Paris, Jehan du Pré, 1488 (BnF, rés. Y² 47), f^o 366r : *La mort au roy Artus*.

- Édition de référence (Frappier 1964).

beauma11652 – Ms. Paris, BnF, fr. 11652, (ca. 1300), f^o 4r-5r, Philippe de Beaumanoir, *Coutumes de Beauvaisis*.

- Édition de référence (Salmon 1899).

chasse619 – Ms. Paris, BnF, fr. 619 (vers 1440), f^o 1r-4v : Gaston Phébus, *Livre de chasse*.

- Édition de référence (Tilander 1971).

chasse765 – Ms. Bibliothèque municipale de Lyon, 765 (XV^e s.), f^o 100r et 102v : Gaston Phébus, *Livre de chasse*.

- Édition de référence (Tilander 1971).

Bibliographie

dictier840 – Ms. Paris, BnF, fr. 840 (ca. 1400), 394r-396v : Eustache Deschamps, *Art de dictier*.

- Édition de référence (Queux de Saint-Hilaire 1878)

dolet1540 – Imprimé à Lyon, Estienne Dolet, 1540 (BnF, Gallica), p. 12-14 : Estienne Dolet, *La maniere de bien traduire d'une langue en aultre*.

- Pas d'édition de référence.

egipte769 – Ms. Lyon, Bibliothèque municipale, 769 (XV^e s.), f^o 4v-5r : Fragment d'une description de l'Orient.

- Inédit.

froissart869 – Ms. Vatican, Reg. Lat. 869 (entre 1399 et 1405), f^o 1v – 2r : Jean Froissart, *Chroniques* (livre premier).

- Édition de référence (Diller 1972).

gerson13318 – Ms. Paris, BnF, fr. 13318 (XV^e s.), f^o 165v-170r : Jean Gerson, *Sermon pour la fête de la Sainte Trinité*.

- Édition de référence (Mourin 1946b).

im1495a – Imprimé à Paris, Jehan Treperl, entre 1492 et 1498 (BnF rés. D. 4722), f^o 2r-2v : Gossouin, *Image du monde* (en prose).

- Édition de référence (Prior 1913).

im1522a – Imprimé à Paris, Jehan Janot, ca. 1522, p. 8-9 : Gossouin, *Image du monde* (en prose).

- Édition de référence (Prior 1913).

im574 – Ms. BnF, fr. 574 (ca. 1315-1320), f^o 4v-6r : Gossouin, *Image du monde* (en prose).

- Édition de référence (Prior 1913).

im587 – Ms. Paris, Sainte-Geneviève, 587; (fin XIII^e / début XIV^e s.), f^o 172v : Gossouin, *Image du monde* (en prose).

- Édition de référence (Prior 1913).

joachim768 – Ms. Lyon, Bibliothèque municipale, 768 (XIV^e s.), f^o 97v-98r : *La prophécie de Mestre Joachin l'astronome*.

- Inédit.

joinville13568 – Ms. Paris, BnF, fr. 13568 (ca 1335), p. 1-7 : Jean de Joinville, *Vie de saint Louis*.

- Édition de référence (Monfrin 1998).

lancelot77 – Ms. Lyon, Bibliothèque municipale, fonds du Palais des arts 77 (XIII^e s.), f^o 8v-9r : *Lancelot du lac*.

- Édition de référence (Micha 1979).

monstre2682 – Ms. Paris, BnF, fr. 2682, (XV^e s.), f^o 233r-233v : Enguerrand de Monstrelet, *Chroniques* (livre 2).

- Édition de référence (Douët d'Arcq 1862).

phares1357 – Ms. Paris, BnF, fr. 1357 (1494-1498), f^o 1v-5r : Simon de Phares, *Recueil des plus celebres astrologues*.

- Édition de référence (Boudet 1997).

qgraal1119 – Ms. Paris, BnF, n.a. fr. 1119 (XIII^e – début XIV^e s.), f^o 138r-141v : *Queste del saint graal*.

- Édition de référence (Pauphilet 1923)

qgraal77 – Ms. Lyon, Bibliothèque municipale, fonds du Palais des arts 77 (XIII^e s.), f^o 160v-164r : *Queste del saint graal*.

- Édition de référence (Marchello-Nizia 2008).

qjm1052 – Ms. Rouen, Bibliothèque municipale, 1052 (daté 1464), f^o 87r-89r : *Les .xv. joies de mariage*.

- Édition de référence (Rychner 1958).

quadrilogue126 – Ms. Paris, BnF, fr. 126 (1^{ère} m. XV^e s.), 191r-192v : Alain Chartier, *Quadrilogue invectif*.

- Édition de référence (Droz 1950).

Bibliographie

strasb9768 – Ms. Paris, BnF, lat. 9768 (ca. 1000), f° 13r : *Les serments de Strasbourg* (*Chronique* de Nithard).

- Édition de référence (Wagner & Collet 1995).

tyr828 – Ms. Lyon, Bibliothèque municipale, 828 (XIV^e s.), f° 1r et 5r : Guillaume de Tyr, *Le liures deracles*.

- Édition de référence (Paris 1879).

ursins2701 – Ms. Paris, BnF, fr. 2701 (3^e q. XV^e s.), f° 121r-121v : Jean Juvenal des Ursins, *Proposition...*

- Édition de référence (Lewis & Hayez 1978).

villehardouin12203 – Paris, BnF, ms. fr. 12203 (XIII^e s.), f° 69r-70v : Geoffroi de Villehardouin, *Conqueste de constantinople*.

- Édition de référence (Faral 1961).

villehardouin587 – Ms. Oxford, Bodleian, Laud. Misc. (XIV^e s.), f° 69r-70v : Villehardouin, *Conqueste de constantinople*.

- Édition de référence (Faral 1961).

Listes et indexes

Liste des tableaux

Tableau 1 : Fonctions des marques segmentales et graphiques dans les manuscrits médiévaux et dans les textes modernes.....	91
Tableau 2: Variables applicables à l'ensemble du corpus.....	152
Tableau 3: Variables applicables à un sous-corpus de textes caractérisés plus précisément.....	153
Tableau 4 : Dimensions du corpus BFM-MSS.....	158
Tableau 5 : Synthèse de la représentation des « données sensibles » aux différents niveaux de la transcription de manuscrits médiévaux.....	210
Tableau 6 : Conventions de la présentation des citations du corpus.....	216
Tableau 7 : Taux général et force de la ponctuation dans l'ensemble du corpus.....	436
Tableau 8 : Fréquences des formes des marques de ponctuation dans l'ensemble du corpus....	440
Tableau 9: Traitement de la ponctuation scribale dans les éditions scientifiques.....	448

Liste des figures

Fig. 1 : Typologie formelle des marques linguistiques.....	77
Fig. 2 : « Pied-de-mouche » (artu1488, p. 50).....	86
Fig. 3 : Début du chapitre 6 de Don Quichotte (Madrid, Juan de la Cuesta, 1605, f. 15).....	100
Fig. 4 : Début d'un chapitre dans l'image du mode (im574, 4v).....	101
Fig. 5 : Mise en forme de l'identification d'un personnage parlant (« rubrique énonciative ») (quadrilogue126, 192b).....	102
Fig. 6 : Crochets adlinéaires (gerson13318, 166r et 168v).....	121
Fig. 7 : « Force 3 » de la ponctuation : lettrine au milieu de ligne (joinville13568, p. 2).....	126
Fig. 8 : « Grande minuscule » (qgraal77, 161d).....	127
Fig. 9 : Paramètres de la distance communicative (Koch & Osterreicher 2001, p. 186).....	148
Fig. 10 : Facsimilé des Serments de Strasbourg (ms. Paris, BnF lat. 9768, 13r, cliché BnF).....	172
Fig. 11 : Transcription des Serments de Strasbourg (Koschwitz 1920, p. 2).....	172
Fig. 12 : Transcription des Serments de Strasbourg (Lauer 1926, p. 104).....	174
Fig. 13 : Transcription des Serments de Strasbourg (Wagner 1949, p. 5).....	175
Fig. 14 : Transcription des Serments de Strasbourg (Henry 1953, p. 1).....	175
Fig. 15 : Transcription des Serments de Strasbourg (Wagner & Collet 1995, p. 9).....	176
Fig. 16 : Le mystère d'Adam (Sletsjöe 1968, p. 6-7).....	179
Fig. 17 : Différents types d'éditions sur un axe de normalisation.....	183
Fig. 18 : Types de transcriptions précisés.....	188
Fig. 19 : Représentation d'une abréviation dans les trois niveaux de transcription.....	194
Fig. 20 : Série de jambages dans le mot 'livre' (im574, 172r).....	198
Fig. 21 : Ambigüité du groupe de lettres ct/tt (dictier 840, 395r).....	199
Fig. 22 : Transcription normalisée (« orthographique ») des Serments de Strasbourg.....	211
Fig. 23 : Transcription diplomatique (graphématique) des Serments de Strasbourg.....	211
Fig. 24 : Transcription imitative (allographique) des Serments de Strasbourg.....	212
Fig. 25 : Exemple de balisage d'un mot « diffracté ».....	226
Fig. 26 : Balisage d'une agglutination selon les recommandations du projet Menota.....	227
Fig. 27 : Balisage d'une agglutination dans notre corpus.....	227
Fig. 28 : Balisage d'une déglutination dans notre corpus.....	228
Fig. 29 : Un pied-de-mouche dans la Mort au roy artus (1488).....	230

Fig. 30 : Doubles barres obliques dans le ms. Paris, BnF fr. 2862 (XVe s.).....	230
Fig. 31 : Punctus interrogativus à la fin d'une incise (qgraal77, 3b).....	245
Fig. 32 : Ponctuation du discours direct (qgraal77, 161a).....	246
Fig. 33 : Manuscrit Lyon, BM, 828, f° 1r (fragment).....	253
Fig. 34 : Ponctuations « incertaines » (tyr828, 5b:15 et 1b:9).....	255
Fig. 35 : Barre oblique légère (im587, 172v).....	264
Fig. 36 : « Fausses » barres obliques (im587, 172d).....	264
Fig. 37 : Marques de fin de ligne (im587, 172d).....	265
Fig. 38 : Initiale en couleur au milieu d'une ligne (joinville13568, 2d).....	282
Fig. 39 : « Item » avec une initiale en couleur (joachim768, 97d).....	290
Fig. 40 : Pied-de-mouche (joachim768, 97c).....	290
Fig. 41 : Litterae oblongatae (joachim768, 97d).....	292
Fig. 42 : Ponctuation « décorative » (joachim768, 97d).....	293
Fig. 43 : Ponctuation incertaine (ou trace de grattage) (joachim768, 98a).....	294
Fig. 44 : Ponctuation en forme de trois points (dictier840).....	302
Fig. 45 : Pied-de-mouche en marge (dictier840).....	302
Fig. 46 : Lettre H majuscule et minuscule (dictier 840).....	303
Fig. 47 : Petite barre oblique en haut de ligne (froissart869, 2v).....	314
Fig. 48 : Absence de ponctuation en fin de ligne (quadrilogue126, 192a).....	321
Fig. 49 : Pied-de-mouche (quadrilogue126, 191a).....	321
Fig. 50 : Lettrine et pied-de-mouche (quadrilogue126, 192b).....	322
Fig. 51 : Aperçu de la décoration (quadrilogue126, 191v et 192r).....	324
Fig. 52 : Majuscules L avec des points (chasse619, 1r).....	330
Fig. 53 : Pied-de-mouche et majuscule surlignée (chasse765, 102v).....	337
Fig. 54 : Petite barre oblique devant une perluète (chasse765, 100r).....	340
Fig. 55 : P minuscule « ordinaire » et « élargie » (gerson13318, 166v).....	344
Fig. 56 : N minuscule « ordinaire », « élargie » et majuscule (gerson13318, 168v, 169v et 170r).....	345
Fig. 57 : Crochets adlinéaires (gerson13318, 166r et 168v).....	346
Fig. 58 : Espace blanc allongé (gerson13318, 170r).....	347
Fig. 59 : Majuscule décorée en début d'une page (egypte769, 5r).....	357
Fig. 60 : Marque de « fin de section » (artois 11610, 4r).....	361
Fig. 61 : Blanc allongé (ursins2701, 121b).....	369
Fig. 62 : A minuscule et majuscule (ursins2701, 121r).....	370

Listes et indexes

Fig. 63 : Mise en forme du titre (ursins2701, 121a).....	371
Fig. 64 : Une « parenthèse » devant le discours direct (qjm1052, 88r).....	379
Fig. 65 : Lettre < d > majuscule et minuscule (anglure15217, 27v).....	391
Fig. 66 : Mise en forme des paragraphes (anglure15217, 28v).....	392
Fig. 67 : Majuscules « décoratives » à la dernière et à la première ligne d'une page (anglure15217, 27v-28r).....	397
Fig. 68 : « Petite majuscule » (anglure15217, 28r).....	397
Fig. 69 : Point suivi d'une barre verticale (phares1357, 3v).....	401
Fig. 70 : Losange de quatre points (phares1357, 1v).....	401
Fig. 71 : Longueur de séquences délimitées par des ponctuations fortes (phares 1357).....	407
Fig. 72 : Point « haut » (artu1488, l.4516, 4476).....	411
Fig. 73 : Caractère < m > retourné (artu1488, l.317).....	412
Fig. 74 : Espace blanc allongé (artus1488, l.828).....	413
Fig. 75 : Distribution des marques de ponctuation dans im1495a et im1522a.....	423
Fig. 76 : Analyse factorielle : distribution des textes et des types de ponctuation.....	452
Fig. 77 : Analyse factorielle : « triangles de force » par type de frontière ponctuable.....	455
Fig. 78 : Analyse factorielle : groupements de textes.....	457